

# Chapter 11

## Spatially Structured Communities



### 11.1 Introduction

Biodiversity is the variety of life. It is fundamental to all aspects of ecology and conservation biology. Biodiversity can be measured at different levels of organization and at different scales (Noss 1990; Magurran 2003). For instance, the number species in a local area, or species richness, is commonly measured in field investigations (Myers et al. 2000; Gotelli and Colwell 2001), whereas genetic diversity is a fundamental concept in evolutionary biology (Ellstrand and Elam 1993; Keller and Waller 2002). Over the last two decades, there has also been a great deal of interest in functional and phylogenetic diversity (Petchey and Gaston 2002; Cadotte et al. 2009; Cavender-Bares et al. 2009; Devictor et al. 2010; Mouchet et al. 2010).

To understand patterns of biodiversity over space and time and to effectively implement biodiversity conservation, concepts from community ecology are essential. Community ecology focuses on the importance of interactions among species and how communities assemble to drive variation in biodiversity (Mittelbach 2012). The importance of space in the outcomes of species interactions and community assembly has long been emphasized (Huffaker 1958; Diamond 1975). As a consequence, modern theory and concepts for community ecology emphasize the role of space (Vellend 2010; Leibold and Chase 2017), and many problems in conservation emphasize spatial issues for protecting or maintaining biodiversity (Moilanen et al. 2009; Rands et al. 2010).

Here, we provide an overview of how space influences biological communities, why space is important for biodiversity conservation, and we illustrate some common approaches for modeling communities over space and time. Space is generally important for understanding biodiversity for at least three reasons. First, some measures of diversity, such as beta diversity, are inherently spatial, focusing on the change in diversity across spatial and/or environmental gradients (Soininen et al. 2007; Anderson et al. 2011). Second, space can provide a mechanism for biodiversity patterns by altering community assembly and disassembly processes (Leibold et al. 2004). Third, incorporating spatial

issues into conservation strategies aimed at promoting biodiversity can provide new insight and can help solve some problems (Karp et al. 2012). We illustrate these issues through the use of spatial modeling of biological communities.

## 11.2 Key Concepts and Approaches

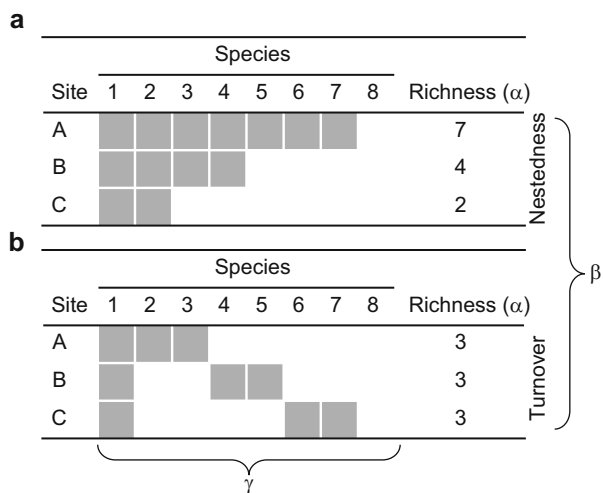
### 11.2.1 Spatial Community Concepts

Spatially structured communities can be described in several ways. For instance, diversity can be measured based on different types of variation (e.g., species-level, genetic) and it can be measured at different scales. In addition, there have been a wide range of concepts and theoretical developments on spatially structured communities. We first provide some terms and definitions regarding the ways in which diversity and communities are quantified across space. We then briefly provide an overview of some key spatial ecology concepts for communities, starting with early work on species–area relationships and moving to more contemporary concepts regarding metacommunities and hierarchies in community assembly.

#### 11.2.1.1 A Diversity of Diversities

Species diversity is a major component of biodiversity. Species diversity is often partitioned into three types: alpha diversity, beta diversity, and gamma diversity (Fig. 11.1 and Table 11.1). *Alpha diversity* is the number of species residing at a locality and is often referred to as species richness. *Beta diversity* has been contextualized in a variety of ways, but it generally focuses on the turnover or change in

**Fig. 11.1** A diversity of diversities illustrated through species-site matrices for two scenarios. In both, eight species occur in the species pool ( $\gamma = 8$ ). In (a), species richness ( $\alpha$  diversity) across sites varies, with beta diversity showing nestedness. In (b), species richness is constant across sites, but spatial turnover occurs. Adapted from Baselga (2010)



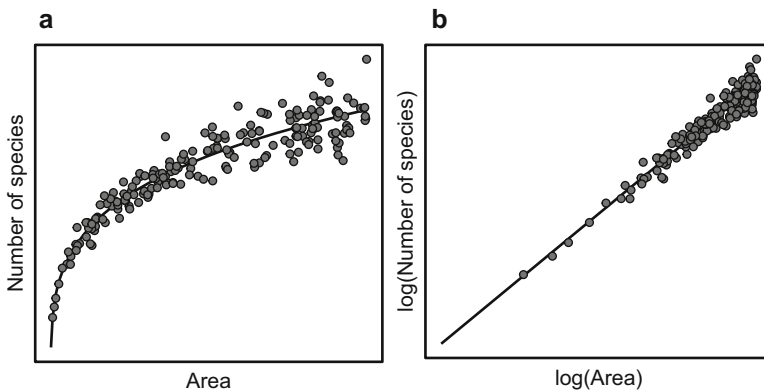
**Table 11.1** Terms and concepts frequently considered in spatial community ecology

| Term                              | Definition  |
|-----------------------------------|---|
| Alpha diversity                   | The number of species residing at a locality.   |
| Beta diversity                    | The change in species across environmental, spatial, or temporal gradients.   |
| Competition-colonization tradeoff | When species that are good colonizers are poor competitors, and species that are good competitors are poor colonizers. This tradeoff can facilitate coexistence across landscapes.  |
| Community assembly rules          | Rules that make predictions for what species will occur in a location, given the species pool.  |
| Gamma diversity                   | The total number of species in a region.  |
| Habitat diversity hypothesis      | One proposed mechanism for the species–area relationship. This hypothesis predicts that variation in the environment (habitat, resources, etc.) increases with area, leading to potentially more species.   |
| Limiting similarity               | A category of community assembly rules whereby a community is assembled via interspecific competition, such that species that coexist have low similarity in key traits of relevance to local resources.  |
| Mass effect                       | When community structure is driven by species source–sink dynamics.   |
| Metacommunity                     | A set of local communities linked by dispersal.   |
| Ordination                        | Analyses that aim to summarize the variation of multivariate data into reduced dimensions or space by arranging species and/or locations along gradients. Typically based on eigen-analysis.  |
| Nestedness                        | The change in species between two (or more) locations that arises from the loss of certain species.   |
| Neutral community models          | Models that assume community assembly is driven entirely by stochastic forces.  |
| Passive sampling                  | A mechanism for species–area relationships that is based solely on sampling effort. As sample effort increases in larger patches or islands, the number of species detected increases.  |
| Patch dynamics                    | When community assembly is viewed through the lens of variation in colonization–extinction dynamics, wherein better dispersers are frequently assumed to be poorer competitors.   |
| Rank-abundance curve              | Relative abundance of species plotted as a function of their rank abundance (most abundant species = rank 1). Used to visualize skewness in community composition.  |
| Species–area relationship         | The pattern whereby the number of species increases with habitat area.  |
| Species pool                      | The total number of potential species in a region that are capable of colonizing localities.  |
| Species sorting                   | When community structure over space is driven by individual species responses to environmental heterogeneity, such that certain local conditions may favor certain species and not others. In this concept, dispersal is not assumed to be a limiting factor.     |
| Target effect                     | A mechanism for species–area relationships, whereby immigration is greater to larger habitats simply because larger areas are bigger targets for dispersers. This relationship typically scales with the linear diameter of habitats rather than the area per se. |
| Turnover                          | The change in species between two (or more) locations based on species replacement.   |

species across environmental, spatial, or temporal gradients (Anderson et al. 2011). Beta diversity is sometimes split into its nestedness and turnover components (Baselga 2010). *Nestedness* between two locations refers to the change in species based on the loss of species, where one location may be “nested” within another; that is, it is a nested subset of the location that contains more species (Wright and Reeves 1992). The idea of nestedness in biological communities has received a great deal of interest over the years (Wright et al. 1998; Mac Nally and Lake 1999; Kerr et al. 2000; Fernandez-Juricic 2002; Driscoll 2008), in part because it was a fundamental issue involved in the SLOSS (single-large versus several small) debate (Wright and Reeves 1992). *Turnover*, on the other hand, refers to a change in species via species replacement (not loss) (Williams 1996). *Gamma diversity* typically refers to the *species pool* in the region, or the species that are potentially available for colonizing local sites or communities (Karger et al. 2016). Understanding the interplay and dependence of each of these components of species diversity is of long-standing interest to community ecologists (Ricklefs 1987; Partel et al. 1996; Caley and Schluter 1997; McPeck and Brown 2000; Koleff and Gaston 2002; Podani and Schmera 2011; Lessard et al. 2012; Fukami 2015).

### 11.2.1.2 Species–Area Relationships

One of the few laws in ecology is the *species–area relationship* (Lawton 1999): the number of species increases with area (island area, patch area, etc.) (Fig. 11.2). This relationship has been documented throughout the world. There is a long history of exploring why this relationship exists and using this relationship to forecast changes in species diversity with ongoing environmental change (Gonzalez 2000; Seabloom et al. 2002; Thomas et al. 2004; Dobson et al. 2006).



**Fig. 11.2** The species–area relationship, shown on the (a) raw (original) scale, and (b) a log–log scale ( $\log_{10}$  scale)

Arrhenius (1921) was one of the first scientists to formally quantify the species–area relationship, describing the relationship as a power function. Preston (1962) further developed this idea. He defined this relationship as:

$$S = cA^z, \quad (11.1)$$

where  $S$  is the number of species,  $A$  is area, and  $c$  and  $z$  are constants that describe the shape of the relationship of species with area. This relationship describes a pattern where the number of species quickly increases with area and then the rate of change slows (a power function relationship). It can be linearized when transformed to a log–log (base 10) scale as:

$$\log(S) = \log(c) + z\log(A). \quad (11.2)$$

There has been interest in understanding variation in  $z$ , because it describes the magnitude of the species–area relationship. Often  $z$  values tend to range from 0.10 to 0.25 (Drakare et al. 2006). We note that in practice, there are actually several types of species–area relationships that have been documented, where different types of sampling designs and functional forms (e.g., power, logistic) have been used to interpret species–area relationships (Scheiner 2003).

Given the ubiquity of the species–area relationship, the immediate question that arises is why this relationship occurs. Understanding why the relationship occurs is essential for understanding the importance of this pattern. Several hypotheses have been put forward to explain SAR; here we focus on a few common ones. First, the *habitat diversity* hypothesis suggests that as area increases, habitat diversity increases, such that the increase in number of species simply reflects an increase in the diversity of habitat or resources. Second, the *target effect* hypothesis posits that larger areas are more likely to be colonized, even simply by chance (or passive diffusion) due to an increasing circumference of the area (Bowman et al. 2002). Third, the *passive sampling* hypothesis states that SAR relationships are simply a reflection of greater sampling effort as area increases (Coleman et al. 1982), such that 10, 10 ha sites would yield the same number of species as 1, 100 ha site. Consequently, this hypothesis implies that there is nothing special about habitat area per se, and that the number of species per unit area sampled will not increase with increasing patch or island area. This hypothesis has similar rationale to the habitat amount hypothesis (Fahrig 2013). Finally, a great deal of interest and effort has focused on the hypothesis that SAR relationships arise from a balance of immigration and extinction effects that may change as a function of area. This hypothesis underlies the Equilibrium Theory of Island Biogeography (MacArthur and Wilson 1967).

### 11.2.1.3 Equilibrium Theory of Island Biogeography

Arguably, the most important conceptual development to our understanding of communities across space was the development of the Equilibrium Theory of Island

Biogeography (ETIB). MacArthur and Wilson (1967, 1963) developed this theory in detail to understand and predict the number of species residing on an island and the turnover rate of species on islands. The underlying premise of this development is that the number of species in an area is a balance between recurrent immigration of new species and recurrent extinction of species in a local area. When immigration and extinction are balanced, the number of species and the rate of species turnover is at equilibrium. This model is a neutral model (Caswell 1976), in the sense that species identity does not inform the model and expectations in the model are driven entirely by stochastic forces.

Immigration rates of new species per unit time are assumed to decline as the number of species increases on the island, eventually reaching 0 when the number of species on the island is equal to  $P$ , the species pool, or the mainland source pool of species. Simply put, as fewer species are able to immigrate from the mainland, immigration rates should decline. Often, this immigration curve is drawn as being non-linear (~exponential decline), to reflect the idea that some species might be better at dispersing than others, where good dispersers will immigrate rapidly, while poor dispersers will be slow to immigrate. Extinction rates are assumed to increase as the number of species on the island increases, where the abscissa is zero (no extinctions can occur when no species inhabit the island). Extinction rates are assumed to occur stochastically, such that extinction rates increase with the number of species simply because there is a greater number of potential species to go extinct. Again, this relationship is also frequently drawn as a non-linear (exponential) relationship, where interspecific competition may increase the extinction rate as the number of species increases (a linear rate would assume that all species behave independently of each other).

This theory received the most attention when MacArthur and Wilson invoked island area and isolation from mainland as critical factors that may alter immigration and extinction rates. MacArthur and Wilson (1967) assumed that as island area increased, the extinction rate should decline relative to smaller islands. The rationale for this assumption is that larger islands will harbor larger populations, where population size is proportional to island size (note: density is assumed to be constant, or in some cases decline; MacArthur 1972), such that demographic stochasticity may play a smaller role in extinction risk of individual species. This component of ETIB leads to specific predictions regarding species–area relationships—that species number increases with area and that turnover rates decline. MacArthur and Wilson also assumed that island isolation would alter immigration rates, where increasing isolation should reduce immigration. Note that since their seminal work, area effects have also been considered to alter immigration rates, where target effects occur (larger islands lead to greater immigration rates) (Lomolino 1990), and isolation has also been considered to alter extinction rates, where less isolated islands are expected to have lower extinction rates via rescue effects (immigration rates preventing extinction) (Brown and Kodric-Brown 1977). While much of the focus of this work has been on area and isolation, MacArthur and Wilson (1967) also developed a variety of related issues, such as the role of corridors, stepping stones, and island aggregation on expected numbers of species.

This theory has been instrumental in community ecology and conservation (Whitcomb et al. 1976; Whittaker et al. 2005). Nonetheless, it is now known that this framework does not capture many of the pressing issues influencing biodiversity across space and over time, such as edge and matrix effects, landscape complementarity, species interactions, and situations where no “mainlands” occur (Haila 2002; Laurance 2008). Importantly, this theory does not predict the distribution of individual species nor species identity (and related traits) in the community. Since this seminal work, several extensions have been made to accommodate some of these issues (Holt 1992; Cook et al. 2002; Gravel et al. 2011). One major advancement has been the development of metacommunity theory (Holyoak et al. 2005).

#### 11.2.1.4 Metacommunities

The metacommunity concept extends ideas from metapopulation ecology (Chap. 10) and community ecology to explicitly understand variation in communities across space (Wilson 1992; Leibold et al. 2004; Holyoak et al. 2005; Leibold and Chase 2017). This concept aims to unite several processes that have been hypothesized to be critical to community structure over space (Vellend 2010). At its core, a *metacommunity* consists of local communities (i.e., communities residing at a particular locality, such as a patch) that are linked spatially through dispersal. Leibold et al. (2004) identified four paradigms that have been applied to understanding metacommunities: the patch-dynamics paradigm, the species-sorting paradigm, the mass effects paradigm, and the neutral paradigm.

The *patch-dynamics* paradigm is a direct extension of two-species metapopulation models to  $N$  species. This paradigm emphasizes that species diversity may be limited by species interactions (e.g., competition) and dispersal. The focus is on colonization-extinction dynamics of  $N$  species, where it is often assumed that patches are similar and each patch is capable of containing populations of each species. In this paradigm, competition-colonization tradeoffs among species are often assumed (Levins and Culver 1971). In the *competition-colonization tradeoff*, it is assumed that poor dispersers are dominant competitors, while good dispersers tend to be poor competitors, which has been observed when contrasting some annual and perennial plants. This tradeoff provides a stabilizing mechanism for species coexistence across landscapes or regions. Tilman et al. (1994) popularized this general framework when modeling communities under scenarios of habitat destruction (see also Neuhauser 1998). This paradigm generally emphasizes that community structure is limited by variation in dispersal limitations.

In contrast to the patch-dynamics paradigm, the *species-sorting* paradigm emphasizes that environmental gradients drive variation in species diversity, while dispersal is less of a limiting force but rather dispersal allows species to track variation in resource gradients across landscapes. It assumes that diversity is driven by spatial niche separation above and beyond spatial dynamics arising from variation in dispersal and colonization (Holyoak et al. 2005).

The *mass-effects* paradigm is largely an extension of source-sink population dynamics (Chap. 10) to community assembly. In this paradigm, variation in immigration and emigration rates across landscapes and their impact on local population dynamics are emphasized. Variation in immigration and emigration rates can generate rescue effects (Brown and Kodric-Brown 1977) and can thereby offset competitive exclusion. In this paradigm, the role of dispersal is emphasized in being a key factor driving variation in local densities and it is assumed that patches vary in their suitability, leading to variation in immigration/emigration rates.

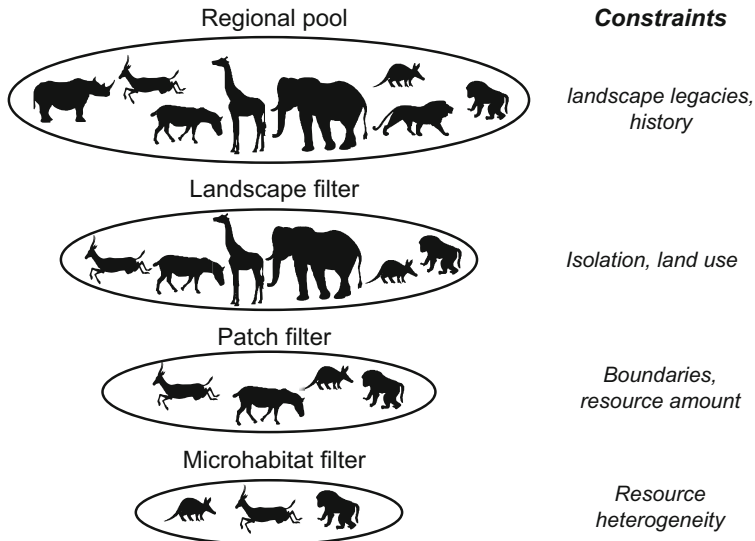
Finally, the *neutral* paradigm assumes that all species are similar in the competitive abilities, dispersal abilities, and fitness. This paradigm assumes that stochastic processes of species loss and gain drive variation in diversity. One of the first popular neutral models for species diversity was the Equilibrium Theory of Island Biogeography (MacArthur and Wilson 1967). The neutral paradigm has been emphasized by Hubbell and colleagues (Hubbell 2001) to explain community structure. Thus, dispersal and spatial dynamics are highly relevant to the neutral paradigm, although these dynamics are assumed to be driven by stochastic forces (Economio and Keitt 2008; Lowe and McPeck 2014; Guichard 2017).

#### 11.2.1.5 Hierarchies from Regional Pools to Local Assemblages

There has been a great deal of interest in scaling from regional species pools to local assemblages by using community assembly rules to interpret how species may coexist. *Community assembly rules* are rules that make predictions for what species will occur in a location, given the regional species pool (Keddy 1992). Diamond (1975) was the first to consider the problem of assembly rules by considering how species traits (e.g., body size) could explain species composition of birds on islands. Predictions for assembly have also been made based on *limiting similarity* of key species' traits and for the role of environmental filtering in wetlands (Van der Valk 1981). *Environmental filtering* occurs when local environmental (abiotic) conditions "filter out" species from the regional pool (i.e., the environment selects against certain species), such that some species do not occur at certain localities due to the poor environmental conditions for that species (Cadotte and Tucker 2017). At its core, the assumption for environmental filtering is that species absence is not driven by biotic interactions (Kraft et al. 2015).

Poff (1997) took the general concept of environmental filtering and applied it in a hierarchical, landscape context (Fig. 11.3). In this framework, environmental filters operate at different scales, placing constraints on local communities. Different spatial constraints (e.g., spatial isolation, resource heterogeneity) operate at different spatial scales, and species traits (e.g., dispersal mode, foraging breadth) will lead to selective filtering of certain species based on these traits. It is often envisioned that environmental filters operate at relatively broad scales, while biotic interactions govern constraints at local scales (akin to ideas in species distribution modeling; see Chap. 7). This idea comes out of applications of hierarchy theory in landscape





**Fig. 11.3** Space-time hierarchies of environmental filtering. Shown are hierarchical filters and some examples of constraints and factors that operate at each scale to drive community structure

ecology (O'Neill et al. 1989; Urban et al. 1987), with Poff (1997) placing an emphasis specifically on local community assembly.

### 11.2.1.6 Communities and Conservation

Components of biodiversity are often used as targets for conservation. For instance, species richness is frequently considered as a key indicator of biodiversity across landscapes, albeit an imperfect one. Beta diversity is increasingly emphasized in conservation (Karp et al. 2012; Socolar et al. 2016), in part because of concerns of *biotic homogenization*: where environmental change causes communities to be more similar across space due to an increase in generalist and exotic species (Olden and Rooney 2006). At a larger scale, Identifying bioregions, or biographic regions that harbor similar communities, has been helpful for interpreting ecological dynamics and developing broad-scale conservation strategies (Vilhena and Antonelli 2015). At a global scale, identifying and mapping biodiversity hotspots across the planet has been central to conservation initiatives (Myers et al. 2000; Brooks et al. 2002; Orme et al. 2005).

These components of diversity are frequently integrated in spatial conservation planning through the identification of sites with high local diversity and how a collection of sites or protected areas combine to reach conservation goals through several conservation concepts (Kukkala and Moilanen 2013). For instance, *comprehensiveness* refers to the objective of capturing the full spectrum of biodiversity in the region of interest while *representativeness* describes the extent to which a

collection of sites (or protected areas) meets that goal (Kukkala and Moilanen 2013). Concepts that directly capture alpha and beta diversity are irreplaceability and complementarity. *Irreplaceability* describes the importance of a potential site for conservation, in terms of its unique contribution to the overall biodiversity goal, such that if the site is lost the ability to reach conservation goals is hampered (Ferrier et al. 2000). *Complementarity* in conservation planning represents the degree to a site (or group of sites) contributes to unrepresented features—typically species—to an existing set of protected sites (Margules and Pressey 2000). A site has higher complementarity when it contains species not protected by existing sites. Thus, when high turnover occurs between one or more protected areas and another site being considered for conservation, there is high complementarity for that site.

The role of species interactions across space is also increasingly considered in conservation strategies. In particular, certain types of interactions, such as plant-pollinator interactions are important for maintaining ecosystem services, and such interactions can vary across landscapes (Winfree et al. 2009). Trophic interactions are also important in some conservation planning, particularly in marine and freshwater environments (Baskett et al. 2007; Decker et al. 2017). As a consequence, spatial modeling of communities and related ecosystem services has increased over the years and is essential for these types of conservation efforts (Brosi et al. 2008; Moilanen et al. 2009; Kaiser-Bunbury and Bluthgen 2015).

## **11.2.2 Common Approaches to Understanding Community–Environment Relationships**

Predicting and mapping communities over space is challenging. There are several types of modeling frameworks for communities. Frameworks that focus on environmental filtering and species sorting as the primary drivers of (meta) communities are most commonly used, likely in part because these frameworks are more feasible to implement than other frameworks that emphasize other metacommunity processes (e.g., variation in dispersal). Ferrier and Guisan (2006) classified community-level models into three categories: (1) predict first, assemble later; (2) assemble first, predict later; and (3) assemble and predict simultaneously (see also D’Amen et al. 2017). Here, we follow this categorization to illustrate several themes regarding the spatial modeling of communities.

### **11.2.2.1 Predict First, Assemble Later**

One way in which models for communities have been developed is simply to model each species separately (see, e.g., Chaps. 6 and 7) and then with this multi-species

information, aggregate or pool across species to predict communities across space. In the species distribution modeling literature, this general approach is commonly referred to as “stacked species distribution models”, or S-SDM (Guisan and Rahbek 2011). This approach implicitly emphasizes that species may respond individualistically to environmental relationships, a “Gleasonian” perspective for community structure.

Predictions for models of individual species can be combined in a variety of ways. For example, probabilities of occurrence can be truncated to expected presence–absence and then summed across species to derive species richness. Alternatively, individual model outputs could be used to interpret spatial variation in community composition by applying predictions to similarity or distance-based metrics (see below). Consequently, this approach uses model predictions as inputs for community classification and summary metrics, rather than the raw data.

### 11.2.2.2 Assemble First, Predict Later

In this approach, communities are first summarized in some way without reference to the environment. For instance, species richness may be quantified, community types (e.g., number of foraging guilds) may be summarized, or community (dis)similarity may be estimated.

Species richness can be estimated in a variety of ways. Typically, the raw count of species is a biased estimator of species richness. Instead, community ecologists attempt to adjust raw counts in at least three different ways. First, rarefaction is commonly used (Gotelli and Colwell 2001). Rarefaction acknowledges that the number of species observed will be a function of the number of individuals detected: as the number of individuals detected increases, we expect that that number of species detected will also increase. This relationship is typically asymptotic, such that rarefaction curves can be used to interpret the point at which sampling for the community was sufficient for interpreting species richness. When using rarefaction curves estimated at different localities, the species richness estimate is often truncated to the site with the lowest number of individuals, thereby allowing less biased comparison among locations regarding species richness. We note that rarefaction approaches have also been extended to account for spatial dependence in species data (Bacaro et al. 2016). Second, some estimators adjust counts of species based on the number of “singletons” (i.e., number of species detected once) or “doubletons” (i.e., the number of species detected twice) in the data (Palmer 1990; Nichols et al. 1998). This is sometimes referred to as species richness estimation through extrapolation, rather than through truncation, as in rarefaction (Colwell and Coddington 1994). The idea here is that if singletons and/or doubletons are rare in the data, then it is likely that few species have been missed. In contrast, if singletons and/or doubletons are frequent, then it is likely that many species have been missed and sampling was not sufficient. The jackknife estimator and Chao estimators for species richness

are both based on this general idea (Palmer 1990). The third approach is to formally estimate species-specific detectability and subsequently derive species richness once species-specific detectability is estimated. This approach is an extension of occupancy modeling (MacKenzie et al. 2002), termed “multi-species occupancy modeling” (Dorazio et al. 2006; Royle and Dorazio 2008; Kery and Royle 2016).

Summarizing community composition typically involves the use of (dis)similarity matrices. These matrices quantify the pairwise (dis)similarity between all sampling locations (i.e., they are square matrices). Similarity can be quantified in several ways (Koleff et al. 2003; Barwell et al. 2015). For abundance data, a Bray–Curtis index is frequently used:

$$\beta_{ij} = \frac{B + C}{2A + B + C}, \quad (11.3)$$

where  $A$  is the sum of the minimum abundance of species between site  $i$  and  $j$  (i.e., the number of individuals occurring at both sites),  $B$  is the number of individuals unique to site  $i$  and  $C$  is the number of individuals unique to site  $j$ . For binary data of species occurrence, a common approach is to use the Sørensen dissimilarity index (another common measure is the Jaccard Index):

$$\beta_{\text{sor},ij} = \frac{b + c}{2a + b + c}, \quad (11.4)$$

where  $a$  is the number of species common to sites  $i$  and  $j$ ,  $b$  is the number of species in site  $i$  that are not in site  $j$  and  $c$  is the number of species in site  $j$  that do not occur in site  $i$ .

The Sørensen and Bray–Curtis indices are functionally very similar but work with binary and count data, respectively. Both of these metrics range from 0 to 1. Dissimilarity is simply  $1 - \text{similarity}$  and can sometimes be considered a distance metric (note, however, that some dissimilarity matrices do not satisfy the “triangle inequality” and are thus not measures of ecological distance). With this approach, we may be interested in only considering the nestedness and turnover components of beta diversity (Fig. 11.1; Baselga 2010). For the Sørensen index, turnover between two sites is:

$$\beta_{\text{turn},ij} = \frac{\min(b, c)}{a + \min(b, c)}. \quad (11.5)$$

Nestedness can then be described as the fraction of  $\beta_{\text{sor},ij}$  not explained by  $\beta_{\text{turn},ij}$ :

$$\beta_{\text{nest},ij} = \beta_{\text{sor},ij} - \beta_{\text{turn},ij}. \quad (11.6)$$

With these newly assembled matrices and community summary statistics, we can then proceed to predict changes in communities across space.

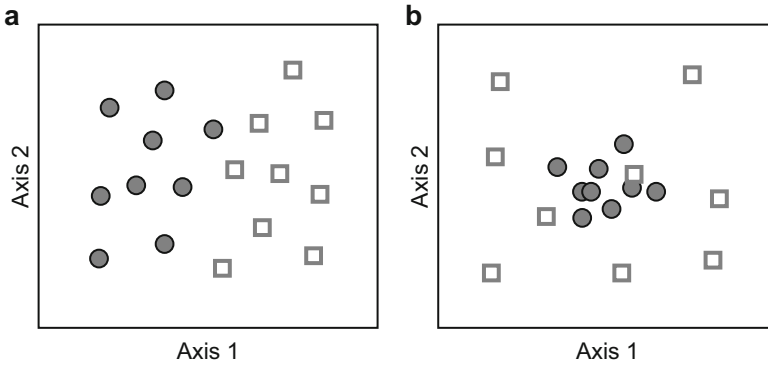
### 11.2.2.3 Predict and Assemble Together

Rather than treating the assembly of communities and their prediction over space as separate components, several modeling approaches integrate these steps formally into a single modeling framework. Multivariate regression (Ovaskainen et al. 2010), constrained gradient ordination techniques (e.g., canonical correspondence analysis) (Palmer 1993), multi-species occupancy modeling (Dorazio et al. 2006; Iknayan et al. 2014), and joint community models (Warton et al. 2015a) are just some techniques that approach the problem in this way. In this case, modeling frameworks typically provide predictions for each species, thereby honoring species identity, such that species richness is typically a derived parameter from this modeling framework (blurring the lines between predict first and assemble later approaches and predict and assemble together).

### 11.2.3 *Spatial Models for Communities*

Depending on the framework considered for community modeling, there are a variety of modeling approaches that could be considered. Here, we provide a brief overview of those approaches that are commonly used, with a focus on approaches that have not been considered elsewhere in the book. We first describe these common approaches and then explicitly address how the problem of space can be accommodated with these models and related issues for communities.

Spatial community models typically either work with summaries of species, such as species richness (Rahbek and Graves 2001), distance-based matrices regarding similarity in community composition (Ferrier et al. 2007), or work directly with species-level variation in occurrence or abundance (Rahbek and Graves 2001; Ovaskainen et al. 2010). Recently, it has been argued that the latter shows better properties than using distance-based summary statistics (Warton et al. 2012), because distance-based analyses can conflate dispersion versus location effects (Fig. 11.4). Some have organized these approaches into algorithmic models and statistical, model-based approaches (Warton et al. 2015b). Algorithmic models are those that are defined based on a set of algorithmic steps taken to interpret the data and these typically do not take fully into account the statistical properties of the data; examples include several techniques based on *ordination* (see below). Model-based approaches focus on explicit, multivariate statistical models that attempt to capture the statistical properties of the data (Warton et al. 2015b); most approaches are extensions of the generalized linear model.



**Fig. 11.4** An illustration of (a) location versus (b) dispersion effects for two groups observed in ordination techniques that use distance-based approaches. Adapted from Anderson et al. (2008) and Warton et al. (2012)

### 11.2.3.1 Multivariate Regression Analysis

Regression models can be extended to simultaneously model multiple species in a community. In this case, there are multiple response variables and as such, these models are referred to as multivariate regression techniques (rather than “multiple regression,” which refers to situations where there is  $>1$  explanatory variable).

Multivariate regression can be implemented in a variety of ways (Legendre 1993; Lichstein 2007; Wang et al. 2012). Traditionally, this approach used distance matrices of response and explanatory variables and used permutation tests to assess significance, because of the lack of independence of site pairs in the matrix formulation. In this approach, matrix regression can be described as:

$$d_{ij} = \alpha + \beta |x_i - x_j|, \quad (11.7)$$

where  $d_{ij}$  is the distance (e.g., compositional dissimilarity) between locations  $i$  and  $j$  and  $|x_i - x_j|$  is the absolute value of the difference in environmental variable  $x$  between locations.

More recently, generalized linear models (GLMs) have been advocated (Wang et al. 2012; Warton et al. 2012) as a way to analyze community data. This GLM approach may be useful for community data because it can be applied to non-normal response variables typically used in community-level modeling without resorting to summarizing species composition based on distance matrices (Warton et al. 2015a). In this case, GLMs are fit to each species separately, similar to a “predict first-assemble later” strategy; however, multi-species (community-level) inference is made based on the suite GLM models fit. Statistical tests have been developed to account for correlations among species (via permutation tests) as well as using combined summary statistics (e.g., sums of squares across models) to make community-wide inference.

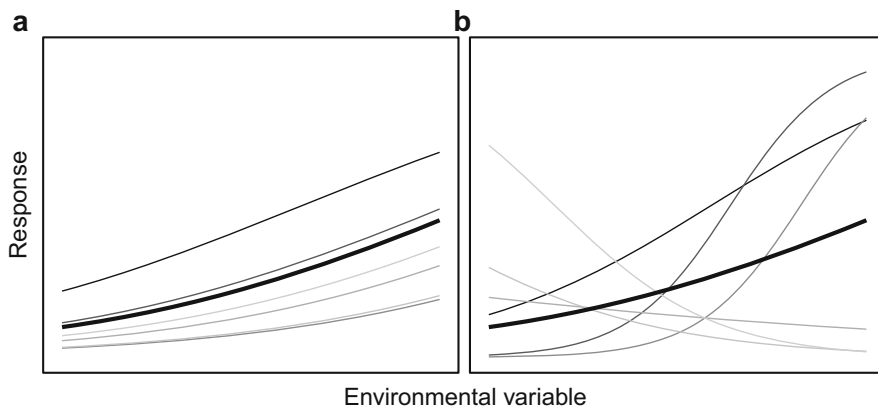
The GLM approach can also be extended to generalized linear mixed model (GLMM) formulations, allowing to account for potential dependencies between species. GLMMs can also reduce the need for permutation tests for inferences (as used in some multivariate regression approaches). In Chap. 6, we specified a generalized linear model for the presence–absence of a single species as:

$$\text{logit}(p_i) = \alpha + \beta x_i, \quad (11.8)$$

where  $p_i$  is the expected value for the probability of occurrence for sampling unit  $i$ ,  $\alpha$  is the intercept,  $\beta$  is the slope (coefficient),  $x_i$  is the explanatory variable measured at  $i$ . Multivariate GLMMs extend this idea to  $K$  species as:

$$\text{logit}(p_{ik}) = \alpha + \beta x_i + \gamma_k + \delta_k x_i, \quad (11.9)$$

where now the response variable is the presence–absence of species  $k$  at location  $i$ . We can account for different species prevalence by adding a species-level random intercept,  $\gamma_k$ , and for variation environment relationships among species through the use of species-specific random coefficients (aka random slopes),  $\delta_k$ , for an environmental variable  $x$  (Bates et al. 2015; Warton et al. 2015a) (Fig. 11.5). In these cases, random effects are typically assumed to be distributed as  $\sim \mathcal{N}(0, \sigma^2)$ . This general approach lies at the heart of several advances in community-level analyses and has been extended to account for imperfect detection (Dorazio et al. 2006), metacommunity colonization-extinction dynamics (Dorazio et al. 2010), hierarchical spatial scaling effects on communities (Ovaskainen et al. 2016a), the potential for



**Fig. 11.5** An illustration of the difference between (a) random intercepts and (b) random coefficients (or random slopes) in generalized linear mixed models. Note that for (b), both random intercept and coefficients are shown. Grey lines are species-specific responses, while black line is the average response across species

biotic interactions between pairs of species by altering the variance–covariance matrices of random effects (Ovaskainen et al. 2010), as well as trait-based dependencies (Dorazio and Connor 2014).

### 11.2.3.2 Canonical Ordination: Redundancy and Canonical Correspondence Analysis

Direct gradient analysis, also known as canonical ordination or constrained ordination analysis, is often used by community ecologists to interpret how communities respond to environmental gradients. In this case, rather than only considering the species community in the *ordination* (as in principal components analysis, PCA, and correspondence analysis, CA; see Legendre and Legendre 2012), the community is related to environmental and/or spatial data in the context of the ordination. Two of the most common approaches are redundancy analysis (RDA) and canonical correspondence analysis (CCA).

Redundancy analysis is a method that effectively combines regression-like techniques with ordination (specifically, PCA). The general idea of RDA is that it is a multivariate linear regression where the fitted values are then subjected to PCA, which provides eigenvectors of the fitted values (Borcard et al. 2011). RDA then takes these eigenvectors and computes new orthogonal (i.e., independent) axes that are linear combinations of all explanatory variables, where the first axis explains the most variation in the response variables, the second axis explains the next most, and so on. This aspect of RDA is reflected in the decrease in eigenvalues for each axis (similar to PCA). The analysis can then be summarized based on species scores, site/location scores (summarizing species scores for each site), and site constraints (the linear combinations of environmental variables for each site). This approach is appropriate when one expects linear environmental relationships, although similar to linear regression (see Chaps. 6 and 7), polynomial terms can be added when warranted to capture some types of non-linear relationships.

Canonical correspondence analysis is similar to RDA but rather than using PCA in its formulation, it uses correspondence analysis (CA). It captures Gaussian relationships of species responses to environmental gradients. Because niche theory often envisions species responses across environmental gradients as hump-shaped, Gaussian curves, CCA has had major appeal since its introduction in the 1980s (Ter Braak 1987). However, there are known limitations of CCA, particularly its use of a  $\chi^2$  distance among sites. This distance measure is known to be a poor distance metric for community composition analyses. Consequently, there is currently a greater focus on the use of RDA for direct gradient ordination analyses (Borcard et al. 2011), and we focus on RDA below.



### 11.2.3.3 Generalized Dissimilarity Modeling

Generalized dissimilarity modeling (GDM) is increasingly used to understand and predict beta diversity across space for ecology and conservation problems (Ferrier et al. 2007; Thimassen et al. 2011; Fitzpatrick and Keller 2015; Jewitt et al. 2016; Rose et al. 2016). This approach is a non-linear extension of multivariate regression, where the response variables are measures of community dissimilarity, and predictors often include spatial (e.g., distance matrices) and environmental factors.

The GDM approach was derived to accommodate two forms of non-linearity in community modeling. First, because dissimilarity is constrained to the 0–1 scale, non-linearities of the response variable occur. This non-linearity is addressed by formulating the problem as a generalized linear model with a custom link function and error distribution (Ferrier et al. 2007). The link function,  $\eta$ , used is:

$$\eta = -\log(1 - \mu). \quad (11.10)$$

where  $\mu$  is the expected value. Note that a beta distribution could also be used, which is a continuous distribution bounded to the 0–1 scale. Second, the rate of turnover at different locations on environmental gradients is expected to be non-linear. To address this issue, GDM fits non-linear, monotonic functions directly to the environmental variables, which are referred to as I-spline basis functions (Ferrier et al. 2007). I-splines are similar to the splines discussed in Chaps. 6 and 7, with the general difference being that they are constrained to be non-decreasing functions. This constraint makes sense in this case because we expect a priori that turnover rates should increase with increasing distance across environmental gradients. Similar to standard matrix regression (see above), significance is inferred through permutation tests.

### 11.2.3.4 The Problem of Space

Most of the above mentioned approaches only indirectly account for spatial dependence in community modeling. Spatial dependence in community modeling is often overlooked (Urban et al. 2002), but Dray et al. (2012) argued that it may alter inferences in our understanding and conservation of communities. Furthermore, they argued that spatial dependence only needs to occur in a portion of the community for it to potentially impact inferences.

Partial ordinations have long been used to account for potential spatial dependence via the inclusion of a geographic distance matrix in modeling (Borcard et al. 1992). Such matrices could be based on Euclidean or some other (effective) distance metric (see Chap. 9). The distance matrix (a square matrix of pair-wise distances between sites) is frequently used as a predictor or “controlling” variable (Borcard et al. 1992). Partial ordination is often used to then partition variance based on

different spatial and environmental factors (Cushman and McGarigal 2002). Yet partitioning generally assumes only additivity in the explanatory variables and it can yield negative components of variance due to interactions between variables. As such, partitioning should be used with caution.

Partial Mantel tests have also been frequently used to account for space. Mantel tests are statistical tests of the correlation between two distance matrices of the same rank (i.e., the same dimensions). Mantel tests calculate a correlation coefficient between the two matrices, and significance is inferred via permutation tests. These matrices are symmetric, distance-based matrices, so the number of distances is  $n(n-1)/2$ , or the number of observations in the upper (or lower) triangle of the matrices. Typically, the Pearson correlation coefficient is used (see Chap. 5). To assess significance of the Mantel correlation, the rows and columns of one the matrices are shuffled many times and the Mantel correlation is calculated on these randomized matrices. Significance is then inferred based on the proportion of times the observed correlation is higher than that of the correlations from the randomized matrices.

In a spatial context, Mantel tests can provide a single global test of autocorrelation for community data when comparing a spatial distance matrix (e.g., geographic distance) the community dissimilarity matrix. It is important to note here that the implicit assumption is that autocorrelation is linear gradient (i.e., the Mantel test typically uses a linear correlation coefficient). Also, there has been some criticism of this approach for a variety of reasons (Guillot and Rousset 2013; Legendre et al. 2015).

The Mantel correlogram is a multivariate extension of the correlogram described in Chap. 5, which quantifies spatial autocorrelation as a function of distance (Bjørnstad and Falck 2001; Borcard and Legendre 2012). For each distance bin, the Mantel correlogram simply calculates a normalized correlation coefficient based on comparing the species dissimilarity matrix to a binary matrix, where sites within the distance bin are 0 and all others are 1. Stringing these correlation coefficients together results in a Mantel correlogram. Significance is inferred via permutation in the same manner as with a standard Mantel test.

Multivariate variograms can also be used. Wagner (2003) pioneered the application of multivariate variograms to community data in ecology. In this application, she derived the *variogram matrix*,  $\mathbf{C}(d)$ , for communities, where the diagonal is the semivariance for species  $i$  at distance class  $d$  (see Chap. 5) and the off-diagonals represent the pairwise cross-variograms for species  $i$  and  $j$  at distance class  $d$ . Cross-variograms are similar to variograms except that they quantify the distance-dependent covariance between two types of observations; in this case, two species. A cross-variogram for species  $i$  and  $j$  can be quantified as:

$$\gamma_{i,j}(d) = \frac{1}{2n_d} \sum (z(x_i) - z(x_i + d))(z(x_j) - z(x_j + d)), \quad (11.11)$$

where  $\gamma$  is a measure of covariance,  $n$  is the number of observations at distance bin  $d$ , and  $z$  is the observation at location  $x_i$ . The variogram matrix can be used in a variety

of ways to interpret spatial dependence of communities. Beyond the species-specific variogram for spatial dependence and pairwise cross-variogram for spatial covariance between pairs of species, Wagner (2003) emphasized two other properties. First, the sum of the diagonal of  $\mathbf{C}(d)$ , she referred to as the empirical variogram of complementarity, or the spatial complementarity of species composition at locations. Second, the sum of  $\mathbf{C}(d)$  (diagonal + off-diagonals) can be considered the empirical variogram for sample-level species richness.

Multivariate variograms can be extended to ordination techniques, commonly referred to as multiscale ordination. The idea is similar to that described above for species composition and richness. In a nutshell, the  $\mathbf{C}(d)$  matrices are summed across distance classes to create a global matrix  $\mathbf{C}$  of empirical variance–covariance. This matrix is then subjected to ordination techniques, typically either PCA (Wagner 2003) or CA (Wagner 2004). Eigenvalues from the ordination can then be partitioned among distance classes and plotted as a function of distance, providing an empirical variogram of ordination axes that describe the spatial covariance of complementary in the species assemblage.

While geographic distance matrices are frequently used in Mantel tests and related analyses (e.g., GDM), the use of geographic distance matrices for inferring and controlling for spatial dependence in community-level modeling may be limited (Dray et al. 2012), due to the difficulty of proper interpretation (Legendre et al. 2015) and potential low power in detecting spatial structures (Legendre et al. 2005). Yet Borcard and Legendre (2012) contrasted multivariate variograms and Mantel correlograms using simulations, finding that under the simulated conditions, the power of these multivariate approaches was high and similar to univariate approaches.

An alternative to the use of distance matrices is using spatial weighting matrices, which come in several forms (Dray et al. 2012). Spatial eigenvector mapping (Dray et al. 2006) described in Chaps. 5 and 6 is one technique that is based upon spatial weighting matrices. Like a distance matrix, a spatial weighting matrix is a site-by-site matrix (i.e., a square matrix) that describes the potential pairwise linkages between sites. Weights can be binary or weighted (continuous, non-negative). This weighting matrix can also be directed (links between  $i$  and  $j \neq i$ ) to account for directed flows across landscapes (Blanchet et al. 2008). The subsequent incorporation of spatial weighting matrices can often occur in ways similar to the inclusion of geographic distances in the methods described above. This general approach provides great flexibility in formally capturing the role of space on communities.

## 11.3 Examples in R

### 11.3.1 Packages in R

In R, there are a few libraries that can be used for community-related models. Some common packages include the `vegan` package for ordination techniques

(Dixon 2003), `betapart` for interpreting beta diversity metrics (Baselga and Orme 2012), `gdm` for fitting generalized dissimilarity models (Manion et al. 2018), `mvnabund` and `VGAM` for multivariate GLM models for abundance and occurrence (Wang et al. 2012).

### 11.3.2 *The Data*

We return to the data shown in Chaps. 6 and 7 regarding bird distribution in Montana and Idaho, USA (Hutto and Young 2002). Sampling locations consist of point counts (100-m radius), along a transect (10 points/transect; transects are approximately 3 km long), with transects randomly selected within USFS Forest Regions across Montana and Idaho. Previously, we considered only one species; here, we extend our questions and analysis to the community. To do so, we only consider species adequately sampled by point counts (e.g., we remove waterfowl, raptors, and nocturnal species). In this example, we pool data across 3 years (2000, 2002, and 2004) for each point location. We consider three covariates used in Chap. 7: elevation, precipitation, and canopy cover.

### 11.3.3 *Modeling Communities and Extrapolating in Space*

We first illustrate common ways to approach modeling without explicit focus on incorporating space into the analysis. We then extend these ideas to formal accounting of space.

To begin, we will import a raster layers of elevation and canopy cover, as well as data on species detections at points, using the `raster` package.

```
> library(raster)

> Elev <- raster("elev.gri") #elevation layer (km)
> Canopy <- raster("cc2.gri") #linear gradient, from PCA
> Precip <- raster("precip.gri") #precipitation (cm)

#convert precipitation to meters
> Precip <- Precip / 100
> layers <- stack(Canopy, Elev, Precip)
> names(layers) <- c("canopy", "elev", "precip")

#species data
> birds <- read.csv("birdcommunity.csv")
```

These community data come in a format that is common for data entry purposes, where each row of data reflects a detection of a species at a site. We need to re-format

the data to produce a species by site data frame, where the columns are species and the rows are sites (Fig. 11.1). Also, note that the coordinates for the site data are in WGS84, which is not the same coordinate reference system as the raster data.

We first convert these data to a `SpatialPointsDataFrame` and transform the data to the projection of the raster data.

```
> birds.latlong <- data.frame(x = birds$LONG_WGS84, y =
  birds$LAT_WGS84)
> birds.attributes <- data.frame(transect = birds$TRANSECT,
  point = birds$STOP, species = birds$SPECIES, pres =
  birds$PRES)

#define CRS
> crs.latlong <- CRS("+proj=longlat +datum=WGS84")
> crs.layers <- CRS("+proj=aea +lat_1=46 +lat_2=48 +lat_0=44
  +lon_0=-109.5 +x_0=600000 +y_0=0 +ellps=GRS80 +datum=NAD83
  +units=m +no_defs")

#create SpatialPointsDataFrame
> birds.spdf <- SpatialPointsDataFrame(birds.latlong, data =
  birds.attributes, proj4string = crs.latlong)

#transform CRS for sites to layers CRS
> birds.spdf <- spTransform(birds.spdf, crs.layers)

#data frame with new x,y coordinates
> birds.df <- data.frame(birds.spdf@data, x =
  coordinates(birds.spdf)[,1], y = coordinates(birds.spdf)[,2])

> head(birds.df, 2)

##
transect point species pres x y
1 452511619 5 AMDI 0 59142.22 173151.8
2 452511619 6 AMDI 0 58834.36 173185.7
```

Now, we reformat the data to a wide format to create a format of a site  $\times$  species data frame with the `reshape2` package (Wickham 2007).

```
> library(reshape2)
> species.site <- dcast(birds.df, transect + point + x + y ~
  species, value.var = "pres")

#no attributes (species only)
> spp.matrix <- species.site[, -c(1:4)]
```

Finally, we will check for very rare species, creating a vector of names of species with detections on  $>20$  points:

```
#subset based on frequency of occurrence
> prevalence <- colSums(spp.matrix)
> prevalence.20 <- prevalence[prevalence > 20]
> species.20 <- names(prevalence.20)
```

We use this list of species to subset the data for species with detections on >20 points (out of the 1145 points). To do so, we use the `%in%` command to select columns with column names matching our vector of species to retain:

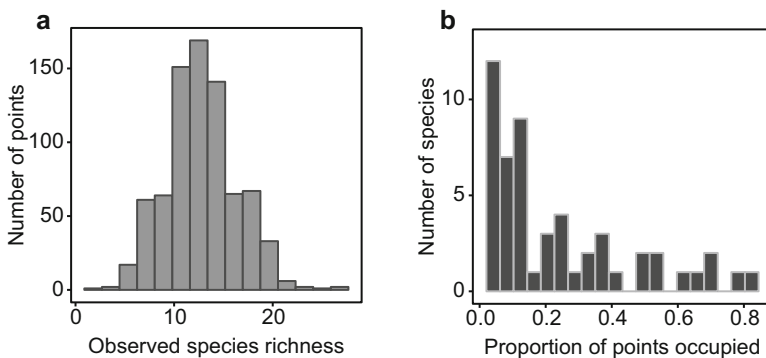
```
> species.matrix <- spp.matrix[,colnames(spp.matrix) %in% species.20]
```

We can summarize this species matrix in a variety of ways. For instance, we can calculate the observed number of species per point (see below), as well as how prevalent species are across the points sampled (Fig. 11.6). In this case, observed (uncorrected) species richness varies considerably, with 12.8 species being observed on average. The prevalence of most species is low: only a few species are commonly observed across most points (Fig. 11.6b). Community composition and similar can also be calculated. For instance, the Sorenson index for community dissimilarity can be calculated as:

```
> sorenson <- vegdist(species.matrix, method = "bray")
> sorenson.mat <- as.matrix(sorenson)
```

Where the method "bray" reduces to the Sørensen with binary data (as is the case here). With these data, we extract relevant environmental information from the points and check for correlations among environmental variables:

```
> site.cov <- extract(layers, species.site[,c("x", "y")])
> cor(site.cov)
```



**Fig. 11.6** Summarizing community-data for the Landbird Monitoring Program. Data come from 782 points sampled during 3 years (2000, 2002, and 2004). Shown are point-level observations of (a) species richness (for species >20% point locations) and (b) species prevalence

```
##
canopy elev precip
canopy 1.00 0.01 0.13
elev 0.01 1.00 -0.04
precip 0.13 -0.04 1.00
```

These environmental variables are not strongly correlated at sampling locations, so in modeling we can consider each of these variables without any substantial concerns regarding collinearity (Dormann et al. 2013). With this information, we can now proceed to spatial modeling of communities.

### 11.3.3.1 Predict First, Assemble Later

First, we model each species separately and combine them to make predictions across the region, sometimes referred to as “stacked species distribution models,” or S-SDMs (Dubuis et al. 2011; D’Amen et al. 2015). To do so, we will illustrate with similar approaches as described in Chaps. 6 and 7, where we use a logistic regression framework to model each species individually, and store relevant output for each species in lists for post-processing. Note, if we were interested in community-level inferences, such as if the community overall changed with elevation or canopy cover, we could use the `mvabund` package (Wang et al. 2018) to automatically model each species as a function of covariates (using the `manyglm` function). The primary benefit of that package is that there are summary inference-related tests that can be applied to the species-specific logistic regression models. Here, we illustrate applying these models manually, which could provide more flexibility on the types of models considered for each species.

```
> pred.map <- list() #stores map predictions
> pred.coef <- list() #stores coefficients
```

Storing in a list format can be particularly helpful when each species has different amounts of summary statistics, such as if different covariates are used for each species (e.g., species-specific model selection). Below we consider each of the environmental covariates and include a potential non-linear effect of elevation on species distribution.

```
> Nspecies <- ncol(species.matrix)
> Nsites <- nrow(species.matrix)

#create covariate vectors for simpler processing
> canopy <- site.cov[, "canopy"]
> elev <- site.cov[, "elev"]
> precip <- site.cov[, "precip"]

#Run a GLM for each species
> for (i in Nspecies){
```

```

species.i <- glm(species.matrix[,i] ~ canopy + poly(elev,2) +
  precip, family = binomial)

#coefficients from model
pred.coef[[i]] <- coef(species.i)

#predictions for mapping
logit.pred <- predict(model = species.i, object = layers, fun = predict)
prob.pred <- exp(logit.pred) / (1 + exp(logit.pred))
pred.map[[i]] <- prob.pred
}

#convert list to a multi-layered raster stack
> prob.map.stack <- stack(pred.map)
> names(prob.map.stack) <- colnames(species.matrix)

```

Note that initially predictions are on the link (logit) scale, but we back-transform predictions to the probability scale for mapping. We can check out maps for any species and the estimated coefficients. Here, we check for the first species in the species matrix, the American robin (*Turdus migratorius*; map not shown):

```

# Plot prediction for species 1
> plot(prob.map.stack$AMRO, xlab = "Long", ylab =
  "Lat",
  main="AMRO - Predict first, assemble later")
> pred.coef[[1]]

##
(Intercept) canopy poly(elev, 2)1 poly(elev, 2)2 precip
-1.4362196 -0.1011538 1.3322123 6.8103904 0.3742213

```

Now, with these species-specific maps, we can assemble the predicted community in a variety of ways. We could convert the probability maps to binary maps of presence–absence based on some sort of threshold (see Chap. 7) (Algar et al. 2009; Dubuis et al. 2011). For example, Dubuis et al. (2011) created binary predictions by selecting a species-specific threshold that maximizes the sum of sensitivity and specificity (D’Amen et al. 2015). With models where predictions are probabilities of occurrence, a more natural way may be to create binary maps based on realizations from the binomial distribution because we assume in the logistic model that observations come from this distribution. Here, we illustrate the use of random deviates (realizations) from the binomial distribution and contrast this to using a simple threshold of species prevalence, which has been shown to be a useful threshold technique for single species models (Liu et al. 2005).

We first illustrate one realization from the binomial distribution. To do so, we create a function for generating binary maps from the predicted probabilities:

```

> binary.map <- function(map) {
  values.i <- values(map)

```



```

binom.i <- rbinom(length(values.i), prob = values.i, size = 1)
map <- setValues(map, binom.i)
return(map)
}

```

This function takes a single map, extracts the values on the map and uses `rbinom` to generate one realization (random deviate) based on the predicted probability. We can then implement this function on the raster stack, where the function will execute on each layer of the stack individually:

```
> binary.map.stack <- binary.map(prob.map.stack)
```

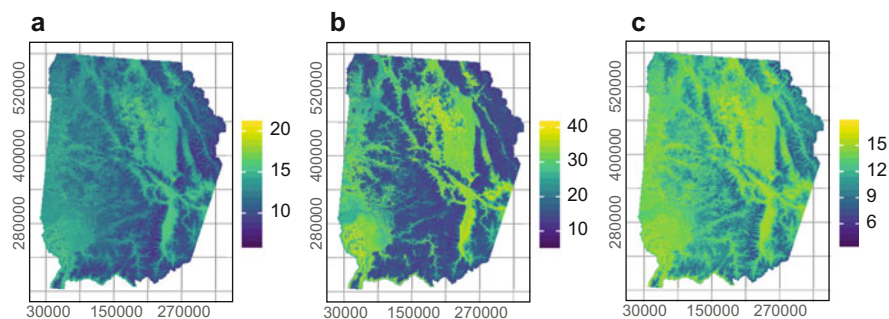
With these predicted, binary maps, we can assemble a variety of summaries at the community level. For instance, we calculate species richness as:

```

#Species richness from predictions
> spp.binomial.map <- sum(binary.map.stack)

```

This map illustrates how using a single random deviate from a binomial distribution based on the predicted probability of occurrence can lead to a great deal of noise in the predictions, where the map shows little spatial pattern. If we do this several times and then plot the mean or median predicted richness, we get a different perspective, where we observe spatial pattern in predicted richness across the region based on the covariates considered (Fig. 11.7a). Below we repeat the `binary.map` function, and with each realization from the binomial distribution we add it to our raster stack using the `addLayer` function (this loop is relatively slow to run):



**Fig. 11.7** Contrasting maps for species richness based on predict then assemble and assemble then predict approaches. For predict then assemble, maps differ when using (a) random deviates from the binomial distribution, or (b) species-specific thresholds based on prevalence. (c) Map based on assemble then predict, which tends to predict lower overall species richness

```
#richness from 19 total random deviates:
> for (i in 1:18) {
  binary.map.i <- binary.map(prob.map.stack)
  richness.i <- sum(binary.map.i)
  spp.binomial.map <- addLayer(spp.binomial.map, richness.i)
  print(i)
}

#summarize distributions
> spp.mean.map <- mean(spp.binomial.map)
> plot(spp.mean.map)
```

We can contrast the above approach to the use of thresholding probabilities, which is more commonly applied. We illustrate the application of thresholds by using the prevalence of each species to truncate probabilities to 0, 1 (Fig. 11.7b).

```
> spp.t.map <- pred.map
> for (i in 1:Nspecies) {
  thresh.i <- sum(species.matrix[,i]) / Nsites
  spp.t.map[[i]][which(spp.t.map[[i]][] > thresh.i)] =
  1
  spp.t.map[[i]][which(spp.t.map[[i]][] <= thresh.i)] =
  0
}
```

Note that other approaches could be used to stack predictions from models. For example, probabilities from species-specific models could be summed across species (Distler et al. 2015). However, when data used in S-SDMs are presence-only data (e.g., herbarium collections), summing probabilities is problematic, as these probabilities are not probabilities of occurrence, but are instead relative probabilities assumed to be proportional to occurrence (Hastie and Fithian 2013; Yackulic et al. 2013). Pearson et al. (2004) argued that in presence-only situations, false negative error rates (i.e., omission errors) should be minimized, such that thresholds are selected that minimize predicting absences (or unsuitable habitat) where observed presence locations occurred. For instance, using presence-only data, Newbold et al. (2009) used a threshold that resulted in a sensitivity to 95% (see also Mateo et al. 2012). More recently, Liu et al. (2013) argued that maximizing the sum of sensitivity and specificity is most appropriate. Other approaches that have been used with presence-only data include selecting thresholds that maximize agreement with independent data on species richness or other variants of sensitivity thresholds (Pineda and Lobo 2009, 2012; Milanovich et al. 2012; Zhang et al. 2016).

### 11.3.3.2 Assemble First, Predict Later

In contrast to modeling each species separately and then compiling model predictions, we can assemble the community first and then model it. The simplest scenario is to compile the total number of species detected/site (or something similar,

such as the use of rarefaction) and then model species richness directly using a GLM like before, but in this case it would be a count-based GLM, such as Poisson regression. Other approaches include assembling measures of species composition and similarity.

**Modeling Species Richness.** Poisson regression is a natural GLM for count-based data. Poisson regression assumes that the data come from a Poisson distribution, or integer data  $\geq 0$ . However, the Poisson distribution assumes that the mean = variance, which often is not the case. Instead, it is more common in ecological data to observe that the variance increases with the mean. When this pattern occurs, GLM models based on the Poisson distribution can be over-dispersed, which can lead to inferences being too liberal (i.e., more likely to commit a type I error; Zeileis et al. 2008). In such cases, quasi-Poisson or negative binomial regression are natural alternatives. Quasi-Poisson models tend to estimate the same coefficients as a Poisson, but will also estimate a scale parameter from the data, which is then used to adjust for over-dispersion in inferences. Quasi-Poisson models can be used with the `glm` function, but specifying `family = quasipoisson`. More commonly, negative binomial regression is used when data are over-dispersed. Negative binomial regression does not make the assumption that the mean = variance, but rather estimates an additional scaling parameter, typically referred to as theta. Negative binomial models can be run with the `glm.nb` function in the MASS package (Venables and Ripley 2002). Quasi-Poisson and negative binomial models have the same number of parameters but make different assumptions regarding the relationship of the variance as a function of the mean: the quasi-Poisson assumes a linear relationship, while the negative binomial assumes a quadratic relationship (Hoef and Boveng 2007).

We can diagnose the potential for over-dispersion in several ways. A crude approach is to look at the ratio of the residual deviance to the degrees of freedom in the model. If this ratio,  $c$ , is  $\gg 1$ , that suggests over-dispersion. An alternative is to use the `dispersiontest` function in the AER package (Kleiber and Zeileis 2008). This test is based on asking whether the mean equals the variance in the Poisson model (Cameron and Trivedi 1990).

We fit the Poisson model and then determine if there is any evidence for over-dispersion. We first assemble richness data and then fit a Poisson GLM.

```
> richness <- rowSums(species.matrix)

> pois.rich <- glm(richness ~ canopy + poly(elev, 2) + precip,
family = poisson)

> summary(pois.rich)

##
glm(formula = richness ~ canopy + poly(elev, 2) + precip, family =
poisson)

Deviance Residuals:
Min 1Q Median 3Q Max
-3.6603 -0.6006 -0.0088 0.6149 2.9355
```

```

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) 2.72432 0.03549 76.770 < 2e-16 ***
canopy -0.19808 0.02889 -6.856 7.06e-12 ***
poly(elev, 2)1 -1.52544 0.29129 -5.237 1.63e-07 ***
poly(elev, 2)2 -1.14501 0.29472 -3.885 0.000102 ***
precip -0.18532 0.04015 -4.615 3.92e-06 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for poisson family taken to be 1)

```

Null deviance: 775.20 on 781 degrees of freedom
Residual deviance: 658.86 on 777 degrees of freedom
AIC: 4077.3

```

For this model, the residual deviance is 658.86 on 777 df, or  $c = 0.85$ , which suggests over-dispersion is absent. We can test this more formally with the AER package:

```
> dispersiontest(pois.rich, trafo = 1)
```

This tests confirms that over-dispersion is absent ( $p = 1.0$ ). If there was a signal for over-dispersion, we could contrast this model to the quasi-Poisson and negative binomial models, such as:

```

> qpois.rich <- glm(richness ~ canopy + poly(elev, 2) + precip,
  family = quasipoisson)
> nb.rich <- glm.nb(richness ~ canopy + poly(elev, 2) + precip)

```

We contrast predictive maps from these “assemble first” models to the prior map created based on “assemble later”:

```

#map the Poisson model
> pois.raster <- predict(pois.rich, layers)
> spp.raster <- exp(pois.raster) #back-transform to count scale

```

To highlight spatial variability between models, we can map their differences:

```
> spp.diff <- spp.mean.map - spp.raster
```

There are several important differences between these approaches. In general, using S-SDMs tend to overpredict species richness (Dubuis et al. 2011; D’Amen et al. 2015), particularly when using species-specific thresholding (Fig. 11.7b). Yet these approaches do tend to be correlated (Newbold et al. 2009). We can check this as:

```

> richness.stack <- stack(spp.mean.map, spp.binomial.thres.map,
  spp.raster)
> names(richness.stack) <- c("binom-rich", "thres-rich", "pois-rich")
> richness.map.corr <- layerStats(richness.stack, 'pearson',
  na.rm = T)
> richness.map.corr

##
$'pearson correlation coefficient'
  binom.rich thres.rich pois.rich
binom.rich 1.0000000 0.4150215 0.8118022
thres.rich 0.4150215 1.0000000 0.5685640
pois.rich 0.8118022 0.5685640 1.0000000

$mean
  binom.rich thres.rich pois.rich
  6.811837 24.379847 6.784896

```

In this case, there is a much stronger positive correlation when using binomial realizations from the logistic model than when thresholding predictions. More fundamentally, by modeling species individually, S-SDMs do not put explicit constraints on the number of species occurring at a site (e.g., S-SDMs assume biotic interactions and available energy are not limiting locally) and implicitly suggests a Gleasonian and species-sorting perspective to community assembly.

Modeling species richness could also be done with subsets of the total number of species, such as the number of endemics or the number of species in a functional group. Related metrics, such as calculating Simpson's or Shannon diversity (Magurran 2004), could also be modeled directly using a similar approach, although the type of GLM used might differ, depending on the distribution of the response variable being considered. Ordination metrics have also been modeled directly in this way (Faith et al. 2003; Chang et al. 2004). For example, we might extract ordination axes from our community data (see below) and then model those axes directly.

**Dissimilarity Modeling.** Another application of assemble first is focused on interpreting beta diversity by first assembling species into a species dissimilarity matrix. With that information, one can use generalized dissimilarity modeling (GDM) (Ferrier et al. 2007), Mantel tests (Legendre et al. 2005), or distance-based redundancy analysis (Legendre and Anderson 1999) to model beta diversity across space. Here, we focus on GDM; see below for some applications of Mantel tests and redundancy analysis.

The `gdm` package can fit generalized dissimilarity models (GDMs). There are several ways in which data can be formatted for the `gdm` package, which is typically accomplished with the `formatsitepairs` function prior to implementation of the GDM algorithm. We illustrate one format that most closely aligns with the data formats used above. Briefly, we take a site-by-species data matrix for the response variables and a site-by-covariates matrix for the explanatory variables. It requires a

site id column, then  $x$ - $y$  coordinates can be specified for each site (coordinates can also be passed into the model function separately), and the remaining site-by-species data. Other formats `gdm` accepts include list formats and passing dissimilarity matrices that have been previously created rather than the raw input data.

```
> library(gdm)
> siteID <- 1:nrow(species.matrix)
> site.utm <- data.frame(x = species.site$x, y = species.site$y)

> gdm.species.matrix <- data.frame(cbind(siteID, site.utm,
  species.matrix))
> gdm.site.matrix <- data.frame(cbind(siteID, site.cov))
```

The `gdm` package uses a dissimilarity matrix of species composition between sites as the response variables. We use a Sørensen dissimilarity matrix. We format the data with the `formatsitepair` function:

```
#get gdm formatted object
> gdm.data <- formatsitepair(gdm.species.matrix, bioFormat = 1,
  dist = "bray", abundance = F, XColumn = "x", YColumn = "y",
  siteColumn = "siteID", predData = gdm.site.matrix)
```

Note that `gdm` passes the raw species data to the `vegan` package to calculate a dissimilarity matrix. In this case, we specify the Bray–Curtis dissimilarity metric and because `abundance = F`, such that this metric collapses to the Sørensen index. With this newly formatted object, we can run the GDM algorithm with the `gdm` function:

```
> gdm.dist <- gdm(gdm.data, geo = T)
> summary(gdm.dist)
```

The `geo = T` command tells the function that our geographic distance matrix should be used as a covariate in the model. The `summary` function provides several key results. First, it provides information on the deviance explained by the model. This can be thought of as a metric similar to metrics of the variation explained ( $R^2$ ). Note that if we re-fit the above model without our geographic information (using `geo = F`), the proportion of deviance explained drops only from 13.2% to 12.7% such that geographic effects explain little variation in dissimilarity overall in this dataset. For each covariate, the `summary` function also provides information on the coefficients fit to the explanatory variables. Recall that one aspect of GDM is the use of I-splines: non-linear, monotonic (non-decreasing) splines for interpreting turnover across environmental or spatial gradients. The `gdm` function defaults to using three knots (see Chap. 6 for discussion of the use of knots in splines) to create I-splines and the summary output provides information on the spline fitting. The user can manually alter the number of knots and their location with `splines` and `knots`

commands. It is also straightforward to interpret the estimated environmental relationships from the model object that is created. For example, we can plot the partial response plots as:

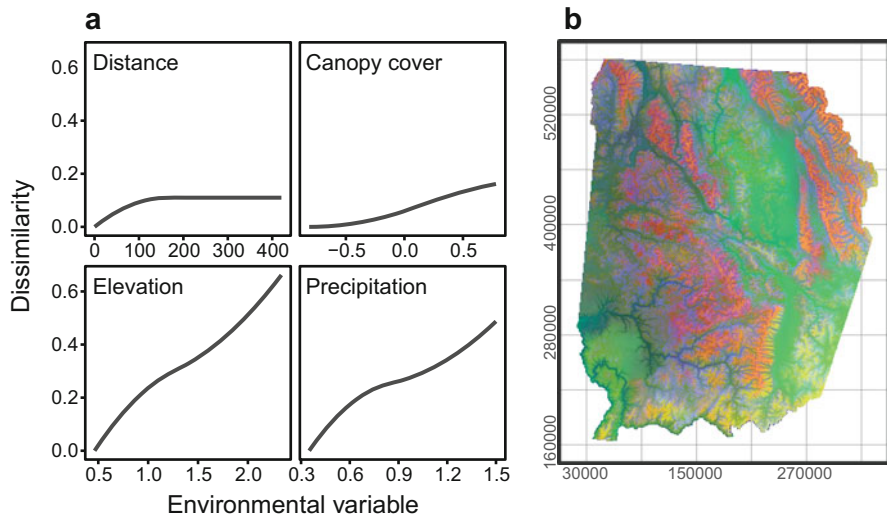
```
> plot(gdm.dist , plot.layout = c(3,2))
```

These plots (Fig. 11.8) provide several insights. First, the maximum height of each spline describes the total magnitude of change along the gradient for the explanatory variable (Manion et al. 2018). In this case, elevation captures the largest change, while the geographic distance captures the least. The shape of the spline provided information on the rate of change (turnover) in the community and where the rate of change is greatest.

We can also make spatial predictions from a `gdm` model object in several ways, but care should be taken. First, we can use the `predict` function to assess model fit, that is, plot the predicted dissimilarity as a function of the observed dissimilarity (output not shown):

```
> gdm.fit <- predict(gdm.dist, gdm.data)
> plot(gdm.data$distance, gdm.fit, xlim = c(0,1), ylim = c(0,1),
      lines(c(0,1),c(0,1)))
```

In this case, the model is a relatively poor fit to the data. This is not too surprising, given the low amount of deviance explained by the model. We can also predict over space. There are several steps to do so. First, we need to transform the raster layers based on the GDM model:



**Fig. 11.8** (a) GDM partial plots and predictions of dissimilarity. (b) Mapping dissimilarity, where similar colors represent similar communities

```
> gdm.trans.data <- gdm.transform(gdm.dist, layers)
> plot(gdm.trans.data)
```

The `gdm.transform` function takes a `gdm` object and transforms the raster layers for further analysis into “biological space” (Manion et al. 2018), such that the values represent the dissimilarity along each gradient where the minimum value for the covariate is zero and the maximum reflects the maximum dissimilarity across the gradient. Consequently, the result of this function is a prediction of dissimilarity for each environmental gradient. Note that the order of the raster layers must be in the same order as specified in the GDM model. To make an overall prediction of dissimilarity based on all variables, we need to combine predictions for each covariate (raster layer). We can either: (1) scale and sum each individual variable when the number of variables are small; or (2) use principal components analysis (PCA) when the number of variables is large (see vignette in Manion et al. 2018). Here we show an example using PCA. PCA is an ordination technique for reducing the dimensionality of multivariate data, wherein PCA produces new variables (i.e., principal components) from multivariate data that are linear combinations of the original multivariate data. We will not focus on the details of how PCA works here; interested readers should see Legendre and Legendre (1998). For relatively large rasters, we can first sample the raster values and then use these values in a PCA.

```
> sample.trans <- sampleRandom(gdm.trans.data, 10000)
> sample.pca <- prcomp(sample.trans)

#inspect
> summary(sample.pca)

##
Importance of components:
 PC1 PC2 PC3 PC4 PC5
Standard deviation 0.1556 0.08892 0.04303 0.02779 0.01519
Proportion of Variance 0.6924 0.22600 0.05292 0.02208 0.00659
Cumulative Proportion 0.6924 0.91841 0.97133 0.99341 1.00000

> round(sample.pca$rotation,2) #eigenvectors

##
 PC1 PC2 PC3 PC4 PC5
xCoord 0.04 -0.13 0.07 0.26 -0.95
yCoord -0.02 0.10 0.03 0.96 0.25
canopy 0.04 0.01 1.00 -0.04 0.06
elev 0.77 -0.63 -0.03 0.04 0.13
precip 0.64 0.76 -0.03 -0.04 -0.09
```



Here, the first two principal components explain 92% of the variation (taken from the “Importance of components” table). The eigenvectors provide information on the linear combinations of the original data that make up the new PC variables. In this case, the first two components are elevation and precipitation effects, while the third focuses on canopy cover. This result makes sense, given the partial predictions (Fig. 11.8a) and correlations among variables. With this PCA, we can then predict the pc scores onto the data. Here, we just focus on the first three principal components, using the `index=1:3` command.

```
> gdm.pca <- predict(gdm.trans.data, sample.pca, index = 1:3)
```

This will provide predictions for each PC. One way to create an integrated prediction is via rescaling the PCs to a 0–1 scale and summing them by using the `plotRGB` function from the `raster` package (see vignette in Manion et al. 2018).

```
#scale to 0-1 range
> gdm.pca <- (gdm.pca - minValue(gdm.pca)) /
  (maxValue(gdm.pca) - minValue(gdm.pca))

> plotRGB(gdm.pca, r = 1, g = 2, b = 3, scale = 1)
```

In the above plot, we specify that the first PC reflects the red channel, the second the green channel, etc.; `scale = 1` notifies the `plotRGB` function that the maximum values are 1 for the PC data. Note that by scaling the PCs in a standardized way and then using the `plotRGB` command, we are implicitly weighting each PC similarly. Taken together, this predictive map (Fig. 11.8b) reflects differences in community composition across space, with similar values reflecting similar communities.

### 11.3.3.3 Assemble and Predict Together

Increasingly, community models assemble and predict together. There are several advantages of doing so. Some frequently used techniques include some types of constrained ordination (Guisan et al. 1999) and multivariate GLM-like models, sometimes referred to as “joint species distribution models” (Dorazio et al. 2006; Ovaskainen et al. 2010; Wang et al. 2012).

**Direct Gradient Analysis.** Direct gradient analysis, such as RDA and CCA, is frequently used in community modeling. The `vegan` package can implement RDA and CCA; here, we focus on RDA. These models use a site-by-species matrix as the response variables. Blanchet et al. (2014) noted that the case of using binary data in an RDA is equivalent to the use of a distance-based RDA (see below; Legendre and Anderson 1999) when dissimilarity is quantified based on the simple matching coefficient:

$$\left(1 - \frac{a + d}{a + b + c + d}\right)^{0.5}, \quad (11.12)$$

where  $a$ ,  $b$ , and  $c$  are defined as in Eq. (11.4), and  $d$  is the number of sites where both species are absent. In this way, a simple RDA can be fit as:

```
> rda.bird <- rda(species.matrix ~ canopy + poly(elev, 2) + precip)
> rda.bird

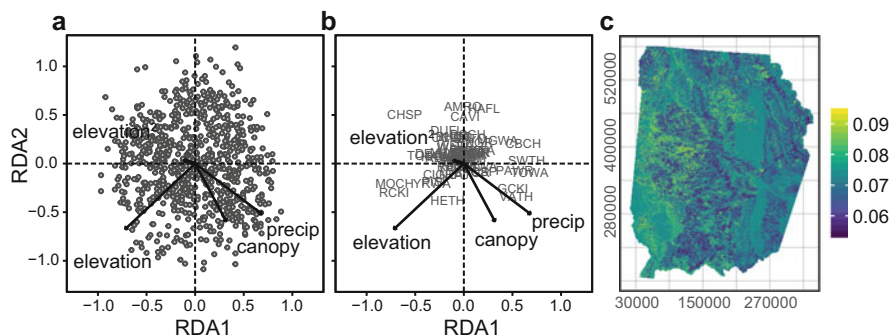
##
Call: rda(formula = species.matrix ~ canopy + poly(elev, 2) + precip)

Inertia Proportion Rank
Total 7.1866 1.0000
Constrained 0.8163 0.1136 4
Unconstrained 6.3703 0.8864 53
Inertia is variance

Eigenvalues for constrained axes:
RDA1 RDA2 RDA3 RDA4
0.4356 0.2517 0.0769 0.0522

Eigenvalues for unconstrained axes:
PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8
0.3771 0.3384 0.3102 0.2809 0.2553 0.2274 0.2111 0.2033
(Shown only 8 of all 53 unconstrained eigenvalues)
```

The output provided by typing `rda.bird` shows the proportion of the total inertia (~variance) explained by the constraining axes. It also provides the eigenvalues for the constraining axes along with eigenvalues for the top unconstrained axes. What does this really mean? Recall that RDA can be thought of as a multivariate regression where the fitted values from separate regressions for each species are then subjected to PCA (Borcard et al. 2011). RDA computes new axes that are linear combinations of all explanatory variables. There will be as many constrained axes as there are explanatory variables, though the first axis will explain more variation than the second, the second more variation than the third, etc. Eigenvalues for each axis are proportional to the variation explained based on the new axes, while eigenvectors reflect the weight of each explanatory variable for explaining that axis. The eigenvalues of the unconstrained (PCA) axes represent the amount of residual variation that is not captured by the explanatory covariates. The summary function provides site and species scores, along with the summary of inertia described above. Specifically, this shows the 'species scores', and the 'site constraints', which reflect where the species and sites (point count locations) fall in this constrained multivariate ordination space. We can extract the site and species scores for plotting and further interpretation with the `scores` function (output not shown):



**Fig. 11.9** Redundancy analysis on the bird community. (a) biplot based on site scores, (b) biplot with species scores (shown are four-letter species codes based on the American Ornithological Society’s standardized codes), and (c) mapping predictions of species scores for the varied thrush

```
> scores(rda.bird, choices = 1:2, display = "sites")
> scores(rda.bird, choices = 1:2, display = "species")
```

These scores can be plotted using “biplots” with the `plot` function, which is a common way to visualize ordination data (Fig. 11.9a, b).

Permutation tests, where rows of the community matrix are shuffled, to assess the significance of the RDA axes can be accomplished with the `anova` function:

```
> anova(rda.bird)

##
Permutation test for rda under reduced model
Permutation: free
Number of permutations: 999

Model: rda(formula = species.matrix ~ canopy + poly(elev, 2) + precip)
Df Variance F Pr(>F)
Model 4 0.8163 24.893 0.001 ***
Residual 777 6.3703
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The term “anova” is a bit of a misnomer here—it is not an analysis of variance test. However, the output of the permutation test can be summarized in a similar way and provided like a standard ANOVA table. In this case, despite the fact that the RDA axes explain little variation in variance (inertia), these axes are considered significant based on the permutation test. Each covariate can also be tested sequentially (by = 'term') or at the same time:

```
> anova(rda.bird, by = 'mar') #marginal tests
```

```
##
Permutation test for rda under reduced model
Marginal effects of terms
Permutation: free
Number of permutations: 999

Model: rda(formula = species.matrix ~ canopy + poly(elev, 2) + precip)
Df Variance F Pr(>F)
canopy 1 0.1291 15.752 0.001 ***
poly(elev, 2) 2 0.4043 24.658 0.001 ***
precip 1 0.2427 29.607 0.001 ***
Residual 777 6.3703
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results from this permutation show that each explanatory variable explains significant variation in the species data. There are several other ways to summarize RDA and we will not cover all of them here. See Borcard et al. (2011) for more details.

Finally, we can map elements of our RDA model across the region of interest (Fig. 11.9). In this case, we can make predictions based on site scores (the linear combinations of environmental covariates) or species scores as:

```
#first convert raster to data frame
> layers.df <- as.data.frame(layers, xy = T, na.rm = T)

#predict site scores onto new data frame
> rda.site.pred <- predict(rda.bird, layers.df, type = "lc")

#predict species scores onto new data frame
> rda.species.pred <- predict(rda.bird, layers.df, type = "response")
```

Predictions can be mapped in space. In Fig. 11.9, we illustrate mapping the species scores for the varied thrush (to compare with mapping in Chap. 7). Species scores could be truncated to 0, 1 data based on optimal thresholds that split occurrence observations (Liu et al. 2005), and then summed for predictions of species richness, similar to the S-SDM approach explained above.

In contrast to using a species occurrence matrix, we could instead use a dissimilarity matrix in RDA, termed distance-based redundancy analysis, or dbRDA (Legendre and Anderson 1999). In that case, we would be assembling first (community dissimilarity) and then interpreting dissimilarity between locations, analogous to generalized dissimilarity modeling. This form of RDA can be accomplished with the `capscale` function in the `vegan` package. See Blanchet et al. (2014) for discussion on the relationship of dbRDA to RDA.

**Multivariate Regression.** Several univariate modeling approaches have been extended to model multiple species simultaneously, sometimes referred to as joint species distribution models (jSDM) (Clark et al. 2014; Pollock et al. 2014;

Ovaskainen et al. 2016b). These include multivariate logistic regression models, multi-species occupancy models, multivariate machine learning methods (e.g., neural networks), and multivariate adaptive regression splines. These methods can be very complex; here, we illustrate some simpler implementations that illustrate these types of models.

We start with a multivariate logistic regression that is formulated through the use of species-level random effects. In Chap. 6, we used random effects as a potential means of accounting for spatial dependence. Specifically, we fit multilevel models by adding random effects (e.g., transects, grids, or watersheds) that capture spatial hierarchies in the data. In doing so, we were adding “random-intercepts” to the modeling framework. Another type of random effect is a “random coefficient” (aka a “random slope”). This type of random effect allows for an effect of a predictor variable to change with the random effect (Fig. 11.5). If we add species as a random coefficient, then this would allow us to consider that species may respond differently to environmental covariates.

To implement a multivariate logistic model, we need to use community data in a long, rather than wide, format. To do so, we will merge the `species.matrix` and `site.cov` objects and then use the `melt` function with the `reshape2` package.

```
> species.matrix.df <- data.frame(cbind(site.cov,
  species.matrix))
> sp.multi <- melt(species.matrix.df, id.vars = c("elev",
  "canopy", "precip"), variable.name = "SPECIES", value.name =
  "pres")
```

With this format, we can start with a random intercept model with the `glmer` function in the `lme4` package:

```
> library(lme4)
> multi.int <- glmer(pres ~ canopy + elev + I(elev^2) + precip
  + (1|SPECIES), family = "binomial", data = sp.multi,
  glmerControl(optimizer = "bobyqa"))
> summary(multi.int)

##
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [
glmerMod]
Family: binomial ( logit )
Formula: pres ~ canopy + elev + I(elev^2) + precip + (1 | SPECIES)
Data: sp.multi
Control: glmerControl(optimizer = "bobyqa")

AIC BIC logLik deviance df.resid
35404.4 35456.1 -17696.2 35392.4 41440
```

```

Scaled residuals:
  Min 1Q Median 3Q Max
-2.4474 -0.4686 -0.2841 -0.1620 6.7555

Random effects:
  Groups Name Variance Std.Dev.
 SPECIES (Intercept) 1.751 1.323
Number of obs: 41446, groups: SPECIES, 53

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.59635 0.24292 -6.572 4.98e-11 ***
canopy -0.35606 0.03881 -9.175 < 2e-16 ***
elev 0.89779 0.23139 3.880 0.000104 ***
I(elev^2) -0.42630 0.08412 -5.068 4.02e-07 ***
precip -0.32776 0.05357 -6.118 9.47e-10 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
  (Intr) canopy elev I(1^2)
canopy 0.041
elev -0.623 -0.048
I(elev^2) 0.595 0.045 -0.986
precip -0.200 -0.121 0.014 -0.006

```

In this case, we are modeling the presence–absence of each species and including canopy, elevation and precipitation as predictors for each species. We include species as a random intercept, but no other effect of species. This model formulation would then account for variation in species occurrence (i.e., prevalence), but it assumes that all species respond to environmental variables similarly. Consequently, the above model is not very helpful in most situations. However, we can extend this model by adding random coefficients of species and their relationships with canopy and elevation:

```

#random coefficient model by species
> multi.coef <- glmer(pres ~ elev + I(elev^2) + canopy + precip +
  (1|SPECIES) + (0 + elev|SPECIES) + (0 + I(elev^2)|SPECIES) + (0 +
  canopy|SPECIES) + (0 + precip|SPECIES), family = "binomial",
  data = sp.multi)

```

In this case, we also include random coefficients for the species effect with environmental variables (e.g.,  $(0 + \text{elev} | \text{SPECIES})$ ). The syntax for random effects structure in `lme4` can be a bit daunting (Bates et al. 2015). Above, the 0 tells `lme4` that there is no random intercept of species (because we have already specified it separately) but `elev|SPECIES` states that there is a random coefficient of species with elevation. This allows species to respond differently to these covariates.

However, it does assume that these random coefficients come from the same normal distribution. In doing so, it implicitly assumes species respond somewhat similarly to these covariates and for species with very little information, it will pull those species' coefficients toward the mean across species (Dorazio et al. 2010). That assumption can be relaxed to some degree using Bayesian methods, which can account for phylogenetic dependence among species and the potential for species interactions altering outcomes (Ovaskainen and Soininen 2011), but we do not cover those methods here. Note that we may want to scale variables (using the `scale` function) prior to fitting this model to help improve model convergence. We can view output from this model as:

```
> summary(multi.coef)

##
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [
glmerMod]
Family: binomial ( logit )
Formula: pres ~ elev + I(elev^2) + canopy + precip + (1 | SPECIES) + (0 +
elev | SPECIES) + (0 + I(elev^2) | SPECIES) + (0 + canopy |
SPECIES) + (0 + precip | SPECIES)
Data: sp.multi
Control: glmerControl(optimizer = "bobyqa")

AIC BIC logLik deviance df.resid
32514.0 32600.3 -16247.0 32494.0 41436

Scaled residuals:
Min 1Q Median 3Q Max
-8.1046 -0.4232 -0.2506 -0.0462 25.7761

Random effects:
Groups Name Variance Std.Dev.
SPECIES (Intercept) 5.0330 2.2434
SPECIES.1 elev 2.7689 1.6640
SPECIES.2 I(elev^2) 0.5695 0.7547
SPECIES.3 canopy 0.9091 0.9534
SPECIES.4 precip 3.3542 1.8314
Number of obs: 41446, groups: SPECIES, 53

Fixed effects:
Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.2549 0.3623 -6.224 4.84e-10 ***
elev 2.2226 0.3625 6.131 8.73e-10 ***
I(elev^2) -1.0261 0.1493 -6.871 6.37e-12 ***
canopy -0.5030 0.1398 -3.598 0.000321 ***
precip -0.5416 0.2609 -2.076 0.037882 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Correlation of Fixed Effects:
Intr) elev I(1^2) canopy
elev -0.381
I(elev^2) 0.331 -0.547
canopy 0.010 -0.008 0.005
precip -0.041 -0.003 0.002 -0.008
```

Parameters estimated from the model can be viewed as (output not shown):

```
> fixef(multi.coef)
> ranef(multi.coef)
> coef(multi.coef) #combines fixed and random effects
> coef(multi.coef)$SPECIES[, "CC2"]
```

Interestingly, in some cases when random coefficients are added, the effect may no longer be significant (when without the random coefficient they are). Why would that be? In general, if there is a great deal of variation in effects of an explanatory variable across species, then, once captured, the marginal effect may vanish. Predictions from the model can be compiled similar to above as:

```
> glmm.map <- list() #stores map predictions

# extract maps for all species
> for (i in 1:Nspecies) {
  logit.raster <- coef(multi.coef)$SPECIES[i,1] +
  coef(multi.coef)$SPECIES[i,"elev"] * Elev +
  coef(multi.coef)$SPECIES[i,"canopy"] * Canopy
  prob.raster <- exp(logit.raster) / (1 + exp(logit.raster))
  glmm.map[[i]] <- prob.raster
}
```

Note that here we are grabbing the species-specific random coefficients to make predictions. We can also make partial predictions of environmental relationships for each species, similar to what we did in Chap. 6. Partial predictions (or response curves) provide a means to visualize how the response variable changes with a covariate, while holding all other covariates constant (typically at their mean value). To do so, we look more closely at the output from this model and contrast it to the individual models we created using the S-SDM approach. Below, we contrast outputs for the black-capped chickadee (*Poecile atricapillus*):

```
> coef(multi.coef)$SPECIES[2,] #GLMM coefficients

##
(Intercept) elev I(elev^2) canopy precip
BCCH -0.2874097 2.089603 -1.322244 -0.2010582 -1.359818
```



```
> pred.coef[[2]] #GLM coefficients from S-SDM
(Intercept) canopy poly(elev, 2)1 poly(elev, 2)2 precip
0.09001473 -0.17531427 -15.42057383 -5.72431042 -1.43894282
```

We find that the multivariate model generally estimates similar coefficients for the variables, except for elevation. The different coefficients can be partly explained by the way in which elevation was modeled using the `poly` function for the S-SDM (which re-scales the variable to ensure each polynomial term is orthogonal) but with the `I` function for the multivariate regression. We can generate partial plots by first creating a new data set for predictions. Below, we focus on elevation effects:

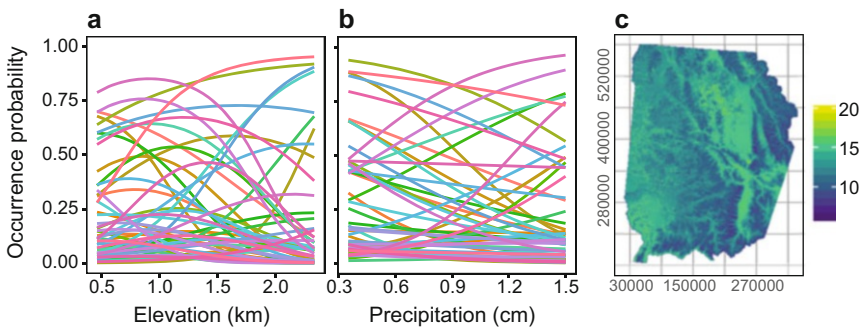
```
> site.cov.df <- data.frame(site.cov)
> elev.range <- seq(min(site.cov.df$elev),
  max(site.cov.df$elev), length = 20)
> precip.mean <- mean(site.cov.df$precip)
> canopy.mean <- mean(site.cov.df$canopy)

> newdata.glm <- data.frame(expand.grid(SPECIES = species.20,
  precip = precip.mean, elev = elev.range, canopy =
  canopy.mean))
```

Note that we set the other variables to their mean values and then use the `expand.grid` function to create a new data frame for predictions. With these data we can use the `predict` function to generate partial plots for each species:

```
> pred <- predict(multi.coef, newdata.glm, type = "response")
> glm.pred <- cbind(newdata.glm, pred)
```

These predictions show that most species are relatively rare and do not respond strongly to elevation and precipitation in the landscape, whereas some of the more common species respond both positively and negatively to these variables (Fig. 11.10). We can use this model in a similar way to our S-SDM to predict



**Fig. 11.10** Partial predictions for elevation effects for each of the 53 species from the random coefficient model for (a) elevation and (b) precipitation, and the (c) resulting prediction for species richness across the region

species richness across the region. We could use a simple thresholding technique or obtain binomial realizations from the model using the `binary.map` function to derive species richness from the `glm.map` output in an identical way as shown in Sect. 11.3.3.1. This model leads to similar predictions for species richness than the approach taken in the S-SDM code above (compare Fig. 11.10c to Fig. 11.7a).

### 11.3.4 *Spatial Dependence in Communities*

While the above modeling frameworks provide a means for predicting and mapping communities across space, most of what we have illustrated does not account directly for spatial dependence (Dray et al. 2012).

We consider several approaches for addressing the problem of space. First, we consider the spatial dependence of the community data via multivariate correlograms and variograms. Multivariate correlograms can be fit with several packages. We start with a Mantel test on geographic distance and follow it with a Mantel correlogram, which uses a distance matrix regarding community composition as the response variable:

```
#calculate distance matrix
> dist.matrix <- as.matrix(dist(site.utm))

#Mantel test
> mantel(sorenson, dist.matrix, method = "pearson", permutations =
  999)

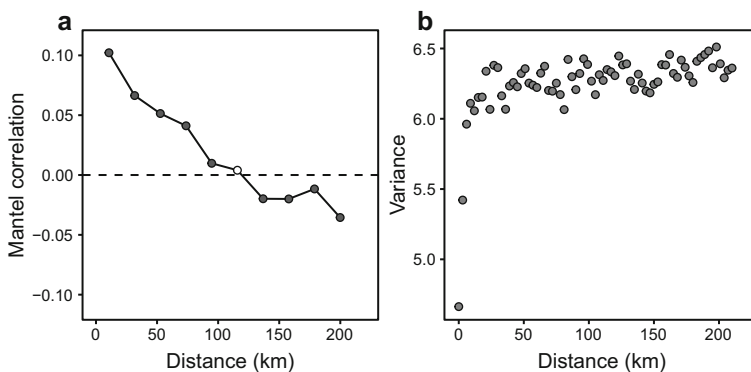
##
Mantel statistic based on Pearson's product-moment correlation

Call:
mantel(xdis = sorenson, ydis = dist.matrix, method = "pearson",
permutations = 999)

Mantel statistic r: 0.1141
Significance: 0.001

Upper quantiles of permutations (null model):
 90% 95% 97.5% 99%
0.0147 0.0191 0.0221 0.0265
Permutation: free
Number of permutations: 999
```

The Mantel test finds significant ( $p < 0.001$ ), but weak ( $r = 0.11$ ), correlation of community dissimilarity with geographic distance, consistent with the geographic effects observed in the GDM (Fig. 11.8). More broadly, we can calculate a Mantel



**Fig. 11.11** Interpreting community-level spatial dependence. (a) The Mantel correlogram based on species dissimilarity, and (b) the multivariate variogram from residuals in the RDA

correlogram with the `vegan` package to better understand spatial dependence in species dissimilarity:

```
#correlogram
> mantel.corr <- mantel.correlog(sorenson, XY = site.utm, cutoff
  = T, r.type = "pearson", nperm = 99)
```

Here, we find modest evidence for spatial dependence in the community data (Fig. 11.11a). Consequently, we may wish to revisit some of the above modeling approaches to explicitly account for spatial dependence.

### 11.3.5 *Community Models with Explicit Accounting for Space*

We can extend direct gradient ordination methods to account for space. In this case, we are doing a partial RDA, in which geographic distance is conditioned or “partialled out” before considering the environmental covariates. Note that it is not possible to pass the entire geographic matrix into a partial RDA; instead, we need to summarize the spatial structure in some way. Consequently, we create a spatial weights matrix from a Euclidean distance matrix. With this distance matrix, we can use a principal coordinates analysis on a truncated distance matrix to capture spatial structure, as we did in Chap. 6. In this case, we can use the `pcnm` function in the `vegan` package:

```
#Principal Components on Neighborhood Matrices
> pcnm.dist <- pcnm(dist.matrix)
```

The `pcnm` function defaults to setting the truncation to the minimum distance that provides a connected network using the minimum spanning tree of the distance matrix. We then fit a partial RDA, controlling for space.

```
#partial RDA
> rda.partial <- rda(species.matrix ~ canopy + elev + precip +
  Condition(scores(pcnm.dist, choices = 1:10)))
```

In the above model, we arbitrarily use the first ten PCNM axes from the PCNM analysis. We could more formally screen `pcnm` variables for their effects using the permutation test described above and then incorporate variables that explain variation in species occurrence:

```
> rda.distance <- rda(species.matrix ~ (scores(pcnm.dist)))
> rda.distance
> anova(rda.distance, by = 'axis', permutations = 200)
```

We find that the first two axes explain the majority of the spatial variation. We then re-fit the partial RDA with only those `pcnm` variables of importance. In this case, once we partial out spatial effects, the environmental covariates are still considered important based on permutation tests. This is consistent with results from the GDM showing that turnover could largely be explained by environmental, rather than geographic, variation.

To understand the spatial dependence in the residuals of community ordination models, we can use “multiscale ordination,” or a multivariate variogram on the residuals of a RDA (or CCA) model (Wagner 2004).

```
> mso(rda.bird, site.utm, grain = 1000, perm = 200)

#variogram based on residuals
> plot(mso.rda$vario$H, mso.rda$vario$CA)
```

In this function, `grain` refers to the bin size for distance classes, so here we choose a 1-km grain (Fig. 11.11b). This analysis finds evidence for spatial dependence at distances <3 km, consistent with what we found in Chap. 6, but it identifies a smaller range in spatial dependence than the correlogram based on raw dissimilarity.

We have already seen how space can be accounted for in GDMs. In addition, Chap. 6 shows how space could be accounted for in S-SDMs through the use of spatial regression models fit to each species. To date, spatial dependence for multivariate regression techniques has been less explored than for univariate spatial regression approaches. In principle, adding spatial weighting functions, such as eigenvector mapping, could be straightforward to implement. Other approaches covered in Chap. 6 may be more difficult to implement.

## 11.4 Next Steps and Advanced Issues

### 11.4.1 *Decomposition of Space–Environment Effects*

Isolating the role of space relative to environmental effects is important for understanding the mechanisms of community assembly and factors that may limit community structure. There has been a long tradition in some areas of community ecology to decompose the spatial effect from environmental effects using variance partitioning methods (Borcard et al. 1992; Peres-Neto et al. 2006). In this approach, several models are fit with and without key covariates (e.g., geographic distance in and out of the model) (Cushman and McGarigal 2002). By quantifying changes in the variance explained (or inertia explained in some modeling approaches, such as RDA), isolating the role of space and environmental effects can be quantified. While such approaches can be helpful, caution must be used because variance partitioning relies on some implicit assumptions, such as the assumption that interactions do not occur between covariates and that biologically these variables are conceptually and empirically independent. Variance partitioning can be implemented several ways in R; see the `varpart` function in the `vegan` package for one approach.

### 11.4.2 *Accounting for Dependence Among Species*

There is increasing interest in explicitly modeling the dependence among species in spatial modeling of communities, what is frequently referred to as joint species distribution models (Clark et al. 2014; Pollock et al. 2014; Warton et al. 2015a). Such dependence can arise for a variety of reasons, including effects of species interactions, phylogenetic dependence, and the fact that species may use similar environmental gradients, such that the distribution of one species may help predict the distribution of another. The generalized linear mixed model approach and its extensions are common ways to address these issues. In the above code, we provided the simplest implementation where we assumed that species came from a common distribution. However, more complex dependences can be addressed. In most of these situations, Bayesian hierarchical models are used to capture potential latent variables that may capture the species dependence in a formal way. See the `boral` and `HMSC` packages that can fit such models (Hui 2016; Ovaskainen et al. 2017).

### 11.4.3 *Spatial Networks*

Communities are often described as networks of interacting species (Bascompte 2007; Ings et al. 2009). In Chap. 9, we considered spatial networks, where nodes were patches and links were reflected movement or flow. In a community context,

nodes typically represent species and links represent interspecific interactions. The network approach to communities has revealed several insights into community structure and stability. In particular, a network approach can potentially capture indirect effects in communities, such as diffuse species interactions and indirect interactions, as well as providing a means to quantify and interpret emergent structure in communities. Increasingly, there has been interest in applying this general approach in space as a means to interpret metacommunity dynamics and related spatial issues (Araújo et al. 2011; Gonzalez et al. 2011). For example, spatial networks of species interactions have been used to determine spatial beta diversity (Poisot et al. 2012, 2017). Community networks are starting to be used to tackle conservation problems as well (Kaiser-Bunbury and Bluthgen 2015).

## 11.5 Conclusions

Spatial modeling of communities is an important and rapidly advancing topic in spatial ecology and conservation (Dray et al. 2012; Warton et al. 2015b; D'Amen et al. 2017). Much of this work has focused on predicting and mapping community structure, including species richness and beta diversity across space. This information is often used in conservation planning both locally and globally (Myers et al. 2000; Brooks et al. 2002; Wilson et al. 2006; Gray et al. 2016; Cardinale et al. 2018).

Despite these rapid advances, there are many challenges to understanding and modeling spatially structured communities. While there has been rapid growth in the theory of spatially structured communities (Gravel et al. 2011; Leibold and Chase 2017), often data are limited for interpreting how different factors govern community structure across space. The use of species co-occurrence data for these questions has a long history (e.g., Diamond 1975), but it can be unclear the extent to which co-occurrence can provide information on limiting factors, such as dispersal limitation and species interactions (Borthagaray et al. 2014; Freilich et al. 2018). Current challenges for spatial modeling of communities include better capturing dependencies among species, spatial dependence, and the potential for dispersal limitation to impact observed outcomes (Cumming et al. 2010; Rota et al. 2016; Clark et al. 2017; Tikhonov et al. 2017). Furthermore, better integration of community theory with empirical modeling of communities (e.g., Dorazio et al. 2010) is needed to interpret why communities assemble and better predict how they may change over time.

## References

- Algar AC, Kharouba HM, Young ER, Kerr JT (2009) Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods. *Ecography* 32(1):22–33. <https://doi.org/10.1111/j.1600-0587.2009.05832.x>

- Anderson M, Gorley RN, Clarke RK (2008) *Permanova+ for primer: Guide to software and statistical methods*. Primer-E Limited
- Anderson MJ, Crist TO, Chase JM, Vellend M, Inouye BD, Freestone AL, Sanders NJ, Cornell HV, Comita LS, Davies KF, Harrison SP, Kraft NJB, Stegen JC, Swenson NG (2011) Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist. *Ecol Lett* 14 (1):19–28. <https://doi.org/10.1111/j.1461-0248.2010.01552.x>
- Araújo MB, Rozenfeld A, Rahbek C, Marquet PA (2011) Using species co-occurrence networks to assess the impacts of climate change. *Ecography* 34(6):897–908. <https://doi.org/10.1111/j.1600-0587.2011.06919.x>
- Arrhenius O (1921) Species and area. *J Ecol* 9:95–99
- Bacaro G, Altobelli A, Cameletti M, Ciccarelli D, Martellos S, Palmer MW, Ricotta C, Rocchini D, Scheiner SM, Tordoni E, Chiarucci A (2016) Incorporating spatial autocorrelation in rarefaction methods: implications for ecologists and conservation biologists. *Ecol Indic* 69:233–238. <https://doi.org/10.1016/j.ecolind.2016.04.026>
- Barwell LJ, Isaac NJB, Kunin WE (2015) Measuring beta-diversity with species abundance data. *J Anim Ecol* 84(4):1112–1122. <https://doi.org/10.1111/1365-2656.12362>
- Bascompte J (2007) Networks in ecology. *Basic Appl Ecol* 8(6):485–490. <https://doi.org/10.1016/j.baae.2007.06.003>
- Baselga A (2010) Partitioning the turnover and nestedness components of beta diversity. *Glob Ecol Biogeogr* 19(1):134–143. <https://doi.org/10.1111/j.1466-8238.2009.00490.x>
- Baselga A, Orme CDL (2012) betapart: an R package for the study of beta diversity. *Methods Ecol Evol* 3(5):808–812. <https://doi.org/10.1111/j.2041-210X.2012.00224.x>
- Baskett ML, Micheli F, Levin SA (2007) Designing marine reserves for interacting species: insights from theory. *Biol Conserv* 137(2):163–179. <https://doi.org/10.1016/j.biocon.2007.02.013>
- Bates D, Machler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67(1):1–48
- Bjørnstad ON, Falck W (2001) Nonparametric spatial covariance functions: estimation and testing. *Environ Ecol Stat* 8(1):53–70. <https://doi.org/10.1023/a:1009601932481>
- Blanchet FG, Legendre P, Borcard D (2008) Modelling directional spatial processes in ecological data. *Ecol Model* 215(4):325–336. <https://doi.org/10.1016/j.ecolmodel.2008.04.001>
- Blanchet FG, Legendre P, Bergeron JAC, He FL (2014) Consensus RDA across dissimilarity coefficients for canonical ordination of community composition data. *Ecol Monogr* 84 (3):491–511. <https://doi.org/10.1890/13-0648.1>
- Borcard D, Legendre P (2012) Is the Mantel correlogram powerful enough to be useful in ecological analysis? A simulation study. *Ecology* 93(6):1473–1481
- Borcard D, Legendre P, Drapeau P (1992) Partialling out the spatial component of ecological variation. *Ecology* 73(3):1045–1055
- Borcard D, Gillet F, Legendre P (2011) *Numerical ecology with R*. Springer, New York
- Borthagaray AI, Arim M, Marquet PA (2014) Inferring species roles in metacommunity structure from species co-occurrence networks. *Proc R Soc B* 281(1792). <https://doi.org/10.1098/rspb.2014.1425>
- Bowman J, Cappuccino N, Fahrig L (2002) Patch size and population density: the effect of immigration behavior. *Conserv Ecol* 6(1):9
- Brooks TM, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Rylands AB, Konstant WR, Flick P, Pilgrim J, Oldfield S, Magin G, Hilton-Taylor C (2002) Habitat loss and extinction in the hotspots of biodiversity. *Conserv Biol* 16(4):909–923. <https://doi.org/10.1046/j.1523-1739.2002.00530.x>
- Brosi BJ, Armsworth PR, Daily GC (2008) Optimal design of agricultural landscapes for pollination services. *Conserv Lett* 1(1):27–36. <https://doi.org/10.1111/j.1755-263X.2008.00004.x>
- Brown JH, Kodric-Brown A (1977) Turnover rates in insular biogeography: effect of immigration on extinction. *Ecology* 58(2):445–449
- Cadotte MW, Tucker CM (2017) Should environmental filtering be abandoned? *Trends Ecol Evol* 32(6):429–437. <https://doi.org/10.1016/j.tree.2017.03.004>

- Cadotte MW, Cavender-Bares J, Tilman D, Oakley TH (2009) Using phylogenetic, functional and trait diversity to understand patterns of plant community productivity. *PLoS One* 4(5). <https://doi.org/10.1371/journal.pone.0005695>
- Caley MJ, Schluter D (1997) The relationship between local and regional diversity. *Ecology* 78(1):70–80
- Cameron AC, Trivedi PK (1990) Regression-based tests for overdispersion in the Poisson model. *J Econ* 46(3):347–364. [https://doi.org/10.1016/0304-4076\(90\)90014-k](https://doi.org/10.1016/0304-4076(90)90014-k)
- Cardinale BJ, Gonzalez A, Allington GRH, Loreau M (2018) Is local biodiversity declining or not? A summary of the debate over analysis of species richness time trends. *Biol Conserv* 219:175–183. <https://doi.org/10.1016/j.biocon.2017.12.021>
- Caswell H (1976) Community structure: neutral model analysis. *Ecol Monogr* 46(3):327–354. <https://doi.org/10.2307/1942257>
- Cavender-Bares J, Kozak KH, Fine PVA, Kembel SW (2009) The merging of community ecology and phylogenetic biology. *Ecol Lett* 12(7):693–715. <https://doi.org/10.1111/j.1461-0248.2009.01314.x>
- Chang CR, Lee PF, Bai ML, Lin TT (2004) Predicting the geographical distribution of plant communities in complex terrain - a case study in Fushian Experimental Forest, northeastern Taiwan. *Ecography* 27(5):577–588. <https://doi.org/10.1111/j.0906-7590.2004.03852.x>
- Clark JS, Gelfand AE, Woodall CW, Zhu K (2014) More than the sum of the parts: forest climate response from joint species distribution models. *Ecol Appl* 24(5):990–999. <https://doi.org/10.1890/13-1015.1>
- Clark JS, Nemergut D, Seyednasrollah B, Turner PJ, Zhang S (2017) Generalized joint attribute modeling for biodiversity analysis: median-zero, multivariate, multifarious data. *Ecol Monogr* 87(1):34–56. <https://doi.org/10.1002/ecm.1241>
- Coleman BD, Mares MA, Willig MR, Hsieh YH (1982) Randomness, area, and species richness. *Ecology* 63(4):1121–1133. <https://doi.org/10.2307/1937249>
- Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. *Phil Trans R Soc B* 345(1311):101–118. <https://doi.org/10.1098/rstb.1994.0091>
- Cook WM, Lane KT, Foster BL, Holt RD (2002) Island theory, matrix effects and species richness patterns in habitat fragments. *Ecol Lett* 5(5):619–623. <https://doi.org/10.1046/j.1461-0248.2002.00366.x>
- Cumming GS, Bodin O, Ernstson H, Elmqvist T (2010) Network analysis in conservation biogeography: challenges and opportunities. *Divers Distrib* 16(3):414–425. <https://doi.org/10.1111/j.1472-4642.2010.00651.x>
- Cushman SA, McGarigal K (2002) Hierarchical, multi-scale decomposition of species-environment relationships. *Landsc Ecol* 17(7):637–646. <https://doi.org/10.1023/a:1021571603605>
- D'Amen M, Dubuis A, Fernandes RF, Pottier J, Pellissier L, Guisan A (2015) Using species richness and functional traits predictions to constrain assemblage predictions from stacked species distribution models. *J Biogeogr* 42(7):1255–1266. <https://doi.org/10.1111/jbi.12485>
- D'Amen M, Rahbek C, Zimmermann NE, Guisan A (2017) Spatial predictions at the community level: from current approaches to future frameworks. *Biol Rev* 92(1):169–187. <https://doi.org/10.1111/brv.12222>
- Decker E, Linke S, Hermoso V, Geist J (2017) Incorporating ecological functions in conservation decision making. *Ecol Evol* 7(20):8273–8281. <https://doi.org/10.1002/ece3.3353>
- Devictor V, Mouillot D, Meynard C, Jiguet F, Thuiller W, Mouquet N (2010) Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. *Ecol Lett* 13(8):1030–1040. <https://doi.org/10.1111/j.1461-0248.2010.01493.x>
- Diamond JM (1975) Assembly of species communities. In: Cody ML, Diamond JM (eds) *Ecology and evolution of communities*. Harvard University Press, Cambridge, MA
- Distler T, Schuetz JG, Velasquez-Tibata J, Langham GM (2015) Stacked species distribution models and macroecological models provide congruent projections of avian species richness under climate change. *J Biogeogr* 42(5):976–988. <https://doi.org/10.1111/jbi.12479>



- Dixon P (2003) VEGAN, a package of R functions for community ecology. *J Veg Sci* 14 (6):927–930. <https://doi.org/10.1111/j.1654-1103.2003.tb02228.x>
- Dobson A, Lodge D, Alder J, Cumming GS, Keymer J, McGlade J, Mooney H, Rusak JA, Sala O, Wolters V, Wall D, Winfree R, Xenopoulos MA (2006) Habitat loss, trophic collapse, and the decline of ecosystem services. *Ecology* 87(8):1915–1924. [https://doi.org/10.1890/0012-9658\(2006\)87\[1915:hltcat\]2.0.co;2](https://doi.org/10.1890/0012-9658(2006)87[1915:hltcat]2.0.co;2)
- Dorazio RM, Connor EF (2014) Estimating abundances of interacting species using morphological traits, foraging guilds, and habitat. *PLoS One* 9(4). <https://doi.org/10.1371/journal.pone.0094323>
- Dorazio RM, Royle JA, Soderstrom B, Glimskar A (2006) Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology* 87(4):842–854. [https://doi.org/10.1890/0012-9658\(2006\)87\[842:esraab\]2.0.co;2](https://doi.org/10.1890/0012-9658(2006)87[842:esraab]2.0.co;2)
- Dorazio RM, Kery M, Royle JA, Plattner M (2010) Models for inference in dynamic metacommunity systems. *Ecology* 91(8):2466–2475. <https://doi.org/10.1890/09-1033.1>
- Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carre G, Garcia Marquez JR, Gruber B, Lafourcade B, Leitaó PJ, Muenkemueller T, McClean C, Osborne PE, Reineking B, Schroeder B, Skidmore AK, Zurell D, Lautenbach S (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36(1):27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Drakare S, Lennon JJ, Hillebrand H (2006) The imprint of the geographical, evolutionary and ecological context on species-area relationships. *Ecol Lett* 9(2):215–227. <https://doi.org/10.1111/j.1461-0248.2005.00848.x>
- Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol Model* 196(3–4):483–493. <https://doi.org/10.1016/j.ecolmodel.2006.02.015>
- Dray S, Pelissier R, Couteron P, Fortin MJ, Legendre P, Peres-Neto PR, Bellier E, Bivand R, Blanchet FG, De Caceres M, Dufour AB, Heegaard E, Jombart T, Munoz F, Oksanen J, Thioulouse J, Wagner HH (2012) Community ecology in the age of multivariate multiscale spatial analysis. *Ecol Monogr* 82(3):257–275. <https://doi.org/10.1890/11-1183.1>
- Driscoll DA (2008) The frequency of metapopulations, metacommunities and nestedness in a fragmented landscape. *Oikos* 117(2):297–309. <https://doi.org/10.1111/j.2007.0030-1299.16202.x>
- Dubuis A, Pottier J, Rion V, Pellissier L, Theurillat JP, Guisan A (2011) Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Divers Distrib* 17(6):1122–1131. <https://doi.org/10.1111/j.1472-4642.2011.00792.x>
- Economu EP, Keitt TH (2008) Species diversity in neutral metacommunities: a network approach. *Ecol Lett* 11(1):52–62. <https://doi.org/10.1111/j.1461-0248.2007.01126.x>
- Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size: implications for plant conservation. *Annu Rev Ecol Syst* 24:217–242. <https://doi.org/10.1146/annurev.es.24.110193.001245>
- Fahrig L (2013) Rethinking patch size and isolation effects: the habitat amount hypothesis. *J Biogeogr* 40(9):1649–1663. <https://doi.org/10.1111/jbi.12130>
- Faith DP, Carter G, Cassis G, Ferrier S, Wilkie L (2003) Complementarity, biodiversity viability analysis, and policy-based algorithms for conservation. *Environ Sci Pol* 6(3):311–328. [https://doi.org/10.1016/s1462-9011\(03\)00044-3](https://doi.org/10.1016/s1462-9011(03)00044-3)
- Fernandez-Juricic E (2002) Can human disturbance promote nestedness? A case study with breeding birds in urban habitat fragments. *Oecologia* 131(2):269–278. <https://doi.org/10.1007/s00442-002-0883-y>
- Ferrier S, Guisan A (2006) Spatial modelling of biodiversity at the community level. *J Appl Ecol* 43 (3):393–404. <https://doi.org/10.1111/j.1365-2664.2006.01149.x>
- Ferrier S, Pressey RL, Barrett TW (2000) A new predictor of the irreplaceability of areas for achieving a conservation goal, its application to real-world planning, and a research agenda for

- further refinement. *Biol Conserv* 93(3):303–325. [https://doi.org/10.1016/s0006-3207\(99\)00149-4](https://doi.org/10.1016/s0006-3207(99)00149-4)
- Ferrier S, Manion G, Elith J, Richardson K (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Divers Distrib* 13(3):252–264. <https://doi.org/10.1111/j.1472-4642.2007.00341.x>
- Fitzpatrick MC, Keller SR (2015) Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecol Lett* 18(1):1–16. <https://doi.org/10.1111/ele.12376>
- Freilich MA, Wieters E, Broitman BR, Marquet PA, Navarrete SA (2018) Species co-occurrence networks: can they reveal trophic and non-trophic interactions in ecological communities? *Ecology* 99(3):690–699. <https://doi.org/10.1002/ecy.2142>
- Fukami T (2015) Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annu Rev Ecol Evol Syst* 46:1–23. <https://doi.org/10.1146/annurev-ecolsys-110411-160340>
- Gonzalez A (2000) Community relaxation in fragmented landscapes: the relation between species richness, area and age. *Ecol Lett* 3(5):441–448. <https://doi.org/10.1046/j.1461-0248.2000.00171.x>
- Gonzalez A, Rayfield B, Lindo Z (2011) The disentangled bank: how loss of habitat fragments and disassembles ecological networks. *Am J Bot* 98(3):503–516. <https://doi.org/10.3732/ajb.1000424>
- Gotelli NJ, Colwell RK (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol Lett* 4(4):379–391. <https://doi.org/10.1046/j.1461-0248.2001.00230.x>
- Gravel D, Massol F, Canard E, Mouillot D, Mouquet N (2011) Trophic theory of island biogeography. *Ecol Lett* 14(10):1010–1016. <https://doi.org/10.1111/j.1461-0248.2011.01667.x>
- Gray CL, Hill SLL, Newbold T, Hudson LN, Borger L, Contu S, Hoskins AJ, Ferrier S, Purvis A, Scharlemann JPW (2016) Local biodiversity is higher inside than outside terrestrial protected areas worldwide. *Nat Commun* 7:12306. <https://doi.org/10.1038/ncomms12306>
- Guichard F (2017) Recent advances in metacommunities and meta-ecosystem theories. *F1000 Res* 6:610
- Guillot G, Rousset F (2013) Dismantling the Mantel tests. *Methods Ecol Evol* 4(4):336–344. <https://doi.org/10.1111/2041-210x.12018>
- Guisan A, Rahbek C (2011) SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *J Biogeogr* 38(8):1433–1444. <https://doi.org/10.1111/j.1365-2699.2011.02550.x>
- Guisan A, Weiss SB, Weiss AD (1999) GLM versus CCA spatial modeling of plant species distribution. *Plant Ecol* 143(1):107–122
- Haila Y (2002) A conceptual genealogy of fragmentation research: from island biogeography to landscape ecology. *Ecol Appl* 12(2):321–334. <https://doi.org/10.2307/3060944>
- Hastie T, Fithian W (2013) Inference from presence-only data; the ongoing controversy. *Ecography* 36(8):864–867. <https://doi.org/10.1111/j.1600-0587.2013.00321.x>
- Hoef JMV, Boveng PL (2007) Quasi-poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology* 88(11):2766–2772. <https://doi.org/10.1890/07-0043.1>
- Holt RD (1992) A neglected facet of island biogeography: the role of internal spatial dynamics in area effects. *Theor Popul Biol* 41(3):354–371. [https://doi.org/10.1016/0040-5809\(92\)90034-q](https://doi.org/10.1016/0040-5809(92)90034-q)
- Holyoak M, Leibold MA, Holt RD (2005) *Metacommunities: spatial dynamics and ecological communities*. University of Chicago Press, Chicago
- Hubbell SP (2001) *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, NJ
- Huffaker CB (1958) Experimental studies on predation: dispersion factors and predator-prey oscillations. *Hilgardia* 27:343–383

- Hui FKC (2016) Boral - Bayesian ordination and regression analysis of multivariate abundance data in r. *Methods Ecol Evol* 7(6):744–750. <https://doi.org/10.1111/2041-210x.12514>
- Hutto RL, Young JS (2002) Regional landbird monitoring: perspectives from the Northern Rocky Mountains. *Wildl Soc Bull* 30(3):738–750
- Ikcnayan KJ, Tingley MW, Furnas BJ, Beissinger SR (2014) Detecting diversity: emerging methods to estimate species diversity. *Trends Ecol Evol* 29(2):97–106. <https://doi.org/10.1016/j.tree.2013.10.012>
- Ings TC, Montoya JM, Bascompte J, Bluthgen N, Brown L, Dormann CF, Edwards F, Figueroa D, Jacob U, Jones JI, Lauridsen RB, Ledger ME, Lewis HM, Olesen JM, van Veen FJF, Warren PH, Woodward G (2009) Ecological networks - beyond food webs. *J Anim Ecol* 78(1):253–269. <https://doi.org/10.1111/j.1365-2656.2008.01460.x>
- Jewitt D, Goodman PS, O'Connor TG, Erasmus BFN, Witkowski ETF (2016) Mapping landscape beta diversity of plants across KwaZulu-Natal, South Africa, for aiding conservation planning. *Biodivers Conserv* 25(13):2641–2654. <https://doi.org/10.1007/s10531-016-1190-y>
- Kaiser-Bunbury CN, Bluthgen N (2015) Integrating network ecology with applied conservation: a synthesis and guide to implementation. *Aob Plants* 7. <https://doi.org/10.1093/aobpla/plv076>
- Karger DN, Cord AF, Kessler M, Kreft H, Kuehn I, Pompe S, Sandel B, Cabral JS, Smith AB, Svenning JC, Tuomisto H, Weigelt P, Wesche K (2016) Delineating probabilistic species pools in ecology and biogeography. *Glob Ecol Biogeogr* 25(4):489–501. <https://doi.org/10.1111/geb.12422>
- Karp DS, Rominger AJ, Zook J, Ranganathan J, Ehrlich PR, Daily GC, Cornell H (2012) Intensive agriculture erodes beta-diversity at large scales. *Ecol Lett* 15(9):963–970. <https://doi.org/10.1111/j.1461-0248.2012.01815.x>
- Keddy PA (1992) Assembly and response rules: two goals for predictive community ecology. *J Veg Sci* 3(2):157–164. <https://doi.org/10.2307/3235676>
- Keller LF, Waller DM (2002) Inbreeding effects in wild populations. *Trends Ecol Evol* 17(5):230–241. [https://doi.org/10.1016/s0169-5347\(02\)02489-8](https://doi.org/10.1016/s0169-5347(02)02489-8)
- Kerr JT, Sugar A, Packer L (2000) Indicator taxa, rapid biodiversity assessment, and nestedness in an endangered ecosystem. *Conserv Biol* 14(6):1726–1734. <https://doi.org/10.1046/j.1523-1739.2000.99275.x>
- Kery M, Royle JA (2016) Applied hierarchical modeling in ecology: analysis of distribution, abundance and species richness in R and BUGS. Academic, San Diego, CA
- Kleiber C, Zeileis A (2008) Applied econometrics with R. Springer, New York
- Koleff P, Gaston KJ (2002) The relationships between local and regional species richness and spatial turnover. *Glob Ecol Biogeogr* 11(5):363–375. <https://doi.org/10.1046/j.1466-822x.2002.00302.x>
- Koleff P, Gaston KJ, Lennon JJ (2003) Measuring beta diversity for presence-absence data. *J Anim Ecol* 72(3):367–382. <https://doi.org/10.1046/j.1365-2656.2003.00710.x>
- Kraft NJB, Adler PB, Godoy O, James EC, Fuller S, Levine JM (2015) Community assembly, coexistence and the environmental filtering metaphor. *Funct Ecol* 29(5):592–599. <https://doi.org/10.1111/1365-2435.12345>
- Kukkala AS, Moilanen A (2013) Core concepts of spatial prioritisation in systematic conservation planning. *Biol Rev* 88(2):443–464. <https://doi.org/10.1111/brv.12008>
- Laurance WF (2008) Theory meets reality: how habitat fragmentation research has transcended island biogeographic theory. *Biol Conserv* 141(7):1731–1744. <https://doi.org/10.1016/j.biocon.2008.05.011>
- Lawton JH (1999) Are there general laws in ecology? *Oikos* 84(2):177–192. <https://doi.org/10.2307/3546712>
- Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology* 74(6):1659–1673
- Legendre P, Anderson MJ (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecol Monogr* 69(1):1–24. [https://doi.org/10.1890/0012-9615\(1999\)069\[0001:dbratm\]2.0.co;2](https://doi.org/10.1890/0012-9615(1999)069[0001:dbratm]2.0.co;2)
- Legendre P, Legendre L (1998) Numerical ecology. Elsevier, Amsterdam

- Legendre P, Legendre L (2012) *Numerical ecology*, 3rd edn. Elsevier, Amsterdam
- Legendre P, Borcard D, Peres-Neto PR (2005) Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecol Monogr* 75(4):435–450. <https://doi.org/10.1890/05-0549>
- Legendre P, Fortin M-J, Borcard D (2015) Should the Mantel test be used in spatial analysis? *Methods Ecol Evol* 6(11):1239–1247. <https://doi.org/10.1111/2041-210x.12425>
- Leibold MA, Chase JM (2017) *Metacommunity ecology*. Princeton University Press, Princeton, NJ
- Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, Holt RD, Shurin JB, Law R, Tilman D, Loreau M, Gonzalez A (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecol Lett* 7(7):601–613. <https://doi.org/10.1111/j.1461-0248.2004.00608.x>
- Lessard JP, Belmaker J, Myers JA, Chase JM, Rabhek C (2012) Inferring local ecological processes amid species pool influences. *Trends Ecol Evol* 27(11):600–607. <https://doi.org/10.1016/j.tree.2012.07.006>
- Levins R, Culver D (1971) Regional coexistence of species and competition between rare species. *Proc Natl Acad Sci U S A* 68(6):1246. <https://doi.org/10.1073/pnas.68.6.1246>
- Lichstein JW (2007) Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecol* 188(2):117–131. <https://doi.org/10.1007/s11258-006-9126-3>
- Liu CR, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28(3):385–393
- Liu C, White M, Newell G (2013) Selecting thresholds for the prediction of species occurrence with presence-only data. *J Biogeogr* 40(4):778–789. <https://doi.org/10.1111/jbi.12058>
- Lomolino MV (1990) The target area hypothesis: the influence of island area on immigration rates of non-volant mammals. *Oikos* 57(3):297–300. <https://doi.org/10.2307/3565957>
- Lowe WH, McPeck MA (2014) Is dispersal neutral? *Trends Ecol Evol* 29(8):444–450. <https://doi.org/10.1016/j.tree.2014.05.009>
- Mac Nally R, Lake PS (1999) On the generation of diversity in archipelagos: a re-evaluation of the Quinn-Harrison ‘saturation index’. *J Biogeogr* 26(2):285–295. <https://doi.org/10.1046/j.1365-2699.1999.00268.x>
- MacArthur RH (1972) *Geographical ecology: patterns in the distribution of species*. Princeton University Press, Princeton, NJ
- MacArthur RH, Wilson EO (1963) Equilibrium theory of insular zoogeography. *Evolution* 17(4):373. <https://doi.org/10.2307/2407089>
- MacArthur RH, Wilson EO (1967) *The theory of island biogeography*. Princeton University Press, Princeton, NJ
- MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83(8):2248–2255
- Magurran AE (2003) *Measuring biological diversity*. Wiley, Chichester
- Magurran AE (2004) *Measuring biological diversity*. Blackwell Science Ltd, Malden, MA
- Manion G, Lisk M, Ferrier S, Nieto-Lugilde D, Mokany K, Fitzpatrick MC (2018) gdm: generalized dissimilarity modeling. R package 1.3.11
- Margules CR, Pressey RL (2000) Systematic conservation planning. *Nature* 405(6783):243–253. <https://doi.org/10.1038/35012251>
- Mateo RG, Felicísimo AM, Pottier J, Guisan A, Muñoz J (2012) Do stacked species distribution models reflect altitudinal diversity patterns? *PLoS One* 7(3). <https://doi.org/10.1371/journal.pone.0032586>
- McPeck MA, Brown JM (2000) Building a regional species pool: diversification of the Enallagma damselflies in eastern North America. *Ecology* 81(4):904–920
- Milanovich JR, Peterman WE, Barrett K, Hopton ME (2012) Do species distribution models predict species richness in urban and natural green spaces? A case study using amphibians. *Landsc Urban Plan* 107(4):409–418. <https://doi.org/10.1016/j.landurbplan.2012.07.010>
- Mittelbach GG (2012) *Community ecology*. Sinauer Associates, Sunderland, MA

- Moilanen A, Wilson KA, Possingham H (eds) (2009) Spatial conservation prioritization: quantitative methods and computational tools. Oxford University Press, Oxford
- Mouchet MA, Villegger S, Mason NWH, Mouillot D (2010) Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. *Funct Ecol* 24(4):867–876. <https://doi.org/10.1111/j.1365-2435.2010.01695.x>
- Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* 403(6772):853–858. <https://doi.org/10.1038/35002501>
- Neuhauser C (1998) Habitat destruction and competitive coexistence in spatially explicit models with local interactions. *J Theor Biol* 193(3):445–463
- Newbold T, Gilbert F, Zalat S, El-Gabbas A, Reader T (2009) Climate-based models of spatial patterns of species richness in Egypt's butterfly and mammal fauna. *J Biogeogr* 36(11):2085–2095. <https://doi.org/10.1111/j.1365-2699.2009.02140.x>
- Nichols JD, Boulinier T, Hines JE, Pollock KH, Sauer JR (1998) Inference methods for spatial variation in species richness and community composition when not all species are detected. *Conserv Biol* 12(6):1390–1398
- Noss RF (1990) Indicators for monitoring biodiversity: a hierarchical approach. *Conserv Biol* 4(4):355–364
- O'Neill RV, Johnson AR, King AW (1989) A hierarchical framework for the analysis of scale. *Landsc Ecol* 3(3–4):193–205. <https://doi.org/10.1007/bf00131538>
- Olden JD, Rooney TP (2006) On defining and quantifying biotic homogenization. *Glob Ecol Biogeogr* 15(2):113–120. <https://doi.org/10.1111/j.1466-822x.2006.00214.x>
- Orme CDL, Davies RG, Burgess M, Eigenbrod F, Pickup N, Olson VA, Webster AJ, Ding TS, Rasmussen PC, Ridgely RS, Stattersfield AJ, Bennett PM, Blackburn TM, Gaston KJ, Owens IPF (2005) Global hotspots of species richness are not congruent with endemism or threat. *Nature* 436(7053):1016–1019. <https://doi.org/10.1038/nature03850>
- Ovaskainen O, Soinen J (2011) Making more out of sparse data: hierarchical modeling of species communities. *Ecology* 92(2):289–295. <https://doi.org/10.1890/10-1251.1>
- Ovaskainen O, Hottola J, Siitonen J (2010) Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology* 91(9):2514–2521. <https://doi.org/10.1890/10-0173.1>
- Ovaskainen O, Abrego N, Halme P, Dunson D (2016a) Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods Ecol Evol* 7(5):549–555. <https://doi.org/10.1111/2041-210x.12501>
- Ovaskainen O, Roy DB, Fox R, Anderson BJ (2016b) Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. *Methods Ecol Evol* 7(4):428–436. <https://doi.org/10.1111/2041-210x.12502>
- Ovaskainen O, Tikhonov G, Norberg A, Blanchet FG, Duan L, Dunson D, Roslin T, Abrego N (2017) How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecol Lett* 20(5):561–576. <https://doi.org/10.1111/ele.12757>
- Palmer MW (1990) The estimation of species richness by extrapolation. *Ecology* 71(3):1195–1198. <https://doi.org/10.2307/1937387>
- Palmer MW (1993) Putting things in even better order: the advantages of canonical correspondence analysis. *Ecology* 74(8):2215–2230. <https://doi.org/10.2307/1939575>
- Partel M, Zobel M, Zobel K, vanderMaarel E (1996) The species pool and its relation to species richness: evidence from Estonian plant communities. *Oikos* 75(1):111–117. <https://doi.org/10.2307/3546327>
- Pearson RG, Dawson TP, Liu C (2004) Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography* 27(3):285–298. <https://doi.org/10.1111/j.0906-7590.2004.03740.x>
- Peres-Neto PR, Legendre P, Dray S, Borcard D (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* 87(10):2614–2625. [https://doi.org/10.1890/0012-9658\(2006\)87\[2614:vposdm\]2.0.co;2](https://doi.org/10.1890/0012-9658(2006)87[2614:vposdm]2.0.co;2)

- Petchey OL, Gaston KJ (2002) Functional diversity (FD), species richness and community composition. *Ecol Lett* 5(3):402–411. <https://doi.org/10.1046/j.1461-0248.2002.00339.x>
- Pineda E, Lobo JM (2009) Assessing the accuracy of species distribution models to predict amphibian species richness patterns. *J Anim Ecol* 78(1):182–190. <https://doi.org/10.1111/j.1365-2656.2008.01471.x>
- Pineda E, Lobo JM (2012) The performance of range maps and species distribution models representing the geographic variation of species richness at different resolutions. *Glob Ecol Biogeogr* 21(9):935–944. <https://doi.org/10.1111/j.1466-8238.2011.00741.x>
- Podani J, Schmera D (2011) A new conceptual and methodological framework for exploring and explaining pattern in presence-absence data. *Oikos* 120(11):1625–1638. <https://doi.org/10.1111/j.1600-0706.2011.19451.x>
- Poff NL (1997) Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. *J N Am Benthol Soc* 16(2):391–409. <https://doi.org/10.2307/1468026>
- Poisot T, Canard E, Mouillot D, Mouquet N, Gravel D (2012) The dissimilarity of species interaction networks. *Ecol Lett* 15(12):1353–1361. <https://doi.org/10.1111/ele.12002>
- Poisot T, Gueveneux-Julien C, Fortin MJ, Gravel D, Legendre P (2017) Hosts, parasites and their interactions respond to different climatic variables. *Glob Ecol Biogeogr* 26(8):942–951. <https://doi.org/10.1111/geb.12602>
- Pollock LJ, Tingley R, Morris WK, Golding N, O'Hara RB, Parris KM, Vesik PA, McCarthy MA (2014) Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods Ecol Evol* 5(5):397–406. <https://doi.org/10.1111/2041-210x.12180>
- Preston FW (1962) The canonical distribution of commonness and rarity: part I. *Ecology* 43(2):185–215, 431–432
- Rahbek C, Graves GR (2001) Multiscale assessment of patterns of avian species richness. *Proc Natl Acad Sci U S A* 98(8):4534–4539. <https://doi.org/10.1073/pnas.071034898>
- Rands MRW, Adams WM, Bennun L, Butchart SHM, Clements A, Coomes D, Entwistle A, Hodge I, Kapos V, Scharlemann JPW, Sutherland WJ, Vira B (2010) Biodiversity conservation: challenges beyond 2010. *Science* 329(5997):1298–1303. <https://doi.org/10.1126/science.1189138>
- Ricklefs RE (1987) Community diversity: relative roles of local and regional processes. *Science* 235(4785):167–171. <https://doi.org/10.1126/science.235.4785.167>
- Rose PM, Kennard MJ, Sheldon F, Moffatt DB, Butler GL (2016) A data-driven method for selecting candidate reference sites for stream bioassessment programs using generalised dissimilarity models. *Mar Freshw Res* 67(4):440–454. <https://doi.org/10.1071/mf14254>
- Rota CT, Wikle CK, Kays RW, Forrester TD, McShea WJ, Parsons AW, Millsbaugh JJ (2016) A two-species occupancy model accommodating simultaneous spatial and interspecific dependence. *Ecology* 97(1):48–53. <https://doi.org/10.1890/15-1193.1>
- Royle JA, Dorazio RM (2008) Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations, and communities. Academic, San Diego, CA
- Scheiner SM (2003) Six types of species-area curves. *Glob Ecol Biogeogr* 12(6):441–447. <https://doi.org/10.1046/j.1466-822X.2003.00061.x>
- Seabloom EW, Dobson AP, Stoms DM (2002) Extinction rates under nonrandom patterns of habitat loss. *Proc Natl Acad Sci U S A* 99(17):11229–11234. <https://doi.org/10.1073/pnas.162064899>
- Socolar JB, Gilroy JJ, Kunin WE, Edwards DP (2016) How should beta-diversity inform biodiversity conservation? *Trends Ecol Evol* 31(1):67–80. <https://doi.org/10.1016/j.tree.2015.11.005>
- Soininen J, McDonald R, Hillebrand H (2007) The distance decay of similarity in ecological communities. *Ecography* 30(1):3–12. <https://doi.org/10.1111/j.2006.0906-7590.04817.x>
- Ter Braak CJF (1987) The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio* 69(1–3):69–77. <https://doi.org/10.1007/bf00038688>
- Thomas CD, Cameron A, Green RE, Bakkenes M, Beaumont LJ, Collingham YC, Erasmus BFN, de Siqueira MF, Grainger A, Hannah L, Hughes L, Huntley B, van Jaarsveld AS, Midgley GF,

- Miles L, Ortega-Huerta MA, Peterson AT, Phillips OL, Williams SE (2004) Extinction risk from climate change. *Nature* 427(6970):145–148. <https://doi.org/10.1038/nature02121>
- Thomassen HA, Fuller T, Buermann W, Mila B, Kieswetter CM, Jarrin P, Cameron SE, Mason E, Schweizer R, Schlunegger J, Chan J, Wang O, Peralvo M, Schneider CJ, Graham CH, Pollinger JP, Saatchi S, Wayne RK, Smith TB (2011) Mapping evolutionary process: a multi-taxa approach to conservation prioritization. *Evol Appl* 4(2):397–413. <https://doi.org/10.1111/j.1752-4571.2010.00172.x>
- Tikhonov G, Abrego N, Dunson D, Ovaskainen O (2017) Using joint species distribution models for evaluating how species-to-species associations depend on the environmental context. *Methods Ecol Evol* 8(4):443–452. <https://doi.org/10.1111/2041-210X.12723>
- Tilman D, May RM, Lehman CL, Nowak MA (1994) Habitat destruction and the extinction debt. *Nature* 371(6492):65–66
- Urban DL, Oneill RV, Shugart HH (1987) Landscape ecology. *Bioscience* 37(2):119–127
- Urban D, Goslee S, Pierce K, Lookingbill T (2002) Extending community ecology to landscapes. *Ecoscience* 9(2):200–212
- van der Valk AG (1981) Succession in Wetlands: A Gleasonian Approach. *Ecology* 62 (3):688–696
- Vellend M (2010) Conceptual synthesis in community ecology. *Q Rev Biol* 85(2):183–206. <https://doi.org/10.1086/652373>
- Venables WN, Ripley BD (2002) Modern applied statistics with S, 4th edn. Springer, New York
- Vilhena DA, Antonelli A (2015) A network approach for identifying and delimiting biogeographical regions. *Nat Commun* 6. <https://doi.org/10.1038/ncomms7848>
- Wagner HH (2003) Spatial covariance in plant communities: integrating ordination, geostatistics, and variance testing. *Ecology* 84(4):1045–1057. [https://doi.org/10.1890/0012-9658\(2003\)084\[1045:scipcj\]2.0.co;2](https://doi.org/10.1890/0012-9658(2003)084[1045:scipcj]2.0.co;2)
- Wagner HH (2004) Direct multi-scale ordination with canonical correspondence analysis. *Ecology* 85(2):342–351. <https://doi.org/10.1890/02-0738>
- Wang Y, Naumann U, Wright ST, Warton DI (2012) mvabund - an R package for model-based analysis of multivariate abundance data. *Methods Ecol Evol* 3(3):471–474. <https://doi.org/10.1111/j.2041-210X.2012.00190.x>
- Wang Y, Naumann U, Eddelbuettel D, Warton D (2018) mvabund: statistical methods for analysing multivariate abundance data. R version 3.13.1
- Warton DI, Wright ST, Wang Y (2012) Distance-based multivariate analyses confound location and dispersion effects. *Methods Ecol Evol* 3(1):89–101. <https://doi.org/10.1111/j.2041-210X.2011.00127.x>
- Warton DI, Blanchet FG, O'Hara RB, Ovaskainen O, Taskinen S, Walker SC, Hui FKC (2015a) So many variables: joint modeling in community ecology. *Trends Ecol Evol* 30(12):766–779. <https://doi.org/10.1016/j.tree.2015.09.007>
- Warton DI, Foster SD, De'ath G, Stoklosa J, Dunstan PK (2015b) Model-based thinking for community ecology. *Plant Ecol* 216(5):669–682. <https://doi.org/10.1007/s11258-014-0366-3>
- Whitcomb RF, Lynch JF, Opler PA, Robbins CS (1976) Island biogeography and conservation: strategy and limitations. *Science* 193(4257):1030–1032
- Whittaker RJ, Araújo MB, Paul J, Ladle RJ, Watson JEM, Willis KJ (2005) Conservation biogeography: assessment and prospect. *Divers Distrib* 11(1):3–23. <https://doi.org/10.1111/j.1366-9516.2005.00143.x>
- Wickham H (2007) Reshaping data with the reshape package. *J Stat Softw* 21(12):1–20
- Williams PH (1996) Mapping variations in the strength and breadth of biogeographic transition zones using species turnover. *Proc R Soc B* 263(1370):579–588. <https://doi.org/10.1098/rspb.1996.0087>
- Wilson DS (1992) Complex interactions in metacommunities, with implications for biodiversity and higher levels of selection. *Ecology* 73(6):1984–2000. <https://doi.org/10.2307/1941449>
- Wilson KA, McBride MF, Bode M, Possingham HP (2006) Prioritizing global conservation efforts. *Nature* 440(7082):337–340. <https://doi.org/10.1038/nature04366>

- Winfree R, Aguilar R, Vazquez DP, LeBuhn G, Aizen MA (2009) A meta-analysis of bees' responses to anthropogenic disturbance. *Ecology* 90(8):2068–2076
- Wright DH, Reeves JH (1992) On the meaning and measurement of nestedness of species assemblages. *Oecologia* 92(3):416–428. <https://doi.org/10.1007/bf00317469>
- Wright DH, Patterson BD, Mikkelsen GM, Cutler A, Atmar W (1998) A comparative analysis of nested subset patterns of species composition. *Oecologia* 113(1):1–20
- Yackulic CB, Chandler R, Zipkin EF, Royle JA, Nichols JD, Grant EHC, Veran S (2013) Presence-only modelling using MAXENT: when can we trust the inferences? *Methods Ecol Evol* 4(3):236–243. <https://doi.org/10.1111/2041-210x.12004>
- Zeileis A, Kleiber C, Jackman S (2008) Regression models for count data in R. *J Stat Softw* 27(8):1–25. <https://doi.org/10.18637/jss.v027.i08>
- Zhang MG, Slik JWF, Ma KP (2016) Using species distribution modeling to delineate the botanical richness patterns and phytogeographical regions of China. *Sci Rep* 6. <https://doi.org/10.1038/srep22400>