American Journal of Botany

American Journal of Botany 98(3): 481–502. 2011.

LINKING BIODIVERSITY PATTERNS BY AUTOCORRELATED RANDOM SAMPLING[1]

BRIAN J. MCGILL[2]

School of Biology and Ecology, Sustainability Solutions Initiative, Deering 303, University of Maine, Orono, Maine 04469 USA

Biodiversity macroecology deals with the commonly measured variables of abundance, distribution, occupancy, and range size across two scales: the local (or α) and regional (γ). There are ca. 15 patterns consisting of the frequency distributions of the variables, variables as a function of area or sample size, and interrelationships between variables that appear to be very general if not close to universal. A number of links can be drawn between these patterns. In particular, I show that local communities can be seen as random samples of the regional pool, but only as a special form of sampling that is autocorrelated due to the spatial clumping of individuals within a species. I describe two distinct sets of mathematical machinery that can start with the regional species abundance distribution and then predict local species richness, local species abundance distributions, and β-diversity (in the form of species area relationships or decay of similarity with distance). I conclude by examining some of the implications of the fact that biodiversity patterns are linked by autocorrelated sampling.

Key words: biodiversity; local; macroecology; regional; species abundance distribution; species–area relationship.

There are many accounts of the stages through which science proceeds. The traditional teaching of the scientific method emphasizes hypothesis then test or multiple hypotheses and test in the modern hypothetico-deductive approach (Platt, 1964). Historians of physics note the progression from data to pattern to theory such as found in the chain from Brahe to Kepler to Newton. Wiegert (1988) suggests that we proceed from discovery of new facts, to organization of facts, to mechanism.

But another account of scientific progress with some ties to Wiegert's organization of facts emphasizes the discovery of the links between patterns to create unified theories or theories that make many predictions from few assumptions (e.g., Lakatos, 1978). Indeed, the seminal achievement of Newton was having one theory (inverse square law of gravitation) that explained the motion of the planets and the motion of bodies on earth. Mendeleev's periodic table, the quark theory making sense of a bestiary of different subatomic particles, and Pasteur's germ theory of disease all fit this paradigm of linking together previously disparate phenomenon.

The study of biodiversity broadly defined to include the study of the abundance and distribution of many species appears to have hit this phase of linking and seeking to make sense of many patterns. For much of the 20th century, ecologists measured and provided quantitative patterns and rules that described biodiversity. In 1989, a distinct discipline around these patterns was identified and named (Brown and Maurer, 1989). In the last decade of the 20th century, several authors wrote books summarizing these patterns and began to seek the underlying links (Brown, 1995; Rosenzweig, 1995; Gaston and Blackburn, 2000). Then in the first decade of the 21st century, a number of theories emerged that unified the different patterns of biodiversity. These theories each invoked one central driving force or mechanism, including stochastic drift of populations (Bell, 2000; Hubbell, 2001), occupancy (He and Legendre, 2002), Gaussian abundance across space (McGill and Collins, 2003), generalized fractals (Storch et al., 2008), and maximum entropy (Harte et al., 2008). From this central assumption, each paper then proceeded to show that it can explain a number of biodiversity patterns such as species abundance distributions and species–area relationships.

In a recent paper (McGill, 2010a), I noted that despite the fact all of these theories started from very different assumptions, they shared some common features including an emphasis on the clustering of individuals within a species and the highly unequal distribution of abundance between species, and I asserted that these few assumptions were what actually produced the links to other patterns such as the species abundance distribution. But I did not provide details of how. Here I go one step further and seek to eliminate the primary assumptions (such as neutral drift) and proceed directly from the two core assumptions of intraspecific clustering of individuals and highly uneven distribution of abundance to explain other patterns such as the species abundance distribution, the species–area relationship and the decay of similarity with distance. Unlike my previous paper (McGill, 2010a), I provide explicit mathematical formulas. The math that links these patterns together is sampling, albeit not traditional independent samples, but rather correlated samples due to the spatial clustering of individuals. Many will find it disappointing that one can demonstrate that cherished biodiversity patterns are explained by sampling

---

[1] Manuscript received 17 December 2010; revision accepted 1 February 2011.

First and foremost, I thank the professionals and volunteers that collected the data that made this paper possible. Large-scale ecological studies such as found in this paper would be impossible without them and their funders. In particular, the USFIA data were sponsored by the U. S. Forest Service, and the BCI data were sponsored by NSF grants to Stephen Hubbell, the Center for Tropical Forest Science, the Smithsonian Tropical Research Institute, the John D. and Catherine T. MacArthur Foundation, the Mellon Foundation, and the Small World Institute Fund. I also acknowledge NSF grant EPS-0904155 to Maine EPSCoR at the University of Maine for funding and thank the editors of the special issue for inviting me to write this paper. Finally, these ideas have been improved by discussions with many, many people but especially Brian Enquist, John Harte, Fangliang He, Nick Isaac, Julie Messier, Jeff Nekola, David Storch, and Ethan White.

[2] Author for correspondence (e-mail: mail@brianmcgill.org)

doi:10.3732/ajb.1000509

(McGill and Nekola, 2010). However, I argue at the end of this paper that this fact simply provides a touchstone for determining which aspects of biodiversity patterns contain biology and are likely truly unique to ecology vs. those aspects that are probably quite general and apply to colored marbles, atoms, and any other entities that can be imagined to be sampled.

In the rest of this paper, I will (1) set the stage by describing the ideas of local community vs. regional pool, define the variables of interest in biodiversity, and give examples of two different kinds of data collected in biodiversity theory; (2) list the biodiversity patterns ecologists have identified and found to be general; and (3) review work describing how sampling theory links together and explains biodiversity patterns.

SETTING THE STAGE

The scales involved—One thing all the recent unified theories have in common is a distinction between two scales: the regional pool and the local community. For much of the latter half of the 20th century, the focus was squarely on the local community (MacArthur, 1968; Tilman, 1988) (however imprecise that term may be: Fauth et al., 1996). Near the end of that period though, there were growing calls for focus on the regional pool (Ricklefs, 1987; Ricklefs and Schluter, 1993). The first decade of the 21st century has seen a growing and fruitful emphasis on the interplay between the local and the regional processes (Loreau, 2000; Hubbell, 2001; McGill, 2003; Zillio and Condit, 2007). Four decades ago, Whittaker (1972, 1975) foreshadowed this development specifically in the area of biodiversity by making a fundamental distinction between different kinds of biodiversity according to their spatial scale and denoted them from smallest to largest as α-, β-, and γ-diversity: γ-diversity is the biodiversity of an entire region; α-diversity is at the other extreme and describes the biodiversity at a single location, often known to ecologists as a "community". I find Whittaker's formulation especially useful as it highlights the intermediate (β) scale—the processes that make local communities vary from one another. Whittaker gave a precise mathematical formula relating these three types of diversity: $\beta = \gamma/\alpha - 1$ (sometimes "-1" is omitted). Thus, if α-diversity is equal to γ-diversity, then β-diversity would be zero. However, in practice, there *is* variation between communities, and β-diversity is always greater than zero in natural ecosystems.

Can we be more precise about what exactly constitutes, for example, the α-scale? In a word, no. The α, β, γ terminology is inherently a relative statement (γ is bigger than α). The single most important aspect is that the regional scale needs to be large enough to include many local communities. To go beyond just the relative nature of the scales, I find it more useful to define the scales not by exact measurements but by the processes that are dominant at each scale (Shmida and Wilson, 1985; Rosenzweig, 1995; McGill, 2010b). Thus, the γ-scale is that at which evolutionary processes (specifically speciation and global extinction) determine biodiversity and variation is driven by broad-scale gradients such as climate (Currie, 1991; O'Brien et al., 2000; Hawkins et al., 2003; Currie et al., 2004), area (Terborgh, 1973; Rosenzweig, 1995), and biogeographical, historical contingencies (Latham and Ricklefs, 1993; Qian and Ricklefs, 1999). In contrast, α-diversity is the scale at which processes traditionally studied in community ecology such as dispersal limitation, microclimate, and species interactions predominate (Diamond, 1986; Vellend and Agrawal, 2010). Even a process-based definition remains imprecise because there are not precise scale breaks at which different processes dominate, the scales vary by organisms, and different ecologists may reasonably define the boundaries differently.

The players: S, N^*, N_i, s, n^*, n_i, O_i/R_i—What are the most important variables to measure in biodiversity? Arguably, we can turn to the definition of ecology as "the study of the distribution and abundance of organisms" (Krebs, 1972; Andrewartha and Birch, 1986). Thus variables that measure distribution and abundance would seem central. Here I introduce a common notational framework used throughout the rest of the paper (also see Table 1). I start with the regional or γ-scale. This is assumed to cover an area denoted by A. This area will have S species and N^* individuals. N^* and S describe community level or emergent properties. One can also track facts about individual species such as the number of individuals N_i of species i within the region (where i ranges from $1 \ldots S$ and $N^* = \Sigma_i N_i$). N_i refers to the abundance across the entire region and is sometimes referred to as the global abundance of a species (Gregory, 2000; McGill, 2003; White et al., 2007). Typically in ecological communities, there is great variation in N_i between species (McGill et al., 2007). It is often convenient to present this information in an alternative form of $M(j)$, giving the proportion of species with abundance j (i.e., number of species with $N_i = j$ divided by S). This form contains identical information as the N_i but matches the probability distribution form commonly used to describe species abundance distributions such as lognormal, log-series, etc. (Fisher et al., 1943; Preston, 1948; Marquet et al., 2003; McGill et al., 2007). Turning to the distribution of species, species are not found everywhere. Thus, it is also meaningful to study what fraction of the total area A is occupied. There are two ways to do this, and which one is used depends on the type of data collected. For one type of data, the percentage of total area A occupied, O_i, is used and is called occupancy. For the other type of data, one measures the range size R_i, which is in the same units as A (e.g., km^2). O_i is usually used when the area A is divided into a lattice or grid and the scales are small relative to the geographic range of a species, while R_i is used for transect data across large spatial scales. More details on these two types of data are given below. These variables are all at the regional or γ scale.

Table 1. A list of the variables and the notation of those variables used in this paper.

Category	Definition	Regional scale (upper case)	Local scale (lower case)
Species level	No. individuals of species *i*	N_i	n_i
	Proportion of species with abundance *j*	$M(j)$	$\mu(j)$
Community level	Total no. individuals in area (all species)	N^*	n^*
	Species richness	S	s
	Occupancy	$O_i(n_i \mid N_i, a/A)$	na
	Range size	R_i	na
Definitional	Area covered	A (divided into C cells in gridded data)	a

Note: na = not applicable.

At the local or α scale, we can imagine a parallel set of variables. To help with keeping scales straight, I will use lower case variables for the α-scale, and the same variables in upper case for the γscale. First, let the area of a local community be denoted by *a*. Then the total number of species is denoted by *s* and the total number of individuals as n^*. Let n_i be the total number of individuals in species *i* at the α scale [or $\mu(j)$ = the fraction of locally present species where $n_i = j$]. There are many local communities, so denote the total number of local communities by *C*. There is no analog of O_i/R_i at the α-scale.

Note that macroecology is sometimes defined to address abundance, distribution, mass, and energy (Brown and Maurer, 1989). However, to date, there has been a significant split of macroecology into two subdisciplines, with one discipline focusing on abundance and distribution. I call this biodiversity theory, and this is what is studied here. The other discipline has focused on mass and energy and has experienced recent exciting innovation under the name of the metabolic theory of ecology (West et al., 1997; Brown et al., 2004). To date, there have been surprisingly few links between the biodiversity and metabolic subdisciplines of macroecology. Damuth's energetic equivalence rule and the more general mass-abundance-relation provides one such link, but it has proven surprisingly hard to use this to unify the two fields, in part because of the large amount of noise and variability in the relationship (Russo et al., 2003; White et al., 2007; McGill, 2008). Studying abundance in units of biomass or energy is also promising (Connolly et al., 2005; Morlon et al., 2008). But much more work is needed in linking the biodiversity and metabolic branches of macroecology, and this is not the subject of this paper.

The data—Broadly, one finds two types of data containing these variables. One type of data is measured on a grid or lattice. Each grid cell is a local (α) community, and the entire square containing *C* grid cells is the regional (γ) community. As already mentioned, usually in this case, O_i is measured instead of R_i. Also in this case, n_i may or may not be known. One example of this data is known as atlas data. Many countries have atlas projects, which record the presence or absence of a species on a 10 km × 10 km grid (e.g., Australia and Britain). Another example of the gridded type of data is stem-mapped data, where the precise geographic coordinates, size, and species of every individual tree is measured. One can ignore the precise spatial coordinate information by aggregating up to a grid placed over the area giving the counts n_i of each species in each grid cell. The number of cells where $n_i > 0$ divided by the number of cells *C* gives O_i. Note that we can calculate $C_i = CO_i$ giving the number of cells where the species is found.

The second main type of data is obtained when spatially dispersed transects or plots are used. Thus, transects are taken across the landscape, but are not directly adjacent to each other, leaving gaps in between the measured transects. These transects usually have no obvious spatial structure and may appear random. The data are sometimes called a geospatial data set. Because of the gaps between transects, one tends to measure range size, R_i, rather than occupancy, O_i. To measure R_i, one draws a bounding polygon (e.g., convex hull) around the transects where the species is observed and calls the area of this bounding polygon the range size (R_i). If desired, one can also calculate a value for occupancy, O_i,for geospatial transect data by calculating how many transects were taken inside the range of the species (CRange_i) and the number of transects where the species is observed (C_i) and using $O_i = C_i/\text{CRange}_i$. In practice, gridded data approaches are often used on smaller areas, and geospatial data are used on larger areas approaching whole continents. It is usually not practical to cover completely edge-to-edge a whole continent (but see the aforementioned atlas projects). As a result in geospatial data the variables *S*, N^*, R_i, and N_i are therefore less than what they would be if we truly measured the entire region. Techniques for extrapolating to the entire region exist (Chao, 1987; Gotelli and Colwell, 2001, 2011) but do not concern us here. If there are enough local communities (i.e., a large enough sample), then the measured variables will be strongly correlated with the unknown true global values, and the measured values can successfully be used as proxies of the true values for the entire region.

THE PATTERNS

Now that the variables of interest, the scales at which they are measured, and the types of data used to measure these data are defined, I turn to the patterns found in these variables. It is fairly simple to produce a list of these patterns (Table 2). These patterns are all well documented (in most cases in hundreds of papers, but always in at least dozens of papers). These patterns have all had significant meta-analyses and review papers written about each individual pattern and been summarized collectively in several books on macroecology (Brown, 1995; Rosenzweig, 1995; Gaston and Blackburn, 2000). There can be little doubt about the generality (dare I say universality) of most of these patterns. So I do not attempt to prove the generality of these patterns here and instead provide entry points into the literature.

These patterns tend to be rather abstract and quickly develop into an alphabet soup of acronyms, so rather than describe them, I will provide concrete examples of the patterns, using two well-known botanical data sets: (1) the trees of a 50-ha plot in Barro Colorado Island (BCI), Panama (Condit, 1998; Hubbell et al., 1999, 2005; Condit et al., 2000; Hubbell, 2001); and (2) the trees of the U. S. Forest Inventory Analysis (USFIA) (USDA, 2010). See Table 3 and the legend of Fig. 1 for more details. The BCI data matches the first kind of data (gridded), while the USFIA data matches the second (isolated transects). Both data sets have natural definitions of local communities (1 ha for BCI, 1 site of about 0.4 ha for USFIA) and regional pools (the 50-ha plot for BCI, eastern North American forests for USFIA). Clearly, the scales are very different. It is doubtful the 50-ha plot in BCI is large enough to even be considered a true regional pool in the process-defined sense given earlier (Ricklefs, 2003), but larger areas have not been fully sampled, and it is convenient and common to treat this scale as a regional pool (Hubbell, 2001; Harte et al., 2008). The fact that each data set can be used to demonstrate each of the patterns in Table 2, does not in itself prove the generality of the patterns (the much more extensive studies mentioned in the previous paragraph do that). But on another level, the fact that I can pick two data sets more or less arbitrarily and then find all of the expected patterns in those data sets is rather profound. Neither data set was seminal in the development of or initial hypothesizing of the patterns (with the possible exception of clumping in the BCI data), so it is not circular to find the patterns in this data. Rather, it is indeed a strong indicator of how general these patterns are. And it contradicts the common assertion that ecology does not have any general laws.

TABLE 2. A list of the biodiversity patterns discussed and illustrated in this paper. It is unfortunate that macroecologists usually describe their central research phenomena by acronyms, but they do. It is even more unfortunate that the acronyms are not consistent across the literature. In this table, I have tried to enforce a consistent nomenclature. Acroynms ending in "D" describe the frequency distribution of a variable. Distributions are taken across some entity; this is denoted in the first letter; in most cases here, I look at distributions of species properties across a set of species and hence start with S (e.g., SAD), but in one case, I look at the distribution of abundance across the local communities and hence start with C (i.e., CND). Although not discussed in this paper, many distributions are measured across individuals and would start with I (e.g., IMD would give the individual mass/bodysize distribution) (also see White et al., 2007). Acronyms ending in R describe a relationship (correlation or function) between two variables. I have consistently used N for abundance, S for species richness, O for occupancy, and R (or RS) for range size). Since the R for range size never occurs as the last letter and the R for relationship occurs only as the last letter, this is not ambiguous. I use A to indicate sampling extent as defined by area. Where necessary, l and g distinguish the local (α) and global (regional or γ) scale. I have violated the rules by using A for abundance instead of N in the case of the SAD because the SAD acronym is very strongly established in the literature. Likewise, three relatively lesser known patterns that I labeled as the NAR, ONR, and the SDR are better known as the IAR, the OAR, and DSD but have not yet gained universal recognition, so I have chosen to relabel these for consistency but tried to use the pre-existing acronyms parenthetically. The collector's curve or rarefaction curve has not yet been given an acronym but fits well into this system as SNR, which is a new coinage.

Scale	Acronym	Name	Formula
Regional (γ)		Species–environment relation	$S = f$(environment)
		Abundance–environment relation	$N = f$(environment)
	gSAD	Global species abundance distribution	~N_i or $M(n)$ log-normal–log-series
	SOD	Species occupancy distribution	~o_i U-shaped
	SRSD or SRD	Species range size distribution	~R_i log-normalish
	ONR	Occupancy–abundance relation	$o_i \propto_+ N_i$
	RSNR	Range-size–abundance relation	$R_i \propto_+ \mathrm{N_i}$
	RONR	Range-size–occupancy–abundance relation	$\mathrm{o_i} \propto_+ R_i \propto_+ \mathrm{N_i}$
Local (α)	lSAD	Global species abundance distribution	~n_i or $\mu(n)$ log-normal–log-series
	SNR	Species–abundance relation (collector's curve)	$s = f(n^*)$
Beta (β)	SAR	Species–area relation	$S = cA^z$
	NAR or IAR	Abundance (individuals)–area relation	$n^* = a/AN^*$
	OAR	Occupancy–area relation	$O_i = f(N_i)$ with f a positive function
	SDR or DSD	Similarity–distance relation	similarity(d) = exp(−cd)
	CND	Community abundance distribution	~n_j log-normal/log-series

Regional (γ) level patterns—S and N* and its interspecific allocation to N_i, O_i, and R_i—First, I examine community patterns known at the regional or γ-level data (Table 2). Recall that the area A was arbitrarily chosen by the sample design (although less so in the USFIA analysis where an effort was made to define a biogeographic province). The BCI data has $S = 225$ species and $N^* = 21\,457$ individuals (>10 cm dbh) in 50 ha. The USFIA data had $S = 123$ species spread across $N^* = 9412$ individual trees (>5 in dbh) at $C = 226$ sites (transects). Neither S nor N^* are directly comparable between the two data sets since they come from different sized plots. These numbers are determined by γ-level processes—trade-offs between speciation and extinction as driven by climate, productivity, and biogeography. It is well known that richness S is highest in the tropics (latitude = 0) (Pianka, 1989; Mittelbach et al., 2007) (Fig. 2A). The reasons for this are not fully understood because several factors influencing richness are collinear with latitude. Specifically, it is known that the areas available in the tropics are larger (Terborgh, 1973; Rosenzweig, 1995). Moreover, climate variables related to energy and water balance (i.e., increasing precipitation and temperature, especially when both occur together) tend be positively related with productivity and both S and N^* in trees (O'Brien et al., 2000; Hawkins et al., 2003) (Fig. 2B, 2C). Slightly different climatic factors are most important for other taxa (Currie, 1991). Regions where the ability to migrate to evade glaciers was limited (e.g., Europe) also tend to have lower diversity (Latham and Ricklefs, 1993), and regions with more seasonal variability have fewer species (MacArthur, 1975) (Fig. 2D). Our understanding of what controls N^* is even more limited. But broadly we know that water balance determines leaf area index (LAI) and thus plant height (Walter, 1973; Gholz, 1982; Woodward, 1987; Stephenson, 1990). Leaf area index then drives plant size and total percent cover, which then are the major drivers of N^* (grasslands have higher N^* than forests, while deserts with bare ground have lower). So in the framework considered here S and N^* are essentially inputs determined by the environmental context broadly defined.

TABLE 3. The empirical values of regional (γ) variables for two example data sets from Barro Colorado Island, Panama (BCI) and the U. S. Forest Inventory Analysis (USFIA). Average values for local (α scale) communities (n^* and s) as well β-diversity are also given.

Variable	BCI	USFIA
A	50 ha (0.5 km^2)	300 000 km^2
a	1 ha	~0.4 ha
C	50	226
$S = \gamma$	225	123
N^*	21 457	9412
$\bar{n}^*$	429.14	41.6
$\bar{s} = \alpha$	90.78	6.8
β	225/90.78 = 2.48	123/6.8 = 18.1

I turn now to two patterns at the regional (γ) scale measured on a species-by-species basis. First, N^* is allocated across the S species to give the N_i. When one plots a histogram of the N_i, one is effectively examining the $M(j)$ representation of the N_i (Fig. 3A, 3D). These allocations are highly uneven. Indeed, any histogram of N_i is strongly right skewed and is sometimes called a hollow curve (McGill et al., 2007). In any community, there are only a few species with high abundances (N_i) and a multitude of species with very low abundances. If abundance is plotted on a log scale (Fig. 3B, 3E), then something closer to a normal

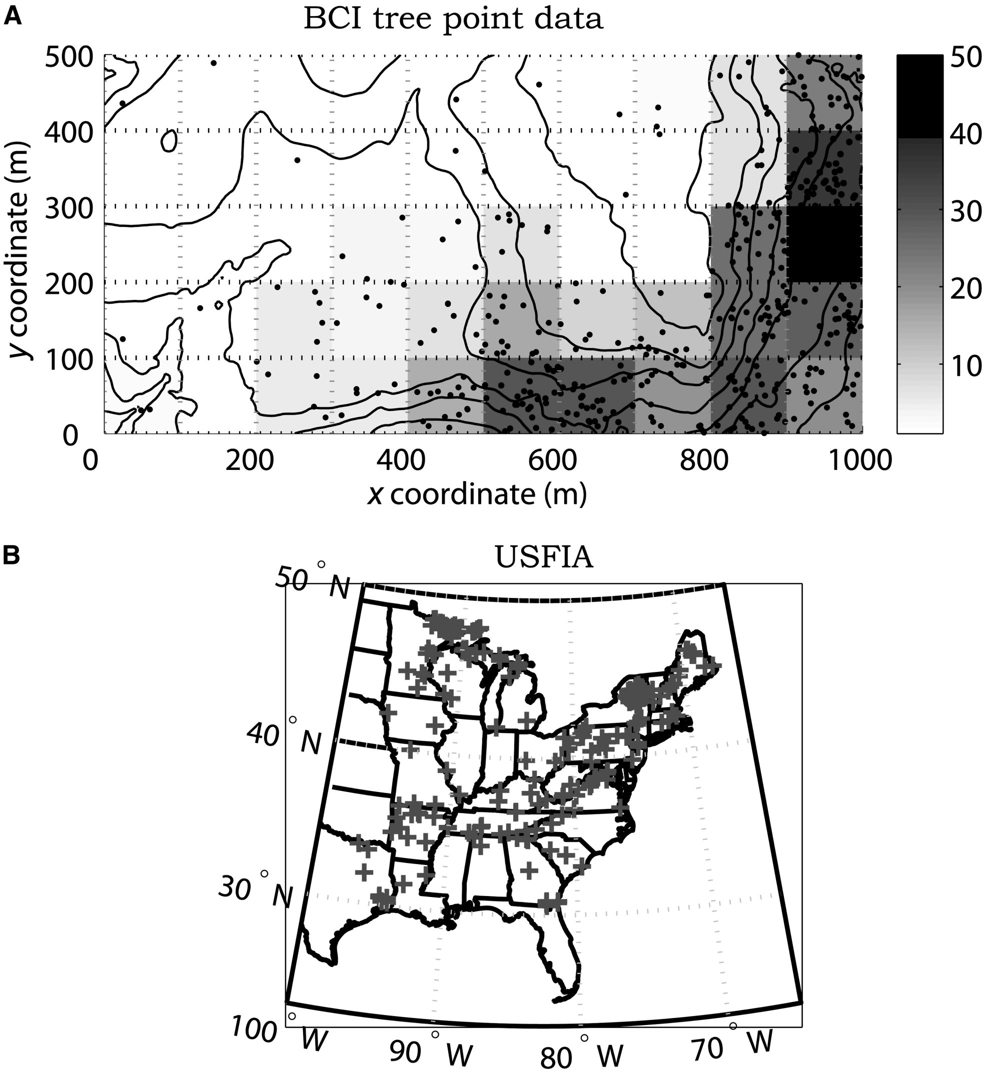


Fig. 1. Two illustrative data sets used throughout the main text. (A) The data from a 50-ha plot on Barro Colorado Island (BCI) in the Panama Canal is one of the best-known examples of a gridded data set. Every tree >1 cm dbh has been surveyed and measured by the Smithsonian for several decades (Fig. 1A). This plot shows the locations of individuals of the species *Jacaranda copaia* with shading showing the number of individuals in each cell (hectare). B) The U. S. Forest Inventory Assessment (denoted in this paper as the USFIA) is a good example of a geospatial data set. The USFIA involves four circular plots arranged in a triangle and totaling ca. 0.4 ha. The data used in this paper include trees >5 in dbh (allowing comparison with BCI) and limited to the forested regions in the eastern United States (all forests east of 100° W longitude), thereby avoiding the boreal forests and montane regions of the Rockies and western United States and keeping to a single biogeographic province. Every site that was on protected land and estimated to be old growth was used. This figure shows the location of each of the sites used.

distribution appears, although there often can still be an "excess" of the rarest species (McGill, 2003). The distribution of the N_i is called the global species abundance distribution (gSAD). The fact that the gSAD is strongly skewed and is roughly log-normal-like seems for now to be a general pattern (Gregory, 1994, 2000; Gaston and Blackburn, 2000; McGill and Collins, 2003; McGill, 2003), but this is one of the more weakly documented patterns listed in Table 2, and more work is needed.

As already noted, the fact that the BCI is gridded data and the USFIA data are from geospatial transects requires us to measure distribution (i.e., spatial extent) differently so I use occupancy (O_i = percentage of grid cells species is found in) for the BCI data and range size (R_i = the area in square kilometers covered by the species) for the USFIA data. In turn, the two measures show very different patterns. The distribution of ranges sizes, R_i, or SRSD (species range size distribution) is again log-normal-like (Fig. 3F, 3G) with many species with very small ranges and a few widely ranging species. Again, this appears to be a general pattern (Gaston, 1990, 2003; Brown, 1995). However, the distribution of occupancy, O_i, or SOD (species occupancy distribution) shows a bimodal distribution (Fig. 3C) with the very smallest and very largest levels of occupancy being common (effectively a U-shaped histogram). This is a consequence of occupancy being measured on a finite scale (0–1). In a certain sense, the long right tail of R_i (Fig. 3F) gets "crunched" up and leads to a high number of species with high occupancy. This phenomenon is also known as area saturation (Šizling and Storch, 2004). The U-shaped SOD is arguably one of the oldest known macroecological patterns (Raunkiaer, 1909) and is found repeatedly (Gaston and Kunin, 1997; Gaston and He, 2011).

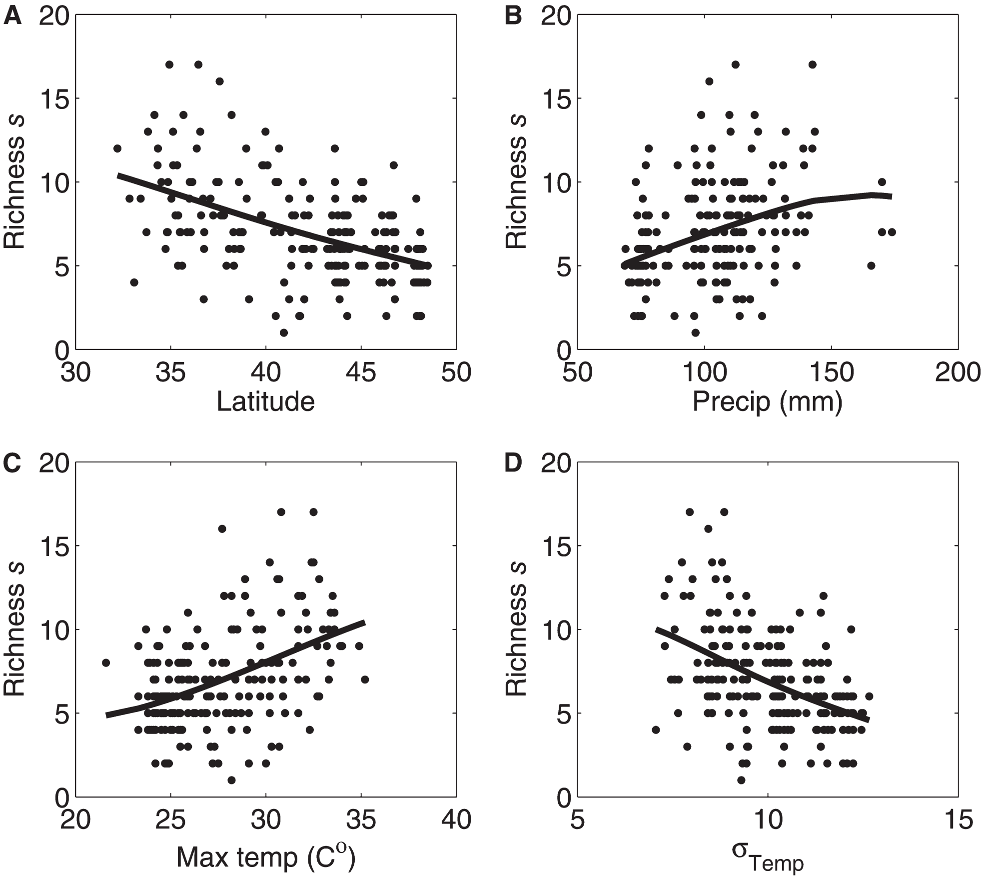


Fig. 2. A number of environmental factors appear to affect diversity in the USFIA. Data below 32° latitude are omitted due to the peninsula effect on species richness in Florida. LOESS lines are shown through the data but r^2 and P values reported are from a linear regression. (A) An indirect factor affecting richness is latitude ($r^2 = 0.13$). (B) Mean annual precipitation ($r^2 = 0.09$), (C) maximum temperature in C° ($r^2 = 0.12$); and (D) the standard deviation of average monthly temperatures (a measure of seasonality) ($r^2 = 0.10$) also all affect richness. Note that $P < 10^{-6}$ for all four plots. These plots show much more variation around the trend line than is typical because of the small grain size (≈0.4 ha), which results in a strong sampling effect in the USFIA data. Larger grain sizes (e.g., 1° × 1°) show much less variation and higher r^2 (e.g., Currie and Paquin, 1987).

Local (α) patterns—s and n* (and the allocation to n_i)—I turn now to patterns in the local community (Fig. 4). Physicists, especially in the field of statistical mechanics, draw a distinction between two types of measurements based on how they behave when two separate regions (usually volumes of gas but here local communities) are combined (McQuarrie and Allan, 2000). For extensive variables, the measurement of the combined system is simply the sum of the measurements on the separate systems. The number of molecules or the total energy in the system are good examples. Intensive variables combine as a weighted average. Temperature and density of molecules are examples of intensive variables.

The total number of individuals in a plot (n^*) behaves extensively. By definition, $N^* = \Sigma n^*$ (where the sum is across all plots). Conversely (see Fig. 5A, B),

$$n^*(a) \approx N^*a/A \tag{1}$$

The approximation is necessary because there is some stochastic variation. But this variation is usually fairly small (CV = 9.9% for BCI) although it can be large when local area, a, itself is small introducing more stochastic variation (CV = 54.8% for USFIA). In either case, the variation in n_i appears to be distributed normally (Fig. 4A, B). More generally, area seems to be the best predictor for abundance across scales (Fig. 5A). In the BCI data, for every scale between the local community and the regional pool, Eq. 1 seems to be an excellent ($r^2 = 0.9999$) predictor of abundance (Fig. 5A). In short, abundance is purely extensive. This relationship between abundance (N) and area (NAR for the abundance area relationship, often known as the IAR) is an extremely useful relationship for moving across scales and also appears to be quite general (Hubbell, 2001). It has been noted that in some special situations the NAR is not so simple. A more general relationship $N(a) = ca^z$ still holds (with $z = 1$ being the special case of Eq. 1). On islands, z is often >1 due to nonlinear feedback of island (or habitat patch) size effects—an island twice is big is more than twice as good. This is sometimes called the PIAR (patch individual area relationship (Pautasso and Gaston, 2006)). Studies have also found $z < 1$ (around 0.8) on nonpatchy environments (Pautasso and Gaston, 2006) and call this the GIAR (general IAR). However, this appears to apply at scales where habitat structure changes and the taxonomic group studied cannot be reasonably expected to be found in parts of the area studied. More work on the contrast between Eq. 1 and the GIAR ($z \approx 0.8$) is needed.

If the total abundance in a local community (n^*) is determined by Eq. 1, then how is the abundance allocated across species (the n_i)? This distribution of n_i is called the local species abundance distribution or lSAD. When n^* (or equivalently, a)

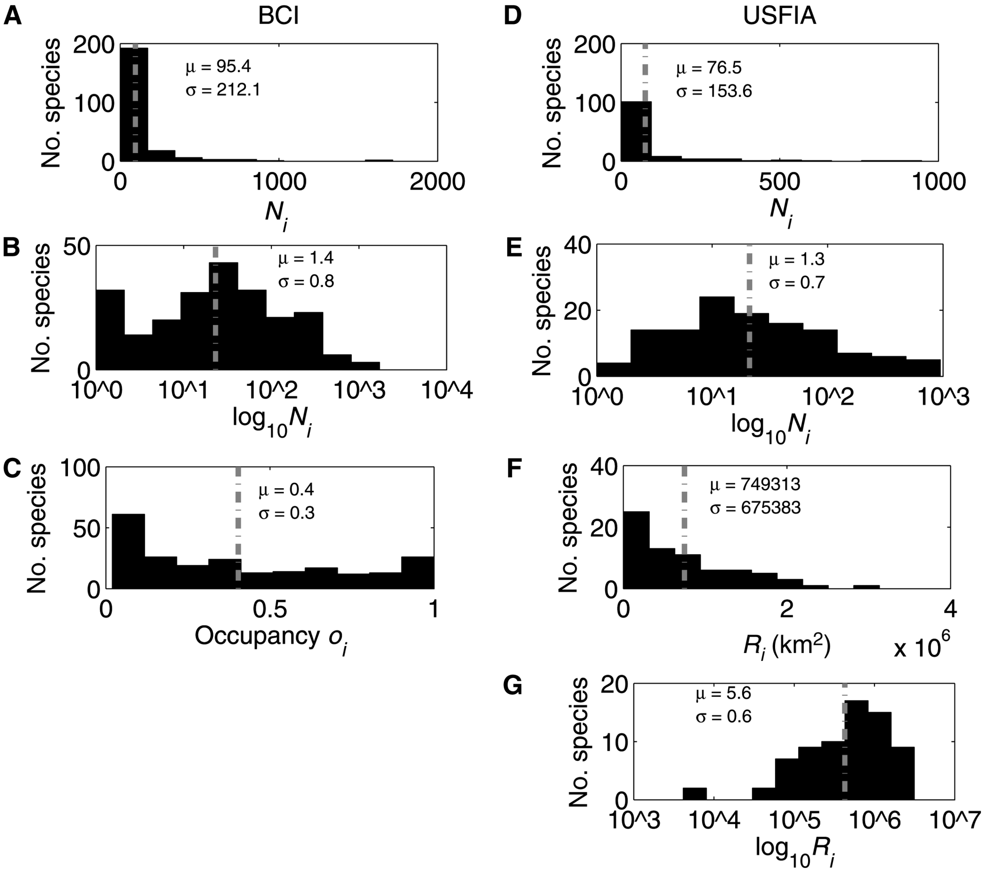


Fig. 3. The frequency distribution of regional (α) scale properties across species. (A, D) abundance (i.e., gSAD) on an arithmetic scale is highly right skewed; (B, E) the same data on a logarithmic scale becomes log-normal-like; (C) the distribution of occupancy (i.e., SOD) is on a finite (0–1) scale and U-shaped. (F) The distribution of range sizes (i.e., SRSD) is also heavily right skewed; (G) the same data on a log scale again looks log-normalish. Graphs A–C are for BCI; D–G are for the USFIA.

is small, there can be a fair amount of variability or stochastic noise in the shape of the lSAD. And because of the discrete nature of n_i (you cannot have 1/2 an individual), it is common to find many singletons in lSADs. But ignoring these details, it turns out the lSAD is usually quite similar to the gSAD, being strongly right skewed on an arithmetic scale and with no left skew on a log-scale (Fig. 4E, F). Indeed, mathematical theory (described in more detail later) suggests that the shape of the lSAD should be the same as the gSAD with just the x-axis (n) rescaled (Green and Plotkin, 2007).

One of the unique features of ecology is that a variable of central interest, species richness, s, is neither intensive nor extensive but somewhere in-between. For BCI, if richness were intensive, then we would expect $s = S = 225$, whereas if it were extensive, we would expect ($s = S^*a/A = 4.5$). Note that treating s as extensive is equivalent to assuming β-diversity is high (no species found in more than one local community), while intensive is equivalent to β-diversity being low (all species found in all communities). In fact, we see an intermediate value, $s = 90.8$ (SD = 7.0). Similarly, for USFIA, we have $s = 6.8$ (SD = 2.9), which is intermediate between the intensive expectation of $s = 123$ and the extensive expectation of $s = 0.54$. This intermediate nature gives ecologists challenges that physicists do not face.

One way of describing the intermediate nature is to plot a collectors curve (Pielou, 1975) or species–individual relationship (SNR), which plots the number of species vs. the number of individuals sampled (Fig. 5B), i.e., the function $s(n^*)$. If plotted literally as encountered in the field, this curve will be a jagged stair-step curve. But in practice, the individuals are ordered randomly several times, and the average SNR is plotted, which is then smooth (Fig. 5B). This process is also known as rarefaction (Sanders, 1968) and can be done either by computer randomization (Sanders, 1968) or by an analytical formula giving the expected value and variance of $s(n^*)$ (Hurlbert, 1971; Simberloff, 1972). The net result is that species richness is an increasing, decelerating function of the number of individuals. Empirically, the data are often well fit by a function of the form $s(n^*) = c(n^*)^z$, where $0 < z < 1$ and often $z \approx 0.5$ (Siemann et al., 1999) or by $s(n^*) = c\log(1 + n^*)$. The exact shape of $s(n^*)$ depends heavily on the evenness of the n_i (i.e., the shape of the lSAD). Rarefaction cannot be performed accurately on presence/absence data—the abundance must be known. This is because having a few highly abundant species that are likely to be encountered over and over again makes the rarefaction curve less steep. In fact, the slope of the rarefaction curve $s(n)$ at $n = 1$ is equal to the evenness of the community (Olszewski, 2004). If one has a mathematical formula for the distribution of the n_i or lSAD [i.e., the probability function $\mu(n)$], then $s(n^*)$ can be derived as a sampling from this distribution. May (1975) derives the analytical form of $s(n^*)$ for a number of different theoretical SADs. The log-series is $s(n^*) = \alpha \log(1 + n^*/\alpha)$, where α here is Fisher's α from the log-series (Fisher et al., 1943; May 1975).

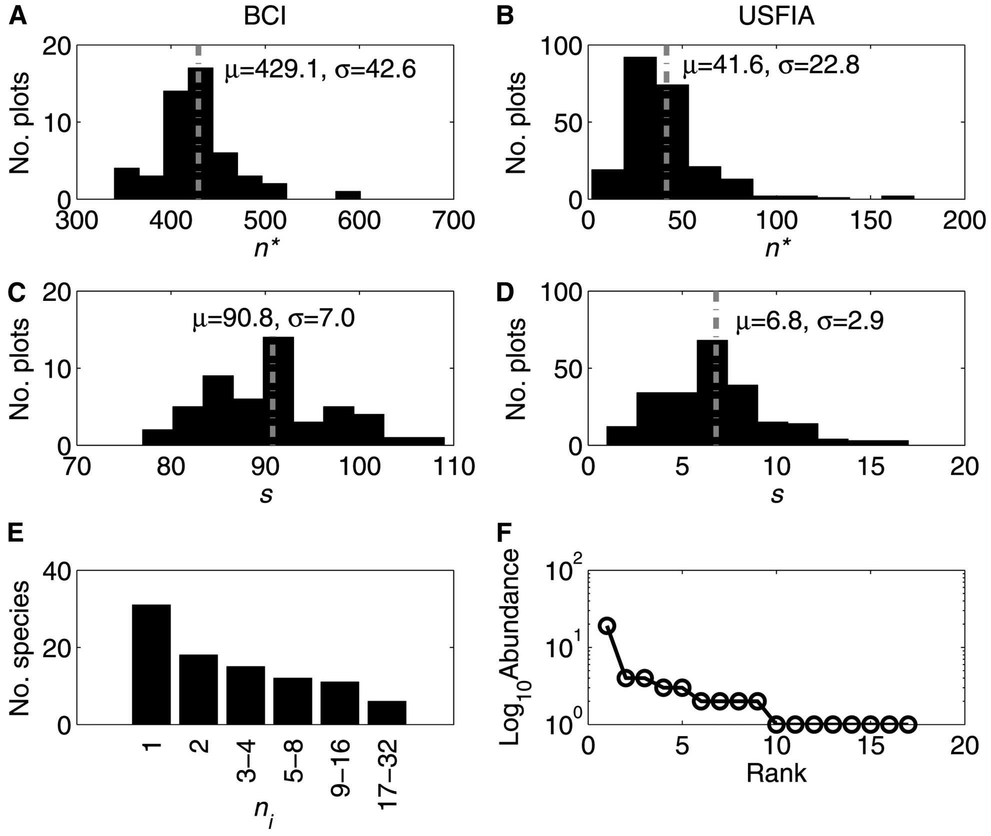


Fig. 4. The frequency distribution of local community variables n^*, s and n_i. (A, B) n^* is roughly normal across local communities. The mean is equal to N^*/C since n^* is extensive; (C, D) local richness, s, is also roughly normal but the mean is intermediate between what an intensive and extensive variable would predict; (E) the local species abundance distribution (lSAD) on a logarithmic scale for one randomly chosen hectare; (F) another lSAD for one randomly chosen site plotted as a rank–abundance diagram (McGill et al., 2007). Subfigures A, C, E for BCI; subfigures B, D, F for USFIA.

As can be seen (Fig. 5B), the theoretical log-series SNR fits the empirical SNR (i.e., rarified or Monte Carlo sampling of the observed N_i) quite well. However, if the distribution of the n_i is wrong (even if still capturing the strong skewness of the data), then the rarefaction curve will be wrong (see the broken stick example in Fig. 5B).

Spatial structure and across scale (β-diversity) patterns—So far I have only looked at patterns at a single scale: the local (α) or the regional (γ). The one exception was the NAR or IAR, which did not show any changes in scale (i.e., as one increased area). I now turn to patterns that cut across scales. Specifically, I looked at patterns that describe community structure and diversity over area. This can broadly be described as the study of β-diversity (Whittaker, 1975; Koleff et al., 2003; Anderson et al., 2011; Jost et al., 2011). There are three common ways of studying β-diversity, which are interrelated but not identical.

(1) Pairwise—One can compare the actual content of communities (which species are present and how abundant they are) at increasingly large distances apart. To do this, one needs a way to measure the similarity or difference between communities. A number of famous metrics have been devised to do this. Some are based only on which species are present in each community. Two well-known examples of this are the Sorensen and the Jaccard indices. The Sorensen index measures the number of shared species divided by the average number of species $(s_1 + s_2)/2$, while the Jaccard measures the number of shared species divided by the total species in the system. Since species richness is between intensive and extensive we have $s_1, s_2 <$ number of shared species $< s_1 + s_2$. The problem with these indices is that a species can be dominant (the most common species) in one community and rare (only one individual) in the other, yet these measures ignore this. To address this limitation, there are also a number of measures that use the relative abundance (n_i) of the species in the two communities including the Morisita–Horn and the Bray–Curtis. In practice, there are dozens of different indices (Legendre and Legendre, 1998; Koleff et al., 2003; Anderson et al., 2011). But once a metric is chosen, one can take many different pairs of plots and compare the similarity with the distance apart, d (Fig. 6A, B). This is called the similarity–distance relationship (SDR), sometimes known as the decay of similarity with distance or DSD (Nekola and White, 1999). Necessarily, such a plot will start with a similarity = 1 at a distance of zero and will decrease with increasing distance, eventually reaching zero. The decrease appears to be fairly smooth and must asymptote to the x-axis. Both the exponential function, similarity(d) = $\exp(-d/k)$, and a power law form, similarity(d) $= ad^{-k}$, are commonly used.

(2) Nested—The original definition of β-diversity was given by Whittaker and was based on comparing two nested spatial scales, in fact the local and regional scales. One can measure β-diversity by $\beta = \gamma/\alpha - 1$ (or in my notation $\beta = S/\overline{s}_j - 1$).

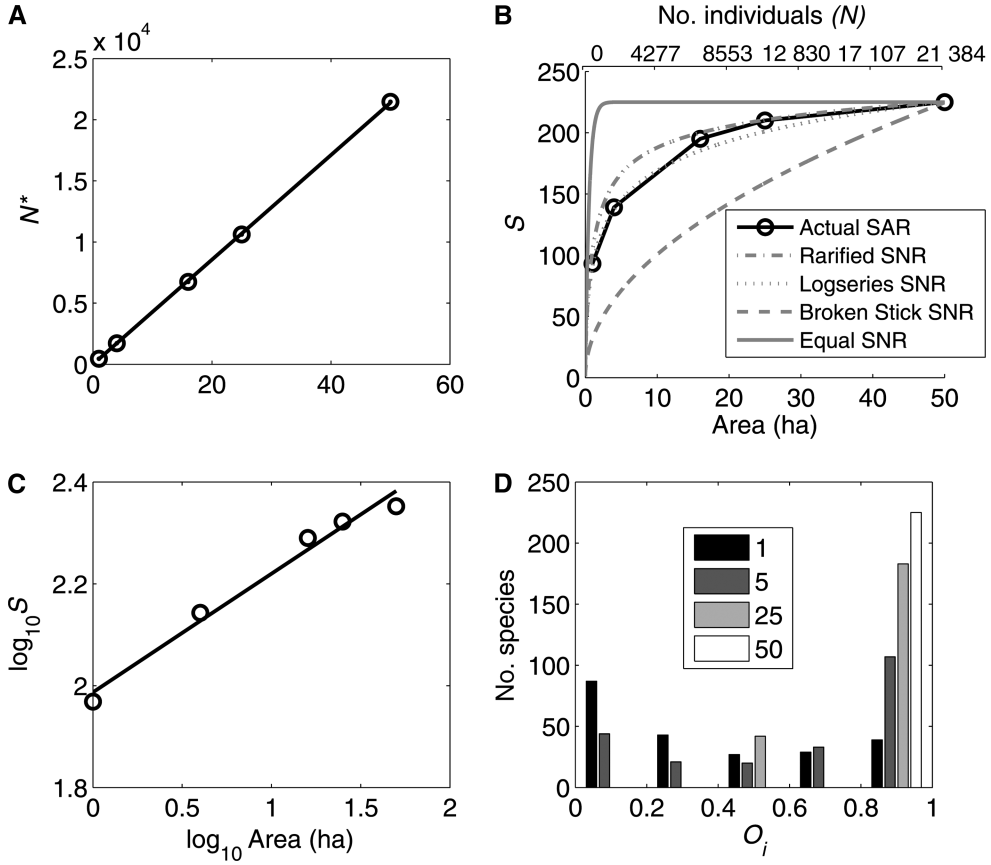


Fig. 5. Plots of how variables change with area for the BCI data. (A) The abundance–area relationship or NAR (or IAR) is nearly perfectly linear ($r^2 = 0.99992$ when forced to pass through zero) and gives $N = 427.667A$. (B) Species richness, *s*, is neither intensive (horizontal) nor extensive (linearly increasing). It is somewhere in between. Given the NAR, we can map area directly to individuals and plot number of individuals across the top. This allows one to compare the species–area relationship (SAR) and the species–individual relationship (SNR). The rarified curve based on observed N_i (i.e., collectors curve or SNR) comes close, as does the SNR based on a theoretical log-series distribution. But if the N_i are wrong either by assuming completely equal N_i or an excessively uneven broken stick lSAD, then the resulting SNR poorly fits the SAR. (C) The SAR is close to linear on a log-log plot ($r^2 = 0.98$, $\log S = 1.99 + 0.23 \log A$, giving $z = 0.23$). (D) The species–occupancy distribution (SOD) gradually shifts from the classic U-shaped histogram to all 1's as the grain size increases. This has been called the area saturation effect.

When β is 0, every species in the region is found in every local community (there is no change in species composition across space). When $\beta = S - 1$, then a different species is found in each local community. It has been suggested that β-diversity should be defined additively instead of multiplicatively ($\beta = \gamma - \alpha$ or $\alpha + \beta = \gamma$) (Lande, 1996; Crist et al., 2003; Crist and Veech, 2006). The additive form makes β harder than the multiplicative form to interpret ($\beta = 3$ has very different interpretations if $\alpha = 1$ or $\alpha = 100$), but it does allow for the partitioning of diversity across more than just two scales (e.g., Fig. 6D, see legend for an explanation).

(3) Accumulative—One can imagine asking if *s* species are found on a plot of area *a*, how much larger an area does one need to see $s + 1$ species, or more generally, what is the function $s(A)$ (the number of species given an area *A*). The function $s(A)$ is known as the species–area relationship or SAR (Table 2, Figs. 5C, 6C). The SAR is highly relevant to conservation and can inform questions like how many species will a reserve of a given size protect or what are the consequences of destroying part of a habitat patch. The slope of $S(A)$, i.e., the derivative of $S(A)$ with respect to *A* [i.e., $S'(A)$] is a measure of β-diversity. It is common to approximate $S(A)$ by using the power law form, $S(A) = cA^z$ (Arrhenius, 1921; Connor and McCoy, 1979; Rosenzweig, 1995). This implies plotting log *S* vs. log *A* should give a straight line ($\log S = \log c + z \log A$). An immediate implication if the power law form holds is that β-diversity is constant over scales. The power law is a good approximation to linear over several of orders of magnitude of scale (Fig. 5C, $r^2 = 0.98$, $z = 0.23$), but over a very large range of scale, *z* clearly changes. For example, in the USFIA (Fig. 6C), the SAR has an inflection and begins curving upward again, eventually approaching $z = 1$. This sigmoidal (S-shaped) curve is sometimes known as the triphasic SAR, which appears to be general (Williams, 1943; Preston, 1960; Shmida and Wilson, 1985; Rosenzweig, 1995; Hubbell, 2001). Even at smaller ranges of scales, there can be detectable variations in *z* (i.e., curvilinearities in a log-log plot, see Fig. 5C) (Harte et al., 2009). There is a relationship between the nested (no. 2) and accumulative (no. 3) forms of β-diversity, and explicit formulas can be derived (Koleff et al., 2003).

Although the mathematical formulas linking between the three forms of β-diversity are often not known, the three forms are all describing the same aspect of reality—namely, that local communities differ from each other and this difference increases with distance. Moreover, the second and third forms of β-diversity (nested, accumulative) make it clear that β-diversity

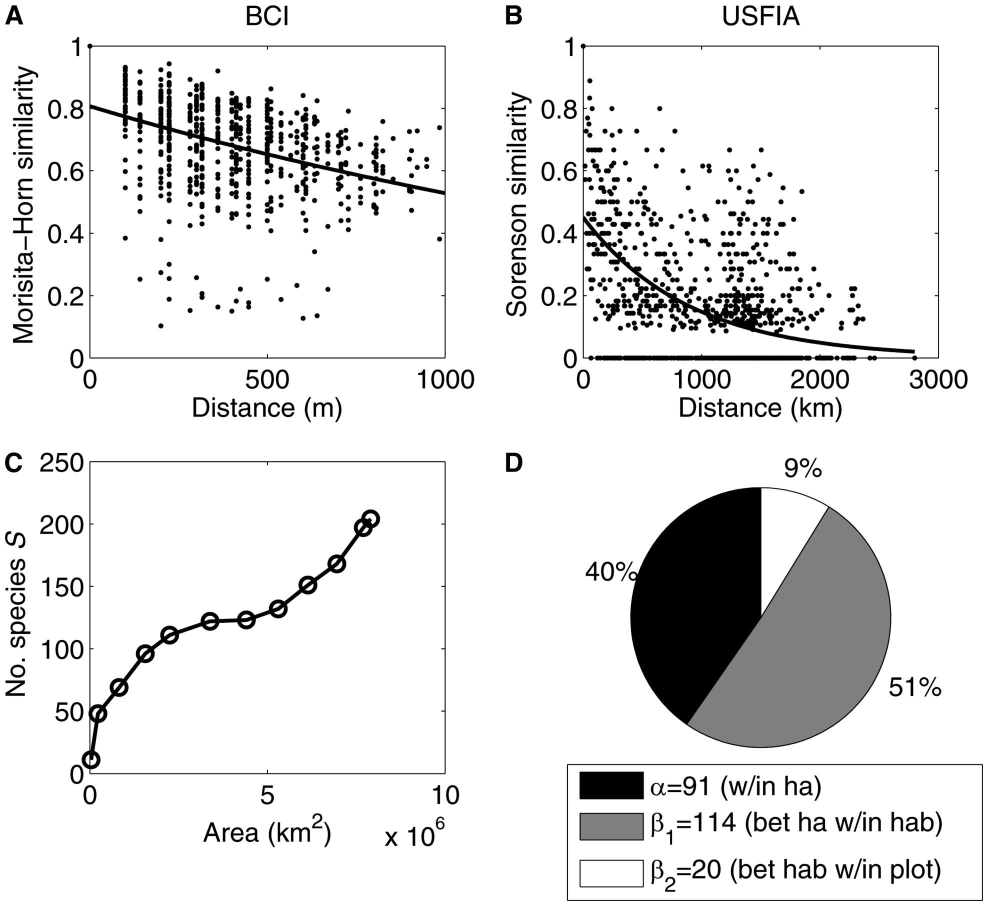


Fig. 6. Different ways to plot β-diversity. (A, B) Pairwise species–distance relationship (SDR or DSD) plots for BCI and USFIA, respectively. Both show decreasing similarity with distance. (C) The SAR for the USFIA. When plotted over a large range of scales as here, the SAR typically is described as triphasic (steep increase, flat shallow increase, then accelerating to steep increase again). (D) A decomposition of the BCI data into three nested levels. The lowest is 1-ha plot (within [w/in] ha), the next highest is between (bet) upland and lowland habitat (hab), and the last is the whole plot. The diversity between plots within the same habitat is extremely high and accounts for 51% of all the diversity, and diversity within a single hectare is also high with relatively little additional diversity added by changes in habitat.

is neither intensive nor extensive in statistical mechanical terms. A benefit of the first form of β-diversity (pairwise) is that it studies not just the accumulation of species but the loss of species between communities, something completely ignored in the second and third forms. Understanding the patterns of β-diversity remains an important challenge in ecology.

From the previous section on local (α) patterns, we know that richness is a function of abundance (SNR), and abundance is a function of area (NAR or IAR, Eq. 1), so one should be able to combine these to calculate richness as a function of area. If $s(n) = cn^z$ and $n = a(N^*/A)$ then one has $s(a) = c(N^*/A)^z a^z = c'a^z$ where c' is a constant independent of a [$c' = c(N^*/A)^z$], giving the power-law form of the species–area relationship. This merger describing the SAR as just a rarefaction curve converted to area seems to accurately describe the SAR up to the scale where the SAR begins to accelerate upward again at ca. 10^4 km^2 (Fig. 6C).

Spatial species level: Clumping—I have repeatedly emphasized that species richness is not intensive. Another way of saying this is that individuals within a species are not spread equally across the whole area A. Instead, species are found in subregions of A. One possibility is that this is just a consequence of the random placement of individuals, but if this were true, then the number of individuals in one species in a subarea of size a would be constant across the local communities with only stochastic variation (specifically, Poisson distributed variation). We can study this by plotting a histogram of the abundances for one species across all the sites (Fig. 7A, C). Here we hold i constant and take n_i for each local community. This pattern has been studied before (Brown et al., 1995; Gaston, 2003; Harte et al., 2008) but has not been given a name, so I label it the CND (community abundance distribution). Empirically, the CND is nowhere close to Poisson but is a log-normal/log-series-like distribution. As shown in Fig. 7A and 7C, the data are overdispersed relative to a Poisson distribution having too many low (mostly 0) abundances and too many very high abundances in contrast to the greater number of intermediate abundance species in the Poisson (Pielou, 1977; He and Gaston, 2000c; He and Legendre, 2002; Harte et al., 2008). This signifies that the processes of placing individuals in space has memory (or shows nonindependence or autocorrelation)—once it has placed one individual in a cell, it is much more likely to place another individual in that cell. But the signature of clumping is even stronger than having memory on a site by site basis; it also works between nearby sites (Figs. 1A, 7B, 7D). If one site has a zero

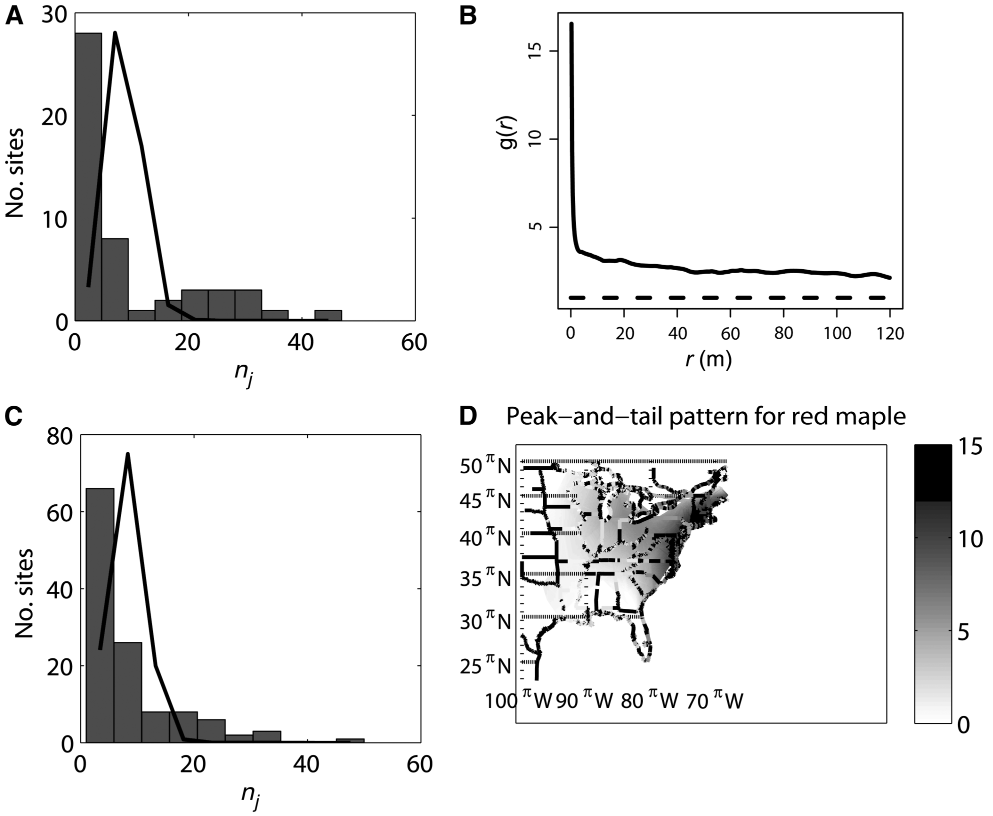


Fig. 7. Spatial structure of individuals. Individuals are spatially clumped within a species. (A) A histogram of n_i for a single species at BCI taken across all local communities (i.e., 1-ha plots) called the community–abundance distribution or CND. The dashed line shows the expected Poisson distribution. The actual data are overdispersed (more low and high values) relative to the Poisson. (B) A pair correlation function for one species (same as Fig. 1A) showing probability of encountering a conspecific relative to probably of encountering other species at different distances r (measured in meters). Values greater than one (indicated by the dashed line) suggest clumping. Note that clumping occurs at all scales for which the function can be estimated. (C) is the same as (A) for USFIA. (D) Plot of the abundance surface for red maple from the USFIA. Abundance peaks in southern New England (with a secondary peak in the coastal mid-Atlantic) but extends at low abundances much further west. This has been called a peak-and-tail pattern and is typical of abundance surfaces.

abundance, then nearby sites are also more likely to have a zero abundance, and likewise if one site has a high abundance, then nearby sites will be more likely to as well. In short, there is spatial autocorrelation. Or alternatively, the abundance is not spatially random, but is organized. There are a number of ways to formally measure this. If we know the location of each point, we can calculate a pair-correlation function or pcf (Fig. 7B), often denoted $g(d)$ (Stoyan and Stoyan, 1994; Condit et al., 2000; Wiegand and Moloney, 2004; McGill, 2011). Variograms, correlograms, and Moran's I (Fortin and Dale, 2005; McGill, 2011) can be used if we don't know the location of each individual but have abundances (Brown et al., 1995; McGill, unpublished data). Although more measurements are needed, it appears to be a very general pattern that nearly all species are clumped in space (Condit et al., 2000; Plotkin et al., 2002). Interestingly, this pattern has been found at many different spatial scales, but we do not understand how clumping varies with scale.

Interactions among the variables—So far, I have only presented univariate patterns that look at single variables at particular scales, single variables across scales and single variables vs. sample size (as measured by area or number of individuals). There has not been any examination of interrelationships among variables. Examining pairwise relationships (Table 4A, 4B, Fig. 8A, 8B), the first thing to note is that total abundance N_i of a species is strongly positively correlated with other measures associated with species commonness such as range size (r = 0.38 for USFIA) and occupancy (r = 0.94 for BCI). This leads to a suggestion of what one might think of as "common is common". This superficially seems to stand in contrast to the classic paper by Rabinowitz and colleagues (1986). They describe rarity as occurring along three axes: small range size, low abundance, or habitat specialization (patchy distribution within range). If one breaks each axis into two levels (common or rare), there are eight possible locations (2 × 2 × 2), only one of which is common in all aspects and seven of which are rare in at least one axis. It has been shown in birds and mammals that because of the positive correlations between these three factors, the common on all three axes and rare on all three axes are the most usual combinations (Brown and Maurer, 1987; Yu and Dobson, 2000). Table 4 and Fig. 9 suggest that "common is common" or what might be called a range-size–occupancy–abundance relationship (RONR) also holds in plants although possibly to a lesser degree than animals. This does not imply that species occur only in the extreme (all common and all rare) combinations, but it does say that they have a disproportionate share of species. There seems to be some evolutionary or ecological constraint that causes these three aspects of rarity to covary.

One problem with the RONR relationship is that total abundance N_i is not independent of range size or occupancy. Indeed, if we let $\overline{n}_i$ be the average abundance across sites where the species is found (i.e., not including zeros), and let C_i be the number of sites where the species is observed, and CRange_i be the number of sites where measurements were taken inside the range, then we have:

$$\text{CRange}_i \propto R_i \tag{2A}$$

$$C_i = O_i\text{CRange}_\text{i} \tag{2B}$$

$$N_i = \overline{n}_i\, C_i. \tag{2C}$$

Thus, total abundance N_i is mathematically necessarily a positive function of O_i and R_i. Also note that in data that occurs on smaller spatial scales over a grid (like BCI), we have no information on total range size R_i and CRange_i is a constant (denoted by C where $C = 50$ in the case of BCI), so Eqs. 2A–2C collapse to $N_i = \overline{n}_i{*}O_i{*}C$. Thus in either case, the RONR pattern is to some degree a mathematical necessity.

This mathematical component can be removed if we look at different measures of abundance, namely average abundance across sites where $n_i > 0$ ($\overline{n}_i$) or the maximum observed abundance (max $n_i = \max_j n_{ij}$). While N_i is clearly dependent on $\overline{n}_i$, there is no inherent reason why $\overline{n}_i$ should be related mathematically to O_i or R_i. But empirically in most cases, it is still positively related, supporting the idea that there is some underlying biology in the RONR or common-is-common pattern. Note though that in at least one case (R_i vs. $\overline{n}_i$ for the USFIA), the relationship is negative, showing that (1) this set of variables can in principle vary independently, and (2) we need more work on studying the measures of abundance, the RONR pattern and Eqs. 2A–C. Given how extreme events work, the maximum abundance (max n_i) should be a function of both mean abundance $\overline{n}_i$ and the number of sites where species i is found (C_i) since the more samples and the higher the mean, the higher the extreme event (maximum). Thus mean abundance ($\overline{n}_i$) is more appropriate when seeking a variable that is independent of range size (R_i) and occupancy (O_i). But one should not throw out maximum abundance (max n_i) because max n_i often has stronger correlations with occupancy and range size than average ($\overline{n}_i$) abundance (Table 4), suggesting it may be closer to the mechanistic cause. It has been hypothesized that the range size can in fact be driven by maximum abundance (McGill and Collins, 2003). While important to disentangle these three abundances for better understanding of mechanism, it is convenient that all three measures of regional/global abundance (N_i, $\overline{n}_i$ and max n_i) are strongly positively correlated (r = 0.8–0.9 in most cases in Table 4), a fact only previously reported once for birds (McGill and Collins, 2003) to my knowledge.

TABLE 4. Correlation coefficients for species properties for (A) Barro Colorado Island (BCI) and (B) U. S. Forest Inventory Analysis (USFIA).

Variable	$\log_{10} N_\text{i}$	$\log_{10}$ avg n_i	$\log_{10}$ max n_i	O_i	$\log_{10} C_\text{i}$	$\log_{10}$ Crange_i	$\log_{10} RS_\text{i}$
A) BCI							
$\log_{10} N_\text{i}$	1.00	0.83	0.92	0.94			
$\log_{10}$ avg n_i	0.83	1.00	0.93	0.76			
$\log_{10}$ max n_i	0.92	0.93	1.00	0.85			
O_i	0.94	0.76	0.85	1.00			
B) USFIA							
$\log_{10} N_\text{i}$	1.00	0.66	0.92	0.23	0.90	0.63	0.38
$\log_{10}$ avg n_i	0.66	1.00	0.85	0.37	0.28	0.10	−0.10
$\log_{10}$ max n_i	0.92	0.85	1.00	0.25	0.69	0.42	0.18
O_i	0.23	0.37	0.25	1.00	0.07	−0.48	−0.52
$\log_{10} C_\text{i}$	0.90	0.28	0.69	0.07	1.00	0.81	0.59
$\log_{10}$ Crange_i	0.63	0.10	0.42	−0.48	0.81	1.00	0.80
$\log_{10} RS_\text{i}$	0.38	−0.10	0.18	−0.52	0.59	0.80	1.00

The one interrelationship that is strong and that has been well studied is the link between occupancy and abundance (the occupancy abundance relationship or ONR in the notation of this paper although sometimes called the OAR). Occupancy, O_i is strongly positively related to total abundance N_i. But unlike the relationships between different measures of abundance and range size, which is just a correlation cloud, this relationship is tight enough to suggest fairly specific mathematical forms (Fig. 8A). O_i vs. N_i with N_i on an arithmetic scale typically produces a Monod (Michaelis–Menten)-like curve that saturates at $O_i = 1$ for high enough abundances, while plotting N_i on a logarithmic scale often gives a sigmoidal (logistic) relationship with $O_i \approx 0$ for low abundances and $O_i \approx 1$ for high abundances with a ramping up in between (Hanski and Gyllenberg, 1997; Gaston et al., 2000; Holt et al., 2002; Gaston and He, 2011). As already noted, there is a circularity in that N_i must be mathematically linked to O_i (Eq. 2A–C). Thus, it is common to plot O_i vs $\overline{n}_i$ instead, which typically reveals a similar strongly positive relationship of the same shape, but slightly less tight. If we use $\overline{n}_i$ averaged not only over sites where present (as done above), but over the zeros (giving a true mean abundance across all sites), then a simple Poisson sampling assumption gives us a mathematical formula similar to the Monod form: $O_i = 1{-}\exp(\overline{n}_i)$, or, since the dependency of O_i on i comes only through $\overline{n}_i$ we can just write $O(\overline{n}_i) = 1{-}\exp(\overline{n}_i)$. In practice, this approximation does not work very well, because individuals are clumped rather than spatially random. A number of more accurate forms can all be subsumed (He and Gaston, 2000b; He et al., 2002; Holt et al., 2002; Gaston and He, 2011) in the general equation:

$$O(\overline{n}_i) = 1 - \left(1 + \frac{a\overline{n}_i^{\,\beta}}{k}\right)^{-k}. \tag{3}$$

This model includes the Poisson ($k = \pm\,\infty$, $\alpha = \beta = 1$), the negative binomial distribution ($\alpha = \beta = 1$), the logistic model ($k = 1$) (Hanski, 1992; Hanski and Gyllenberg, 1997), and the Nachman model ($k = \pm\,\infty$) (Nachman, 1981). For a good review of this topic, see Holt and colleagues (2002). One closely related pattern is the occupancy–area relationship or OAR, which predicts occupancy as a function of the area sampled. Occupancy ultimately increases to 100% ($O_i = 1.0$) for large enough areas (Fig. 5D). This is known as the "area saturation effect" (Storch et al., 2003).

There are at least two other relations similar to the ONR. In common with the species–area relationship (SAR), the OAR can be derived from the ONR (occupancy–abundance relationship) and the NAR (abundance–area relationship). Because the NAR is of the form $N = cA$, the OAR must take the same form as the ONR (Eq. 3) with area a substituted for $\overline{n}_i$ (He and Gaston, 2000a; Gaston and He, 2011). A second pattern closely related to the ONR, is when a single population is followed over time, and if the population size varies widely, then an *intraspecific* occupancy–abundance relationship or temporal occupancy–abundance relationship (tONR) can be calculated (Gaston and He, 2011). Among other applications, the tONR can be useful for predicting the infection rate of plants by pests (Nachman, 1981).

Summary of patterns—In summary, there are over a dozen well-known, well-documented patterns (Table 3) relating to biodiversity metrics of abundance, species richness, occupancy, and area (Brown, 1995; Rosenzweig, 1995; Gaston and Blackburn, 2000). As shown in this paper, it is highly likely to be able to take any appropriate data set and find these patterns in the chosen data, making these patterns very general and close to universal in nature. I now turn to showing how these patterns are linked together through the notion of correlated sampling and summarize the mathematical theory developed in this direction.

EXPLANATIONS I: RANDOM (UNCORRELATED) SAMPLING

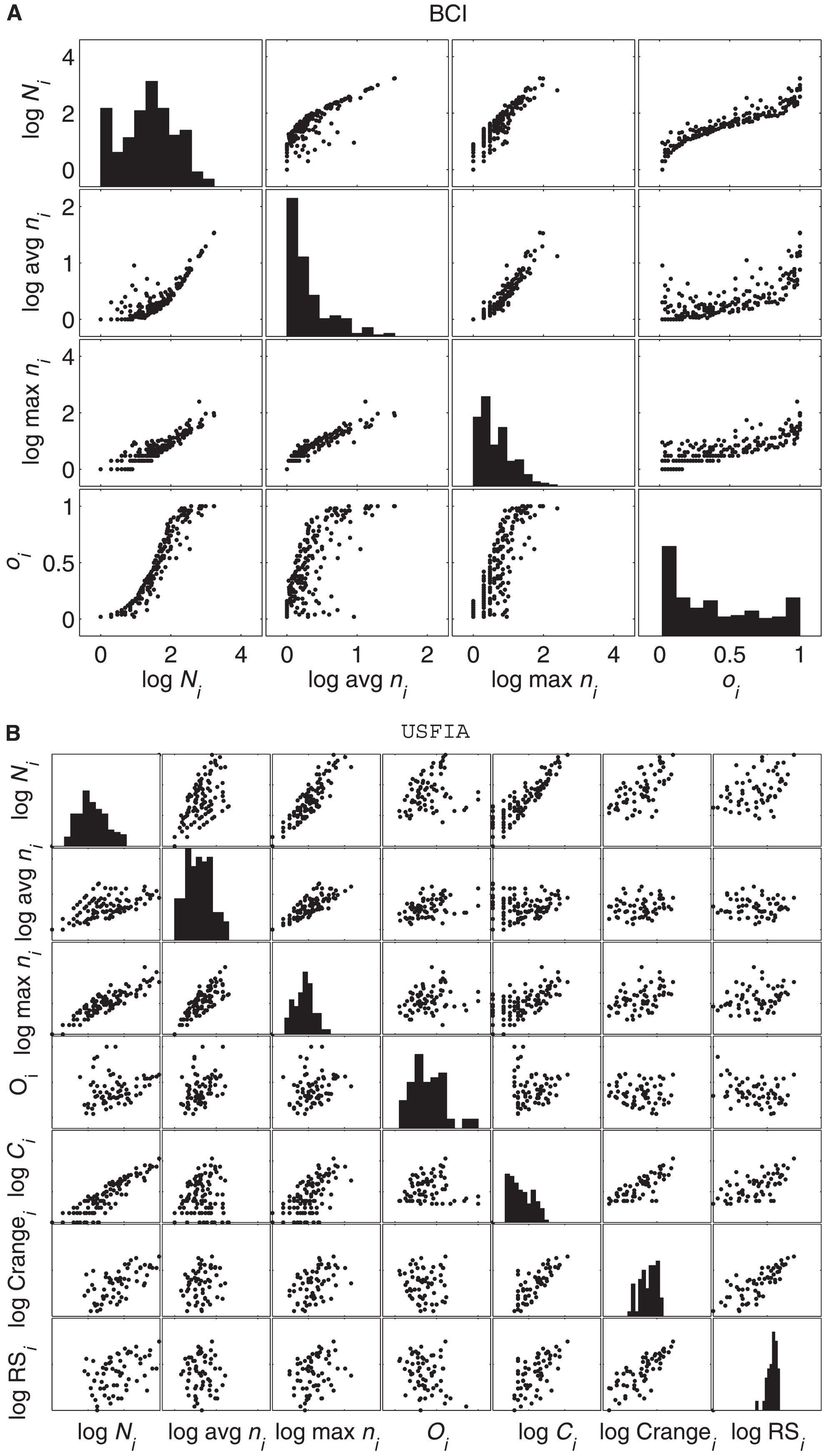
The simplest model of sampling is where each sampling event is independent from the others. There have already been several indications that this model is inadequate for biodiversity data (e.g., the need to go beyond the Poisson-derived exponential form of the ONR or occupancy–abundance relationship to the more general Eq. 3). But mathematical parsimony demands a thorough analysis of the simplest form of sampling—namely sampling where events are independent or uncorrelated. One canonical model for such independent sampling is the placement of marbles into urns. To create an exact analogy to the biodiversity patterns, suppose one has a large tray of area A packed with N^* marbles with S different colors. For each color, one knows the number of marbles, N_i, of that color on the big tray. Now suppose one has C smaller trays, each of area a ($a = A/C$) such that it will hold about $n^* = N^*a/A$ marbles. Suppose one randomly distributes the marbles on the big tray into the small trays. One could then count how many marbles of each color i occurred on tray j ($n_{i,j}$) and count the number of colors on each tray (s_j). One could then calculate all of the patterns that were described in the previous section. Next, one could start combining adjacent trays to calculate the SAR, one could take individuals one by one to make SNRs, one could look at the global and local SADs (lSAD, gSAD) or distributions of N_i and n_i, and so on. Clearly, if this worked, then the patterns listed in Table 2 and studied broadly would not have much biology behind them. Such a pure random sampling model is an obvious one and has been tried before (Gaston and Blackburn, 2000; McGill, 2003; Dolman and Blackburn, 2004; Zillio and Condit, 2007). How well does this most parsimonious sampling model work?

I first briefly examine the most extreme parsimonious version of this where the N_i are all equal (i.e., $N_i = N^*/S$). This requires only S and N^* as inputs. Clearly, this will fail to produce a gSAD and probably an lSAD similar to empirical patterns (Figs. 3, 4), but does it affect the other patterns? In fact, many patterns beyond the two SAD patterns also fail to match empirical data in this case. The local richness, s, is too low (Fig. 9B), and the species–area relationship shows much too fast an accumulation of species (Fig. 5B). In summary, the α-diversity (local community) is too low, and the β-diversity (between community) is too high. Understanding the causes of unequal abundances appears critical to understanding the causes of overall biodiversity structure.

What about the slightly less parsimonious model where not only N^* and S are taken as inputs but also the log-normal-like distribution of N_i [or equiavelntly $M(j)$] is taken as a constraint or input assumption? This can be studied by taking observed values of N_i and then using computer Monte Carlo techniques to randomly sample from this regional "tray" of N^* individual marbles to create C smaller local "trays" of marbles. Is this sufficient to reproduce the biodiversity patterns one observes? It is certainly a marked improvement over the assumption of equal abundances. On some fronts, the random null model works adequately. In particular, it produces lSADs that are reasonably realistic (McGill, 2003; Dolman and Blackburn, 2004; Green and Plotkin, 2007). But Monte Carlo simulations where one samples from a regional pool defined by the N_i from BCI identify several problems. For one thing, they tend to over-predict the number of species in a local community, s (Fig. 9B) to such a degree that the hypothesis that random sampling drives local richness is highly significantly rejected ($P < 10^{-8}$). In contrast, the individual communities are also too similar (Fig. 9D) with the average Morisita–Horn similarity index being much higher in the randomly sampled communities than the observed community. In a similar finding on β-diversity, the collector's curve or species individual relationship (SNR) overpredicts richness for smaller areas relative to the observed species area relationship (SAR) (Fig. 5B), implying there is more autocorrelation of individuals and thus less accumulation of new species in the real world (SAR) than a purely random, independent sample (i.e., SNR). In short, α-diversity is too high and β-diversity is too low in the randomly (uncorrelated) sampled communities. (One necessarily follows from the other because $\beta = \gamma/\alpha$.) Thus, using equal N_i causes α-diversity to be too low and β-diversity to be too high, while using realistically unequal N_i overcorrects, leaving α-diversity too high and β-diversity too low. There must be an additional missing factor that returns results to the middle so as to match empirical reality (see Fig. 9B, D). What is it?

Both models so far have used random or uncorrelated or Poisson sampling which assumes that individuals within a species are placed spatially randomly with respect to each other (and thus sampled independently), despite knowing that intraspecific individuals are in fact clumped together in space. This could easily explain the observation of too high α-diversity and too low β-diversity. If individuals are intraspecifically clumped, this will tend to make the individuals in any one local community more likely than chance to come from a single species (thus driving down α-diversity), while making it more likely than chance to encounter a new clump representing a new species as one moves across space (thus increasing β-diversity). Thus, intraspecific clumping has the potential to scale back the overcorrection found in using unequal N_i, thereby finally producing intermediate results that match reality. Does it actually work? It is not easy to do Monte Carlo computer simulations that clump individuals spatially while simultaneously observing the other constraints we need to obey (e.g., $\Sigma n_i = n^*$). Fortunately, sufficient analytical theory has been developed which matches the Monte Carlo spatially random placement of N_i individuals while providing an easy method to then include clumping.

In fact, there are two completely independent lines of theoretical development that achieve this goal of realistically modeling sampling of spatially clumped individuals. As already discussed in the Setting the stage section, there are two types of data: data collected on a grid at usually smaller spatial scales and data collected at transects over larger spatial scales. It turns out that these two types of data parallel the two different models


A
BCI
log N_i
log avg n_i
log max n_i
o_i
0
2
4
1
0.5
B
USFIA
log N_i
log avg n_i
log max n_i
O_i
log C_i
log Crange_i
log RS_i

for incorporating clumping. I turn to the first such line of theory/type of data in the next section.

EXPLANATIONS II: SPATIALLY IMPLICIT OCCUPANCY (SIO)

In this section, I address the models that have grown up around the first type of data where one has a lattice or grid laid out over an area. In this case, the regional community (γ-scale) is the total area and the local community (α-scale) is a single grid cell. The representation of geographic extent is given by occupancy, O_i (in the next section, I will address the other type of data with a focus on range size R_i). Although the grids in typical data are spatially explicit (one knows the distance between any two grid cells), the modeling approach to this data has remained largely spatially implicit—each grid cell is modeled in isolation. Thus, I call this the spatially implicit occupancy or SIO model.

Analytical theory development—The key step to developing an analytical approach in the SIO model is to specify a sampling function, which I will denote by σ to avoid confusion with species richness *s*. The sampling function gives the probability that *n* individuals of a species are present in a local community of area *a*, given that it has abundance N_i in the regional pool of area *A*. The most general form is $\sigma(n_i|N_i,a,A,\theta)$, where θ is a vector of parameters. Often times, the function can be simplified to have no parameters and to depend only on the fraction of area in the local community—$\sigma(n_i|N_i,a/A)$ or even to depend only on the total number of individuals of that one species expected in the local community of $\sigma(n_i|a/AN_i)$.

Once the sampling function is known, it is easy to calculate a number of properties. The lSAD is given (Pielou, 1977; Dewdney, 1998; Green and Plotkin, 2007) by

$$\mu(j)=\sum_{k=1}^{\infty}M(k)\sigma(j|k,a,A,\theta) \tag{4}$$

Thus one sums over the different possible regional abundances ($k = 1\ldots\infty$) and multiplies the probability, $M(k)$, that a species has regional abundance $N_i = k$ multiplied by the probability that a species with that regional abundance has a local abundance of $n_i = j$ in a local community of size *a* (given by the function σ).

A key special case of the sampling function is the odds that at least one individual is present (i.e., occupancy). This is most easily calculated as one minus the probability of 0 individuals:

$$\begin{aligned}O_i&=O(N_i|a,A,\theta)=\sum_{j=1}^{\infty}\sigma(j|N_i,a,A,\theta)\\&=1-\sigma(0|N_i,a,A,\theta)\end{aligned} \tag{5}$$

This has immediate application to calculating α-diversity (local community richness) and the species area relationship (SAR). To wit, in area *a*, the expected number of species is (Arrhenius, 1921; Coleman, 1981; He and Legendre, 2002):

$$\begin{aligned}s(a)=E(s|a)&=\sum_{i=1}^{S}O(N_i|a,A,\theta)\\&=\sum_{i=1}^{S}\sum_{j=1}^{\infty}\sigma(j|N_i,a,A,\theta)=\sum_{i=1}^{S}1-\sigma(0|N_i,a,A,\theta)\\&=S-\sum_{i=1}^{S}\sigma(0|N_i,a,A,\theta).\end{aligned} \tag{6}$$

Storch and colleagues (2003) gave not just the expected value of *S* but the full probability distribution of $s(a)$ and the variance of $s(a)$. Thus, if one takes α-diversity as occurring at the scale *a*, then Eq. 6 gives a formula for *s*, and if one varies *a* anywhere from the α-scale *a* to the γ-scale *A*, then Eq. 6 gives the species–area relationship (SAR), giving *s* as a function of *a*.

Finally, one can calculate pointwise β-diversity such as would be used for a similarity–distance relation (SDR or DSD) calculation. Two limitations occur in this scenario. First, the measure of similarity used must be based on presence/absence only (not abundance) because the SIO model focuses on occupancy. Typically, the Sorenson index (Koleff et al., 2003) is used. Second, the β-diversity calculation represents the turnover in species composition between two local (α) communities drawn from the same regional community, but the SIO model is spatially implicit—there is no concept of how far apart those two local communities are. They are in effect just two different draws from the same regional pool. Thus, the similarity is a function of area, *a*, but not distance. The SIO formula (Plotkin and Muller-Landau, 2002; Morlon et al., 2008) for Sorenson similarity is

$$\begin{aligned}\beta_{\text{Sorenson}}(a)&=\frac{\text{no. species shared}}{\text{mean no. species in one site}}\\&=\frac{\sum_{j=1}^{\infty}M(j)\sigma(j|a,A,\theta)^2}{\sum_{j=1}^{\infty}M(j)\sigma(j|a,A,\theta)}\end{aligned} \tag{7}$$

Thus, knowing the regional parameters N^* and *S*, the regional gSAD [N_i or equivalently $M(j)$] and the sampling function, σ, explains many other biodiversity patterns. These include the lSAD (Eq. 4), the occupancy–abundance relation (ONR) (Eq. 5), the species–area relationship (SAR) (Eq. 6) and the similarity–distance relation (SDR or DSD) (Eq. 7).

SIO model with random (uncorrelated) sampling function σ—The obvious question is what to use for the sampling function, σ? The exact analog of the modeling of random placement of marbles or Monte Carlo simulations used in the previous section is to use the Poisson or binomial distributions.

←

Fig. 8. Scatter plots for relations between pairs of variables. Top grid is for BCI, lower grid is for USFIA. The scatter plot for the relation between any two variables can be found at the intersection of those variables along the horizontal and vertical axes. The histogram of distributions is shown on the diagonal. All logarithms are base 10. In particular, the ONR (occupancy–abundance relationship) is shown in the fourth row, first column in both graphs. The RNR (range–size–abundance relationship) is shown in the last row, first column in the second (USFIA) graph. The distribution of log abundances (log N_{i}) or gSAD is shown in the first row, first column of both graphs. The distribution of occupancies (SOD) is shown in the fourth row, fourth column of both graphs. The distribution of range sizes (SRSD) is shown in the bottom right corner of the second (USFIA) graph.

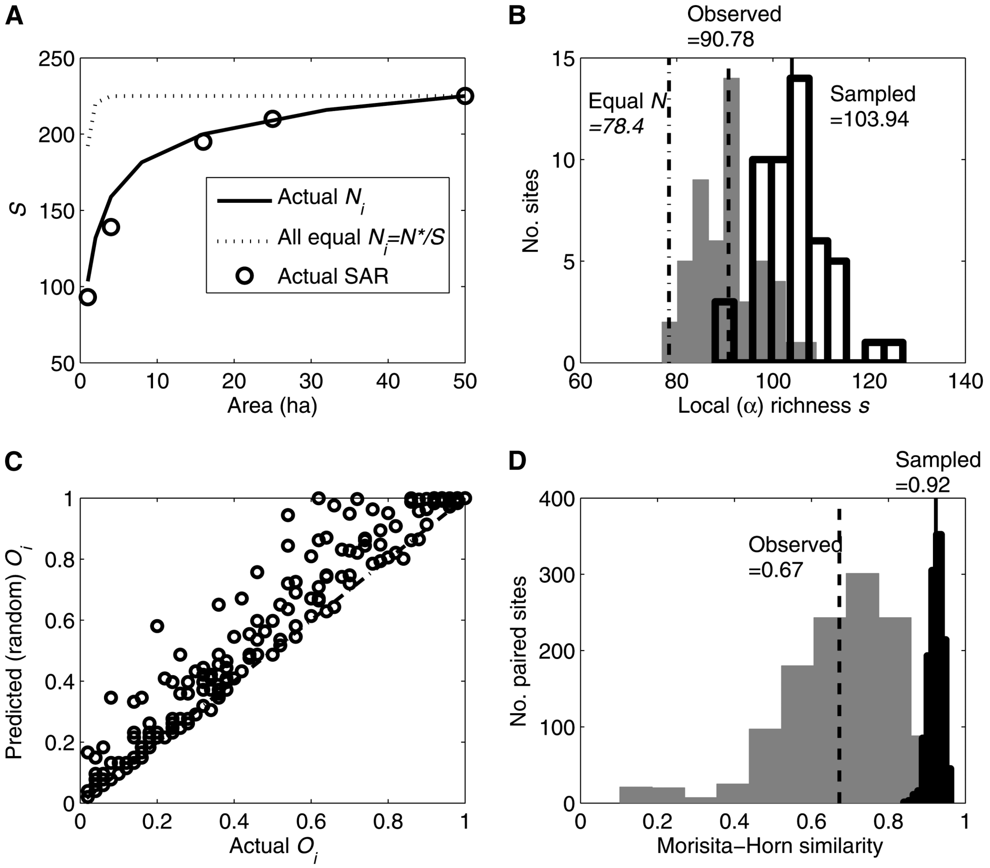


Fig. 9. Tests of random sampling null model. All graphs use BCI data. (A) Plot of the actual species–area–relationship SAR (dots) for BCI vs. that predicted by Eq. 6 using observed values of N_i and that predicted by Eq. 6 if all species are equally abundant. Note how the plot using observed values of N_i is very similar to the rarefaction line shown in Fig. 5B. This demonstrates Monte Carlo and SIO analytical approaches do converge. (B) Local communities of the same sample size (n*) as the observed communities were drawn without replacement. The histogram for species richness, s, in 50 Monte Carlo replicates is significantly to the right of (higher α-diversity) than the histogram for the 50 actually observed values of s. The value of s obtained under an assumption of equal abundances is lower than the observed values. (C) Plot of the observed and predicted (Eq. 5 using Poisson sampling) values of O_i vs. a 1-1 line. Observed values are consistently lower than predicted. (D) The same 50 random and actual communities used in (B) were generated. The Morisita–Horn similarity between pairs of communities was calculated across many pairs, and the resulting histograms are given. Random plots have very high similarity of community structure. Panels (A) and (C) show the shortcomings of the analytical SIO model with Poisson sampling, while (B) and (D) show shortcomings of the equivalent Monte Carlo version.

Both represent randomness, but they are useful in different contexts (Zillio and He, 2010). Specifically, the Poisson takes a rate (mean number of species per unit area) and is applicable when the area rather than the sample size is known and allows the area to be continuous (e.g., a = 4.5723). The binomial is applicable when the finite sample size is known (and discrete, e.g., 4 or 5 but not 4.5723) and the precise probability of the event of interest occurring in a single sample is known. The idea of binomial and Poisson sampling functions have actually been in the ecological literature for a long time (Arrhenius, 1921; Pielou, 1977), but have become a central construct in theoretical ecology recently (Plotkin and Muller-Landau, 2002; Green and Plotkin, 2007; Harte, 2008; Morlon et al., 2009; Zillio and He, 2010).

As one example, it has been shown that under Poisson sampling, the lSAD is just a rescaled version of the gSAD (Green and Plotkin, 2007) with the abundance scale rescaled by N^*/n^*. As another example, to reproduce the SAR, one can use a binomial sampling formula, so the probability of having n_i individuals is

$$\sigma\left(n_i \middle| N_i, a, A\right) = \binom{N_i}{n_i}\left(\frac{a}{A}\right)^{n_i}\left(1-\frac{a}{A}\right)^{N_i-n_i} \tag{8}$$

So specifically, the probability of having 0 individuals in an area a is

$$\left(1-\frac{a}{A}\right)^{N_i},$$

and so the occupancy or probability of having at least 1 individual is

$$O_i = O\left(N_i \middle| a, A\right) = 1-\left(1-\frac{a}{A}\right)^{N_i}.$$

Then the expected number of species s, $E(s)$, in a local community is simply

$$E(s) = \sum_{i=1}^{S} O_i = S - \sum_i \left(1-\frac{a}{A}\right)^{N_i}. \tag{9}$$

In principle, these analytical approaches using a random (uncorrelated) σ sampling function should work no better and no worse than the Monte Carlo approaches, which is in fact true. The performance of these analytical calculations using realistically unequal abundances [i.e., $M(j)$ is log-normal/log-series-like], but with spatially random positioning of intraspecific individuals (i.e., Poisson or binomial σ function) can be evaluated (Fig. 9A, C) and compared to earlier results from Monte Carlo simulations. Equations 6 and 9 produce a realistically shaped SAR that actually falls quite close to the empirically found SAR (Fig. 9A), although just as in the Monte Carlo, it consistently over predicts species richness for smaller areas (cf. Fig. 5B). Similarly, Eq. 5 (also see Eq. 3) provides a formula predicting the ONR (occupancy O_i as a function of abundance N_i), that clearly is related to the actual data (Fig. 9D) but consistently overpredicts occupancy, O_i. So as expected, the analytical model with random (Poisson or binomial) σ and the Monte Carlo approach behave identically—producing results that are close to empirical observations (so long as a realistic distribution of abundance N_i is used), but which appears to systematically overestimate α-diversity while underestimating β-diversity.

SIO model with clumped (correlated) sampling function σ—So what form of σ can be used to realistically incorporate the spatial clumping and correlated sampling? The simplest way to do this is to use the negative binomial distribution (NBD) (Bliss and Fisher, 1953; Boswell and Patil, 1971; Evans et al., 1993). While the Poisson distribution gives values of $\sigma(n_i)$ that match randomness, the NBD gives values for $\sigma(n_i)$ that show clumping by having both a greater number of zeros and of high abundances but fewer intermediate abundances relative to the Poisson. Multiple authors have shown that biodiversity patterns are more accurately predicted when a clumped (almost always NBD) sampling function is used. This includes the species–area relationship (SAR) (He and Legendre, 2002), the lSAD (Green and Plotkin, 2007) and the similarity–distance relationship (SDR or DSD) (Plotkin and Muller-Landau, 2002; Morlon et al., 2009). While the NBD serves as an analogue of the Poisson, no obvious clumped analogue of the binomial existed until Zillio and He (2010) recently provided one. Thus, a model using a log-normal-like gSAD and intraspecific spatial clumping expressed by a NBD sampling function can quite accurately reproduce many well-known biodiversity patterns.

In answer to the question set at the end of the section Explanations I, it appears that using a log-normal-like regional pool of abundances (N_i) in combination with intraspecific clumping producing autocorrelated samples seem to balance each other out in terms of α- vs. β-diversity in just such a fashion that they produce very empirically realistic results.

Moving the SIO beyond spatial implicitness—Despite its successes, the SIO with a clumped sampling function remains spatially implicit—there is an ability to tell us what any one grid cell will look like but not how two adjacent grid cells will be similar or different. This is particularly problematic if one wants to look at β-diversity. There have been several recent attempts to address this. Two models become more spatially explicitly by looking at neighbor correlation—i.e., some measure of similarity between strictly adjacent cells. Conlisk and colleagues (2009) examined the occupancy–abundance relationship (ONR). They showed that not only does using a clumped distribution within one cell such as the NBD improve results, but providing a correction term for spatially explicit clumping across neighbors (specifically Moran's I value for adjacent grid cells) further improves the fit and in total explains much of the variance found in the ONR relationship. Hui and colleagues (2009) developed a spatially explicit occupancy model with extensions similar in spirit to those of Conlisk by looking at autocorrelation with neighboring cells and showed that in general such spatially explicit extensions to occupancy models improve the ability of the models to produce a wide variety of biodiversity patterns.

Two other models implement global (i.e., not just neighbor) models of spatial structure. Morlon and colleagues (Morlon et al., 2009) showed that the most general way to fix this (in the context of β-diversity) is by extending the sampling function σ to focus on probability of finding an individual at a distance d from a focal individual. This has direct ties to the pair correlation function (Stoyan and Stoyan, 1994; Wiegand and Moloney, 2004; McGill, 2011). In another approach to becoming more spatially explicit, Kunin and colleagues (Kunin, 1998) built a model centered on the occupancy abundance (ONR) relationship but added the assumption of fractal structure across scales, creating a form of spatial explicitness. All of these more spatially explicit occupancy models building on the occupancy-focused SIO model show promise and should be pursued further. However, I now turn to a completely different way to build a spatially explicit model.

EXPLANATIONS III: SPATIALLY EXPLICIT RANGES (SER)

The model for spatial structure of a species in the SIO models discussed in the previous section focused almost entirely on occupancy of small grid cells and makes the assumption that a species is equally likely to appear in each grid cell. This allowed the SIO to be spatially implicit in nature. This is not useful at larger spatial scales where species have more or less spatially coherent geographic ranges and the probability of finding a species is essentially zero outside of the geographic range. Several authors noticed that the equally likely everywhere assumption of the SER model (Coleman, 1981; He and Legendre, 2002) for SARs was unrealistic and had limited applicability to large spatial scales (Hui et al., 2009), so they started to model spatially contiguous ranges. I turn now to an alternative to the SIO modeling framework that is spatially explicit and focuses on ranges (the SER model) and is useful at scales up to entire continents. In essence, the SER approach replaces the correlated σ with the species range.

In the SER model, species ranges are assumed to be circular and vary in size. I have so far used R_i to denote the area of a geographic range, but in this section I will use R_i to denote the radius and switch to using RS_i for range area [i.e., $R_i = \sqrt{(RS_i/\pi)}$]. The circularity assumption is more for convenience than anything and does not materially affect results (Leitner and Rosenzweig, 1997 used rectangular ranges). But of course the variation in size is critical. I do not bother here to develop a model assuming all RS_i are equal as was done with N_i for SIO model, but it should come as no surprise that if one did, such a model would work badly. Instead, the SER model starts with what is already known about species ranges: RS_i is roughly log-normally distributed and is positively correlated with N_i. Thus one can either start with log-normal N_i and derive the distribution of RS_i (McGill and Collins, 2003) or just directly assume a log-normal-like distribution of the RS_i (Allen and White, 2003). The pattern receiving the most attention in the SER model is the species–area relationships (SAR) derived by computer simulations (Leitner and Rosenzweig, 1997; Maurer, 1999; Ney-Nifle and Mangel, 1999; McGill and Collins, 2003). More recently, an analytical model has been provided (Allen and White, 2003), and I will summarize this analytical approach here.

Let $f(R)$ be the probability density function for range sizes (e.g., the log-normal density). Then the average range radius is

$$\bar{R} = \int f(R) R dr,$$

and the average range size is

$$\bar{A}_R = \pi \int f(R) R^2 dr.$$

Assume the centers of the S species ranges are placed at random (spatial Poisson process) and ignore edge effects including mid-domain effects (Colwell and Lees, 2000). Let the average species density (i.e., the intensity parameter in the Poisson process) be $S_D = S/A$. Now focus on a target point (transects are usually treated as points in the SER model). Due to spatial randomness of the ranges, the odds that a given range for species i covers that point is given by the ratio of the area over which that species could cover the point to the total area (i.e., $RS_i/A = \pi R_i^2/A$). The expected total number of species covering that point will be the sum of the probability that each range covers that point, i.e.,

$$s = \sum_{i=1}^{S} \frac{\pi R_i^2}{A}.$$

If range size were constant, then

$$s = S\pi R^2 / A = S_D \bar{A}_R,$$

so α-diversity increases with γ-diversity (S) and with average range size. Now imagine one wants to know how many species ranges will overlap a circular area of radius P having area $a = \pi P^2$ (i.e., $P = \sqrt{\pi a}$). Then if a species range i with radius R_i has a center that falls in the circle of radius $R_i + P$ around the center of our target area, then it will intersect the target area. So, as before, the probability of falling in this circle is equal to the fraction of the total area this circle covers, giving us

$$s = \sum_{i=1}^{S} \frac{\pi (R_i + P)^2}{A}.$$

Returning to the assumption of nonconstant range sizes, where one knows only the distribution of range sizes $f(R)$, one has

$$s = \sum_{i=1}^{S} \int_0^\infty \frac{\pi (R_i + P)^2}{A} dr = \sum_{i=1}^{S} \int_0^\infty \frac{\pi \left(R_i^2 + 2PR_i + P^2\right)}{A} dr.$$

By passing the integral through the sum and recalling the definitions for $\bar{R}$ and $\bar{A}_R$, the SER models gives

$$s(a) = S_D \left(\bar{A}_R + 2\bar{R}\sqrt{\pi a} + a\right).$$

This gives us the species–area relationship (SAR). By taking the derivative with respect to a, one can get the z-value or slope of the SAR. It is clearly nonconstant as a varies. In fact, the SAR derived in the SER model looks like the middle and right phases of the triphasic species area curve (Fig. 6C). The SAR starts flat and then accelerates until it reaches a slope of $z = 1$ at the largest spatial scales. Computer simulations show that this model is a good fit to empirical data for birds of North America above scales of about 100 × 100 km^2 (the same break point found in Fig. 6C for trees). As already noted, at smaller scales the increasing, decelerating function portion of the triphasic SAR is explained by the species–abundance–relationship (SNR) or collectors curve. This suggests that the triphasic SAR is really just biphasic with small scales explained by sampling from the SAD and over large scales explained by the SER model and with the two meeting smoothly in the middle.

To date most development in the literature of the spatially explicit range (SER) model on β-diversity has focused on the SAR. However, this same image of placing down circles at random and using stochastic geometry (Stoyan and Stoyan, 1994) to check how many ranges overlap a point can be used to calculate Sorenson's index as a function of distance between the two points, leading to the decay of similarity with distance curve. To my knowledge, this has only been done by computer simulation for birds (McGill and Collins, 2003). An analytical solution can be written (B. McGill, unpublished), but the formulas are even more complex than for the SAR and do not simplify.

The SER model is also quite successful at explaining the third major pattern, the lSAD. To do this, the SER needs to bring abundance back into the picture. Specifically, the SER model replaces the circular range with a circular abundance surface where the height of the surface indicates the abundance. And in particular, the circular range is replaced by a circular, bivariate Gaussian bell curve. Gauch and Whittaker (1972) explored by computer simulation the placing of such abundance surfaces randomly in space. The gSAD (global SAD) determined the variation in the height of the peaks (i.e., the gSAD described the distribution maximum abundances). Then a sample taken at a point takes abundances from all the ranges that cover the point and the abundances vary depending on where in the range (i.e., height of the abundance surface) the species is found. They show that this reproduces realistic lSADs (e.g., see Fig. 3 in McGill, 2010a). Hengeveld and colleagues first derived analytical proofs of this (Hengeveld et al., 1979; Hengeveld and Haeck, 1981). Later, McGill and Collins (2003) independently derived an analytical solution. McGill and Collins also used Monte Carlo simulations to show that the assumption of a centered Gaussian bell curve is not essential—the model is robust to variation in shape of the abundance surface so long as the abundance surface has small high peaks surrounded by large tails of low abundance.

Thus, just like the SIO model, the SER model can take as input S, N^*, and the unequal global abundances [gSAD or N_i or $M(j)$] and clumping (here implemented by contiguous, compact circular ranges with abundance peaks) and produce three classic biodiversity patterns: the lSAD, the SAR, and the SDR (or DSD).

CONCLUSIONS AND FUTURE DIRECTIONS

I have set forth two mathematical approaches (SIO and SER) that are capable of producing and explaining multiple biodiversity patterns (Fig. 10) including the lSAD (n_i), local species richness (s), and the two most common reflections of β-diversity, the SAR and SDR. The spatially implicit occupancy theory

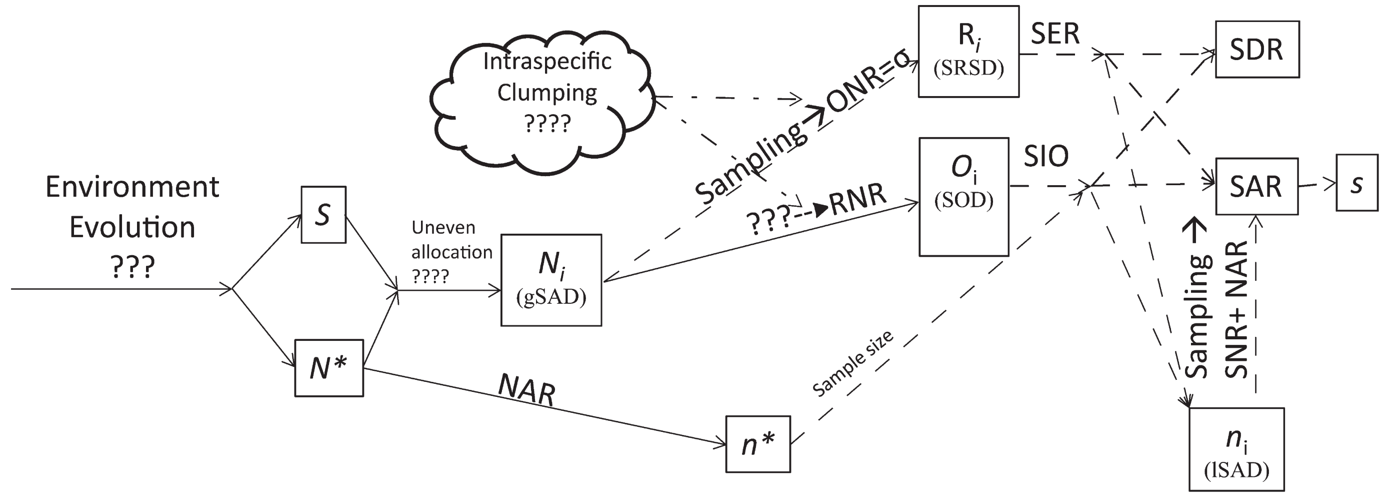


Fig. 10. A pictorial representation of the theory presented in this paper. Variables are drawn in boxes. Patterns linking variables (the interactions of the main text) are shown as solid lines. Links based on known mathematical processes (in all cases here, random processes related to sampling) are shown with a dashed line. Regional richness, S, and total individuals, N^*, are set by the environment (e.g., climate) and evolutionary/historical processes. The N^* individuals are allocated across the S species by unknown processes but in a highly uneven fashion, resulting in a log-normal/log-series-like gSAD (N_i). Given the regional abundance (N_i) of a species, one can predict the occupancy of the species, O_i, from a Poisson sampling process. One can also predict the range size, R_i from N_i as an empirical pattern, but we do not know what processes are involved. Finally, one can predict the total population size of the local community, n^*, using the NAR (abundance area relationship or IAR). Both processes (the sampling of the ONR and the unknown process behind the RNR) are modified from a simple random, independent sample by the fact that individuals within a species tend to be clumped together. Relatively little is known about the patterns of clumping or how exactly they affect the sampling process, so this is shown by the cloud and the dash-dot line. Once this clumped (autocorrelated sampling) is described, multiple patterns can be derived under the SER (given R_i) or SIO (given O_i and n^*). These include the lSAD (local species abundance distribution or n_i) and local species richness, s. The species–area relationship (SAR) can be derived in one of three ways: (1) as a sum of O_i as summarized in the spatially implicit occupancy (SIO) model, (2) as a stochastic geometry problem in the intersection of plots with ranges in the spatially explicit range (SER) model, (3) as a collector's curve or rarefaction merely as a sample of individuals from the lSAD. Models 1 and 3 of the SAR produce a decelerating function and are appropriate for the left (smaller spatial scale) side of the triphasic SAR (Fig. 6C). Model 2 produces an accelerating function and is an appropriate model for the right (large spatial scale) side of the triphasic SAR. Although there is not usually a precise, one-to-one mapping between the SAR and the decay–of similarity with distance (DSD), both the SAR and SDR (similarity–distance relation or DSD) are manifestations of the same underlying spatial structure (i.e., β-diversity) of species and individuals and thus can be derived through the same mathematical models as the SAR.

(SIO) focuses on occupancy and is more appropriate for small spatial scales, while the spatially explicit range (SER) theory focuses on geographic ranges and is more appropriate for large spatial scales.

Both of these theories have several key features in common (Fig. 10) including:

(1) The emphasis on local communities as samples from a regional pool.

(2) The regional pool takes S and N^* as exogenously determined by the environment and history.

(3) The regional pool must be modeled to have the N^* individuals allocated highly unequally over the S species.

(4) The sampling from the regional pool may be approximated as random (independent) but is much more accurately modeled as a highly correlated sample resulting from spatial clumping.

(5) The sample size is determined by the fact that number of individuals in local community scales isometrically (i.e., is proportional) to the area.

None of these five commonalities, the mathematical machinery of the SIO or SER, nor the links in Fig. 10 are novel to this paper. Indeed, most of them have been independently derived multiple times in the literature. And several authors have presented theories that capture a significant subset of what is presented here (Gaston and Blackburn, 2000; He and Legendre, 2002; McGill and Collins, 2003; Harte et al., 2008; Storch et al., 2008; McGill, 2010a). But it is my hope that collecting the entire approach in one location will advance the field.

One desirable direction for advancement would be further mathematical development. Mathematical development might take the form of developing a cross-scale approach that incorporates both the SIO and SER models. I also hope that mathematical developments will also lead to quantitative tools helpful to ecologists and conservation biologists. Beyond the math, I hope that the ideas in this paper will also help to focus and advance future macroecological research. This paper draws strong attention to the importance of the regional–local framing of community ecology and in particular to the top-down (the regional driving the local) view. It also draws attention to the importance of understanding regional controls of S (well studied but still poorly understood) and N (hardly studied). It highlights the unique nature of richness as intermediate between extensive and intensive. And finally, it highlights three patterns as "primary": the gSAD, the intraspecific clumping, and the NAR. Ironically, these three patterns are the least well studied of all those listed in Table 2.

Further thought is also needed on the sampling-based approach described here vs. more mechanistic models mentioned in the beginning such as the population drift of neutral theory or fractal approaches. I argue that parsimony is an important criteria for choosing between theories and that the sampling of the SER/SIO models are maximally parsimonious, but it is reasonable for somebody to counter argue that a single assumption like neutral drift or maxent is more parsimonious. In the end, it is probably too early to decide which approach is more parsimonious while also being an accurate depiction of reality. In particular, we just don't know enough about the nature of intraspecific clumping to feel confident that we have accurately captured this process be it in the SIO σ sampling function or generalized fractals/maxent/netural dispersal limitation/etc. Theoreticians need much more empirical data about clumping including: empirical measurement across many organisms (a strong bias to plants is found in the current measurements), an understanding of how clumping changes with spatial scales, and precise enough measurements that we can begin to develop precise mathematical models.

I believe the ultimate goal of macroecology and the study of biodiversity patterns should be to develop a quantitative predictive theory useful to conservation. This is not only socially utile but is a litmus test for how well we understand the basic processes. I hope this paper is a contribution in this direction, although significant additional work is needed.

LITERATURE CITED

Allen, A. P., and E. P. White. 2003. Effects of range size on species–area relationships. *Evolutionary Ecology Research* 5: 493–499.

Anderson, M. J., T. O. Crist, J. M. Chase, M. Vellend, B. D. Inouye, A. L. Freestone, N. J. Sanders, et al. 2011. Navigating the multiple meanings of β diversity: A roadmap for the practicing ecologist. *Ecology Letters* 14: 19–28.

Andrewartha, H., and L. Birch. 1986. The ecological web: More on the distribution and abundance of animals. University of Chicago Press, Chicago, Illinois, USA.

Arrhenius, O. 1921. Species and area. *Journal of Ecology* 9: 95–99.

Bell, G. 2000. The distribution of abundance in neutral communities. *American Naturalist* 155: 606–617.

Bliss, C., and R. Fisher. 1953. Fitting the negative binomial distribution to biological data. *Biometrics* 9: 176–200.

Boswell, M. T., and G. P. Patil. 1971. Chance mechanisms generating the logarithmic series distribution used in the analysis of number of species and individuals. *In* G. P. Patil, E. C. Pielou, and W. E. Waters [eds.], Statistical ecology, vol. I, Spatial patterns and statistical distributions, 99–130. Pennsylvania State University Press, University Park, Pennsylvania, USA.

Brown, J. H. 1995. Macroecology. University of Chicago Press, Chicago, Illinois, USA.

Brown, J. H., J. F. Gillooly, A. P. Allen, V. M. Savage, and G. B. West. 2004. Toward a metabolic theory of ecology. *Ecology* 81: 1771–1789.

Brown, J. H., and B. A. Maurer. 1987. Evolution of species assemblages: Effects of energetic constraints and species dynamics on the diversification of the North American avifauna. *American Naturalist* 130: 1–17.

Brown, J. H., and B. A. Maurer. 1989. Macroecology: The division of food and space among species on continents. *Science* 243: 1145–1150.

Brown, J. H., D. H. Mehlman, and G. C. Stevens. 1995. Spatial variation in abundance. *Ecology* 76: 2028–2043.

Chao, A. 1987. Estimating the population size for capture–recapture data with unequal catchability. *Biometrics* 43: 783–791.

Coleman, B. 1981. Random placement and species area relations. *Mathematical Biosciences* 54: 191–215.

Colwell, R. K., and D. C. Lees. 2000. The mid-domain effect: Geometric constraints on the geography of species richness. *Trends in Ecology & Evolution* 15: 70–76.

Condit, R. 1998. Tropical forest census plots. Springer-Verlag, Berlin, Germany, and R.G. Landes Georgetown, Texas, USA.

Condit, R., P. S. Ashton, P. Baker, S. Bunyavejchewin, S. Gunatilleke, N. Gunatilleke, S. P. Hubbell, et al. 2000. Spatial patterns in the distribution of tropical tree species. *Science* 288: 1414–1418.

Conlisk, E., J. Conlisk, B. J. Enquist, J. Thompson, and J. Harte. 2009. Improved abundance prediction from presence–absence data. *Global Ecology and Biogeography* 18: 1–10.

Connolly, S. R., T. P. Hughes, D. R. Bellwood, and R. H. Karlson. 2005. Community structure of corals and reef fishes at multiple scales. *Science* 309: 1363–1365.

Connor, E. F., and E. D. McCoy. 1979. The statistics and biology of the species–area relationship. *American Naturalist* 113: 791–833.

Crist, T. O., and J. A. Veech. 2006. Additive partitioning of rarefaction curves and species–area relationships: Unifying α, β and γ diversity with sample size and habitat area. *Ecology Letters* 9: 923–932.

Crist, T. O., J. A. Veech, J. C. Gering, and K. S. Summerville. 2003. Partitioning species diversity across landscapes and regions: A hierarchical analysis of α, β and γ diversity. *American Naturalist* 162: 734–743.

Currie, D. J. 1991. Energy and large scale patterns of animal- and plant-species richness. *American Naturalist* 137: 27–49.

Currie, D., G. Mittelbach, H. Cornell, R. Field, J. Guegan, B. Hawkins, D. Kaufman, et al. 2004. Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters* 7: 1121–1134.

Currie, D. J., and V. Paquin. 1987. Large-scale biogeographical patterns of species richness in trees. *Nature* 329: 326–327.

Dewdney, A. K. 1998. A general theory of the sampling process with applications to the "veil line". *Theoretical Population Biology* 54: 294–302.

Diamond, J. 1986. Community ecology. Harper & Row, New York, New York, USA.

Dolman, A. M., and T. M. Blackburn. 2004. A comparison of random draw and locally neutral models for the avifauna of an English woodland. *BMC Ecology* 4: 8. doi:10.1186/1472-6785/4/8.

Evans, M., N. Hastings, and B. Peacock. 1993. Statistical distributions, 2nd ed. John Wiley, New York, New York, USA.

Fauth, J. E., J. Bernardo, M. Camara, W. J. Resetarits Jr., J. V. Buskirk, and S. A. McCollum. 1996. Simplifying the jargon of community ecology: A conceptual approach. *American Naturalist* 147: 282–286.

Fisher, R. A., A. S. Corbet, and C. B. Williams. 1943. The relation between the number of species and the number of individuals in a random sample from an animal population. *Journal of Animal Ecology* 12: 42–58.

Fortin, M. J., and M. R. T. Dale. 2005. Spatial analysis: A guide for ecologists. Cambridge University Press, Cambridge, UK.

Gaston, K., T. Blackburn, J. Greenwood, R. Gregory, R. Quinn, and J. Lawton. 2000. Abundance–occupancy relationships. *Journal of Applied Ecology* 37: 39–59.

Gaston, K. J. 1990. Patterns in the geographical ranges of species. *Biological Reviews of the Cambridge Philosophical Society* 65: 105–129.

Gaston, K. J. 2003. The structure and dynamics of geographic ranges. Oxford University Press, Oxford, UK.

Gaston, K. J., and T. M. Blackburn. 2000. Pattern and process in macroecology. Blackwell, Oxford, UK.

Gaston, K. J., and F. He. 2011. Species occurence and occupancy. *In* A. E. Magurran and B. J. McGill [eds.], Biological diversity: Frontiers in measurement and assessment, 141–151. Oxford University Press, Oxford, UK.

Gaston, K. J., and W. E. Kunin. 1997. Rare–common differences: an overview. *In* W. E. Kunin and K. J. Gaston [eds.], The biology of rarity: Causes and consequences of rare-common differences, vol. 17, 12–29. Chapman & Hall, London, UK.

Gauch, H. G. J., and R. H. Whittaker. 1972. Coencline simulation. *Ecology* 53: 446–451.

Gholz, H. 1982. Environmental limits on aboveground net primary production, leaf area, and biomass in vegetation zones of the Pacific Northwest. *Ecology* 63: 469–481.

Gotelli, N. J., and R. K. Colwell. 2001. Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 4: 379–391.

Gotelli, N. J., and R. K. Colwell. 2011. Estimating species richness. *In* A. E. Magurran and B. J. McGill [eds.], Biological diversity: Frontiers in measurement and assessment, 39–54. Oxford University Press, Oxford, UK.

Green, J. L., and J. B. Plotkin. 2007. A statistical theory for sampling species abundances. *Ecology Letters* 10: 1037–1045.

Gregory, R. D. 1994. Species abundance patterns of British birds. *Proceedings of the Royal Society of London, B, Biological Sciences* 257: 299–301.

Gregory, R. D. 2000. Abundance patterns of European breeding birds. *Ecography* 23: 201–208.

Hanski, I. 1992. Inferences from ecological incidence functions. *American Naturalist* 139: 657–662.

Hanski, I., and M. Gyllenberg. 1997. Uniting two general patterns in the distribution of species. *Science* 275: 397–400.

Harte, J. 2008. From spatial pattern in the distribution and abundance of species to a unified theory of ecology: The role of maximum entropy methods. *Applied Optimization* 102: 243.

Harte, J., A. Smith, and D. Storch. 2009. Biodiversity scales from plots to biomes with a universal species–area curve. *Ecology Letters* 12: 789–797.

Harte, J., T. Zillio, E. Conlisk, and A. B. Smith. 2008. Maximum entropy and the state variable approach to macroecology. *Ecology* 89: 2700–2711.

Hawkins, B., R. Field, H. Cornell, D. Currie, J. Guégan, D. Kaufman, J. Kerr, et al. 2003. Energy, water, and broad-scale geographic patterns of species richness. *Ecology* 84: 3105–3117.

He, F., and K. J. Gaston. 2000a. Estimating species abundance from occurrence. *American Naturalist* 156: 553–559.

He, F., and K. J. Gaston. 2000b. Occupancy–abundance relationships and sampling scales. *Ecography* 23: 503–511.

He, F. L., and K. J. Gaston. 2000c. Occupancy–abundance relationships and sampling scales. *Ecography* 23: 503–511.

He, F., K. J. Gaston, and J. Wu. 2002. On species occupancy–abundance models. *Ecoscience* 9: 119–126.

He, F. L., and P. Legendre. 2002. Species diversity patterns derived from species–area models. *Ecology* 83: 1185–1198.

Hengeveld, R., and J. Haeck. 1981. The distribution of abundance. II. Models and implications. *Proceedings of the Koninklijke Nederlandse Akadamie Van Wetenschappen C* 84: 257–284.

Hengeveld, R., S. A. L. M. Kooijman, and C. Taillie. 1979. A spatial model explaining species–abundance curves. *In* J. K. Ord, G. P. Patil, and C. Taillie [eds.], Statistical distributions in ecological work, 337–347. International Co-operative Publishing House, Fairland, Maryland, USA.

Holt, A. R., K. J. Gaston, and F. L. He. 2002. Occupancy–abundance relationships and spatial distribution: A review. *Basic and Applied Ecology* 3: 1–13.

Hubbell, S. P. 2001. A unified theory of biodiversity and biogeography. Princeton University Press, Princeton, New Jersey, USA.

Hubbel, S. P., R. Condit, and R. B. Foster. 2005. Forest census plot on Barro Colorado. Website https://ctfs.arnarb.harvard.edu/webatlas/datasets/bci. Center for Tropical Forest Science, Smithsonian Tropical Research Institute, Panama City, Panama.

Hubbell, S. P., R. B. Foster, S. T. O'Brien, K. E. Harms, R. Condit, B. Wechsler, S. J. Wright, et al. 1999. Light-gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. *Science* 283: 554.

Hui, C., M. McGeoch, B. Reyers, P. Roux, M. Greve, and S. Chown. 2009. Extrapolating population size from the occupancy–abundance relationship and the scaling pattern of occupancy. *Ecological Applications* 19: 2038–2048.

Hurlbert, S. H. 1971. The nonconcept of species diversity: A critque and alternative parameters. *Ecology* 52: 577–586.

Jost, L., A. Chao, and R. L. Chazdon. 2011. Compositional similarity and beta diversity. *In* A. E. Magurran and B. J. McGill [eds.], Biological diversity: Frontiers in measurement and assessment, 66–84. Oxford University Press, Oxford, UK.

Koleff, P., K. J. Gaston, and J. J. Lennon. 2003. Measuring beta diversity for presence–absence data. *Journal of Animal Ecology* 72: 367–382.

Krebs, C. 1972. Ecology: The experimental analysis of distribution and abundance. Harper and Row, New York, New York, USA.

Kunin, W. E. 1998. Extrapolating species abundance across spatial scales. *Science* 281: 1513.

Lakatos, I. 1978. Introduction: Science and pseudoscience. *In* J. Worrall and G. Currie [eds.], The methodology of scientific research programs, 1–8. Cambridge University Press, Cambridge, UK.

Lande, R. 1996. Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos* 76: 5–13.

Latham, R., and R. Ricklefs. 1993. Global patterns of tree species richness in moist forests: Energy-diversity theory does not account for variation in species richness. *Oikos* 67: 325–333.

Legendre, P., and L. Legendre. 1998. Numerical ecology. Elsevier, Amsterdam, Netherlands.

Leitner, W. A., and M. L. Rosenzweig. 1997. Nested species–area curves and stochastic sampling: A new theory. *Oikos* 79: 503–512.

Loreau, M. 2000. Are communities saturated? On the relationship between, and diversity. *Ecology Letters* 3: 73–76.

MacArthur, J. W. 1975. Environmental fluctuations and species diversity. *In* M. L. Cody and J. M. Diamond [eds.], Ecology and evolution of communities, 74–80. Belknap Press of Harvard University Press, Cambridge, Massachusetts, USA.

MacArthur, R. H. 1968. The theory of the niche. *In* R. Lewontin [ed.], Population biology and evolution, 159–176. Syracuse University Press, Syracuse, New York, USA.

Marquet, P. A., J. A. Keymer, and C. Hernan. 2003. Breaking the stick in space: Of niche models, metacommunities and patterns in the relative abundance of species. *In* T. M. Blackburn and K. J. Gaston [eds.], Macroecology: Concepts and consequences, vol. 43, Annual Symposium of The British Ecological Society, 64–86. Blackwell, Oxford, UK.

Maurer, B. A. 1999. Untangling ecological complexity. University of Chicago Press, Chicago, Illinois, USA.

May, R. M. 1975. Patterns of species abundance and diversity. *In* M. L. Cody and J. M. Diamond [eds.], Ecology and evolution of communities, 81–120. Belknap Press of Harvard University Press, Cambridge, Massachusetts, USA.

McGill, B. J. 2003. Does Mother Nature really prefer rare species or are log-left-skewed SADs a sampling artefact? *Ecology Letters* 6: 766–773.

McGill, B. J. 2008. Exploring predictions of abundance from body mass using hierarchical comparative approaches. *American Naturalist* 172: 88–101.

McGill, B. J. 2010a. Towards a unification of unified theories of biodiversity. *Ecology Letters* 13: 627–642.

McGill, B. J. 2010b. Ecology: Matters of scale. *Science* 328: 575.

McGill, B., and C. Collins. 2003. A unified theory for macroecology based on spatial patterns of abundance. *Evolutionary Ecology Research* 5: 469–492.

McGill, B., and J. Nekola. 2010. Mechanisms in macroecology: AWOL or purloined letter? Towards a pragmatic view of mechanism. *Oikos* 119: 591–603.

McGill, B. J. 2011. Measuring the spatial structure of biodiversity. *In* A. E. Magurran and B. J. McGill [eds.], Biological diversity: Frontiers in measurement and assessment, 152–171. Oxford University Press, Oxford, UK.

McGill, B. J., R. S. Etienne, J. S. Gray, D. Alonso, M. J. Anderson, H. K. Benecha, M. Dornelas, et al. 2007. Species abundance distributions: Moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters* 10: 995–1015.

McQuarrie, D. A., and D. Allan. 2000. Statistical mechanics. University Science Books, Sausalito, California, USA.

Mittelbach, G., D. Schemske, H. Cornell, A. Allen, J. Brown, M. Bush, S. Harrison, et al. 2007. Evolution and the latitudinal diversity gradient: Speciation, extinction and biogeography. *Ecology Letters* 10: 315–331.

Morlon, H., G. Chuyong, R. Condit, S. Hubbell, D. Kenfack, D. Thomas, R. Valencia, et al. 2008. A general framework for the distance–decay of similarity in ecological communities. *Ecology Letters* 11: 904.

MORLON, H., E. P. WHITE, R. S. ETIENNE, J. L. GREEN, A. OSTLING, D. ALONSO, B. J. ENQUIST, ET AL. 2009. Taking species abundance distributions beyond individuals. *Ecology Letters* 12: 488–501.

NACHMAN, G. 1981. A mathematical model of the functional relationship between density and spatial distribution of a population. *Journal of Animal Ecology* 50: 453–460.

NEKOLA, J., AND P. WHITE. 1999. The distance decay of similarity in biogeography and ecology. *Journal of Biogeography* 26: 867–878.

NEY-NIFLE, M., AND M. MANGEL. 1999. Species–area curves based on geographic range and occupancy. *Journal of Theoretical Biology* 196: 327–342.

O'BRIEN, E., R. FIELD, AND R. WHITTAKER. 2000. Climatic gradients in woody plant (tree and shrub) diversity: Water–energy dynamics, residual variation, and topography. *Oikos* 89: 588–600.

OLSZEWSKI, T. D. 2004. A unified mathematical framework for the measurement of richness and evenness within and among multiple communities. *Oikos* 104: 377–387.

PAUTASSO, M., AND K. GASTON. 2006. A test of the mechanisms behind avian generalized individuals–area relationships. *Global Ecology and Biogeography* 15: 303–317.

PIANKA, E. R. 1989. Latitudinal gradients in species diversity. *Trends in Ecology & Evolution* 4: 223.

PIELOU, E. C. 1975. Ecological diversity. John Wiley, New York, New York, USA.

PIELOU, E. C. 1977. Mathematical ecology. John Wiley, New York, New York, USA.

PLATT, J. R. 1964. Strong inference. *Science* 146: 347–353.

PLOTKIN, J. B., J. CHAVE, AND P. S. ASHTON. 2002. Cluster analysis of spatial patterns in Malaysian tree species. *American Naturalist* 160: 629–644.

PLOTKIN, J. B., AND H. C. MULLER-LANDAU. 2002. Sampling the species composition of a landscape. *Ecology* 83: 3344–3356.

PRESTON, F. W. 1948. The commonness and rarity of species. *Ecology* 29: 254–283.

PRESTON, F. W. 1960. Time and space and the variation of species. *Ecology* 41: 611–627.

QIAN, H., AND R. RICKLEFS. 1999. A comparison of the taxonomic richness of vascular plants in China and the United States. *American Naturalist* 154: 160–181.

RABINOWITZ, D., S. CAIRNS, AND T. DILLON. 1986. Seven forms of rarity and their frequency in the flora of the British Isles. *In* M. E. Soule [ed.], Conservation biology: The science of scarcity and diversity. Sinauer, Sunderland, Massachusetts, USA.

RAUNKIAER, C. 1909. Formationsundersogelse og Formationsstatistik. *Botanisk Tidsskrift* 30: 20–132.

RICKLEFS, R., AND D. SCHLUTER. 1993. Species diversity in ecological communities: Historical and geographical perspectives. University of Chicago Press, Chicago, Illinois, USA.

RICKLEFS, R. E. 1987. Community diversity—Relative roles of local and regional processes. *Science* 235: 167–171.

RICKLEFS, R. E. 2003. A comment on Hubbell's zero-sum ecological drift model. *Oikos* 100: 185–192.

ROSENZWEIG, M. L. 1995. Species diversity in space and time. Cambridge University Press, Cambridge, UK.

RUSSO, S. E., S. K. ROBINSON, AND J. TERBORGH. 2003. Size–abundance relationships in an Amazonian bird community: Implications for the energetic equivalence rule. *American Naturalist* 161: 267–283.

SANDERS, H. L. 1968. Marine benthic diversity: A comparative study. *American Naturalist* 102: 243–282.

SHMIDA, A., AND M. WILSON. 1985. Biological determinants of species diversity. *Journal of Biogeography* 12: 1–20.

SIEMANN, E., D. TILMAN, AND J. HAARSTAD. 1999. Abundance, diversity and body size: Patterns from a grassland arthropod community. *Journal of Animal Ecology* 68: 824–835.

SIMBERLOFF, D. S. 1972. Properties of the rarefaction diversity measurement. *American Naturalist* 106: 414–418.

ŠIZLING, A., AND D. STORCH. 2004. Power law species–area relationships and self similar species distributions within finite areas. *Ecology Letters* 7: 60–68.

STEPHENSON, N. 1990. Climatic control of vegetation distribution: The role of the water balance. *American Naturalist* 135: 649–670.

STORCH, D., A. IZLING, AND K. GASTON. 2003. Geometry of the species–area relationship in central European birds: Testing the mechanism. *Journal of Animal Ecology* 72: 509–519.

STORCH, D., A. SIZLING, L., J. REIF, J. POLECHOVA, E. SIZLINGOVA, AND K. J. GASTON. 2008. The quest for a null model for macroecological patterns: Geometry of species distributions at multiple spatial scales. *Ecology Letters* 11: 771–784.

STOYAN, D., AND H. STOYAN. 1994. Fractals, random shapes and point fields. John Wiley, Chichester, UK.

TERBORGH, J. 1973. On the notion of favorableness in plant ecology. *American Naturalist* 107: 481–501.

TILMAN, D. 1988. Plant strategies and the dynamics and structure of plant communities. Princeton University Press, Princeton, New Jersey, USA.

USDA [U. S. Department of Agriculture]. 2010. National Forest Inventory and Analysis. Forest Inventory Data Online (FIDO) web-application version 1.7.5r1. Website http://fiatools.fs.fed.us/fido/index.html.

VELLEND, M., AND A. AGRAWAL. 2010. Conceptual synthesis in community ecology. *Quarterly Review of Biology* 85: 183–206.

WALTER, H. 1973. Vegetation of the earth in relation to climate and the ecophysiological conditions.English Universities Press, London, UK.

WEST, G. B., J. H. BROWN, AND B. J. ENQUIST. 1997. A general model for the origin of allometric scaling laws in biology. *Science* 276: 122–126.

WHITE, E. P., S. K. M. ERNEST, A. J. KERKHOFF, AND B. J. ENQUIST. 2007. Relationships between body size and abundance in ecology. *Trends in Ecology & Evolution* 22: 323–330.

WHITTAKER, R. 1972. Evolution and measurement of species diversity. *Taxon* 21: 213–251.

WHITTAKER, R. H. 1975. Communities and ecosystems, 2nd ed. Macmillan, New York, New York, USA.

WIEGAND, T., AND K. A. MOLONEY. 2004. Rings, circles, and null-models for point pattern analysis in ecology. *Oikos* 104: 209.

WIEGERT, R. G. 1988. Holism and reductionism in ecology: Hypotheses, scale and systems models. *Oikos* 53: 267–269.

WILLIAMS, C. B. 1943. Area and number of species. *Nature* 152: 264–267.

WOODWARD, F. 1987. Climate and plant distribution. Cambridge University Press, Cambridge, UK.

YU, J., AND F. DOBSON. 2000. Seven forms of rarity in mammals. *Journal of Biogeography* 27: 131–139.

ZILLIO, T., AND R. CONDIT. 2007. The impact of neutrality, niche differentiation and species input on diversity and abundance distributions. *Oikos* 116: 931–940.

ZILLIO, T., AND F. HE. 2010. Modelling spatial aggregation of finite populations. *Ecology*. 91:3698-3701