Between Geometry and Biology: The Problem of Universality of the Species-Area Relationship

Arnošt L. Šizling,[1] William E. Kunin,[2] Eva Šizlingová,[3] Jiří Reif,[4] and David Storch[1,5,]*

1. Center for Theoretical Study, Charles University and the Academy of Sciences of the Czech Republic, Jilská 1, 110 00 Praha 1, Czech Republic;   2. Institute of Integrative and Comparative Biology, LC Miall Building, University of Leeds, Leeds LS2 9JT, United Kingdom; 3. Department of Philosophy and History of Science, Faculty of Science, Charles University, Viničná 7, 128 44 Praha 2, Czech Republic; 4. Institute for Environmental Studies, Charles University in Prague, Benatská 2, 128 01 Praha 2, Czech Republic;   5. Department of Ecology, Faculty of Science, Charles University, Viničná 7, 128 44 Praha 2, Czech Republic

ABSTRACT: The species-area relationship (SAR) is considered to be one of a few generalities in ecology, yet a universal model of its shape and slope has remained elusive. Recently, Harte et al. argued that the slope of the SAR for a given area is driven by a single parameter, the ratio between total number of individuals and number of species (i.e., the mean population size across species at a given scale). We provide a geometric interpretation of this dependence. At the same time, however, we show that this dependence cannot be universal across taxa: if it holds for a taxon composed from two subsets of species and also for one of its subsets, it cannot simultaneously hold for the other subset. Using three data sets, we show that the slope of the SAR considerably varies around the prediction. We estimate the limits of this variation by using geometric considerations, providing a theory based on species spatial turnover at different scales. We argue that the SAR cannot be strictly universal, but its slope at each particular scale varies within the constraints given by species' spatial turnover at finer spatial scales, and this variation is biologically informative.

Keywords: diversity patterns, spatial scaling, beta diversity, maximum entropy, power law, taxon invariance.

Introduction

For more than a century, biologists have been fascinated with the increase in species number (S) with area (A), which can be approached by a power law $S = cA^z$, where z is the slope of the relationship in a log-log plot. This has been observed so regularly that it has challenged many biologists to explain the pattern, resulting in a proliferation of models and explanations (e.g., Preston 1960; Coleman 1981; He and Legendre 1996; Lennon et al. 2002; Matter et al. 2002; Green et al. 2003; Triantis et al. 2003; Šizling

* Corresponding author; e-mail: storch@cts.cuni.cz.

and Storch 2004, 2007; Harte et al. 2005, 2009; for reviews see Dengler 2009; Tjørve 2009). Behind this wide diversity of approaches lies a more fundamental question: does the species-area relationship (SAR) reflect biological processes, a real law of ecology, or is the observed regularity the result of some statistical or mathematical necessity? How we answer these questions colors the way we interpret biodiversity patterns. If the shape of the SAR reflects biological processes, then it carries ecologically relevant information, and we should learn to read it (e.g., Tjørve and Tjørve 2008); alternatively, if SAR shape results from statistical or mathematical necessity, then the universality of such processes would strengthen the argument for using SARs for the estimation of species richness or for forecasting extinction (Magurran 2004). Of course, many possibilities exist between these two extremes. For example, the SAR derived from Hubbell's (2001) neutral theory collapses to a universal function, but the exact shape depends on biologically relevant variables, namely, the dispersal kernel and speciation rate (Rosindell and Cornell 2007).

Despite the claims of universal SAR properties (e.g., Preston 1960; Rosenzweig 1995), the biological literature reveals some repeatedly documented differences in curves between taxa and environments that imply the presence of at least some ecological information. A clear signal that the SAR is influenced by some natural processes, rather than being exclusively governed by mathematical laws, springs from the evidence that the slope z is typically lower in unicellular organisms than in multicellular ones (Finlay et al. 1998; Hillebrand et al. 2001; Finlay 2002; Green and Bohannan 2007) or for more productive environments and coarser spatial scales than for unproductive environments and fine scales (He and Legendre 2002; Storch et al. 2005; Kallimanis et al. 2008; for review see Drakare et al. 2006). However, recently Harte et al. (2009) have provided an argument that even some of these "biological" patterns

can be interpreted as resulting from mathematical necessity. They argue that the major driver of the slope of the SAR is the individuals-per-species ratio, which will tend to be high in small-bodied (e.g., unicellular) taxa and in more productive and larger areas, thus providing a single explanation for all of the biological factors described above. Harte et al. (2009) derived the relationship between the slope of the logarithmically transformed SAR (z; note that hereafter we will use the parameter z to indicate the local slope of the SAR in logarithmic space at a given scale, without necessarily implying a power law SAR, in which the z would be constant across scales) and the logarithm of mean abundance per species (hereafter D) at a given scale. They then showed that this relationship (hereafter the z-D relationship), together with two anchor values (the size of a focal area and the species richness found in that area), fully determines the SAR. Harte et al. (2009) derived their z-D relationship by using the principle of maximum entropy (MaxEnt; Jaynes 1957, 1982), suggesting universality of the relationship across all assemblages regardless of their taxonomical composition and location. As the authors note, the logic of the MaxEnt calculations should apply to any species set, even one delimited by arbitrary criteria (e.g., "plants with yellow flowers"; Harte et al. 2009, p. 796). They, however, used several assumptions such as constraints on total number of species and individuals and total energy consumption, which may be ultimately taken as a definition of assemblages to which their approach applies. The fact that some well-studied natural systems fit the proposed SAR can then be taken as evidence that these systems follow the constraints of Harte et al.'s (2008, 2009) approach.

Here we argue in favour of biologically informative SARs, suggesting that observed z-D relationships do not arise from purely mathematical necessity. We base our argument on the idea of taxon invariance (Storch and Šizling 2008; Šizling et al. 2009a), demonstrating that no realistic z-D relationship can hold simultaneously for both a full set of species and all subsets within it. We show that such a universal z-D relationship would be possible only for small assemblages and that its shape would differ from that commonly observed in nature (and from that predicted by Harte et al. [2009]). We use data on three contrasting z-D relationships to examine their natural variability. Using geometric considerations, we provide a clue to understanding the biological basis of this variability. These purely geometrical constraints will prove narrow enough to produce regularities in SAR shape but wide enough to allow reading assemblage-specific biological signals within each SAR.

Theory

The SAR Is Not Universal for All Taxa

The taxon-invariant z-D relationship can be found by solving an equation that captures the local (scale-specific) slopes of logarithmically transformed SARs, z, of two spatially overlapping assemblages and of their composite set (fig. 1c). Let us consider two scales of samples taken from the same region of interest (e.g., a forest or focal landscape): a set of smaller sample plots each of area A and a set of larger sample plots of area εA ($\varepsilon > 1$, epsilon for enlargement). We examine two subassemblages of species that we will later combine together. At the smaller scale, we denote their species richness as $S_{A,i}$ and $\sigma_{A,i}$ (where index i indicates a particular plot), while their richness in our larger sample areas is denoted $S_{\varepsilon A,i}$ and $\sigma_{\varepsilon A,i}$, respectively. For our first subassemblage, the slope characterizing the increase in species richness when enlarging the sampling area by the factor ε can be written as

$$z_S = \frac{\ln \overline{S}_{\varepsilon A} - \ln \overline{S}_A}{\ln \varepsilon}. \tag{1}$$

The variables $\overline{S}_{\varepsilon A}$ and $\overline{S}_A$ represent mean species richness across all samples of sizes εA and A, respectively. Similarly,

$$z_\sigma = \frac{\ln \bar{\sigma}_{\varepsilon A} - \ln \bar{\sigma}_A}{\ln \varepsilon} \tag{2}$$

in the case of the second assemblage. The slope of the combined assemblage, $z_{S \cup \sigma}$, is

$$z_{S \cup \sigma} = \frac{\ln (\overline{S}_{\varepsilon A} + \bar{\sigma}_{\varepsilon A}) - \ln (\overline{S}_A + \bar{\sigma}_A)}{\ln \varepsilon}. \tag{3}$$

The reason for equation (3) is that the mean species richness of both assemblages combined across all samples is calculated as $[(S_1 + \sigma_1) + (S_2 + \sigma_2) + \ldots + (S_n + \sigma_n)]/n$, where n is the number of samples. It can be rearranged as $(S_1 + S_2 + \ldots + S_n)/n + (\sigma_1 + \sigma_2 + \ldots + \sigma_n)/n$, which equals $\overline{S} + \bar{\sigma}$. Apparently, no rearrangement (e.g., $[(S_1 + \sigma_3) + (S_2 + \sigma_n) + \ldots + (S_n + \sigma_5)]/n)$ affects the result, so equation (3) is valid regardless of the spatial distribution of the two subassemblages.

Putting equations (1)–(3) together and substituting for $\pi_A := \overline{S}_A / (\bar{\sigma}_A + \overline{S}_A)$, we can solve for the relationship between the three slopes, which is as follows:

$$\varepsilon^{z_{S \cup \sigma}} = \pi_A \varepsilon^{z_S} + (1 - \pi_A) \varepsilon^{z_\sigma} \tag{4}$$

(see "Derivation I" in app. A, available online). Now we should ask whether there is a universal relationship between z and D that would hold for all possible assemblages, namely, for the assemblage and both its subassemblages. We proceed by attributing each slope, z_S, z_σ, and $z_{S \cup \sigma}$, to
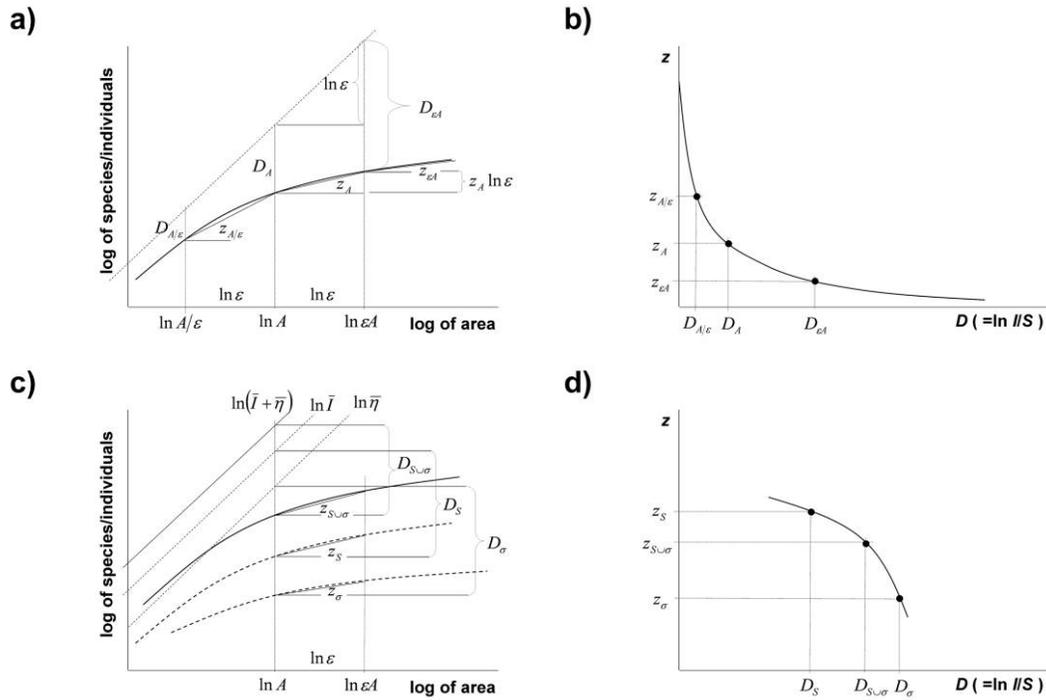
Figure 1: *a*, Species-area relationship (SAR; curved) and individual-area relationship (IAR; linear) in a log-log plot. The distance *D* between IAR and SAR for a given area is equivalent to the logarithm of mean abundance per species, as $\log I - \log S = \log (I/S)$. Note that the SAR cannot be linear, because (1) it has a lower slope than the IAR and (2) it cannot intersect the IAR, as there cannot be more species than individuals. *b*, Corresponding *z-D* relationships based on *D* for the three areas *A*, εA, and A/ε lie on a convex curve. *c*, SARs and IARs for a taxon and its two subtaxa. If the slopes (*z*) and distances (*D*) of two subassemblages (dashed relationships in *c*) and their composite set (full relationship in *c*) were to lie on a universal curve, then this universal curve would be downward accelerating (*d*).

the corresponding vertical distances between the SAR and the individual-area relationship in logarithmic space, D_S, D_σ, and $D_{S \cup \sigma}$ (fig. 1*c*), respectively. The distance *D* is equal to the logarithm of the mean number of individuals per species at a given area, used by Harte et al. (2009), as $\ln \overline{I/S} = \ln \overline{I} - \ln \overline{S} = D$. The relationship between the distances for the whole assemblage and the subassemblages is $D_{S \cup \sigma} = \ln [\pi_A \exp (D_S) + (1 - \pi_A) \exp (D_\sigma)]$ (see "Derivation II" in app. A). To derive this relationship, we assumed that total abundance of all species scales linearly with area (see "Thesis I" in app. A). Consequently, we can derive a functional equation:

$$\varepsilon^{z\{\ln [\pi_A \exp (D_S) + (1 - \pi_A) \exp (D_\sigma)]\}} = \pi_A \varepsilon^{z(D_S)} + (1 - \pi_A) \varepsilon^{z(D_\sigma)}. \quad (5)$$

The only possible solution of equation (5), if it exists, obeys

$$z(D) = \frac{\ln (\alpha e^D + \beta)}{\ln \varepsilon}, \quad (6)$$

where α and β are parameters without any particular meaning (see "Thesis II" in app. A). If $\alpha e^D + \beta \le 0$, then $z(D)$ in equation (6) is undefined, and the universal *z-D*

relationship does not exist. Hence, a universal *z-D* relationship can exist only within a limited range of the variable *D* ($D < \ln (-\beta/\alpha)$ if $\alpha\beta < 0$). Equation (6) also suggests that the shape of the *z-D* relationship depends on ε. Theoretically, we should consider as low a value of ε as possible ($\varepsilon \to 1^+$) to derive a smooth curve. Practically, we need a reasonably large ε (Harte et al. 2009 used $\varepsilon = 2$). Therefore, we will discuss *z-D* relationships for all possible ε values ($\varepsilon > 1$).

Equation (6), or an equivalent formulation, presents the only possible universal (i.e., taxon-invariant) *z-D* relationship. It has some important properties that can be tested against data. The parameters α and β can be any real numbers; however, $\alpha \ge 0$ makes the relationship between *z* and *D* increasing or constant, which is unrealistic, contradicting both the preponderance of reported data sets (Harte et al. 2009; fig. 4*b*) and the MaxEnt solution (Harte et al. 2009). Hence, we will not discuss this scenario, and we hereafter assume that $\alpha < 0$. Since equation (6) is defined only for $\alpha e^D + \beta > 0$, negative values of α make $\beta > 0$. Hence, the relationship between *z* and *D* as de-

scribed in equation (6) is (i) decreasing, (ii) downward accelerating, and (iii) reaches 0 at $D = \ln(-\beta/\alpha)$ (fig. 2*a*; for proofs see "Theses III–VI" in app. A), which is clearly unrealistic, contradicting both the observations (fig. 1*a* in Harte et al. 2009; fig. 4*b*) and the MaxEnt solution.

The taxon-invariant *z-D* relationship has been derived assuming a constant ratio (π) between species richness of the two subassemblages at all scales. In reality, this ratio can vary across scales if, for example, one subassemblage exhibits higher species spatial turnover than the other. This means that, considering an increasing series of sampling areas, A, εA, $\varepsilon^2 A$, ..., $\varepsilon^n A$, each subassemblage's ratio $\pi_{\varepsilon^i A}$ can be different. Nonetheless, if we focus on a particular step of the enlargement of the area (e.g., on the particular slope $z(D)$ between $\varepsilon^i A$ and $\varepsilon^{i+1} A$), only one locally fixed value of $\pi_{\varepsilon^i A}$ is attributed to this step (fig. 1*c*). Hence, in accord with the proof, the three particular points (see fig. 1*d*) $\{D_S, z(D_S)\}$, $\{D_\sigma, z(D_\sigma)\}$, and $\{D_{S \cup \sigma}, z(D_{S \cup \sigma})\}$ would certainly lie on a downward-accelerating curve had the *z-D* relationship been universal. This holds true for any three D_S, D_σ, and $D_{S \cup \sigma}$, and thus the whole universal taxon-invariant *z-D* relationship must be necessarily downward accelerating, reaching 0 (fig. 2). The variation in π can only modify its curvature and shift the point of $z(D) = 0$.

The Expected Shape of z-D Relationships

While we have demonstrated above that all SARs cannot simultaneously follow Harte et al.'s (2009) universal so-

lution, most data show a hollow *z-D* relationship, as Harte et al. (2009) predicted. Why? One way to conceptualize the SAR is to begin by examining the species richness of a set of fine grid cells (e.g., fig. 3, *A1*, *A2*, *B1*, *B2*) and then to successively merge neighboring cells to form pairs (e.g., fig. 3, *A*, *B*), then neighboring pairs to form quartets, and so on, equivalent to setting $\varepsilon = 2$ (see also Harte and Kinzig 1997). We contend that the Jaccard index (the percentage of species shared by two assemblages; Brown and Lomolino 1998) between two adjacent plots of a given area (J_n for the Jaccard index at one stage of this process) will constrain the subsequent Jaccard index between two adjacent plots of twice that area (J_{n+1} for the Jaccard index at the next stage), following the equation

$$J_{n+1} \cong \frac{KJ_n}{2 - KJ_n}, \tag{7}$$

where the parameter K is constrained as

$$1 \leq K \leq \min\left[4, \frac{1}{J_n}\right]. \tag{8}$$

The reason is that (i) the Jaccard index is defined as a proportional number of species shared by two focal plots and (ii) the number of species shared by two adjacent larger plots (e.g., fig. 3, *A*, *B*) cannot be lower than the mean number of species shared by pairs of finer subplots within them (i.e., the number of species shared by *A1* and *B1*, *A1* and *B2*, *A2* and *B1*, and *A2* and *B2*), but it also cannot be more than four times higher than this mean
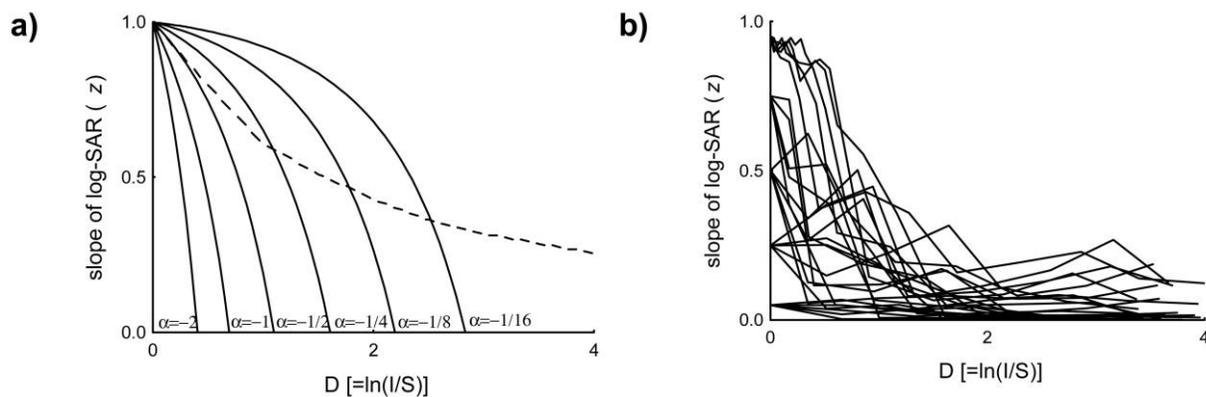
Figure 2: Taxon-invariant (*a*) and consecutive-constraints (*b*) models of *z-D* relationships. *a*, An example of various *z-D* relationships (eq. [6]; solid lines) for various coefficients α (see labels) and $\beta = \varepsilon^{z_0} - \alpha e^{D_0}$ ($z_0 = 0.9999$, $D_0 = 0$). Such a relationship would be followed by the data had the *z-D* relationship been universal and thus independent of the focal taxa. The dashed line shows the MaxEnt solution adopted from Harte et al. (2009). *b*, Various *z-D* relationships resulting from the consecutive-constraints model, with various starting points $\{D(0), z(0)\} = \{0, 0.95\}$, $\{0, 0.75\}$, $\{0, 0.5\}$, $\{0, 0.25\}$, and $\{0, 0.05\}$. The parameter K varies randomly (with regular distribution) between its theoretical limits (eq. [8]). Note that the real distribution of parameter K is not regular; however, the parameter must have been distributed with extreme preference toward low values (strong right-skewed distribution) to break the decreasing hollow shape of the *z-D* relationship.
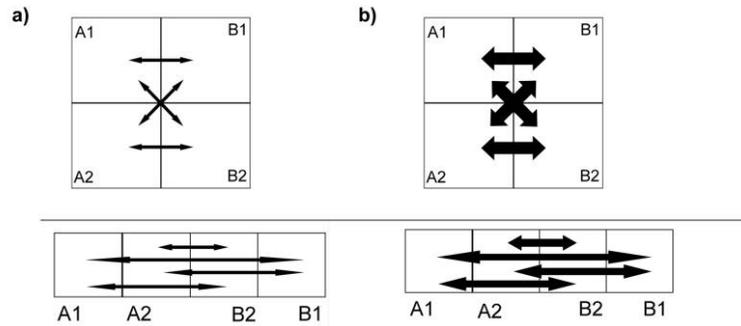
Figure 3: A square-shaped plot (*top*) or a linear plot (*bottom*) can be divided into four adjacent plots $A1$, $A2$, $B1$, and $B2$ (hereafter, small plots). These plots can be grouped into two adjacent plots A ($\equiv A1 \cup A2$) and B ($\equiv B1 \cup B2$; hereafter, large plots). The number of species shared by large plots A and B cannot be lower than the number of species shared by a pair of small plots depicted by one of the four arrows, and at the same time, it cannot be higher than the simple sum of species shared by all the pairs (all four arrows) together. Therefore, the number of species shared by the large plots A and B is K times the average number of species shared by a pair of small plots, where K lies between 1 and 4. These limits are universal, although the lower limit can be reached only if the four arrows represent equal numbers of shared species. K is thus the number by which the mean number of shared species between small plots must be multiplied to get the number of shared species between the two large plots. It is constrained by the relative proportion of species shared by the small plots. If small plots ($A1$, $A2$, $B1$, $B2$) mutually share only a low proportion of their species (narrow arrows in *a*), it is likely that different pairs of the small plots will share different sets of species, so that the number of species shared between A and B will be considerably higher, increasing K. If the Jaccard index approaches 0, the lists of species shared by the smallest plots are almost completely nonoverlapping, and all four lists (arrows) should be involved in calculating the Jaccard index between A and B (K in eq. [7] and eqq. [B1], [B2], available online, approaches anything between 1 and 4). On the other hand, in an extreme case, if the small plots share all their species (Jaccard index $= 1$), the overlap is 100% and the number of shared species between A and B is the same as the number of shared species between any two small subplots (i.e., all the species), so that in equations (7), (B1), and (B2) $K = 1$. We suggest that the relationship between the Jaccard indices of the small plots and K (see eq. [7]) is assemblage-specific; however, the limits imposed on K and the tendency of K to increase with the decrease in the Jaccard index between the smallest plots are universal, as they represent geometric necessity.

number of shared species (fig. 3). The value of $1/J_n$ cannot be exceeded by K because the Jaccard index can never exceed 1 (for details see app. B, available online).

The local slope z and the Jaccard index of assemblage similarity between adjacent plots (averaged across their various locations) are closely related as

$$z \cong 1 - \frac{\ln(1 + J)}{\ln 2} \tag{9}$$

(see Koleff et al. 2003 and "The Link between z and J (Eq. [9])" in app. B; for a review see Gaston et al. 2007). Therefore, equation (7) also governs the z-D plot. Each z_n and D_n in this plot constrains the subsequent z_{n+1} and D_{n+1}, parameterized by K (see "Consecutive Constraints" in app. B). The parameter K varies according to equation (8).

If we make the null assumption that parameter K is distributed regularly or symmetrically between its limits, the hollow shape of the z-D relationship (fig. 2*b*) simply results from the higher probability that z decreases when increasing area due to the disproportionately wide region of parameter space where z declines with increasing scale (fig. 4*c*). This effect is further promoted by the fact that no K can approach its lower limit as long as there is any variability in numbers of shared species between pairs of

adjacent samples, which is very probable (for details see app. B). Realistic (i.e., hollow) z-D relationships are produced by all distributions of K whose median values lie above the line that separates K for decreasing and increasing z values with scale (fig. 4*c*, dotted line). This line lies entirely in the lower half of the space of possible K, which means that all symmetric, all left-skewed, and a large proportion of right-skewed distributions of K produce a hollow shape in the z-D relationship. In sum, the hollow pattern is the most likely one, although other shapes (e.g., increasing z-D relationship for high D, corresponding to the triphasic SAR; Shmida and Wilson 1985; Fridley et al. 2005; Rosindell and Cornell 2009) are also possible.

The considerations of consecutive constraints apply for any starting point $\{z(0), D(0)\}$ in the bivariate z-D space (fig. 2*b*); that is, they do not lead to any particular value of z for a given D, predicting only the overall trend. It is, however, reasonable to assume that all observed z-D relationships start with low D and high SAR slopes ($D(0) \to 0^+$ and $z(0) \to 1^-$, respectively). This is because the SAR is constrained to fall below the individual-area line, as there cannot be more species than individuals (fig. 1*a*, 1*c*; for broader discussion see app. B). This is the reason why all SARs have high slopes for very small areas.
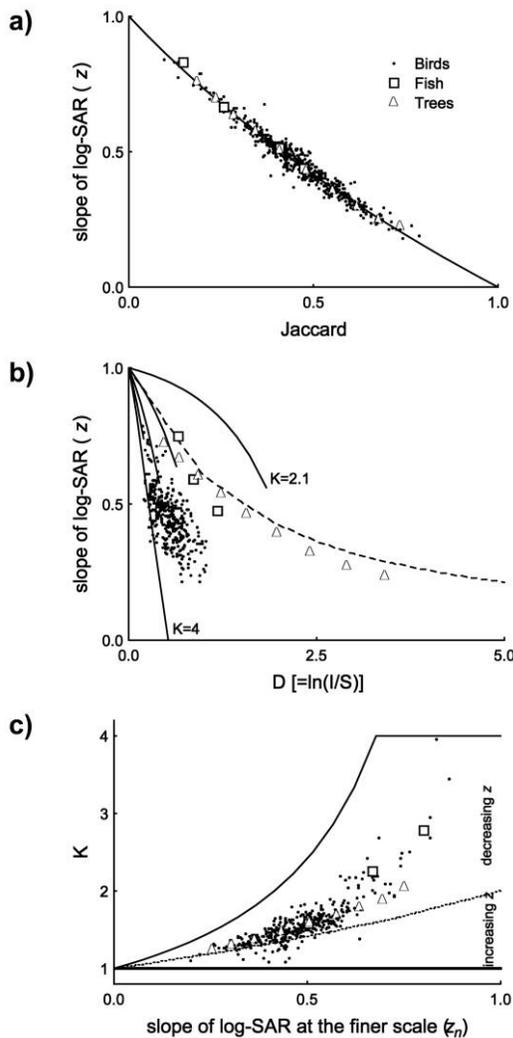
Figure 4: Data on Czech birds (dots), Mediterranean fish (squares), and tropical trees (triangles) plotted together with the theoretical relationships. *a*, Relationship between the average Jaccard index of similarity in species composition across various equal-sized plots and the slope of the species-area relationship in logarithmic space, *z*. The solid line shows the prediction of equation (9). *b*, Observed *z-D* relationships (*z* computed following Harte et al. 2009) plotted against five trends in *z* decay with various values of *K* (*K* = 4, 3.5, 3, 2.5, and 2.1 from left; solid lines); the higher the slope of *z* is, the lower the Jaccard index is and the more the *K* is expected to drive the decay in *z* (fig. 3). The MaxEnt prediction (dashed line) is adopted from Harte et al. (2009). *c*, Lower and upper limits of the parameter *K* (which determines *z* at the next coarser scale; eq. [8]) as a function of the slope (*z*) of the species-area relationship (SAR) at the finer scale (solid lines). The higher the value of *K* is for a given *z*, the lower (relatively) the consecutive (i.e., next-coarser) *z* value must be (see "The Meaning of *K*" in app. B, available online). The dotted line represents values of *K* for which *z* remains constant (*z* at the current scale = *z* at the next coarser scale). For points above the dotted line, consecutive slopes (z_A) are lower than z_A, so that the SAR is decelerating on logarithmic axes.

The Meaning of Parameter K and Biological Effects on the Shape of the SAR

The considerations of consecutive constraints give a clue to understanding biological information beyond the SAR at various scales. The parameter *K* links *z* at the focal scale with *z* at the consecutive scale; *z* at the finer scale determines the limits of *K*, and *K* taken from this envelope determines *z* at the next coarser (i.e., consecutive) scale. Any biological effect acts through affecting the variation in *K* within its limits given by a particular *z* at the finer scale. Higher *K* relates to lower *z* at the coarser scale, relative to *z* at the focal scale, indicating that the number of species shared between plots at the coarser scale is considerably higher than the number shared between plots at finer scales. This can be realized only by the relative increase of occupancy at the coarse scale, which tends to occur when occupied cells at the focal scale are widely scattered. Relationships between *z* and *K* (fig. 4*c*) that are closer to the lower limit thus indicate that most species do not have a tendency to increase their relative occupancy when increasing scale, revealing aggregated distributions at each focal scale.

Species' spatial aggregation at given scale can be driven by various factors, ranging from the distribution of suitable habitat and spatial population dynamics comprising dispersal, colonization, and local extinctions to behavioral processes such as social group formation and conspecific attraction or repulsion. All these effects then translate into particular occupancy patterns (i.e., relative proportion of occupied sites) at each spatial scale, which then drive the relationship between *z* and *K* as explained above. Systematically low *K* would indicate systematic bias toward lower occupancies when changing scale, systematically high *K* would indicate a tendency for increasing occupancies with scale, and *K* values lying approximately in the middle between the two constraints could be interpreted as an absence of bias toward systematic increase or decrease of species' relative occupancy at the coarse scale, given the occupancy at the focal scale. This could result either from a general tendency for moderate levels of aggregation in species spatial distributions across scales or from a wide range of aggregation patterns for different species, resulting in a moderate level of aggregation on average.

Data Tests

We performed three data tests. First, we tested for agreement between *z* (from eq. [1], with $\varepsilon = 2$) and its estimate by using the mean Jaccard index across various locations of sampling plots (eq. [9]). We expect some small departures from the straightforward relationship, as Jaccard index values vary spatially (i.e., for various comparisons of

plots), whereas *z* comprises only an overall value. Then we explored the extent of differences between *z-D* relationships for various data. Finally, we examined distribution of the parameter *K* within the limits of its possible values (eq. [8]) for these data sets.

We used three data sets. (i) The first consisted of Czech bird data, comprising abundances of 137 species collected by skilled volunteers on 123 transects scattered throughout the whole area of the Czech Republic. Each transect contained 20 census points at 300–500-m intervals counted two times per breeding season in one randomly selected year from the period 2006–2008. The maximum count was treated as abundance of a species at a given point. See Reif et al. (2010) for more details on this data set (data are available on request from verm@birdlife.cz or at Dryad, http://dx.doi.org/10.5061/dryad.8733). (ii) The second data set consisted of Barro Colorado Island tree data (Condit 1998; Hubbell et al. 1999; see http://ctfs.si/edu/datasets/bci) overlapped by a grid of 1,030 × 20 cells. (iii) The third data set consisted of data on Mediterranean fish collected by A. L. Šizling (http://dx.doi.org/10.5061/dryad.k9711) along a 20-point, 80-m-long transect located in a bay at the east end of island Brač, Croatia. The space around each point (2 m in radius) was checked for 3 min (spaces around all consecutive points were adjacent); the transect was checked for 3 days at 10 a.m. Data from each day (40, 47, and 44 individuals of 13, 11, and 11 species, respectively; 16 species in total) were analyzed separately, and the presented slopes of logarithmically transformed SARs and Jaccard indices are mean values over all the three data sets.

In computation procedures, all data were sampled as linear series of plots (so that the area of a sample was enlarged by prolonging one dimension only), with consecutively doubled lengths starting with one cell (one cell, two cells, four cells, and so on up to 16 cells in birds and fish and 1,024 cells in tree data). All possible locations of each plot were taken into account. Slopes of logarithmically transformed SARs were computed from the mean number of species across all plots of a given size; similarly, the Jaccard index was a mean value across all pairs of adjacent plots of the given size (for computational tools see http://www.cts.cuni.cz/wiki/ecology:start). The comparability of this computational approach with that of Harte et al. (2009) is demonstrated by the similarity of results for the Barro Colorado Island data used in both articles (cf. figs. 4*b*, 1*a* in Harte et al. 2009).

Observed *z* values followed the predictions given by Jaccard indices (eq. [9]; fig. 4*a*). The regression line between observed and predicted *z* values had parameters as follows: slope, 0.988 (SE = 0.004); intercept, 0.003 (SE = 0.002, df = 622); and correlation coefficient, 0.9946 (*N* = 624).

The *z-D* relationships of Mediterranean fish and Barro Colorado Island tree data differed significantly from the Czech bird data (fig. 4*b*), as the probability of each falling entirely above the 123 bird relationships at random was less than 0.037, with confidence 0.95 (using Wilks's nonparametric test; Wilks 1941; see also app. C in Šizling et al. 2009*b* for a detailed explanation). Unlike in the other two data sets, some examined plots were not adjacent in the bird survey, with gaps of up to 200 m. Longer interplot distances, however, should make the Jaccard index between those plots lower (due to distance decay in community similarity; e.g., Nekola and White 1999; Azaele et al. 2009) and thus make *z* higher. Hence, the true *z* values for birds should be lower than those plotted, which further accentuates the differences between the *z-D* relationships of the three groups, making our test conservative.

The observed relationships between *z* and the coefficient *K* were mostly clustered approximately midway between the constraints imposed by equation (8), at least in the bird and fish data (fig. 4*c*). The observed *K* values continued to rise with *z* after the upper constraint had reached its maximum (*K* = 4), so that at very high *z* values, this upper line was approached (at least in the bird data set). The three data sets differed somewhat in their behavior, especially at very high slope values, with trees displaying lower *K* values than fish or most birds, indicating that these taxa differ in the way their relative occupancies vary with spatial scale.

Discussion

The MaxEnt approach in ecology does not require any specification about the focal group of species, and its results should thus be applicable to any—even arbitrarily chosen—set of species (Harte et al. 2009). We used the principle of taxon invariance (Storch and Šizling 2008; Šizling et al. 2009*a*) to show that the species-area relationship cannot universally follow Harte et al.'s (2009) solution; hence, we need to specify the group of species to which this solution applies, if such a group exists. The main point of our proof was that the *z-D* relationship would have to be downward accelerating, soon reaching 0, had it been universal, which would allow universality only at small scales and contradicts the great mass of SAR patterns observed in nature.

The extent of variability in *z-D* relationships is considerable. The tree, bird, and fish data show strongly diverging *z-D* relationships when increasing *D*, all starting from low *D* and high *z* values (fig. 4*b*). The notion that taxa may differ in SAR properties is further supported by Finlay et al. (1998), who found 48,186 individuals of 85 species (*D* = 6.3) of marine interstitial species and 20,486 individuals of 104 species (*D* = 5.3) of freshwater benthic

ciliate species (Protozoa) in their samples and estimated *z* as 0.04, similar to the low slopes of logarithmically transformed SARs frequently reported in planktonic and microbial assemblages (Finlay et al. 1998; Hillebrand et al. 2001; Finlay 2002; Green and Bohannan 2007). Conversely, Harte et al. (2009) predict *z* of approximately 0.2 for these values of *D* (fig. 4*b*). The crucial evidence for lack of universality in *z-D* relationships, however, comprises the analytical proof (app. A).

Our evidence suggests the existence of assemblage-specific SARs carrying ecologically relevant information. At the same time, the geometrical considerations suggest the prevalence of hollow shapes of *z-D* relationships due to the existence of constraints resulting from mathematical necessities. Once *z* is known at a specific scale, geometrical rules impose constraints on *z* at the next coarser scale, with change in *z* being mediated by the parameter *K*.

The realized distribution of *K* values found in our data sets is not uniform within the constraints imposed by equation (8) and varies between the focal assemblages (fig. 4*c*), which ultimately indicates something about their ecological properties. First, the clustering of observed *K* values in the middle between the constraints indicates that any tendency toward higher relative occupancy in some species is generally balanced by the opposite tendency in other species at each scale (for details see "The Meaning of *K*" in app. B). Second, the bird and fish data follow a similar relationship, their *K* values being higher than the *K* values for tropical trees, namely, for higher *z* at the finer scale ($z \geq 0.5$). Because higher *K* values indicate relatively lower *z* values and thus higher mean occupancies at the next coarser scale, we conclude that our tree species sample was generally more aggregated at finer scales than were the bird and fish species surveyed. At coarser scales, at which $z < 0.5$, the tendency toward aggregation becomes similar for all three assemblages.

We argue against the traditional attempts to derive the SAR from species abundance distributions (SADs; e.g., Preston 1962; Harte et al. 2008, 2009; Tjørve et al. 2008). Our approach was based on consecutive constraints imposed on variation in the Jaccard index of community similarity between two adjacent plots with scale. We did not need information on any particular shape of the SAD to build a SAR under this approach. Indeed, if there is a link between SAD and SAR, it may run the opposite direction, as species' spatial turnover has recently been demonstrated to determine the SAD (Šizling et al. 2009*c*; Kůrka et al. 2010). Hence, the variation in species' spatial turnover between adjacent plots (alternatively, *z* for $\varepsilon = 2$) may prove a master pattern simultaneously driving several macroecological and diversity patterns, including the SAR and the SAD.

Our considerations enable a better understanding of the

potential value of Harte et al.'s (2008, 2009) MaxEnt predictions. Harte et al. (2009) argue that no specification in terms of focal group of species is present in their theory, so that their results are applicable to any group of species. They built their theory with a constraint imposed on energy, which applies only to a conservative system (in the physical sense): one where energy income is at equilibrium with energy emission in each sample. However, ecological assemblages with multiple trophic levels are rather dissipative systems, that is, systems that are not conservative in the physical sense. The reason is that the flow of energy cascades between the trophic levels, with lower trophic levels making up the resources needed by species at higher trophic levels (Thornton 1997). A new prey species thus increases opportunities for a new predator to occur, which is the opposite of what would be expected in a conservative assemblage. The lack of taxon invariance in MaxEnt predictions may reflect an implicit limit in the application of the model to taxonomical levels or ecological guilds where all species compete for a fixed common pool of resources. It is thus possible that the MaxEnt prediction applies only to specific ecological groupings (taxonomical, trophic, or other) where its assumptions are met.

In summary, we have shown that the local slope of the species-area relationship, *z*, cannot be universally determined by the ratio between the total number of individuals and the number of species for a given area, *D*. Still, the MaxEnt prediction can be taken as a reasonable expectation (which may not necessarily be followed by observations; Jaynes 1957) if we have information on only these two variables. Indeed, the observed relationship between *D* and *z* documented here is qualitatively similar to the hollow curve predicted by Harte et al. (2009), although various taxa are free to (and do) deviate from this prediction to various extents. Such deviations carry ecologically relevant meaning (see also Banavar et al. 2010; Haegeman and Etienne 2010); for example, they can reveal a tendency of most species to be widely dispersed or, conversely, spatially aggregated at each particular scale. The existence of geometric constraints and the fact that observed SARs vary somewhat within them underscores not only the mathematical and statistical regularities governing the relationship but also the power of ecological differences between species groups to mold the species-area relationship.

Literature Cited

Azaele, S., R. Muneepeerakul, A. Maritan, A. Rinaldo, and I. Rodriguez-Iturbe. 2009. Predicting spatial similarity of freshwater fish biodiversity. Proceedings of the National Academy of Sciences of the USA 106:7058–7062.

Banavar, J. R., A. Maritan, and I. Volkov. 2010. Applications of the principle of maximum entropy: from physics to ecology. Journal of Physics 22:063101, doi: 10.1088/0953-8984/22/6/063101.

Brown, J. H., and M. V. Lomolino. 1998. Biogeography. Sinauer, Sunderland, MA.

Coleman, D. B. 1981. On random placement and species-area relations. Mathematical Biosciences 54:191–215.

Condit, R. 1998. Tropical forest census plots. Springer, Berlin.

Dengler, J. 2009. Which function describes the species-area relationship best? a review and empirical evaluation. Journal of Biogeography 36:728–744.

Drakare, S., J. J. Lennon, and H. Hillebrand. 2006. The imprint of the geographical, evolutionary and ecological context on species-area relationships. Ecology Letters 9:215–227.

Finlay, B. J. 2002. Global dispersal of free-living microbial eukaryote species. Science 296:1061–1063.

Finlay, B. J., G. F. Esteban, and T. Fenchel. 1998. Protozoan diversity: converging estimates of the global number of free-living ciliate species. Protist 149:29–37.

Fridley, J. D., R. K. Peet, T. R. Wentworth, and P. S. White. 2005. Connecting fine- and broad-scale species-area relationships of southeastern U.S. flora. Ecology 86:1172–1177.

Gaston, K. J., K. L. Evans, and J. J. Lennon. 2007. The scaling of spatial turnover: pruning the thicket. Pages 181–214 in D. Storch, P. A. Marquet, and J. H. Brown, eds. Scaling biodiversity. Cambridge University Press, Cambridge.

Green, J. L., and B. J. M. Bohannan. 2007. Biodiversity scaling relationships: are microorganisms fundamentally different? Pages 129–145 in D. Storch, P. A. Marquet, and J. H. Brown, eds. Scaling biodiversity. Cambridge University Press, Cambridge.

Green, J. L., J. Harte, and A. Ostling. 2003. Species richness, endemism and abundance patterns: tests of two fractal models in a serpentine grassland. Ecology Letters 6:919–928.

Haegeman, B., and R. S. Etienne. 2010. Entropy maximization and the spatial distribution of species. American Naturalist 175:E74–E90.

Harte, J., and A. Kinzig. 1997. On the implications of species-area relationships for endemism, spatial turnover, and foodweb patterns. Oikos 80:417–427.

Harte, J., E. Conlisk, A. Ostling, J. L. Green, and A. B. Smith. 2005. A theory of spatial structure in ecological communities at multiple spatial scales. Ecological Monographs 75:179–197.

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

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

He, F., and P. Legendre. 1996. On species-area relations. American Naturalist 148:719–737.

———. 2002. Species-diversity patterns derived from species-area models. Ecology 83:1185–1198.

Hillebrand, H., F. Watermann, R. Karez, and U. G. Berninger. 2001. Differences in species richness patterns between unicellular and multicellular organisms. Oecologia (Berlin) 126:114–124.

Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, NJ.

Hubbell, S. P., R. B. Foster, S. T. O'Brien, K. E. Harms, R. Condit, B. Wechsler, S. J. Wright, and S. L. de Lao. 1999. Light gap disturbances, recruitment limitation, and tree diversity in a Neotropical forest. Science 283:554–557.

Jaynes, E. 1957. Information theory and statistical mechanics. Physical Review 106:620–630.

———. 1982. On the rationale of maximum entropy methods. Proceedings of the IEEE 70:939–952.

Kallimanis, A. S., A. D. Mazaris, and J. Tzanopoulos. 2008. How does habitat diversity affect the species-area relationship? Global Ecology and Biogeography 17:532–538.

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

Kůrka, P., A. L. Šizling, and J. Rosindell. 2010. Analytical evidence for scale-invariance in the shape of species abundance distributions. Mathematical Biosciences 223:151–159.

Lennon, J. J., W. E. Kunin, and S. Hartley. 2002. Fractal species distributions do not produce power-law species area relationships. Oikos 97:378–386.

Magurran, A. E. 2004. Measuring biological diversity. Blackwell, Oxford.

Matter, S. F., I. Hanski, and M. Gyllenberg. 2002. A test of the metapopulation model of the species-area relationship. Journal of Biogeography 29:977–983.

Nekola, J. C., and P. S. White. 1999. Distance decay of similarity in biogeography and ecology. Journal of Biogeography 26:867–878.

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

———. 1962. The canonical distribution of commonness and rarity. I. Ecology 43:185–215.

Reif, J., F. Jiguet, and K. Šťastný. 2010. Habitat specialization of birds in the Czech Republic: comparison of objective measures with expert opinion. Bird Study 57:197–212.

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

Rosindell, J., and S. J. Cornell. 2007. Species-area relationship from a spatially explicit neutral model in an infinite landscape. Ecology Letters 10:586–595.

———. 2009. Species-area curves, neutral models, and long-distance dispersal. Ecology 90:1743–1750.

Shmida, A. A., and M. V. Wilson. 1985. Biological determinants of species-diversity. Journal of Biogeography 12:1–20.

Šizling, A. L., and D. Storch. 2004. Power-law species-area relationships and self-similar species distributions within finite areas. Ecology Letters 7:60–68.

———. 2007. Geometry of species distributions: random clustering and scale invariance. Pages 77–99 *in* D. Storch, P. A. Marquet, and J. H. Brown, eds. Scaling biodiversity. Cambridge University Press, Cambridge.

Šizling, A. L., D. Storch, J. Reif, and K. J. Gaston. 2009*a*. Invariance in species-abundance distributions. Theoretical Ecology 2:89–103.

Šizling, A. L., E. Šizlingová, D. Storch, J. Reif, and K. J. Gaston. 2009*b*. Rarity, commonness, and the contribution of individual species to species richness patterns. American Naturalist 174:82–93.

Šizling, A. L., D. Storch, E. Šizlingová, J. Reif, and K. J. Gaston. 2009*c*. Species abundance distribution results from a spatial analogy of central limit theorem. Proceedings of the National Academy of Sciences of the USA 106:6691–6695.

Storch, D., and A. L. Šizling. 2008. The concept of taxon invariance in ecology: do diversity patterns vary with changes in taxonomic resolution? Folia Geobotanica 43:329–344.

Storch, D., K. L. Evans, and K. J. Gaston. 2005. The species-area-energy relationship. Ecology Letters 8:487–492.

Thornton, I. W. B. 1997. Krakatau: the destruction and reassembly of an island ecosystem. Harvard University Press, Cambridge, MA.

Tjørve, E. 2009. Shapes and functions of species-area curves. II. A review of new models and parameterizations. Journal of Biogeography 36:1435–1445.

Tjørve, E., and K. M. C. Tjørve. 2008. The species-area relationship, self-similarity, and the true meaning of the *z*-value. Ecology 89:3528–3533.

Tjørve, E., W. E. Kunin, C. Polce, and K. M. C. Tjørve. 2008. Species-area relationship: separating the effects of species abundance and spatial distribution. Journal of Ecology 96:1141–1151.

Triantis, K. A., M. Mylonas, K. Lika, and K. Vardinoyannis. 2003. A model for the species-area-habitat relationship. Journal of Biogeography 30:19–27.

Wilks, S. S. 1941. Determination of sample sizes for setting tolerance limits. Annals of Mathematical Statistics 12:91–96.