Chapter 15
Linking Landscapes and Metacommunities

Joseph R. Bennett* and Benjamin Gilbert*

OBJECTIVES

Landscape ecology is the study of interactions between spatial landscape patterns and ecological processes, typically examining real landscapes at broad spatial scales. Metacommunity ecology focuses more specifically on how spatial processes alter species interactions and typically involves a localized spatial extent and more abstracted spatial landscapes (Bolker, 2004). These disciplines have evolved somewhat independently, despite a shared interest in how organisms respond to and interact with spatial phenomena. In this chapter, we combine perspectives from both disciplines using a suite of multivariate spatial statistical techniques designed to help understand the relative importance of abiotic factors (such as climatic gradients, geologic features, and resource availability) and biotic factors (such as predator territoriality and seed dispersal) in determining the abundances of species in communities. To illustrate these techniques, we use a well-known dataset of tropical trees. This lab will enable students to:

1. Utilize semivariograms to examine and understand spatially autocorrelation in species and environmental data;
2. Model distributions of species and communities along environmental gradients using redundancy analysis;

---

*Both authors contributed equally to this work.

J.R. Bennett (✉)
Department of Biology, Carleton University, Ottawa, ON, Canada
e-mail: joseph.bennett@carleton.ca

B. Gilbert
Department of Ecology and Evolutionary Biology, University of Toronto,
Toronto, ON, Canada

S.E. Gergel, M.G. Turner (eds.), *Learning Landscape Ecology*,
DOI 10.1007/978-1-4939-6374-4_15

3. Gain familiarity with multivariate spatial regression as well as the associated tools of trend-surface analysis and eigenvector analysis; and
4. Use joint modeling of abiotic and biotic factors to determine their influences on species distributions in a metacommunity.

The lab is divided into four parts, which build sequentially. In Part 1, we explore the concepts of spatial processes in metacommunities, linking the perspectives of both community and landscape ecology using tree species information from a long-term vegetation study in Barro Colorado Island, Panama. In Part 2, we introduce how communities are measured in order to represent many species at once. We then show how multiple regression can be applied to model communities along abiotic environmental gradients. In Part 3, we introduce two types of spatial methods, called trend-surface analysis and eigenvector analysis, and apply these to the tropical tree species. Finally, in Part 4 we compare the abiotic and spatial determinants of metacommunity structure. We also use our results (from Part 3) to illustrate the advantages and disadvantages of each method, as well as important caveats in the spatial analysis of communities.

This lab aims to make several spatial statistical techniques accessible and understandable to those without extensive training in statistics. However, familiarity with basic regression, especially multiple regression, is extremely helpful. Familiarity with the concepts covered in Chapter 5 (basic semivariograms) is also assumed. This lab assumes prior familiarity with R (including installation procedures), and thus is not intended as the first exposure to the R environment for instructors or students. That being noted, the provided R code (available on the book web site) is very well documented and numbered according the various Figures, Exercises, and Steps in the lab. The lab requires a computer running R version 2.12.0 or higher and access to the datasets provided with this chapter available from the book web site.

INTRODUCTION

Landscape ecology and metacommunity ecology offer different but complementary world views and approaches. **Metacommunity ecology** generally considers entire communities of species and their interactions in a quantitative and spatially explicit (or spatially implicit) way. Metacommunity theory is important for understanding how factors such as habitat suitability and species-specific dispersal abilities impact community-level responses such as alpha and beta diversity. Landscape ecology provides tools and approaches for understanding how the structure of the landscape alters diversity, thus offering a positive feedback between the disciplines. In this chapter, we combine the benefits of both approaches by using maps of species and the environment in conjunction with multivariate analyses of community composition to model relationships among species, underlying environmental conditions, and spatial locations.

Part 1. Biotic and Abiotic Influences in a Metacommunity

Ecologists know that all species are limited in part by resources, predators, competitors, and diseases, and these items can be collectively considered to influence a species' **niche**. For our purposes, a working definition of a species' niche is *the environmental conditions that allow a species to persist*, with "environmental conditions" referring to the collective factors that influence reproduction and mortality. To understand how biotic and abiotic components of a species' niche influence its distribution, we begin by considering one tree species from a 50 ha plot on Barro Colorado Island, Panama (BCI). *Trichilia tuberculata* is a new world tree species in the mahogany family and is relatively common in this forest. A map of its distribution (Figure 15.1, Panel A) shows that *Trichilia* appears to have a patchy distribution, with some areas having high densities (light shading) and other areas having a low density (dark areas).

The distribution of *Trichilia* may reflect areas of favorable environmental conditions. Plant ecologists have hypothesized that these conditions are mainly based on available resources, such as nitrogen, an important nutrient for plants. Therefore,

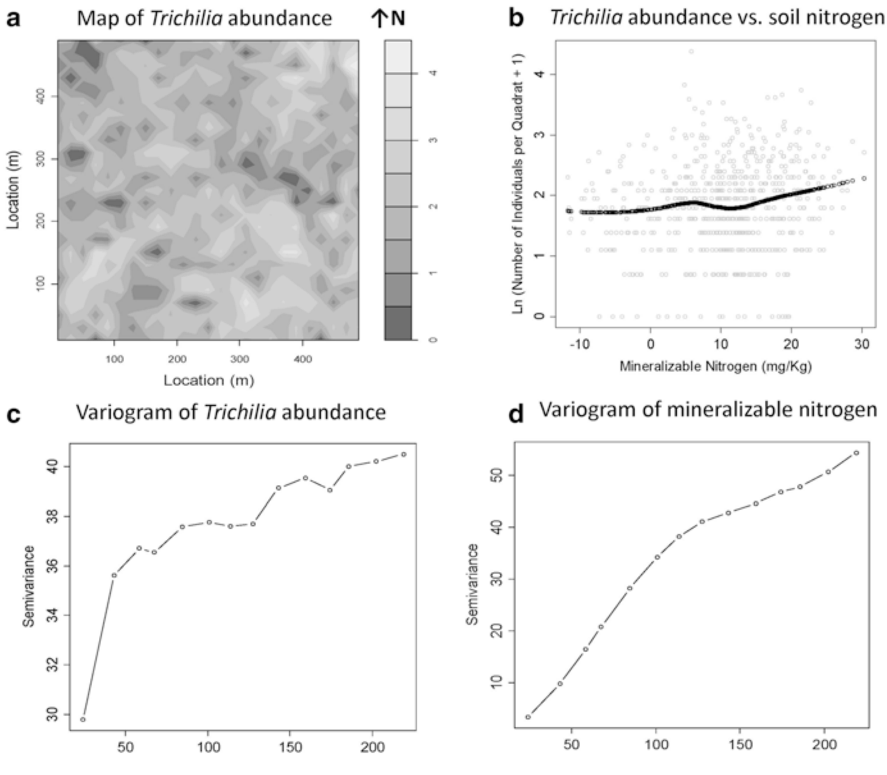

Figure 15.1 Important spatial relationships for *Trichilia tuberculata Panel a*. Abundance in the 500 × 500 m sampling plot. *Panel b*. Relationship to mineralizable nitrogen. *Panel c*. Spatial clustering of *Trichilia. Panel d*. Mineralizable nitrogen

one hypothesis is that the distribution of *Trichilia* is caused by available nitrogen. By plotting the abundance of *Trichilia* against mineralizable nitrogen (Figure 15.1, Panel B), we make a preliminary exploration of this hypothesis. Modeling the relationship between *Trichilia* and mineralizable nitrogen using regression is a more quantitative way of testing this hypothesis.

Q1 How would you describe the distribution of *Trichilia* in the study area?

Q2 How would you describe the relationship between *Trichilia* and mineralizable N?

Q3 What are some other potential causes for *Trichilia* abundance patterns on the landscape in addition to nitrogen availability?

There is an important caveat to simply modeling species and environmental relationships without considering space. Species distributions can also be strongly influenced by spatial processes such as inter- or intraspecific competition from neighbors, dispersal, and foraging patterns of predators or herbivores. For example, windblown seeds, or even seeds such as those of *Trichilia* that are dispersed by large birds and mammals, tend to fall close to their parent. Relatively few seeds travel long distances (Muller-Landau et al. 2008) often creating clumpy patterns of species distributions. Environmental characteristics also often show similar spatial patterns (areas in close proximity share similar characteristics), making inferential tests based on these patterns difficult.

Variograms (explained more fully in Chapter 5) help illustrate this problem. In Figure 15.1, Panels C and D show how predictable *Trichilia* abundances and mineralizable nitrogen are at different lag distances. Recall that low variation (small values on the *Y*-axis) indicate that sample locations closer together tend to have similar values. Because this variation increases with distance between sample points, we deduce that sample locations that are closer together tend to have more similar values of both mineralizable nitrogen and *Trichilia* abundance. Since the pattern of *Trichilia* based on dispersal alone would also be clumped, we cannot be sure of the extent to which mineralizable nitrogen influences *Trichilia* distributions, independent of the effect of dispersal. In the case of *Trichilia*, or any single species, mixed models are well suited to account for spatial autocorrelation that is separate from the effect of abiotic variables (Zuur et al. 2009). However, when considering the responses of an entire community, a more common approach is to use multivariate partitioning methods, which we explore next.

Part 2. RDA for Modeling Abiotic Gradients

To illustrate how multivariate partitioning methods work, we will first model the response of a community to abiotic gradients only. When modeling the entire community, we cannot simply do statistical tests for each species because the large number of tests would inflate the Type 1 error. Testing a large number of environmental variables against each species would make the problem even worse. Therefore, an analytical method that models how the abundances of individual species change along the abiotic gradients and can test this for the entire community at once using a global statistic is extremely useful.

Ordination is a way of projecting relationships among multiple samples, predictor and response variables into mathematical space in more than one dimension, where the first axis represents the greatest amount of variability and each subsequent axis is uncorrelated with the previous ones. In a nutshell, ordination algorithms find a configuration of samples (or species) in mathematical space that best represents differences among them. Mathematically, there are a number of ways in which ordination algorithms can work (see Jongman et al. (1995) and Lepš and Šmilauer (2003) for details). We use two methods we use in this chapter: **principal component analysis (PCA) and principal coordinate analysis (PCoA)** which use an analytic solution to compute the mathematical distances. The resulting **site scores** represent positions of sites relative to each other in mathematical space and will be similar for samples with similar species abundances. Likewise, species commonly found together will have similar **species scores**.

Constrained ordination is simply ordination performed after the response variables (site or species scores) are obtained from a regression on the predictor variables. **Redundancy analysis (RDA)** is a constrained ordination technique that explicitly models linear relationships among multiple predictor variables and uses a randomization approach that statistically tests the strength of these relationships. In our case, we will create an RDA where species scores are predicted by regression on the environmental (or spatial) variables, and the new site scores are those generated by these predictions. The ordination results now correspond to the greatest variability in the dataset *that can be explained by the measured variables*.

EXERCISE 1: Data Preparation

The first step involves reading in the data and creating the two matrices that we need. The data provided (see R code available from the web site) have been organized for multivariate analyses:

- Rows represent sample sites, and columns hold information on species abundances or predictor variables.
- By looking at the subsets of the data matrices provided, you can get an idea of the data characteristics. For example, sample locations are spaced along a 20 m grid, with observations at *x* or *y* locations of 10, 30, 50, and so on.
- The 25 species included in the dataset are represented by six letter codes, with the full species names for these codes given at https://repository.si.edu/handle/10088/20925. For example, *Trichilia* is represented by the code TRI2TU.
- Finally, you can see that 13 soil variables were collected. Details on these soil variables are given at: http://biogeodb.stri.si.edu/bioinformatics/bci_soil_map/.

Since RDA is essentially an extension of linear regression, it is important to note that linear relationships with independent variables are tested; thus, variables may need to be transformed to achieve linearity. In addition, in order for the ordination portion of the RDA to work correctly, species data often need to be transformed to avoid creating a "horseshoe effect" (see Lepš and Šmilauer 2003 for details). For this chapter, we use a transformation called the Hellinger transformation that has been shown to work for a wide range of communities (Legendre and Gallagher 2001). Scientists who are interested in other transformations should consult Legendre and Legendre (1998).

Each environmental variable needs to be transformed if it is related to a nonlinear change in species scores. The choice of transformation is extremely important for inferring the relationship between species and the predictor variables (Gilbert and Bennett 2010), and requires considerable thought. Species or community responses to environmental variables are often nonlinear, so for the soils data we need to build a dataset that includes transformations for each variable that can model nonlinear response. From our initial analysis of the data, using visual inspection of trends, spline graphs, and polynomial regressions, we opted to use a third-order polynomial transformation of each variable. We chose polynomials in this case because they are relatively simple and flexible transformations and opted for the smallest polynomial that appeared to fit the data sufficiently well.

EXERCISE 2: Species-Environment RDA

Now that we have the two data matrices, we can perform the species-environment RDA. The significance of the relationship between the species matrix and the environmental matrix is determined with a Monte Carlo permutation procedure that tests whether the association between the matrices is stronger than expected by chance. This is done by comparing the test statistic from the true data to the test statistic that would be generated if the data were randomly assigned to sample plots. If the test statistic from the true data is higher than 95% of the random ones, the *P*-value is 0.05; if it's higher than 99% of the random ones, then the *P*-value is 0.01, etc.

Selection of significant variables in ordination has similar problems with model selection in regular multiple regression. Here, we use a forward selection method that attempts to control for model selection errors, using an adjusted R^2 that accounts for the number of variables used (Blanchet et al. 2008). The method makes sure the adjusted R^2 from all of the selected variables never exceeds the adjusted R^2 for the full model that includes all of the variables. In particular, we first do an RDA of the entire data soil matrix versus the species matrix to get the output in Table 15.1. The *p*-value on the right (0.005) is the value from the permutation.

Q4 What does significance in the full model mean? What would lack of significance mean?

Because this first test of the full model is significant, we go on to use a constrained forward selection, which is analogous to forward stepwise multiple regression.

Table 15.1 Result from RDA that includes all soil variables as predictors

Permutation test for rda under reduced model					
Model: rda(*X* = species.matrix, *Y* = soil.matrix)					
	Df	Var	*F*	N.Perm	Pr(>*F*)
Model	39	0.058567	6.5182	199	0.005**
Residual	585	0.134776			

Table 15.2 Result from RDA that includes only significant soil variables from forward selection

Model: rda(*X* = species.matrix, *Y* = soil.selected)					
	Df	Var	*F*	N.Perm	Pr(>*F*)
Model	32	0.05659	7.6554	199	0.005**
Residual	592	0.13675			

Q5 What single soil variable explained the most variation? (*HINT:* type "fwd.sel" to see the forward selection results).

It is important to note that the number of significant variables can change from one analysis to another because the RDA uses a randomization procedure; a variable that is marginally significant in one run (say, p value of 0.047–0.053) could change to insignificant in another run, or vice versa. This can be resolved by increasing the number of randomizations, but also reflects the problem of choosing an arbitrary significance level.

We can now do the analysis with the selected variables only, obtaining the output found in Table 15.2. We can see from this output that there is a significant influence of the selected environmental variables on community composition.

EXERCISE 3: Exploring the Results—Variation Explained

An important step in constrained ordination is to determine the amount of variation explained (R^2) in the community by the included variables. The simple R^2 is calculated by dividing the variation accounted for by the environmental variables (termed the "constrained inertia") by the total variation in the species dataset (the "total inertia"). Our uncorrected R^2 is 0.29, meaning that 29% of the variation in species distributions is explained by the selected soil variables. However, a correction is necessary to account of the number of independent variables tested, just as an adjusted R^2 is used in multiple regression. In the corrected model, the explained variation by environmental variables is 0.25 or 25%. We can also find the order of variable selection, from the first (the variable that on its own explains the most variation) to the last significant variable.

Table 15.3 The first 5 variables selected in the forward selection

	Variables	Order	R^2	R^2Cum	AdjR^2Cum	*F*	Pval
1	P1	25	0.084913	0.084913	0.083444	57.80954	0.001
2	Al1	1	0.030145	0.115058	0.112212	21.18773	0.001
3	K1	16	0.019986	0.135043	0.130865	14.34904	0.001
4	Cu1	10	0.013505	0.148548	0.143055	9.833824	0.001
5	N_min1	34	0.013006	0.161554	0.154782	9.601939	0.001

Using the R code in the Appendix, we'll look at the first five variables (Table 15.3). You'll notice in the table that the cumulative R^2 (that includes the variable and all previously selected ones) as well as adjusted cumulative R^2 are presented. Once the last variable is selected, the cumulative adjusted R^2 is equal to the adjusted R^2.

EXERCISE 4: Exploring the Results Graphically

We can also explore our results graphically by using an ordination plot (Figure 15.2), which plots species and site scores on the axes described above. In Figure 15.2, we graph only the five most important soil variables to simplify the presentation, but the R code in the online appendix can be used to plot all variables. The ordination plot includes sample sites, species, and independent variables on the first two ordination axes (Figure 15.2). There are different ways to scale these plots (see R help for CCA. plot and also Legendre and Legendre 1998). Here, we have used "species" scaling, which can be interpreted as follows. The angles between the arrows that point to each environmental variable represent the correlations between those variables. For example, potassium (K1) and aluminum (A1) are strongly negatively correlated, as they point in almost opposite directions. Similarly, if a line were drawn from the plot center to each species, the angles between each pair of lines would represent the correlation between species abundances. For example, abundances of the species labeled DRYPST and HIRTTR are also strongly correlated and are higher at high potassium and low aluminum levels. The length of the arrows along the ordination axes indicate the strength of their relationship to each axis, with phosphorus (P1) being more closely related to the first axis than mineralizable N (N_min1), for example.

Q6 Examine the distribution of *Trichilia* (TRI2TU) along the mineralizable nitrogen (N_min1) gradient in Figure 15.1. Where are TRI2TU and N_min1 in the ordination figure? What does the location of the arrow head and the species tell you?

Q7 Using the same approach as in the question above, name a species that should occur at higher abundances when there are large amounts of copper (Cu1) present. Use R to graph the abundance of this species relative to copper to see if your prediction is correct.

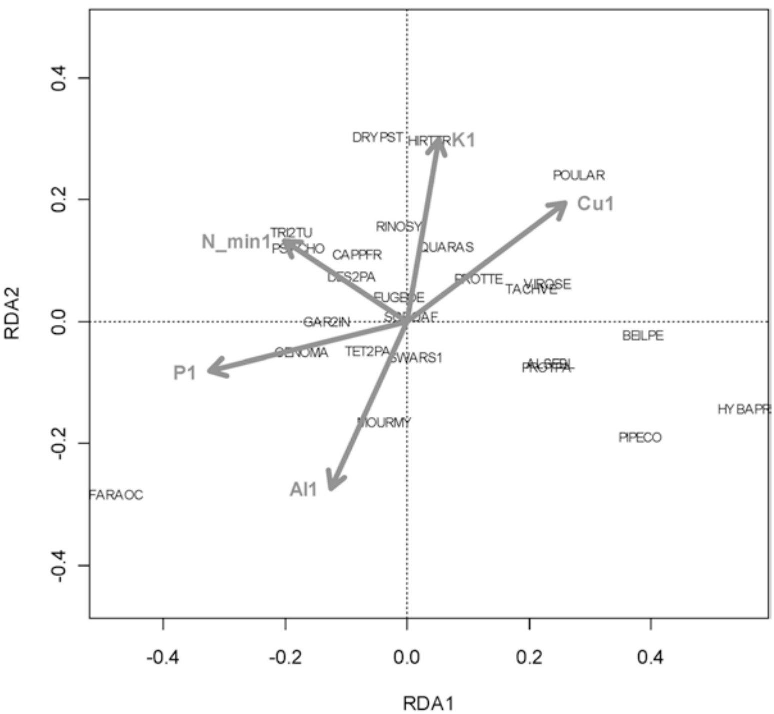

Figure 15.2 Biplot showing species (six letter codes) and soil resources (vectors with letters at end)

Q8 Species can also be negatively correlated to specific resources. Find the species VIROSE. How do you think this species' abundance will change with soil phosphorus (P1)? Use R to graph the abundance of this species relative to copper.

Part 3. Accounting for Space in Community Analyses

Quantifying spatial changes in distributions of species and communities, and then attributing these changes to specific causes, can be quite difficult. As mentioned previously, spatially correlated environmental factors are not the only possible causes of the spatial distributions of organisms. Biotic factors such as dispersal, inter- and intraspecific competition, predation, and disease can also create spatial patterns which may be distinct. For example, intraspecific competition promotes negative spatial autocorrelation among species and communities while dispersal limitation promotes positive spatial autocorrelation. However, some patterns may not be distinct; for example, disease may promote negative spatial autocorrelation (similar to intraspecific competition) since close proximity of individuals may increase risk of infection. The refinement of techniques to attribute spatial change to specific causes is an active area of ecological research (Diniz-Filho et al. 2003; Gilbert and Bennett 2010).

For the multivariate statistics that we are presenting, the common method for modeling spatial patterns is to use regression-based approaches to model the abundances of species or changes in community composition as a function of location. While the many methods used for multivariate spatial analysis are well beyond the scope of this chapter, here we present two commonly used methods: **trend-surface analysis** and **Moran's eigenvector mapping (MEM)**. Both approaches are used within the RDA framework and can help us ask what type of spatial patterns might result from a given process and then test which spatial variables can model these patterns. We outline each approach below, and explain the reasoning behind using one or the other in different situations.

EXERCISE 5: Polynomial Trend-Surface Analysis

Using polynomials to model nonlinear spatial patterns in communities is very similar to using them to model nonlinear changes due to environmental factors. Instead of polynomials of environmental variables, we generate polynomials of the centered *X* and *Y* coordinates of our spatial data. In particular, if all plots have an *x* and *y* location (in meters or UTMs), the trend-surface is a polynomial function of those *x* and *y* locations so that predictor variables are: X, Y, X^2, Y^2, XY, X^2Y, XY^2, X^3, Y^3. We can then use these variables in RDA, in a similar forward-selection process to that used for the environmental polynomials.

Q9 Use the code provided in the Appendix to conduct a forward selection of the spatial polynomials. What variables were selected in the forward-selection process? How does the number of spatial variables compare with the number of environmental variables that were selected in the last section?

Q10 How much variation was explained by the spatial variables in the analyses above? How does that compare with the variation explained by the initial analyses of environmental variables?

Q11 Do you think it is valid to add the variation explained by environmental and spatial variables, as analyzed so far, to get the full variation explained by both? Why or why not? (*HINT:* reexamine Figure 15.1).

EXERCISE 6: Moran's Eigenvector Maps

The second approach that we demonstrate, called **Moran's Eigenvector Maps (MEM)**, is an adaptation of an earlier method, called principal coordinates of neighbor matrices (PCNM) that is still frequently used. The simplest explanation of the MEM approach is that it uses a series of waves to model a spatial pattern to fit complex spatial relationships among samples. This approach is similar to spectral analysis, with a more detailed explanation found in Dray et al. (2006) and Peres-Neto and Legendre (2010). We will use a step-by-step process that generates MEM. This process can take a long time because some of the randomizations are computer intensive. If time is short, your instructor may choose to skip steps 1–5 (see "Shortcut" in appendix R code), or assign them as homework.

The MEM technique has several components:

Step 1: Create a distance matrix. This matrix should represent the spatial distances among all sample sites.

Step 2: Simplify the distance matrix. Ensure that all distances that are greater than a critical value are all assigned the same (large) value.

Step 3: Conduct a principal coordinates analysis (PCOORD) on the simplified distance matrix. PCOORD is another ordination technique that represents the distances among samples in different dimensions. These dimensions are the same as the axes of an ordination plot, as explained above, and the axes are called **eigenvectors**. The technique can be used to represent spatial relationships among samples that are not readily apparent to us, for example, wave patterns of repeating spatial clusters, and other complex, nonlinear relationships.

Step 4: Test the eigenvectors for significant spatial autocorrelation. Some of the eigenvectors are not spatially autocorrelated thus including them is equivalent to including useless predictor variables in a regression. For the purposes of this chapter, we consider only significant positive autocorrelation. We retain only the axes that indicate strong positive spatial structure by only considering those values of Moran's I that are positive and statistically significant. (*NOTE:* Depending on the size of the dataset and available computing power, the randomizations may require a lot of time (up to an hour for this particular dataset).

Q12 The "dim" function gives the dimensions of the dataset (i.e., the number of rows and columns). How many eigenvectors show positive and significant autocorrelation? How does this compare to the number of predictors in the trend-surface analysis and in the environmental analysis?

Step 5: Remove linear trends. The vectors created using MEM are good at detecting nonlinear patterns at a finer scale than the trend-surface analysis. However, they are not good at detecting linear trends, which can cause analysis problems. We therefore do a separate analysis of the linear spatial trends over the sample area and include this in the total spatial signal. In our case, we have already included linear x and y trends in the trend-surface analysis, and we know that they are significant. So, we first remove the effects of these predictors on species distributions, and then test the significance of the MEM axes on the residuals. We use the usual forward selection procedure, plus an additional selection criterion that is designed to help with over-fitting problems that have been observed with eigenvectors like MEMs (see Gilbert and Bennett 2010).

Table 15.4 Results from RDA that include only significant MEMs

Permutation test for rda under reduced model					
Model: rda(*X*=species.matrix, *Y*=MEM.selected)					
	Df	Var	*F*	N.Perm	Pr(>*F*)
Model	61	0.084236	7.1256	199	0.005**
Residual	563	0.109107			

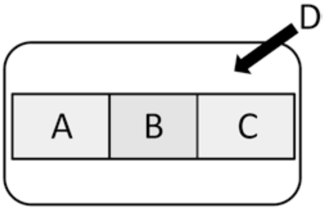

Figure 15.3 Representation of variation explained in variation partitioning. The entire area (*A*+*B*+*C*+*D*) represents the total variation in the dataset, and the *grey rectangles* represent variation explained by each component. Segment *A* represents the component of variation explained solely by environmental variables, while segment *C* represents the component solely explained by spatial variables, and segment *B* represents variation jointly explained by spatial and environmental variables. *D* represents unexplained variation

Step 6: Perform RDA with Species and MEM. We now do the analysis on the selected variables and determine the variation explained, as with the previous examples (Table 15.4).

Q13 How many spatial variables were in this analysis? What is the corrected R^2 for the spatial signal using MEM? Is this very different from the uncorrected R^2? Why is that? How does this compare with the corrected R^2 of the spatial polynomials?

Part 4. Variation Partitioning

As we have noted, the spatial determinants of species composition may be environmentally correlated, which can lead to problems if we draw conclusions from environmental analyses without considering spatial signals, or conversely, from spatial analyses without considering environmental factors. We therefore need a technique that will allow us to divide the signal of community change into separate environmental and spatial components, as well as the component that is shared between them. **Variation partitioning** (Borcard et al. 1992; Figure 15.3) allows different independent components of variation to be allocated. The technique works in steps using simple algebra and requires results from three separate constrained ordinations. Because you will need information from prior Exercise 6, we continue numbering our steps from the previous exercise.

Step 7: Record the Variation Explained by the RDA with Environmental Variables. The corrected R^2 result from Exercise 3 (RDA using forward-selected environmental variables) gives this portion of the variation explained. In Figure 15.3, the variation explained from this ordination is represented by $A+B$.

Step 8: Record the Variation Explained by the RDA with Spatial Variables. The corrected R^2 result from either the trend-surface analysis (Exercise 5) OR the RDA with MEM (Exercise 6, Step 6) gives this portion of variation explained. In Figure 15.3, the variation explained from this ordination is represented by $B+C$.

Step 9: Record the Variation Explained by an RDA with all Significant Spatial and Environmental Variables. This involves creating a dataset that includes the selected environmental variables from Exercise 2 and spatial variables from Exercise 5 or Exercise 6, Step 6. In Figure 15.3, the variation explained form this ordination is represented by $A+B+C$.

EXERCISE 7: Partition the Variation Explained

Variation explained solely by environmental variables is represented by component A; variation explained solely by spatial variables by component C; and shared variation by component B. All of these can be derived using algebra. Depending on the research question, ecologists may be interested in any combination of the components above. The shared spatial, shared environmental, and total explained variation are easily attained using the ordinations described above.

In order to get component A (the independent environmental signal), take the variation explained by all selected variables ($A+B+C$), and subtract the variation

Table 15.5 Results from partitioning spatial (trend-surface) and soil variables

Explanatory tables:						
$X1$: soil.selected						
$X2$: t.s.selected						
No. of explanatory tables: 2						
Total variation (SS): 120.65						
Variance: 0.19334						
No. of observations: 625						
Partition table:						
			Df	R^2	AdjR^2	Testable
$[a+b]$	=	$X1$	32	0.29269	0.25446	TRUE
$[b+c]$	=	$X2$	9	0.20996	0.1984	TRUE
$[a+b+c]$	=	$X1+X2$	41	0.34799	0.30214	TRUE
Individual Fractions						
$[a]$	=	$X1\|X2$	32		0.10374	TRUE
$[b]$			0		0.15072	FALSE
$[c]$	=	$X2\|X1$	9		0.04768	TRUE
$[d]$	=	Residuals			0.69786	FALSE

explained by the spatial and shared component: $(A+B+C)-(B+C)=A$. This is accomplished by subtracting the variation explained in Step 9 from that in Step 8. Similarly, to get component C (the independent, or "pure" spatial signal), take the variation explained by all selected variables ($A+B+C$, Step 9), and subtract the variation explained by the environmental and shared component ($A+B$, Step 7).

The partitioning of these components can also be calculated directly with a specialized function in R. For example, Table 15.5 gives the results from partitioning the environmental and trend-surface components. In this output, the fraction "[$a+b$]" or variable "$X1$", refers to the environmental plus shared variation explained ($A+B$ in the figure above). Likewise, "[$b+c$]" or variable "$X2$" refers to the spatial plus shared variation ($B+C$ in the figure above) while individual fractions refer to the "pure" environmental or spatial variation explained ("[a]" and "[c]", respectively), and the shared variation "[b]". Component "[d]", the residuals, refers to the variation that is NOT explained by the ordination.

Q14 What is the total (adjusted) variation explained by the ordination? What signal appears greater, space, or environment? How does the variation explained by either fraction compare to the shared variation, and what does this mean, in terms of interpretation of results?

Q15 Partition the variation explained by the MEM spatial predictors and the soil variables. How do these results differ from the trend-surface results above? Why do you think these techniques give different results?

The differences in signals using the trend-surface and MEM techniques illustrate an important aspect of spatial analysis: the method of representing the spatial configuration of samples or communities on a landscape can have profound influences on the results. Recall that the polynomial trend-surface model the spatial signal as curves that can be drawn on a map while the MEMs model space as a series of waves with different periodicities. The latter technique is much more flexible in terms of what is considered "spatial," making it difficult to attribute specific biological causes to the signal. Given these differences, it is important to keep in mind one's original research question when choosing a sampling configuration and an analytical technique. Similarly, the sampling design can also influence the outcome of these analyses: sample plots that are surveyed with spatial lags of kilometers are unlikely to show the same patterns as contiguous plots (see Fortin and Dale 2005 for details on this issue). Below, we clarify how these results can be interpreted and incorporated into landscape level models.

EXERCISE 8: Testing the Independent Components

Although we have already tested the significance of shared components ($A+B$, for example), we have not yet tested the independent components. The significance of the independent environment component (A) can only be tested by including the spatial component ($B+C$) as a covariate. Similarly, the independent spatial component (C) is tested by including the environmental component ($A+B$) as a covariate.

Table 15.6 Exploring the spatial structure of important soil explanatory variables

Variable	Component $A+B$ *	Component A	Component B	Percent spatial ($100*B/(A+B)$)
P1	0.085	0.014	0.071	83.079
Al1	0.043	0.006	0.037	85.586
K1	0.033	0.008	0.026	76.908
Cu1	0.067	0.010	0.056	84.753
N_min1	0.042	0.005	0.037	88.407

These numbers differ from Table 15.3 because the variables are tested individually, without considering the correlations with other variables

Using the R code in the Appendix, we can see from testing both of these fractions that when using the spatial trend-surface predictors, each component is significant.

EXERCISE 9: Interpreting and Applying the Partitioning Results

Many studies in spatial ecology have stopped after step 10 and reported the variation partitioning results as evidence for environmental or biotic processes. This approach must be taken with caution. It has been shown that partitioning results can give a good indication of the processes at work; however, they are sensitive to sampling design, unmeasured (especially spatially structured) variables, the spatial model used (MEM or trend-surface), and whether correct transformations of environmental variables were used (Legendre and Legendre 1998; Gilbert and Bennett 2010).

Despite these reservations about strictly interpreting partitioning results, the separate components do contain information that is extremely useful to landscape ecologists. In particular, the component B (Figure 15.3, Table 15.5) represents the spatially structured environment that predicts species distributions. Although component B cannot be tested statistically, it can be compared with other components. A large amount of explained variation in this component indicates that some portion of species' spatial distributions are explained by spatially structured environmental variables. One method of better understanding this component is to examine how much of a variable's predictive power is spatially structured.

We do this for the five most significant variables (Table 15.6). We can see that for the tropical forest studied, the most important predictors are strongly spatially structured (Table 15.6).

The spatial structure of these variables, determined using the methods presented earlier and elsewhere in this book, can be employed in landscape models that explicitly consider how spatial processes may work in conjunction with responses to abiotic variables. The important benefit of the partitioning approach is that it has allowed us to identify the abiotic variables that are significant predictors of species distributions, and also provide an indication of how well spatially explicit landscape models could capture the effects of these predictors.

CONCLUSIONS

When contrasting spatial and environmental signals, it is important to realize that the two signals are almost always somehow intertwined. Figure 15.1 illustrates such a relationship in the BCI dataset. In fact, nearly all environmental variables have a spatial component: hard environmental boundaries such as sheer cliffs are not nearly as common as more gradual ecotones. The analyses we presented are useful tools for understanding the spatial structuring of communities due to both biotic and abiotic influences. If appropriate variables and sampling techniques are employed, these analyses can be used to gain an understanding of how abiotic and biotic influences act independently, and whether one influence tends to overshadow another. Partitioning results also offer a unique opportunity for landscape ecologists to identify variables that both structure communities and that are themselves spatially structured. Incorporating these results into landscape models can allow for quantitative estimates about the relevance of landscape models for predicting species distributions.

SYNTHESIS

Q16 Using the BCI dataset, design a research question that uses both spatial and environmental components. What variables and analyses would you test with this question? What are some potential unknowns and limitations in your analyses?

Q17 Consider a new and different dataset (either a potential hypothetical dataset or data from your own research) and answer Q16. Be sure to explain the type of data you might examine.

REFERENCES AND RECOMMENDED READINGS[1]

Blanchet FG, Legendre P, Borcard D (2008) Forward selection of explanatory variables. Ecology 89:2623–2632

Bolker BM (2004) Continuous-space models for population dynamics. In: Hanski I, Gaggioti OE (eds) *Ecology, genetics, and evolution of metapopulations*. Elsevier Science, San Diego, pp 45–69, Chapter 3

*Borcard D, Legendre P, Drapeau P (1992) Partialling out the spatial component of ecological variation. Ecology 73:1045–1055. *A classic paper outlining some of the approaches covered in this lab in more detail.*

*Borcard D, Gillet F, Legendre P (2011) *Numerical ecology with R*. Springer, New York. *A useful reference for a variety of analyses in R.*

[1] *NOTE:* An asterisk preceding the entry indicates that it is a suggested reading.

Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol Model 196:483–493

Fortin M-J, Dale MRT (2005) *Spatial analysis: a guide for ecologists*. Cambridge University Press, Cambridge, UK

Gilbert B, Bennett JR (2010) Partitioning variation in ecological communities: do the numbers add up? J Appl Ecol 47:1071–1082

Hanski I (1991) Metapopulation dynamics: brief history and conceptual domain. Biol J Linn Soc 42:3–16

Jongman RHG, Ter Braak CJF, Van Tongeren OFR (1995) *Data analysis in community and landscape ecology*. Cambridge University Press, Cambridge, UK

Legendre P, Gallagher ED (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

*Lepš J, Šmilauer P (2003) *Multivariate Analysis of Ecological Data using CANOCO*. Cambridge University Press, Cambridge. *Chapters 1, 3 and 5 provide an excellent, readable introduction into basic ordination techniques.*

Levins R (1970) Extinction. In: Gesternhaber M (ed) *Some mathematical problems in biology*. American Mathematical Society, Providence, Rhode Island, pp 77–107

Macarthur RH, Wilson EO (1967) *The theory of island biogeography*. Princeton University Press, Princeton, NJ

Muller-Landau HC, Wright J, Calderon O, Condit RS, Hubbell SP (2008) Interspecific variation in primary seed dispersal in a tropical forest. J Ecol 96:653–667

Peres-Neto PR, Legendre P (2010) Estimating and controlling for spatial structure in the study of ecological communities. Glob Ecol Biogeogr 19:174–184

Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM (2009) *Mixed effects models and extensions in ecology with R*. Springer, New York