Ecology Letters, (2009) **12**: 789–797 doi: 10.1111/j.1461-0248.2009.01328.x

LETTER

Biodiversity scales from plots to biomes with a universal species–area curve

John Harte,[1*] Adam B. Smith[1] and David Storch[2,3]
[1]Energy and Resources Group, University of California at Berkeley, 310 Barrows Hall, Berkeley, CA 94720, USA
[2]Center for Theoretical Study, Charles University & Academy of Sciences of the CR, Jilská 1, 110 00 Praha 1, Czech Republic
[3]Department of Ecology, Faculty of Science, Charles University, Vinicná 7, 128 44 Praha 2, Czech Republic
**Correspondence: E-mail:jharte@berkeley.edu*

Abstract

Classic theory predicts species richness scales as the quarter-power of area, yet species–area relationships (SAR) vary widely depending on habitat, taxa, and scale range. Because power-law SAR are used to predict species loss under habitat loss, and to scale species richness from plots to biomes, insight into the wide variety of observed SAR and the conditions under which power-law behavior should be observed is needed. Here we derive from the maximum entropy principle, a new procedure for upscaling species richness data from small census plots to larger areas, and test empirically, using multiple data sets, the prediction that up to an overall scale displacement, nested SAR lie along a universal curve, with average abundance per species at each scale determining the local slope of the curve. Power-law behaviour only arises in the limit of increasing average abundance, and in that limit, the slope approaches zero, not ¼. An extrapolation of tree species richness in the Western Ghats to biome scale (60 000 km^2) using only census data at plot scale (¼ ha) is presented to illustrate the potential for applications of our theory.

Keywords

Abundance, biodiversity, macroecology, power law, scaling, species–area relationship.

Ecology Letters (2009) 12: 789–797

INTRODUCTION

The rate of increase in the number of species with area, or species–area relationship (SAR), is a central focus of ecology (Rosenzweig 1995) and has been a subject of contention despite 230 years of measurement and theoretical analysis (Lomolino 2001). Classic theory suggested that the number of species (S) would scale with area (A) as a power law ($S = cA^z$) with an exponent $z \approx$ ¼ (Preston 1962; May 1975). The power-law assumption with this value of z has provided a simple, widely used means of estimating both species richness in areas too large to census thoroughly, and future species losses under habitat loss and climate change (May *et al.* 1995; Brooks *et al.* 1999; Seabloom *et al.* 2002; Thomas *et al.* 2004; Costello & Ward 2006). Nevertheless, empirical evidence indicates a huge variety of SAR shapes, with the reasonableness of the power-law model and the value of the fitted z varying greatly depending on the scale range over which an SAR is analysed, the taxonomic group, and the habitat type (Preston 1962; May 1975; Plotkin *et al.* 2000a,b; He & Legendre 2002; Drakare *et al.* 2006). A systematic understanding of this diversity of observed SAR patterns is lacking.

Two broad categories of SAR can be distinguished. Island SAR are constructed from lists of species (within the taxonomic group of interest) found in spatially disjoint patches of habitat of differing area, such as actual islands or different-sized plots scattered throughout a larger mainland ecosystem. Nested SAR are constructed by averaging species richness across the subplots of specified area, using presence–absence data for each species at some finest scale within a larger plot. This work focuses on nested SAR.

The classic derivation of $z \sim$ ¼ was obtained by sampling larger and larger numbers of individuals drawn from the individual pool, and thus at best applies to island, not nested, SAR. Indeed, the abundance distribution alone cannot determine the shape of a nested SAR; that shape will also depend on the degree of aggregation of the individuals within each of the species (He & Legendre 2002; Harte *et al.* 2008; see Materials and methods). Furthermore, the validity of the power-law model is questionable for many SAR. In particular, when nested SAR data spanning a scale range of several orders of magnitude in area are plotted on log–log axes, the data often clearly exhibit distinct curvature (Rosenzweig 1995;

Plotkin *et al.* 2000a,b; He & Legendre 2002; Drakare *et al.* 2006; Harte *et al.* 2008); such curvature arises even when the R^2-values of linear regressions on log–log scales exceed 0.98 as is typical in SAR analyses.

Here we modify and extend a previously developed application (Harte *et al.* 2008) of the maximum entropy principle from information theory (Jaynes 1982; MaxEnt) to accomplish three closely related goals. We show that: (i) a single quantity, the average species abundance at any specified 'anchor' spatial scale, uniquely and accurately predicts the shape of the SAR at larger or smaller scales and thus the theory permits extrapolation to biome scale of species richness from representative small plots nested within a biome; (ii) excluding the trivial case in which each individual is a unique species, exact power-law behaviour for the SAR emerges only in the limit of increasing average species abundance; (iii) in that power-law limit, z approaches 0, not ¼. We test these predictions against multiple empirical SAR for plant and bird assemblages. Finally, we provide an application of the upscaling inference procedure by successfully extrapolating tree species richness from 48 ¼-ha plots in the Western Ghats to the entire 60 000 km^2 preserve; although this is only a single test of the upscaling procedure over such a huge spatial scale range, it is presented to encourage other tests, which if successful would lead to eventual applications to biome-scale studies of biodiversity.

The theory developed here differs from that in Harte *et al.* (2008) in three essential ways. First, that previous work was a top–down theory that allowed prediction of the SAR at smaller scales from knowledge of species richness and total abundance at some largest scale. Thus, it permitted prediction of information that is frequently available (small-scale species richness) from information that is rarely available (biome-scale species richness). Here, we modify the methods developed in the previous paper in such a way so as to allow the maximum entropy principle to be used to upscale knowledge of species richness at a small scale, where it is often available. This difference is consequential because of widespread interest in having reliable predictions of species richness at large spatial scales. Second, we show that a result hinted at in Harte *et al.* (2008) concerning the limiting case of power-law behaviour for the SAR is stronger than anticipated there. In the previous work we speculated that the power-law limit, in which abundance approaches ∞, corresponded to a value of $z \sim 0.16$. In fact, using the revised upscaling methods developed here, we show that the power-law limit corresponds to $z \to 0$, not 0.16. Third, we use a mix of published and new extensive data sets to demonstrate the universality of the SAR behaviour predicted by our theory.

MATERIALS AND METHODS

Theoretical background

We write the SAR in the form:

$$S(A) = S_0 \sum_{n=1}^{N_0} [1 - P(0|n, A, A_0)]\phi(n|S_0, N_0). \quad (1)$$

Here A_0 is the area at some specified largest spatial scale under consideration, while S_0 and N_0 are the total number of species and individuals, respectively within A_0; A is the area of an arbitrarily selected cell nested within A_0 and $S(A)$ is the expected number of species in A. Equation 1 states that the expected number of species in a cell of area A located within a larger region of area A_0 is the total number of species, S_0, in A_0 times the sum of the probabilities of the occurrences in A of each of those species.

The term $\phi(n|S_0, N_0)$ is the species abundance distribution, which is the probability that a species picked at random from the species pool in A_0 has n individuals. To understand the term $1 - P(0|n, A, A_0)$ in eqn 1, consider the more general function $P(m|n, A, A_0)$, characterizing the individual species. It is a spatial abundance distribution, defined as the probability that a species with abundance n in A_0 has m individuals in a census patch of area A nested within A_0. The expression $1 - P(0|n, A, A_0)$ is thus the probability that a species with n individuals in A_0 is present in such a cell of area A. Explicit expressions for $\phi(n|S_0,N_0)$ and $P(m|n, A, A_0)$ were derived in Harte *et al.* (2008) using methods summarized in Appendix S1.

MaxEnt provides a method for inferring from incomplete knowledge the shape of a probability distribution. In particular, it was proven (Jaynes 1982) that maximizing the information entropy [in its simplest form, information entropy $= -\Sigma P \ln(P)$] of the distribution, P, under the available constraints yields the most likely, or least biased, functional form of the distribution. Here, most likely or least biased means the smoothest possible distribution that satisfies the constraints. Maximization under constraint is carried out using the method of Lagrange multipliers (Stewart 2007). In Appendix S1, we provide a very basic explanation of the MaxEnt method.

In applications of MaxEnt, a critical issue is identifying the constraints. In Harte *et al.* (2008), we assumed an ecosystem characterized by four state variables, in loose analogy with the state variables such as pressure, volume, temperature, and number of moles in thermodynamics. The selected state variables are: ecosystem area (A_0), total species richness (S_0) within A_0, total number of individuals (N_0) in the S_0 species, and total metabolic rate (E_0) of the N_0 individuals. Along with a normalization condition, the ratios

of these state variables provide the needed constraints on a joint probability distribution, $R(n, \epsilon)$, over the abundances of the species and the rates of metabolic energy consumption by the individuals in the species (for details, see Appendix S1). Thus, the mean of the species abundance distribution is given by N_0/S_0 and the mean of the distribution of metabolic rates over the species is E_0/S_0. We denote the spatial scale at which state variables are specified as the 'anchor scale'.

By maximizing the information entropy of that joint probability distribution under the constraints imposed from knowledge of the ratios of the state variables, we showed in Harte *et al.* (2008) that if S_0 is sufficiently large so that $e^{-S_0} << 1$, then the species abundance distribution, ϕ, is the Fisher log series,

$$\phi(n|S_0, N_0) = c_{\phi,0} \cdot e^{-\lambda_{\phi,0} n}/n, \tag{2}$$

where $c_{\phi,0}$, a normalization constant, and $\lambda_{\phi,0}$, a Lagrange multiplier, are given by satisfying the constraints that mean of n over the distribution is equal to N_0/S_0, and that ϕ is normalized to 1.

These constraints lead to the following determining equations for $\lambda_{\phi,0}$ and $c_{\phi,0}$, and thus the species abundance distribution, at the arbitrary spatial scale A_0:

$$\frac{S_0}{N_0}\sum_{n=1}^{N_0} e^{-\lambda_{\phi,0} n} = \sum_{n=1}^{N_0}\frac{e^{-\lambda_{\phi,0} n}}{n} \tag{3}$$

and

$$c_{\phi,0} = \frac{N_0}{S_0}\frac{(1 - e^{-\lambda_{\phi,0}})}{\left(e^{-\lambda_{\phi,0}} - e^{-\lambda_{\phi,0}(N_0+1)}\right)}. \tag{4}$$

As shown in Harte *et al.* (2008), while the constraint of total metabolic energy, E_0 is needed to derive eqn 2, the actual form of the species abundance distribution depends only on S_0 and N_0, not on the numerical value of E_0.

The spatial abundance distribution for each species with known abundance, n, in A_0, can also be derived from MaxEnt. In Harte *et al.* (2008), we showed that

$$P(m|n, A, A_0) = c_P.\exp(-\lambda_P m), \tag{5}$$

where c_P and another Lagange multiplier, λ_P, both depend on n, A, and A_0 and are calculated by satisfying the constraints that P, summed from $m = 0$ to n, is normalized to 1 and that the mean of m over the distribution is nA/A_0. We note that the spatial pattern in the distribution of individuals within species predicted by eqn 5 is not the same as the binomial distribution predicted by the random placement model (Coleman 1981). Relative to the random placement model, eqn 5 predicts aggregation of individuals. Extensive tests of eqn 5 for plant data are presented in Harte *et al.* (2008).

Extending the theory to permit upscaling

At this stage, we depart from Harte *et al.* (2008). There we used a downscaling procedure for calculating the consequences of eqn 1 for areas smaller than the anchor scale A_0. To do that, we fixed the abundance distribution, ϕ, at the anchor scale A_0, where the values of S_0 and N_0 are presumed to be known, and then to implement eqn 1 we calculated $1 - P(0 \mid n, A, A_0)$ from eqn 5 at each scale smaller than A_0. Here, however, we wish to use eqn 1 to upscale, so that we can predict species richness at scales larger than the anchor scale. The derivations in Harte *et al.* (2008) can only be used to downscale from the anchor scale, and S at each scale is presumed to depend solely upon S and N at the anchor scale. Now we use eqn 1 in an iterative procedure that allows us to downscale or upscale because it is re-calculated at each scale with new constraint information, the values of S and N at the halved or doubled spatial scales.

To illustrate this, consider the simpler case of downscaling first. Now, we require only the special case of eqn 5 that is applicable to a scale halving:

$$P\left(0|n, \frac{A_0}{2}, A_0\right) = \frac{1}{n+1}. \tag{6}$$

The apparent discrepancy between eqns 5 and 6 is illusory because for $A = A_0/2$, the Lagrange multiplier λ_P in eqn 5 is identically zero.

Using a notation in which $\lambda_{\phi,A}$ is the Lagrange multiplier appearing in the log-series distribution (eqn 2) at anchor scale A, it is straightforward to show that if eqs 2, 4, and 6 are substituted into eqn 1, and the summation in eqn 1 is carried out, then

$$S(A/2) = S(A)e^{\lambda_{\phi,A}} - N(A)\frac{1 - e^{-\lambda_{\phi,A}}}{e^{-\lambda_{\phi,A}} - e^{-\lambda_{\phi,A}(N(A)+1)}}\left(1 - \frac{e^{-\lambda_{\phi,A}N(A)}}{N(A)+1}\right). \tag{7}$$

If the value of $\lambda_{\phi,A}$ determined by eqn 3 is substituted into eqn 7, then $S(A/2)$ is determined. Because N scales linearly with area in a nested design, $N(A/2) = N(A)/2$, and now with knowledge of $S(A/2)$ and $N(A/2)$ this iterative process can be repeated to determine $\lambda_{\phi,A/2}$ and $S(A/4)$, and so forth. In effect, the approach in Harte *et al.* (2008), in which

the SAD only at the anchor scale is used, discards information whereas this modified approach updates the state variables at each successive scale interval.

More importantly, this modified approach can also be used to upscale species richness data from small plots to allow prediction of biome-scale species richness. In particular, suppose we want to know how many species are in a sampling unit that has area twice that of the anchor scale A at which S and N are known. We can iterate eqn 7 now to obtain:

$$S(A) = S(2A)e^{\lambda_{\phi,2A}} - N(2A) \frac{1 - e^{-\lambda_{\phi,2A}}}{e^{-\lambda_{\phi,2A}} - e^{-\lambda_{\phi,2A}(N(2A)+1)}} \left(1 - \frac{e^{-\lambda_{\phi,2A}N(2A)}}{N(2A)+1}\right). \quad (8)$$

Here there are two unknowns: $S(2A)$ and $\lambda_{\phi,2A}$. $N(2A)$ is of course equal to $2N(A)$, but a second equation relating $S(2A)$ and $\lambda_{\phi,2A}$ is still needed. Equation 3 provides that second relationship if it is rewritten in the equivalent form:

$$\frac{S(2A)}{N(2A)} \sum_{n=1}^{N(2A)} e^{-\lambda_{\phi,2An}} = \sum_{n=1}^{N(2A)} \frac{e^{-\lambda_{\phi,2An}}}{n}. \quad (9)$$

Equations 8 and 9 can be solved numerically for $S(2A)$ and the process can then be iterated up to an arbitrarily large spatial scale that is a multiple of two times the anchor scale area.

SAR metrics for data comparison

To characterize the shape of the predicted and observed SAR, we use two measures that can be calculated at every scale from both theory and SAR data: the local slope, $z(A) = d\log(S(A))/d\log(A)$, and the local curvature, $\eta(A)$, of $\log(S)$ vs. $\log(A)$. We define curvature in the standard way as:

$$\eta(A) = \frac{\left(\frac{d^2\log(S(A))}{d\log(A)^2}\right)}{(1 + z^2(A))^{3/2}}. \quad (10)$$

Because our data and our theoretical predictions are for discrete values of area, we calculate z and η using differences rather than differentials. Hence, $z(A) = 0.5[\ln(S(A)/S(A/2)) + \ln(S(2A)/S(A))]/\ln(2)$ and $\eta(A) = [\ln(S(A/2)) + \ln(S(2A)) - 2\ln(S(A))]/[\ln^2(2)\ (1 + z^2(A))^{3/2}]$.

Data

To test our predictions for the shape of the SAR, we examined 41 empirical species–area curves and 194 values of each of z and η at different N/S values along these curves. These data are from plant censuses in both wet and dry tropical forests, serpentine meadows, temperate montane forest understory, and avian communities in South Africa (details in caption of Fig. 1). They are data sets in which not only are species richness values measured over some range of spatial scales, but also a total abundance estimate is available at the anchor scale. This is essential because knowledge of the state variables, S and N, at that scale is needed to predict the SAR at finer scales. Our demonstration of the potential applicability of our theory for upscaling from small plots to biome-scale species richness values uses tree census data from the Western Ghats Preserve in India (details in caption of Fig. 2).

Total abundance census of plants poses certain unique problems. For grasslands, the main problem, resolving what an individual is, is generally resolved by the convention of counting ramets for clonal species. For the tree censuses, a convention is generally adopted to only count trees with a diameter-at-breast-height above some minimal value (e.g. 10 mm) and as long as that convention is consistently adhered to within a given data set, the MaxEnt procedure can be applied.

RESULTS

The universal SAR and the power-law limit for the widely applicable case in which $e^{-\tilde{S}} << 1$ and $N_i > S_i$, it follows from either eqn 7 or 8 that

$$z(A) \approx \frac{1}{\ln(2)\ln(\frac{1}{\lambda_{\phi,A}})}. \quad (11)$$

The curvature at scale A, $\eta(A)$, can also be expressed solely in terms of $\lambda_{\phi,A}$. But eqn 9 implies that under those same conditions, the Lagrange multiplier, $\lambda_{\phi,A}$ is a function solely of $N(A)/S(A)$, and hence the shape of the SAR, $S(A)$, is a universal function of the ratio N/S, provided that at each scale $e^{-S} << 1$. In the limit $N(A)/S(A) \rightarrow \infty$, it follows from eqn 3 or 9 that $\lambda_{\phi,A} \rightarrow 0$. From eqn 11, it follows that in the $N/S \rightarrow \infty$ limit, $z \rightarrow 0$. But for all practical purposes, this is also the limit of increasing area because N increases linearly with area in a completely nested census design, whereas S increases more slowly. Thus, the power-law model is predicted to be increasingly valid in the limit in which area increases and z decreases towards 0.

Because our theory predicts that the value of N/S at some specified scale determines the shape of the SAR at larger or smaller scales, all SAR have the same shape, differing only in the scale at which they share a common N/S value and the actual number of species at that value. Thus, we predict that all SAR are scale displacements of a universal SAR shape.

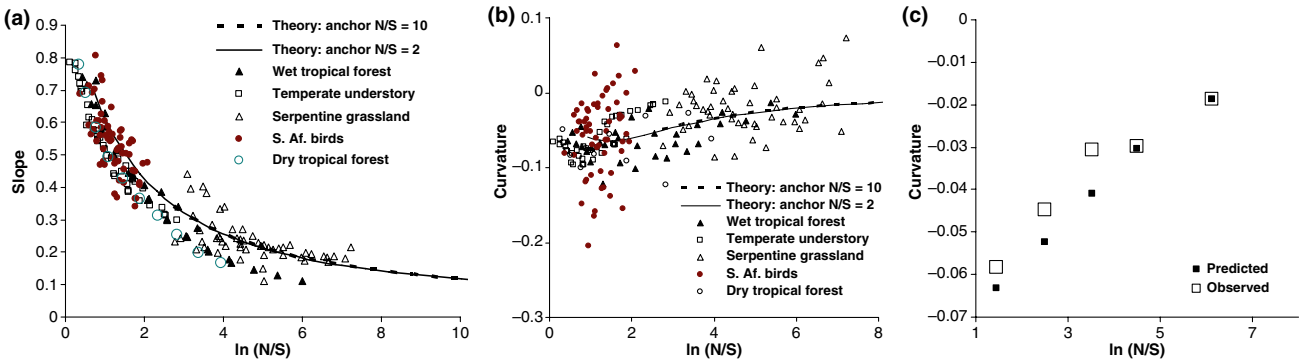

Figure 1 Predicted and observed values for (a) the slope parameter, z, and (b, c) the curvature, η, of nested species–area relationships for log(S) vs. log(A) plots, but expressed here as a function of ln(N/S). (b) All the predicted and observed curvature values are shown here. (c) The results of averaging all the curvature data and the theoretical predictions over binned intervals of N/S are shown here. In (a) and (b), the dashed and solid lines correspond to the theoretical prediction for anchor values of $N_0 = 100$, $S_0 = 10$, and $N_0 = 40$, $S_0 = 20$, respectively. In all the three figure parts, the theoretical predictions are generated by upscaling from the anchor values using eqns 3 and 8. In (a) and (b), filled triangles represent tree census data from three wet tropical forest plots (BCI, Cocoli, Sherman; Condit 1998; Hubbell *et al.* 1999, 2005); open squares represent understory vegetation census data in thirteen 1-m^2 plots in Rocky Mountain conifer and aspen forests (Godinez & Harte, unpublished data); open triangles represent plant census data from five 4-m^2 and one 64-m^2 plots (census repeated for each of the 4 years) in temperate serpentine meadows (Green *et al.* 2003; Smith, unpublished data); filled circles represent data from South African bird censuses carried out during the peak breeding season in 2007 (October–December) at 15 transects located along the west–east productivity gradient, from Kalahari desert in Kgalagadi National Park to humid savannah in Kruger National Park; each transect comprised 32 census points and the distance between the neighbouring points was 1 km. Birds (recorded both acoustically and optically) were censused within 100 m distance around each point during the 20-minute period to ensure equal detectability of all species; open circles represent tree census data from a 10-ha dry tropical forest plot in Costa Rica (Enquist *et al.* 1999).

Numerical confirmation of universality

To demonstrate this numerically, we use eqns 8 and 9 to upscale S starting from the values of S_0 and N_0 at some 'anchor' scale. The solid and dashed lines in Fig. 1a show the predicted z values for $N_0/S_0 = 2$ and 10, respectively. That the curves lie on top of each other is a numerical confirmation that the slope of the SAR is a function of N_0/S_0 alone. As area increases, N increases proportionally to area and thus faster than S, and so increasing N/S corresponds to increasing area. Because the predicted z decreases with increasing N/S, this implies that SAR should exhibit negative curvature. This is confirmed in Fig. 1b, where the predicted curvatures for the same two anchor values of N/S as in Fig. 1a ($N/S = 2$ and 10) again coincide, indicating dependence solely on that ratio.

From Fig. 1b, we see that theory predicts that the power-law limit for the SAR, which is the limit in which the curvature of a log(S) vs. log(A) plot approaches 0, is the limit of increasing N/S. In that limit, Fig. 1a shows that the predicted z does indeed decrease towards zero.

Tests of theory

The central tendency of the observed z values plotted against N/S is in fine agreement with the theoretical prediction (Fig. 1a). Although there is considerable scatter in the observed values of η plotted against N/S (Fig. 1b), the averages of binned curvature values are well predicted by theory (Fig. 1c). We would expect the relatively small observed variability in z-values around the central tendency shown in Fig. 1a to result in the large relative variability in observed η-values because curvatures are small differences of relatively large quantities; plotted on the same vertical scale from 0 to 1, the curvature values appear to be tightly clustered as the slopes. Thus, we note that while the predicted SAR curvatures for the South African bird data appear to deviate considerably from the observed values, the slopes themselves for the bird data appear to be in as good agreement as are the slopes for the plant data sets.

An application of upscaling

The predicted shape of the universal SAR is shown in Fig. 2, where the calculated values of species richness (X in the figure) are obtained by solving eqns 8 and 9 (upscaling) or the equivalent eqns 3 and 7 (downscaling) for the particular case of an anchor value of $S_0 = 32.5$ species and $N_0 = 109$ individuals, at an anchor scale of ¼ ha. These particular anchor values were chosen because they correspond to census data on tree species richness from the 60 000 km^2 Western Ghats Preserve in southern India (Krishnamani *et al.* 2004).

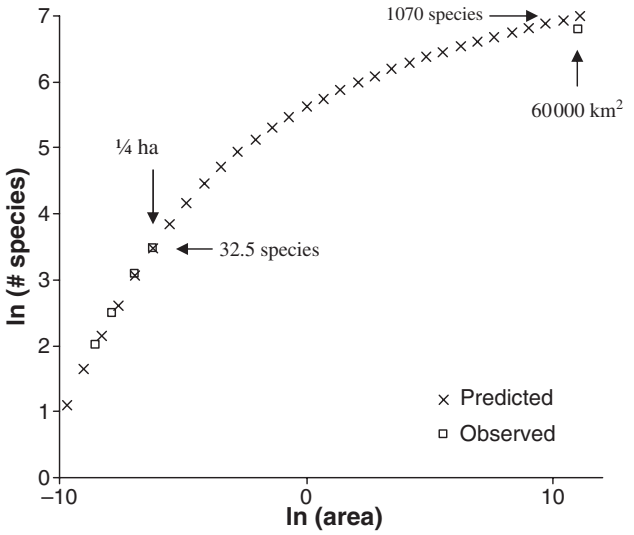

Figure 2 Test of upscaling and downscaling predictions for tree species richness in the Western Ghats. The input information for all the predicted values of species richness shown in the graph are the averages of total abundance and species richness in 48 ¼-ha plots scattered throughout the 60 000 km^2 preserve (Krishnamani *et al.* 2004). In those small plots, abundance and spatial distribution data for every tree species with ≥ 30 mm dbh were obtained and used to construct 48 species–area relationships spanning nested areas within each plot of 0.25, 0.125, 0.05, and 0.025 ha (averages shown by open squares in the graph). Currently *c.* 900 tree species (data point at upper right in the graph) are identified in the preserve, with *c.* 5 new ones reported each year. Area is in units of square kilometre.

We now use these data to demonstrate the potential usefulness of our upscaling procedure. First, we note that downscaling from the anchor scale of ¼ ha predicts accurately the species richness data at these smaller scales (Fig. 2). Then, using eqns 8 and 9, and data only from the anchor scale of ¼ ha, MaxEnt also predicts a total of 1070 tree species in the entire preserve (Fig. 2). Approximately 900 species are currently identified, with several additional ones being reported each year (Krishnamani *et al.* 2004), and thus our extrapolation from ¼ ha to 60 000 km^2, a scale interval greater than 2^{24}, yields a reasonable prediction for the upscaled total tree species richness.

DISCUSSION

The predicted slope, z, of the SAR decreases with increasing N/S and thus increasing area. Because the absolute value of the curvature is predicted to decrease in that limit, the straight-line fit on log–log scales improves as that ratio increases. Indeed, if the predicted SAR over a range of areas spanning three powers of ten (ten doublings, 11 data points) at the largest areas plotted in Fig. 2 is fit to a power-law model, the resulting R^2 is 0.9956 and the slope is 0.10. But distinct curvature is apparent in the predicted log–log SAR over just those 11 data points; with a relatively small number of data points, even an $R^2 > 0.99$ can be associated with distinct and systematic curvature.

Many authors (Rosenzweig 1995; Plotkin *et al.* 2000b; Fridley *et al.* 2006; He *et al.* 2006) have noted that across larger spatial scale ranges (which tend to correspond to larger average abundance per species because N increases roughly in proportion to area and S increases more slowly) fitted z values tended to drop well below ¼. Notably, low slopes ($z < 0.1$) of the SAR have been reported for microbes (Green *et al.* 2004; Horner-Devine *et al.* 2004, Green & Bohannan 2007); such low slopes are in accord with our theory because average abundance per taxonomic unit is likely to be extremely large for micro-organisms. From Fig. 1a, we predict that the average abundance per taxon satisfies $\ln(N/S) > 10$ at the spatial scales at which these data are obtained.

Our MaxEnt-based state-variable theory of macroecology predicts a universal shape for the SAR, exactly as shown in Fig. 2. The variable governing the local slope of the SAR at any scale A is the average abundance per species at that scale, $N(A)/S(A)$. Although average abundance per species will differ from habitat to habitat and from taxa to taxa at any given absolute value of area, all SAR are predicted by our theory to have the same shape when plotted against N/S. Any two SAR can differ, of course, in the magnitude of the area at which they share a common N/S value, which corresponds to a horizontal displacement of the SAR, and they can differ in the actual number of species at that value of N/S, which corresponds to a vertical displacement of the SAR. The data we have examined so far are in reasonable agreement with this prediction.

Applying our theory to a landscape with two distinct biota occupying non-overlapping regions will at the very least require imposition of an additional constraint. But all habitats are heterogeneous to at least some degree, and so the question arises as to what influence this has on the validity of our theory. Because we are looking at completely nested SAR, habitat heterogeneity is effectively averaged over and so we speculate that to the extent the theory reliably predicts SAR, the influence of heterogeneity on the theory's predictions is captured by the influence of heterogeneity on the measured values of the state variables and thus on the constraints.

But in the extreme limit of heterogeneity, in which a landscape is occupied by two distinct biota in non-overlapping regions, the prediction that the slope of the predicted SAR is a monotonically decreasing function of area cannot be valid. In such a landscape, a triphasic SAR is expected. Indeed, it is sometimes observed that at relatively large, often continental, scales the slope of the SAR steepens (Rosenzweig 1995; Fridley *et al.* 2006; Palmer 2007). Is a

MaxEnt approach capable of describing such behaviour? Three possibilities can be identified. First, it is conceivable that if the existence of two distinct biota is incorporated as a new constraint within the MaxEnt framework, the theory might make reasonable predictions. Second, it is possible that even if knowledge of the disjoint species lists is incorporated in the form of an additional constraint, an observed triphasic SAR will still not be predicted by the theory. Finally, at least in some cases it may be that empirical support for a triphasic SAR, at large spatial scales, is a consequence of not using a complete nested design in the construction of the SAR. Further analysis of each of these possibilities is warranted, and in particular we suggest that investigating procedures for incorporating information about non-overlapping species lists between distinct habitats as a constraint is likely to yield interesting insights.

We have not attempted to explain the magnitude of the scatter in the observed z-values in Fig. 1a around their central tendency. It is possible that considerable information about processes not captured in our purely information-theoretic approach can be mined from the deviations from theory, particularly if they are observed to differ from taxa to taxa, from habitat to habitat, and from scale to scale.

Our theory provides a way of estimating species richness at biome scale from data on species richness and total abundance (summed over species) at a much smaller spatial scale. If further tests of our upscaling procedure continue to validate the technique, opportunity would exist to answer long-standing questions concerning, e.g. Amazonian insect species richness from canopy fumigation studies and large-scale microbial diversity from soil cores or water samples.

Traditional methods of upscaling species richness, using nonparametric estimators (Chao & Lee 1992) or extrapolating collectors' curves, are not capable of extrapolating reliably over scale intervals as huge as in the Western Ghats (Krishnamani *et al.* 2004). Indeed only one other previously published method, based on a hypothesized relationship between the dependence of species richness on area and the dependence of a community similarity index on distance between small plots (Krishnamani *et al.* 2004), has yielded a value in the vicinity of 10^3 for the number of tree species in the Western Ghats. Because the MaxEnt application that we use here does not predict the species similarity index, we are unable to establish whether or not the agreement between these two methods is accidental.

We note that our results, and specifically the form of eqn 6, reflect the implicit assumption in our application of MaxEnt that the individuals within a species can be considered indistinguishable; in a previous work (Harte *et al.* 2005, 2008), it was shown that this assumption yields realistically aggregated species-level spatial abundance distributions, whereas the assumption of distinguishability leads to an unrealistic Poisson distribution for the spatial distribution of individuals within species. The implications for spatial structure of the distinction between the assumptions of distinguishability and indistinguishability are discussed in Appendix S1.

The application (Harte *et al.* 2008) of MaxEnt, on which we based our findings here, is by no means the only possible approach of using MaxEnt to understand scaling patterns in ecology. More generally, the best choice of state variables, the related choices of what constraints to assume, and whether to assume distinguishability or indistinguishability (see Appendix S1) of the individuals in the system can only be determined empirically. We have demonstrated here and in Harte *et al.* (2008) that with the state variables and constraints described before, and with our choice of indistinguishability, a realistic set of predictions are obtained; we await seeing if other choices and applications of this powerful approach to inference under limited information in the form of constraints can provide comparably realistic predictions.

This raises the question of the generality of our findings about the dependence of the slope parameter on N/S and thus the absence of power-law behaviour for the SAR. Is non-power-law behaviour an inevitable consequence of applying MaxEnt to macroecology? Here it is essential to understand that MaxEnt is simply an algorithm for finding the least-biased shape of a probability distribution subject to known constraints on that distribution. The ecology is all in the constraints, which in our theory arise from ratios of the defining state variables, A, S, N, and E. Applying MaxEnt to our state variable theory, with the constraints it imposes, does indeed preclude power-law behaviour for the SAR. That is the result illustrated in Fig. 1a or Fig. 2. But applied to different constraints on the two probability distributions, P and ϕ in eqn 1, one could obtain quite different results from the MaxEnt algorithm. For example, if one's prior knowledge of the species abundance distribution happened to be the values of the first and second moments of $\log(n)$, one would obtain a log-normal distribution for ϕ from MaxEnt, and for particular values of those moments, one would obtain the canonical log-normal. In addition, using certain prior information about the spatial distribution of individuals within species, one could obtain from MaxEnt a binomial distribution for $P(m)$, as in classical statistical mechanics. In combination, by the results of Preston and May, referred to in the Introduction, a power-law SAR would result. In that case, the predicted slope values in Fig. 1a would follow a horizontal line with $z \sim 0.25$, rather than falling as a function of N/S. Thus, MaxEnt by itself does not constrain the behaviour of the SAR. Again, the question for ecology is: What set of state variables and constraints derived from them, in combination with MaxEnt, leads to empirically defensible predictions?

Another question arises as to what constitutes the 'correct' unit of analysis in a MaxEnt application. For example, we could have carried out our analyses for genera– or family–area relationships instead of SAR. We also could have confined our analysis to just individuals and species satisfying some perhaps arbitrary criteria (e.g. only plants with yellow flowers). Nothing in the logic of MaxEnt would forbid such choices; if the constraints are selected, the procedure should generate testable predictions. Applications of MaxEnt to such taxonomic categories remain to be tested.

We conclude that MaxEnt illuminates simplicity behind the apparent wide variety of SAR observed in nature at subbiome scales. It informs us that the widely used assumption that SAR generally obeys power-law behaviour with *c.* ¼ power exponents is not likely to be useful in ecology, predicting instead that over limited scale ranges nested SAR approximate power-law behaviour in the limits of $z \to 0$. At a given scale, a single parameter, the ratio *N/S*, governs both the slope and curvature of nested SAR. The empirical success of this theory suggests that the hitherto less-than-successful search for a process-based understanding of the SAR might more productively be directed towards the arguably simpler task of finding a mechanistic explanation for observed values of average abundance per species at a specified scale.

ACKNOWLEDGEMENTS

The authors are grateful to Richard Condit for sharing data from the Smithsonian Tropical Research Institute plot at BCI, to Jessica Green for vegetation data from the UC McLaughlin Reserve, to Jacob Godinez for data collected with J. Harte from Montane forest understory, to David Horak, Ondrej Sedlacek, Jiri Reif and Tomas Albrecht for bird data collected with D. Storch from South Africa, and to R. Krishnamani for both data and extensive insights into biodiversity from the Western Ghats. The project was supported by an NSF grant to J. Harte, an NSF REU grant to the Rocky Mountain Biological Laboratory, and grants from the Czech Ministry of Education (MSM0021620845, LC06073), and the Academy of Sciences of the Czech Republic (IAA601970801) to D. Storch. The authors benefited from conversations with Jacob Godinez, Erin Conlisk, Ken Harte, and Graham Pyke, as well as from the comments by Hugh Possingham.

REFERENCES

Brooks, T.M., Pimm, S.L. & Oyugi, J.O. (1999). Time lag between deforestation and bird extinction in tropical forest fragments. *Conserv. Biol.*, 13, 1140–1150.

Chao, A. & Lee, S.-M. (1992). Estimating the number of classes via sample coverage. *J. Am. Statist. Assoc.*, 87, 210–217.

Coleman, B. (1981). On random placement and species–area relations. *J. Math. Biosci.*, 54, 191–215.

Condit, R. (1998). *Tropical Forest Census Plots*. Springer-Verlag & R.G. Landes Co., Berlin.

Costello, C. & Ward, M. (2006). Search, bioprospecting and biodiversity conservation. *J. Environ. Econ. Manage.*, 52, 615–626.

Drakare, S., Lennon, J.J. & Hillebrand, H. (2006). The imprint of the geographical, evolutionary and ecological context on species–area relationships. *Ecol. Lett.*, 9, 215–227.

Enquist, B.J., West, G.B., Charnov, W.L. & Brown, J.H. (1999). Allometric scaling of production and life history variation in vascular plants. *Nature*, 401, 907–911.

Fridley, J.D., Peet, R.K., van der Maarel, E. & Willems, J.H. (2006). Integration of local and regional species–area relationships from space-time species accumulation. *Am. Nat.*, 168, 133–143.

Green, J.L. & Bohannan, B.J.M. (2007). Biodiversity scaling relationships: are microorganisms fundamentally different? In: *Scaling Biodiversity* (eds Storch, D., Marquet, P.A. & Brown, J.H.). Cambridge University Press, Cambridge, pp. 129–149.

Green, J.L., Harte, J. & Ostling, A. (2003). Species richness, endemism, and abundance patterns: tests of two fractal models in a serpentine grassland. *Ecol. Lett.*, 6, 919–928.

Green, J.L., Holmes, A.J., Westoby, M., Oliver, I., Briscoe, D., Dangerfield, M. *et al.* (2004). Spatial scaling of microbial eukaryote diversity. *Nature*, 432, 747–750.

Harte, J., Conlisk, E., Ostling, A., Green, J.L. & Smith, A.B. (2005). A theory of spatial-structure at multiple spatial scales. *Ecol. Monogr.*, 75, 179–197.

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

He, F. & Legendre, P. (2002). Species-diversity patterns derived from species–area models. *Ecology*, 83, 1185–1198.

He, Z., Zhao, W., Chang, X., Chang, X. & Fang, J. (2006). Scale dependence in desert plant diversity. *Biodiver. Conserv.*, 15, 3055–3064.

Horner-Devine, M., Lage, M., Hughes, J. & Bohannan, B.J.M. (2004). A taxa–area relationship for bacteria. *Nature*, 432, 750–753.

Hubbell, S.P., Foster, R.B., O'Brien, S., Wechsler, B., Condit, R., Harms, K.E. *et al.* (1999). Light gaps, recruitment limitation and tree diversity in a neotropical forest. *Science*, 283, 554–557.

Hubbell, S.P., Condit, R. & Foster, R.B. (2005). Barro Colorado Forest Census Plot Data; URL http://ctfs.si/edu/datasets/bci.

Jaynes, E.T. (1982). On the rationale of maximum entropy methods. *Proc. Instit. Elec. Electron. Eng.*, 70, 939–952.

Krishnamani, R., Kumar, A. & Harte, J. (2004). Estimating species richness at large spatial scales using data from small discrete plots. *Ecography*, 27, 637–642.

Lomolino, M.V. (2001). The species–area relationship: new challenges for an old pattern. *Prog. Phys. Geogr.*, 25, 1–21.

May, R.M. (1975). Patterns of species abundance and diversity. In: *Ecology and Evolution of Communities* (eds Cody, M.L. & Diamond, J.M.). Harvard University Press, Cambridge, pp. 81–120.

May, R.M., Lawton, J.H. & Stork, N.E. (1995). Assesing extinction rates. In: *Extinction Rates* (eds Lawton, J.H. & May, R.M.). Oxford University Press, Oxford, pp. 1–24.

Palmer, M.W. (2007). Species-area curves and the geometry of nature. In: *Scaling Biodiversity* (eds Storch, D., Marquet, P.L. & Brown, J.H.). Cambridge University Press, Cambridge, pp. 15–31.

Plotkin, J.B., Potts, M.D., Yu, D.W., Bunyavejchewin, S., Condit, R., Foster, R. *et al.* (2000a). Predicting species diversity in tropical forests. *Proc. Natl. Acad. Sci. USA*, 97, 10850–10854.

Plotkin, J.B., Potts, M.D., Leslie, N, Manokaran, N., LaFrankie, J. & Ashton, P.S. (2000b). Species–area curves, spatial aggregation, and habitat specialization in tropical forests. *J. Theor. Biol.*, 207, 81–99.

Preston, F.W. (1962). The canonical distribution of commonness and rarity: Part I. *Ecology*, 43, 185–215, 431, 432.

Rosenzweig, M.L. (1995). *Species Diversity in Space and Time*. Cambridge University Press, Cambridge.

Seabloom, E.W., Dobson, A.P. & Stoms, D.M. (2002). Extinction rates under nonrandom patterns of habitat loss. *Proc. Natl. Acad. Sci. USA*, 17, 1129–1134.

Stewart, J. (2007). *Calculus*. Brooks Cole, Florence.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C. *et al.* (2004). Extinction risk from climate change. *Nature*, 427, 145–148.

SUPPORTING INFORMATION

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

Appendix S1 The MaxEnt method and its application to ecology.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Editor, Howard Cornell
Manuscript received 2 January 2009
First decision made 2 February 2009
Second decision made 29 March 2009
Third decision made 10 April 2009
Manuscript accepted 19 April 2009