

On the origin and robustness of power-law species–area relationships in ecology

Héctor García Martín*[†] and Nigel Goldenfeld**^{‡§}

*Department of Physics and [†]Institute for Genomic Biology, University of Illinois at Urbana–Champaign, 1110 West Green Street, Urbana, IL 61801

Edited by Simon A. Levin, Princeton University, Princeton, NJ, and approved May 12, 2006 (received for review January 14, 2006)

We present an explanation for the widely reported power-law species–area relationship (SAR), which relates the area occupied by a biome to the number of species that it supports. We argue that power-law SARs are a robust consequence of a skewed species abundance distribution resembling a lognormal with higher rarity, together with the observation that individuals of a given species tend to cluster. We show that the precise form of the SAR transcends the specific details of organism interactions, enabling us to characterize its broad trends across taxa.

clustering | lognormal

There are an estimated 5–10 million nonmicrobial species on earth (1). This estimate is based on the extrapolation of sampled data rather than on a thorough count. Understanding the spatial distribution of biodiversity is of practical importance, providing information regarding the probability of species extinction due to loss of habitat (2) and the design of reserves that can protect the greatest amount of biological diversity (3, 4). In addition, the spatial distribution of species is an important factor governing a wide range of key biological phenomena, including competition (5), division of niche space (6), and the effect of disturbances (7).

One of the main tools for extrapolation is the species–area relationship (SAR), relating area, A , to the number of species it supports, S . Several functional forms have been demonstrated for the SAR (8–10), but the power-law form, in which the number of species is a constant power of the area $S = cA^z$, is most commonly cited (10, 11). This form is reported (3) for climates ranging from temperate to tropical and for such different organisms as plants, birds, insects, mammals, and fish; this functional dependence also was recently extended to bacteria (12) and microbial eukaryotes (13). The exponent z has been reported to depend on habitat, scale, and taxa (3), with values in the range 0.1–0.4 for plants and birds in island groups [the most frequent values being in the range 0.2–0.4 (14)], values near unity for intercontinental scales, and a value as small as 0.05 for microbes (12). The widely reported occurrence of this functional form is indicative of a very robust mechanism, transcending the specific details of organism interactions. Explanations for the occurrence of this power-law SAR focusing on abundance distributions (15, 16), the allocation of individuals (17–23), or population dynamics (24–26) have been proposed. Nonetheless, there remains no consensus as to the common occurrence of the power-law SAR itself nor the systematics of the SAR exponents z . Modern theoretical work assuming self-similarity (27) permits neither a calculation of the SAR exponent nor its variation across taxa. Thus, the key questions (11) are (i) why is the power-law relationship so commonly reported? and (ii) what factors account for the observed distribution of exponent values?

In this paper, we show that the SAR of an ecosystem depends primarily on the distribution of mean minimum distances between conspecific individuals and is, therefore, determined by the statistics of extremes. We then proceed to show that the power-law SAR is a very good description whenever two conditions are fulfilled: (i) individuals of each species cluster; and (ii) the abundance distribution for the species is similar to

Preston's lognormal (15) [i.e., the fraction of species with n individuals is a Gaussian function of $\log(n)$] but with a higher rarity. Clustering or aggregation is reported for most species (28, 29) and lognormal abundance distributions are commonly reported (24, 27), although for many ecosystems other functional forms provide better fits (16, 30).

Our findings follow purely from statistical geometry and do not involve ecosystem internal dynamics processes, such as competition, selection, immigration, speciation, or the effects of landscape variability, which does not mean that they are immaterial but rather that they are relevant only in as much as they affect the two properties above. A similar separation of ultimate and immediate causes for the SAR has been suggested by He and Legendre (31). We find that the appearance of an effective power-law SAR is robust to small changes in the details of the clustering and the abundance distribution and give some insight into the factors that may determine systematic trends in the variation of the SAR exponent z across taxonomic groups.

Although the importance of clustering and abundance distributions equal or close to lognormals has been discussed in the past (16, 17, 19–23, 27, 32), the present paper links both facts and extreme statistics to explain the robustness and the prevalence of the power-law SAR.

Results and Discussion

A Continuum Definition of the Species Area Rule and Its Relation to Extremal Statistics. The SAR is commonly measured in field settings by imposing a grid on a given area and averaging the number of species found throughout all possible groups of nonoverlapping grid cells. We will refer to the number of species found in an area A averaged in this way as $S_G(A)$. For the purpose of analysis, it is more convenient to use a continuum version of this concept, $S_C(A)$, defined to be the number of species found in a circular area around a given origin averaged over all possible origins (see Fig. 6, which is published as supporting information on the PNAS web site). S_C and S_G are quantitatively equivalent, with the former automatically ensuring nested sampling within the biome, which is particularly relevant because it is known that when the sampling grid is not nested, the SAR is not a power law but exhibits downward curvature (3).

Assume that there are S species with n_s individuals ($s = 1, 2, 3 \dots S$) in a biome Ω with area A . To obtain the statistic measured by the SAR, we must count how many species are within a circular area A of radius R centered on a particular point \vec{r} within the biome, and then average over all points \vec{r} . The contribution of a single species to the total number of species in the area A will only depend on the closest individual of species s to \vec{r} ; we denote the distance of this individual by r_m^s . This

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviation: SAR, species–area relationship.

[†]Present address: Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598.

[§]To whom correspondence should be addressed. E-mail: nigel@uiuc.edu.

© 2006 by The National Academy of Sciences of the USA

contribution will be 1 if the closest individual is within a radius R of \bar{r} and 0 otherwise. Thus, there is a connection between the SAR and the statistics of extremes (33, 34), although classic results (35) are not applicable here because of strong correlations between individuals.

Expressed mathematically, this reasoning leads to a formal expression for $S_C(A)$ (see *Supporting Text*, which is published as supporting information on the PNAS web site):

$$S_C(A) = \sum_s F^s(R) \quad [1]$$

$$F^s(R) \equiv \frac{1}{A} \int_{\Omega} d\bar{r} \Theta(R - r_m^s), \quad [2]$$

where $\Theta(x)$ denotes the Heaviside function, which is unity for positive values of its argument, and zero for negative or null values of its argument. We refer to the function $F^s(R)$ as the proximity function; it equals the fraction of the area Ω for which the closest individual of species s is within radius R . Alternatively, it can be interpreted as the probability that upon selecting a random point from Ω , the closest individual of species s is within a radius R .

This result is exact and valid for any $S_C(R)$, not only power-law ones. The functions $F^s(R)$ can be calculated for a given distribution of individuals, and their sum exactly gives $S_C(R)$. This link between SAR and extremal statistics is valid for any functional dependence of the SAR, although we will concentrate on the power-law case, because it is the one most often reported (10).

Note the importance of the extremal character of the SAR. Bulk changes in the distribution of individuals may be irrelevant for the SAR, as long as the minimum distance individuals stay the same. Conversely, changes in a small group of well selected individuals can change the SAR dramatically. This nonlinear characteristic will prove to be important when considering the robustness of the SAR below.

Modeling the Distribution of Individuals. Data from thoroughly sampled plots of several tropical forests (28, ¶) show that individuals cluster in such a way that their correlation function $\rho_s(r)$ [equivalent to the relative neighborhood density $\Omega_s(r)^\ddagger$] is close to a power law. The correlation function $\rho_s(r)$ is defined as the probability that two individuals of species s are at a distance r from each other.

To properly mimic individual distributions, we will therefore use a bisecting tree algorithm that creates individual distributions whose correlation functions are power laws with exponents ω_s (see Fig. 7, which is published as supporting information on the PNAS web site). The algorithm is depicted in Figs. 8 and 9, which are published as supporting information on the PNAS web site, to which the reader is referred. It is a continuous and randomized version of the one proposed by Harte *et al.* (27) and avoids some of the problems pointed out by Maddux (36) (see Fig. 10, which is published as supporting information on the PNAS web site).

A pool of 6,000 species (500 for each value of ω_s) was created. We will see in the following sections which selections of species from this pool produce a power-law SAR. The individuals' distribution for each species s is characterized by its cover c . Cover, or occurrence or range as it is also known, is defined as the number of squares that a species is present in if a square grid was to be superimposed on the area (37). Cover is the relevant quantity for the SAR: It does not matter how many individuals there are in the minimum area as long as there is at least one.

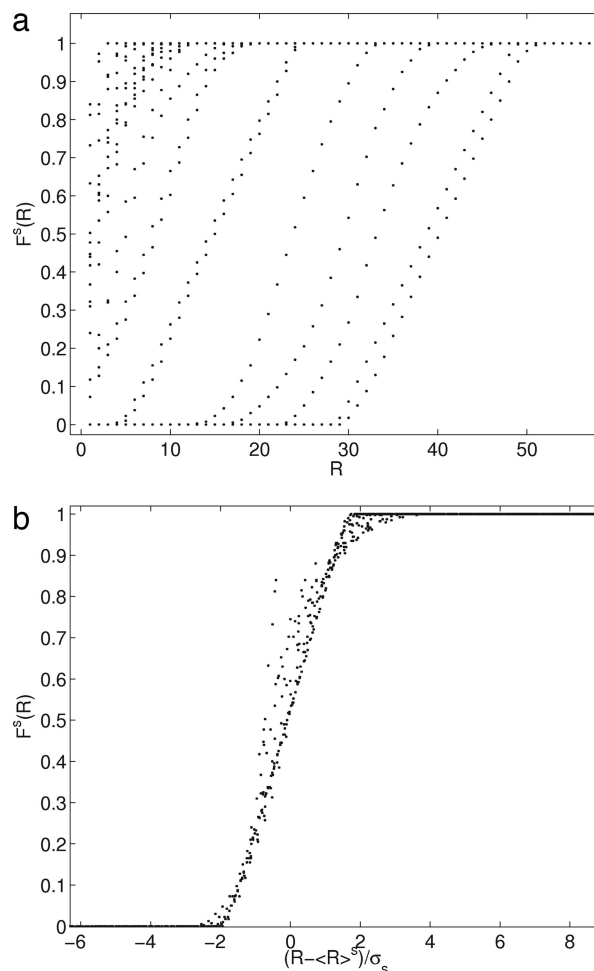


Fig. 1. Data collapse of $F^s(R)$ for the case of power-law-correlated individuals. (a) Plot of $F^s(R)$, the fraction of the area Ω for which the closest individual of species s is within radius R , without scaling. (b) The same data approximately collapsed onto a universal curve after the functions have been scaled by the standard deviation σ_s and centered around the mean minimum distance $\langle R \rangle^s$. For purposes of clarity, only two species for each value of ω_s have been displayed.

In the end, we will be taking the length of the minimum displacement in the algorithm explained in Fig. 8 as the size of the minimum area, so cover and abundance will be the same. We will use them indistinguishably in the rest of the paper. Abundance distributions and cover distributions will therefore be equivalent.

We are now in position to calculate the proximity functions $F_s(R)$ from these allocations of individuals. $F_s(R)$ is the distribution of minimum distances from points in Ω to individuals of species s . For the case of uncorrelated, independent, identically distributed variables, the statistics of extremes (maximum or minimum values) have been heavily studied. In such a case, the extremal distribution is a universal function, one of the three classes of Fisher–Tippett distribution (33–35).

This classical result is not applicable here. The distances from a point in the area Ω to each of the individuals are not independent and identically distributed because they are clustered. If an individual is at a distance R from \bar{r} , it is very likely that other individuals are at a distance similar to R : They are strongly correlated. In the same way, the set of distances from a point \bar{r} is correlated with the set of distances from a nearby point \bar{r}' .

[¶]Ostling, A. & Harte, J. (2000) *Science* 290, 671 (technical comment).

Our key result, however, is that it is still possible to reduce the extremal distributions for highly correlated individuals to a scaling function of the kind:

$$F^s(R) \approx F\left(\frac{R - R^s}{\sigma_s}\right), \quad [3]$$

where F is a universal scaling function (Fig. 1) that is apparently only weakly dependent on ω_s , $\langle R \rangle^s$ is the average of $f^s(r) \equiv (dF^s(R))/(dR)_{R=r}$, and σ_s is the corresponding standard deviation (see *Supporting Text* for a formal definition). The symbol \approx indicates that the equality is only approximate, as can be observed in Fig. 1.

The function $f^s(r)$ represents the fraction of area for which the closest individuals are at a distance r . Thus, $\langle R \rangle^s$ is the average minimum distance to an individual of species s from points in the averaging area Ω .

The mean minimum distance $\langle R \rangle^s$ and the standard deviation σ_s are not independent but follow a relationship well known for extremal statistics (35): They are linearly related until the former is comparable to half the side of the square containing the biome (see Fig. 6; see also Fig. 11, which is published as supporting information on the PNAS web site). Therefore, the only piece of information necessary from each species that is relevant regarding the SAR is $\langle R \rangle^s$. The distribution of these mean minimum distances will characterize the SAR.

We have used a specific algorithm to obtain power-law correlation functions, but we will see in *The Importance of Clustering* that the specific shape of the correlation function will not be significant as long as it reproduces aggregation, which is a common characteristic of populations (21).

Parametrization of Mean Minimum Distances as a Function of Cover.

Intuitively, it is clear that the average minimum distance to an individual of species s , $\langle R \rangle^s$, should depend on its cover, c . For higher covers, the probability of finding an individual of that species within a radius R of \bar{r} is greater than for smaller values of c . The degree of conspecific clustering also influences $\langle R \rangle^s$. For extremely aggregated species, the average minimum distance from an averaging point will increase for a given cover, whereas, for a spatially random or regular distribution, the average minimum distance will tend to be smaller. For the practical purpose of describing the ensemble of species characterized by power-law correlation functions, the distribution of $\langle R \rangle^s$ can be represented as the sum of a cover-dependent average value plus a noise term (see Fig. 2):

$$\frac{1}{R_s} = 0.070043 + 0.000167c + \delta(c), \quad [4]$$

where the cover-dependent noise term $\delta(c)$ represents the stochasticity of the algorithm and has a probability distribution derived from the density of points with a given $\langle R \rangle^s$ and cover c . Each species in the pool will thus be characterized exclusively by its cover c .

This analysis does not ignore the importance of the spatial distribution of individuals (38). The fact that we can parametrize the mean minimum distances as a function of cover is a consequence of using a specific type of clustering represented by power-law correlation functions. Nonetheless, we will show below that for highly clustered, individual distributions, the actual clustering structure at constant cover only contributes very weakly to the SAR.

Abundance Distributions Compatible with the Power-Law SAR and Correlation Functions. Because the proximity functions are determined by the average minimum distance and the distribution of these can be written as function of cover, it is possible to write Eq. 1 in a linear form with respect to the fraction of species $p(c)$ with cover c (see *Supporting Text*):

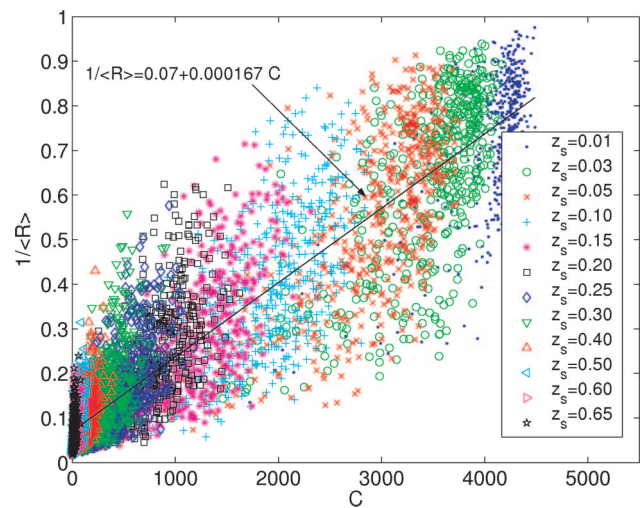


Fig. 2. The dependence of average minimum distance $\langle R \rangle^s$ with the cover c , as expected, decreases with c . The ensemble of $\langle R \rangle^s$ can be described as a stochastic distribution around a cover-dependent average value. Each species will therefore be characterized by its cover.

$$S_C(R) = \int_0^{c_{\max}} h(R, c) p(c) dc, \quad [5]$$

where $h(R, c)$ is the contribution to the number of species inside a radius R by species of cover c . Bear in mind that it is only possible to cast $S_C(R)$ in this form because we have chosen a particular [but well supported by experimental evidence (28, 1)] type of clustering. We will, nonetheless, show that, regarding the highly averaged statistic that the SAR is (averaged over species and over origin points \bar{r}), the actual cluster structure is unimportant for highly clustered distributions, such as those represented by power-law correlation functions. Therefore, the function $h(r, c)$ will be approximately the same for all highly clustered distributions (Fig. 12, which is published as supporting information on the PNAS web site).

Eq. 5 is a linear system solvable through standard methods (39). For any functional dependence of $S_C(R)$, it is possible to find the abundance/cover distributions $p(c)$ compatible with it. Because it is an underdetermined system, there is always at least one solution. Furthermore, this solution is not unique; there exists a set of cover distributions $p_k^*(c)$ (the kernel) that added to any solution of Eq. 5 produces the same SAR. This wide range of solutions is a consequence of standard deviations growing linearly with $\langle R \rangle^s$ (see *Field Data* for an example), which, in turn, is a repercussion of the extremal nature of the SAR and the dimensionality (see *Supporting Text* for the mathematical details). Note that these same considerations and the linear system-solving algorithm can be applied to non-power-law SARs. Different SAR patterns would give rise to different compatible abundance distributions, but the properties of the solution and the procedure to obtain it would be the same.

Not all solutions to Eq. 5, however, are ecologically acceptable. Because $p(c)$ represents the fraction of species with cover c , we must demand that it remains positive and that its sum is equal to the total of species. Additionally, one would expect the fraction of species to have a smooth dependence of c for a stable ecosystem. These conditions will limit the range of solutions compatible with power-law SARs for each exponent z and determine which values of the exponent are more likely to be observed.

The resulting SARs can be seen in Fig. 3 *Left*. The corresponding abundance distributions are shown in Fig. 4. Instead of

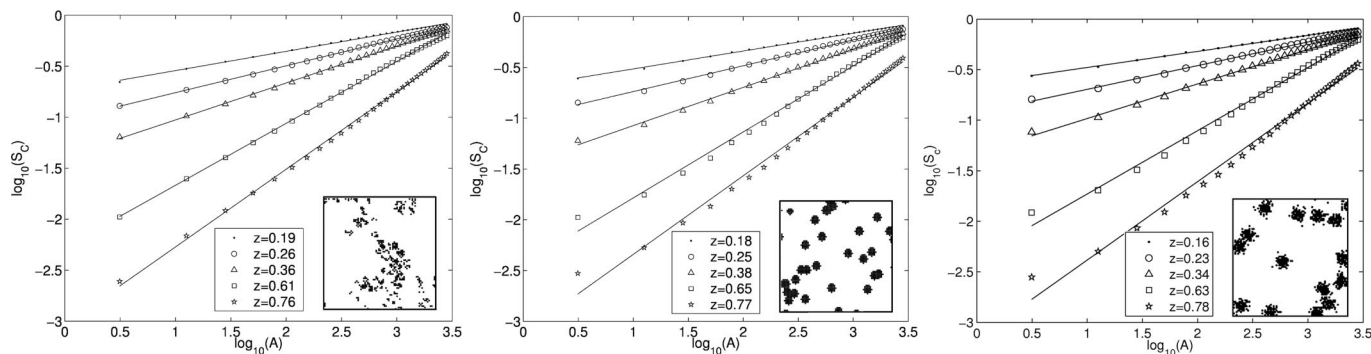


Fig. 3. Number of species as estimated by the continuum counting method described in the text, $S_C(A)$, for self-similar (*Left*), compact cluster (*Center*), and Gaussian cluster distributions (*Right*) of individuals, plotted as a function of sampling area, A . *Insets* show examples of the individual distributions for a randomly chosen species. The same cover distribution $p(c)$ as for the power-law correlation function case was used for the compact and Gaussian case. The total cover for each species was divided in compact groups and randomly spread over the area. The Gaussian case was identically managed, except that individuals in each cluster were distributed by following a Gaussian distribution (plots shown have a clustering coefficient of $\mu = 1.5$ and use the same abundance distributions as the compact cluster case). In all three cases basically the same SARs are obtained, showing that differences in cluster structure only produce lower-order corrections to the SAR. All SARs display a linear correlation coefficient $r^2 > 0.99$.

a continuous function $p(c)$, the solutions have been calculated for discrete abundance distributions binned according to a logarithmic scale of doubling intervals: [1; 2–3; 4–7; 8–15; 16–31; 32–63; 64–127 ...]. For the current case, cover and abundance are equivalent, and this binning is roughly equivalent to the octave classification proposed by Preston (15) and is commonly used in ecology for abundance distributions. Because populations tend to increase geometrically, the natural variable to be considered is the logarithm of the abundance (16).

Note the variation with the SAR exponent z of the mode location and spread. For small z , large covers are preferred, and the distribution spreads over all possible covers. Such a large cover would be expected to be applicable to microorganisms, which are widely accepted to be homogeneously dispersed, with abundance distribution reflecting selection by the local environment (40).

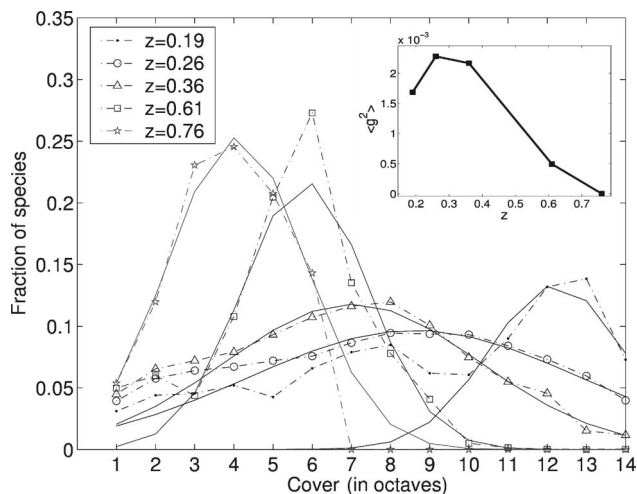


Fig. 4. Abundance distributions corresponding to the SARs shown in Fig. 3 *Left*. Data are plotted as a function of cover (equivalent to abundance in our case), following Preston's octave system (15). Lognormal fits to the cover distributions that produce power-law SARs are shown by the solid line. The fit is satisfactory, except at low cover, where the lognormal underestimates the fraction of species, as is typical of observed abundance distributions (24, 27). Multiple abundance distributions are compatible with the same power-law exponent z . (*Inset*) This fact is portrayed through the average degeneracy (g^2) (see main text). The average degeneracy drops with growing values of the exponent z , indicating that high values of z are less robust to changes in the abundance distribution.

Conversely, for high exponents z , the mode is centered around low cover, and the spread of the distribution is much smaller. The distributions fit rather well to lognormals, except for the fact that they exhibit more rarity, which is a well known characteristic of realistic abundance distributions (24, 27).

The Robustness of Scale-Free SARs. The appearance of an effective power-law SAR is quite robust and arises from a wide range of species abundance distributions of the form shown in Fig. 4. Moreover, as stated before, for each exponent z , there are several cover distributions for Eq. 5 that added to any solution produce exactly the same SAR. The range of these equivalent solutions is shown in the Fig. 4 *Inset* as plot of the average degeneracy (g^2) versus z . The average degeneracy is found by calculating the standard deviation for all of the degenerate (same value of z) abundance distributions with respect to the plotted (i.e., mean) one and averaging its square over all abundances. Taking into account these two factors, the occurrence of an effective power-law SAR is not mandatory but is very likely to arise with the type of abundance distributions commonly reported. Finally, we found that even the addition of cover distributions to the solution of Eq. 5 not belonging to its kernel produces acceptable power laws, given the amount of noise that these plots usually display (3).

These considerations suggest two reasons for why power laws with exponents $z = 0.2-0.4$ are more commonly reported (14). First, their corresponding species–abundance distributions are spread relatively evenly over a wide range of cover, and, secondly, the range of compatible cover distributions $p(c)$ is higher for lower values of z , as can be seen in the inset of Fig. 4. The explanation for the latter consideration is that the addition of cover distributions from the kernel of Eq. 5 is more likely to violate the positivity condition ($p(c) > 0$) for high values of z , because the corresponding abundance distributions are almost null for high cover.

The Importance of Clustering. So far, we have shown that the appearance of the power-law SAR is quite robust to changes in the abundance distribution: As far as it resembles a lognormal with extra rarity, everything seems to be in qualitative and even semi-quantitative agreement with observations. We have not, however, checked the robustness with regard to changes in the distribution of individuals. Which characteristic of the individual distributions is important for the power-law SAR to appear? We now show that the key factor is clustering, which does not mean that the individuals must be inhomogeneously distributed; rather, they must form sets of compact blocks in which individuals are at minimum cover grid

distance from each other (so abundance and cover stay equivalent), as graphically shown in Fig. 3 *Center Inset* (a quantitative measure of the level of clustering required is given below). With this criterion, distributions with a power-law correlation function of exponent close to zero (see Fig. 7) are still effectively clustered despite being homogenous: The whole area is covered by a single cluster. As long as individuals cluster in this fashion and their species' relative abundance distribution is consistent with the behavior reported above, a power-law SAR results.

To justify this assertion, consider what would happen to the distribution of average minimum distances $\langle R \rangle^s$ if a fraction χ of the individuals for each species were to be randomly relocated in the area. As emphasized above, only the closest individuals are relevant when calculating $\langle R \rangle^s$. When individuals are redistributed randomly, the effect is that they become more evenly distributed over the area and for every position \vec{r} (see Fig. 6) the closest individual tends, on average, to be closer. The minimum distances $\langle R \rangle^s$, therefore, decrease for a given cover, as can be seen in Fig. 13, which is published as supporting information on the PNAS web site. As χ increases, the fraction of large $\langle R \rangle^s$ is depleted, and reasonable abundance distributions (without an inordinate amount of low cover species) produce an excess of $\langle R \rangle^s$ in the low and middle end. From Eqs. 1 and 3, it follows that these distributions generate convex log-log plots (negative second derivative), as reported for random distributions (14), instead of power-law SARs.

Moreover, the clustering does not need to be self-similar. The structure of the cluster has little bearing on the final SAR. As shown in Fig. 3, self-similar clusters, compact clusters, or Gaussian-distributed clusters all give basically the same results. The reason is that, as stated above, only the minimum distances to the averaging points \vec{r} matter regarding the contribution to the SAR. The minimum distance to the points in the cluster depends only on the closest individual and, as long as the individuals are highly clustered, will be very weakly dependent on the cluster structure.

This property is better explained by considering how the average minimum distance $\langle R \rangle^s$ changes when a new individual is added. To define $d\langle R \rangle^s/dc$ unambiguously, let us assume that all individuals of species s are removed from the considered area and are added back randomly one by one to their previous positions. $\bar{R}(c)$ is then defined as the average minimum distance with cover c , and $\bar{R}(C^s) = \langle R \rangle^s$, where C^s is the cover of species s . The derivative $d\bar{R}/dc$, or change of \bar{R} when a new individual is added, then determines $\langle R \rangle^s$ through its integration. In the case of clustered distributions, when a new individual is added, it will on average lie close to one of the individuals already present. The distribution of minimum distances will therefore scarcely change, and the change will be largely independent of the cluster structure (see Fig. 5). This distribution of minimum distances will change only greatly when a new cluster, separated from the rest, is created. Therefore, $\langle R \rangle^s$ is more strongly dependant on the number of clusters than the cluster structure. In the case of nonclustered distributions, the new individuals are more likely to appear in the void spaces between individuals. In this case, the correlation with the position of present points has a greater effect on the change of $\langle R \rangle^s(c)$.

The considerations above suggest that cluster structure has little effect on the SAR in the case of highly clustered distributions. This fact can be demonstrated by showing that the same SAR can be obtained for the given cover distribution by using the crudest description of clustered distributions: A randomly placed set of compact blocks of individuals (see Fig. 3 *Center Inset*). The same cover distribution $p(c)$ as for the power-law correlation function case was chosen. The number of groups n_G was randomly chosen between a minimum m_g and the maximum possible amount of groups M_G given the species cover c and the group minimum size m_a ($M_G = c/m_a$). The leftover cover units were clustered in groups of size equal to m_a and distributed randomly. A further test that the cluster structure does not have a great influence on the SAR is obtained by using Gaussian

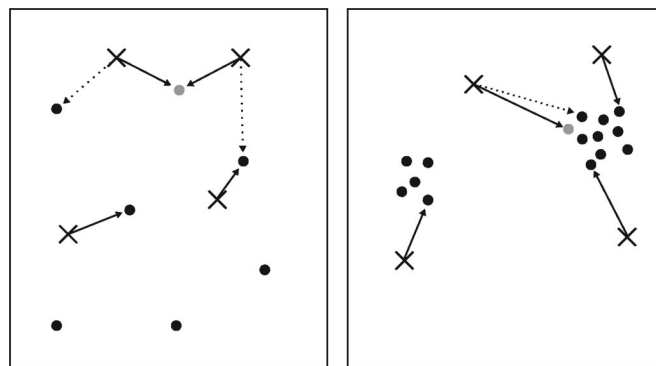


Fig. 5. Clustered (*Right*) versus nonclustered (*Left*) distributions of individuals. Shown is the change in the distribution of minimum distances when a new individual is introduced. Black points represent species individuals, and the gray point represents the new individual. The crosses are a subset of averaging sites for the minimum distances (average is done over all points in the averaging area, see Fig. 6). Solid arrows represent minimum distances after the additional individual is added. Dotted arrows show minimum distances before the additional individual is added. For clustered distributions, the cluster structure is less relevant for determining minimum distances than the actual number of clusters. For nonclustered distributions, the relative position of the new point with respect to the rest has great influence in determining the minimum distances.

distributed clusters (21) instead of compact clusters (see Fig. 3 *Right*). As far as the SAR is concerned (clustering details are doubtlessly important for other statistics), any of these coarse approximations of individual clustering suffices, with cluster structure mattering only for lower-order effects. This property explains why very different clustering patterns display the power-law SAR and, in particular, how fractally distributed species with different fractal exponents can still yield a power-law SAR (41, 42). The average over species with different fractal exponents is dominant over the average over species with the same exponent.

In any of these cases, the proximity functions exhibit good data collapse, σ_s and $\langle R \rangle^s$ are proportional, power-law SARs are obtained, and the species-abundance distributions are essentially unchanged from the power-law-correlated case (Fig. 3; see also Figs. 14 and 15, which are published as supporting information on the PNAS web site). The only difference seems to be that standard deviations tend to be somewhat smaller for the compact cluster case. For $z = 0.2-0.4$, we obtain very good fits to power-law SARs for the new distributions of mean minimum distances, which is not the case for $z = 0.61, 0.76$: A curvature remains in log-log plots (see Fig. 3 and *Supporting Text* for details).

Thus, we conclude that there is another reason for the commonness of values of z in the range $z = 0.2-0.4$. The power-law SAR for these assemblages of species is more robust to variations in clustering originating in the species' intrinsic dynamics or the effect of landscape heterogeneity.

In the case of the Gaussian distribution, a clustering coefficient ν can be defined (based on the ratio of the Gaussian distribution variance to its minimum value; see *Supporting Text*) and its variation allows us to quantitatively investigate the minimum level of clustering necessary to obtain a power-law SAR. For the same abundance distribution, a value of $\mu > 5$ destroys the power law (i.e., $r_1^2 < 0.99$; see Fig. 16, which is published as supporting information on the PNAS web site).

The considerations above may seem to be in contradiction with previous studies (43, 44), which indicate that the degree of conspecific aggregation has large effects on the shape of the SAR. However, there is no contradiction: Our claim is not that the degree of clustering is unimportant (it is not, as shown in Figs. 13 and 16) but that highly clustered distributions with the

same cover but different cluster structure (i.e., compact, Gaussian, or self-similar clusters) yield the same SAR.

Field Data. To check the validity of our approach with field data, we used Californian serpentine grassland data collected and analyzed by Green *et al.* (45) at the Donald and Sylvia McLaughlin Reserve at the University of California, Davis, Natural Reserve System. The particular site (referred to hereafter as the Blue Ridge site) was 64 m² in area and included 37,182 individuals and 24 plant species. The plot was divided into a square, 16 × 16 grid, and the sampling was performed in 1998 from early May to late July. These data were chosen because they are one of the best-characterized in the literature: The site was completely sampled, and a good fit to a power-law SAR was reported (see Fig. 17, which is published as supporting information on the PNAS web site).

By using the raw field data, we have generated the proximity functions and have verified that the results exhibit the same characteristics as for the computer-generated data. The functions $F^s(R)$, scaled by the mean and standard deviation, collapse to a single function; the standard deviation σ_s and the mean $\langle R \rangle^s$ are proportional up to finite-size effects; and both S_C and S_G yield power-law SARs with very similar exponents (Fig. 17; see also Fig. 18, which is published as supporting information on the PNAS web site). Our theory could have been falsified if the proximity functions did not scale in the necessary way, although a power-law SAR based on S_G has been well documented in this case.

The distribution of covers is not particularly close to a lognormal in this example (Fig. 19, which is published as supporting information on the PNAS web site), but still an excellent power law SAR is measured, showing the robustness of the power-law SAR to changes in the cover distribution. This example perfectly highlights the importance of large standard deviations σ_s to explain the robustness of the power-law SARs, as shown in Fig. 20, which is published as supporting information on the PNAS web site. Fig. 20 *Left* shows the actual distribution of the average minimum distances $\langle R \rangle^s$ against the distribution that should be present to obtain a power-law SAR with the

reported exponent, if all standard deviations σ_s were zero. Fig. 20 *Right* shows how the contribution of each species is spread over two standard deviations, therefore compensating for the inadequate distribution of $\langle R \rangle^s$.

Conclusion

We have shown that the widely reported scaling of number of species with area is a consequence of the combination of statistical geometry, extremal statistics, and the generically observed form of species–abundance distributions. The exponent z is determined by the shape of the abundance distribution.

Furthermore, we have shown that the appearance of power-law SARs is somewhat independent (robust) of the specific details of either the abundance distribution or the clustering mechanism, especially for exponents $z = 0.2$ – 0.4 , which may explain its predominance in gathered data (3, 14, 24). One should therefore expect to find power-law SARs whenever clustering and lognormal abundance distribution with extra rarity are observed. The inevitability of the form of the SAR under these general conditions underscores the need for more sensitive measures of individual and species-level correlations to characterize usefully complex ecological systems.

Methods

A detailed explanation for the calculations connecting the SAR and proximity functions (Eq. 1), the algorithms used to produce individual distributions for the different cluster structures, and the mathematical details for the solution of Eq. 5, which provides the abundance distributions compatible with power-law SARs, are provided in *Supporting Text* (see also Figs. 21 and 22, which are published as supporting information on the PNAS web site).

We thank John Harte, Jessica Green, Rachel Whitaker, and Guillermo Gea Izquierdo for valuable discussions and Jessica Green for making her raw data available (45). This work was supported in part by National Science Foundation Grant NSF-EAR-02-21743.

- May, R. (2000) in *Nature and Human Society: The Quest for a Sustainable World* (Natl. Acad. Sci., Washington, DC), pp. 30–45.
- Rosenzweig, M. L. (1999) *Science* **284**, 276–277.
- Rosenzweig, M. (1995) *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, U.K.).
- Huston, M. (1993) *Science* **262**, 1676–1680.
- Leps, J. (1990) in *Spatial Processes in Plant Communities*, eds. Krahulec, F., Agnew, A. D. Q., Agnew, S., & Willems, J. H. (Academia, Prague), pp. 1–11.
- Sugihara, G. (1980) *Am. Nat.* **116**, 770–787.
- Lawrey, J. (1991) *Bryologist* **94**, 377–382.
- Plotkin, J. B., Potts, M. D., Yu, D. W., Bunyavejchewin, S., Condit, R., Foster, R., Hubbell, S., LaFrankie, J., Manokaran, N., Lee, H.-S., *et al.* (2000) *Proc. Natl. Acad. Sci. USA* **97**, 10850–10854.
- Condit, R., Hubbell, S. P., LaFrankie, J. V., Sukumar, R., Manokaran, N., Foster, R. B. & Ashton, P. S. (1996) *J. Ecol.* **84**, 549–562.
- Tjorve, E. (2003) *J. Biogeogr.* **30**, 827–835.
- May, R. M. & Stumpf, M. P. H. (2000) *Science* **290**, 2084–2086.
- Horner-Devine, M. C., Lage, M., Hughes, J. B. & Bohannan, B. J. M. (2004) *Nature* **432**, 750–753.
- Green, J. L., Holmes, A. J., Westoby, M., Oliver, I., Briscoe, D., Dangerfield, M., Gillings, M. & Beattie, A. J. (2004) *Nature* **432**, 747–750.
- Williamson, M. (1988) in *Analytical Biogeography*, eds. Myers, A. A. & Giller, P. S. (Chapman & Hall, New York), pp. 91–115.
- Preston, F. (1948) *Ecology* **29**, 254–283.
- May, R. (1975) in *Ecology and Evolution of Communities* (Harvard Univ. Press, Cambridge, MA), pp. 81–120.
- Sizling, A. L. & Storch, D. (2004) *Ecol. Lett.* **7**, 60–68.
- Coleman, B. D. (1981) *Math. Biosci.* **54**, 191–215.
- Leitner, W. A. & Rosenzweig, M. L. (1997) *Oikos* **79**, 503–512.
- Picard, N., Karembé, M. & Birnbaum, P. (2004) *Ecoscience* **11**, 45–54.
- Plotkin, J. B., Potts, M. D., Leslie, N., Manokaran, N., LaFrankie, J. & Ashton, P. S. (2000) *J. Theor. Biol.* **207**, 81–99.
- Williams, M. R. (1995) *Ecology* **76**, 2607–2616.
- Ney-Nifle, M. & Mangel, M. (1999) *J. Theor. Biol.* **196**, 327–342.
- Hubbell, S. P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton).
- Durrett, R. & Levin, S. (1996) *J. Theor. Biol.* **179**, 119–127.
- Bastolla, U., Lassig, M., Manrubia, S. C. & Valleriani, A. (2001) *J. Theor. Biol.* **212**, 11–34.
- Harte, J., Kinzig, A. & Green, J. (1999) *Science* **284**, 334–336.
- Condit, R., Ashton, P. S., Baker, P., Bunyavejchewin, S., Gunatilleke, S., Gunatilleke, N., Hubbell, S. P., Foster, R. B., Itoh, A., LaFrankie, J. V., *et al.* (2000) *Science* **288**, 1414–1418.
- He, F., Legendre, P. & LaFrankie, J. (1997) *Science* **8**, 105–114.
- Tokeshi, M. (1993) *Adv. Ecol. Res.* **24**, 111–186.
- He, F. & Legendre, P. (2002) *Ecology* **83**, 1185–1198.
- Preston, F. (1962) *Ecology* **43**, 185–215.
- Castillo, E. (1988) *Extreme Value Theory in Engineering* (Academic, London).
- Gumbel, E. (1958) *Statistics of Extremes* (Columbia Univ. Press, New York).
- Fisher, R. & Tippett, L. (1928) *Proc. Cambridge Phil. Soc.* **24**, 181–191.
- Maddux, R. D. (2004) *Am. Nat.* **163**, 616–626.
- Kunin, W. E. (1998) *Science* **281**, 1513–1515.
- Cressie, N. A. C. (1993) *Statistics for Spatial Data* (Wiley, New York).
- Strang, G. (1988) *Linear Algebra and Its Applications* (Harcourt–Brace–Jovanovich, San Diego).
- Baas-Becking, L. G. M. (1934) *Geobiologie of Inleiding tot de Milieukunde* (van Stockum & Zoon, The Hague, The Netherlands).
- Lennon, J., Kunin, W. E. & Hartley, S. (2002) *Oikos* **97**, 378–386.
- Ostling, A., Harte, J., Green, J. L. & Kinzig, A. P. (2003) *Oikos* **103**, 218–224.
- He, F. & Gaston, K. J. (2000) *Am. Nat.* **156**, 553–559.
- Plotkin, J. B. & Muller-Landau, H. C. (2002) *Ecology* **83**, 3344–3356.
- Green, J., Harte, J. & Ostling, A. (2003) *Ecol. Lett.* **6**, 919–928.