Predicting species diversity in tropical forests

Joshua B. Plotkin\textsuperscript{a,b}, Matthew D. Potts\textsuperscript{c}, Douglas W. Yu\textsuperscript{d}, Sarayudh Bunyavejchewin\textsuperscript{e}, Richard Condit\textsuperscript{f}, Robin Foster\textsuperscript{g}, Stephen Hubbell\textsuperscript{i}, James LaFrankie\textsuperscript{j}, N. Manokaran\textsuperscript{k}, Hua-Seng Lee\textsuperscript{l}, Raman Sukumar\textsuperscript{m}, Martin A. Nowak\textsuperscript{a}, and Peter S. Ashton\textsuperscript{m}\textsuperscript{1}

\textsuperscript{a}Institute for Advanced Study and Princeton University, Princeton, NJ 08540; \textsuperscript{b}Division of Engineering and Applied Sciences, and \textsuperscript{c}Department of Organismal and Evolutionary Biology, Harvard University, Cambridge, MA 02138; \textsuperscript{d}Centre for Population Biology, Imperial College, Ascot, Berkshire SLS 7PY, United Kingdom; \textsuperscript{e}Silvicultural Research Division, Royal Forest Department, Chutuchak, Bangkok 10900, Thailand; \textsuperscript{f}Center for Tropical Forest Science, Smithsonian Tropical Research Institute, Balboa, Republic of Panama Box 2072; \textsuperscript{g}Botany Department, The Chicago Field Museum, Chicago, IL 60605; \textsuperscript{h}Botany Department, University of Georgia, Athens, GA 30602; \textsuperscript{i}Center for Tropical Forest Science, National Institute of Education, 1025 Singapore; \textsuperscript{j}Forest Research Institute of Malaysia, Kepong, Malaysia 52109; \textsuperscript{k}Sarawak Forestry Department, Kuching, Sarawak, Malaysia 9W 93750; and \textsuperscript{m}Centre for Ecological Science, Indian Institute of Science, Bangalore, India 560 012

Communicated by John W. Terborgh, Duke University, Durham, NC, July 26, 2000 (received for review April 15, 2000)

A fundamental question in ecology is how many species occur within a given area. Despite the complexity and diversity of different ecosystems, there exists a surprisingly simple, approximate answer: the number of species is proportional to the size of the area raised to some exponent. The exponent often turns out to be roughly 1/4. This power law can be derived from assumptions about the relative abundances of species or from notions of self-similarity. Here we analyze the largest existing data set of location-mapped species: over one million, individually identified trees from five tropical forests on three continents. Although the power law is a reasonable, zeroth-order approximation of our data, we find consistent deviations from it on all spatial scales. Furthermore, tropical forests are not self-similar at areas \( \leq 50 \) hectares. We develop an extended model of the species-area relationship, which enables us to predict large-scale species diversity from small-scale data samples more accurately than any other available method.

A primary motivation for modern ecological research is the effort to save as many species as possible from the sixth great mass extinction that currently threatens them (1, 2). How does habitat loss and destruction of tropical forests relate to species extinction? How many tree species must remain in an exploited forest if primate species are to survive in it? What is the best possible design of a natural reserve that maximizes the exploited forest if primate species are to survive in it? What is species extinction? How many tree species must remain in an area to prevent species extinction? The power law fails is often poorly recognized in the ecological literature (3, 18, 21). Previous research has uncovered the power law’s failure for small areas, but has downplayed its deviations for areas larger than 2 ha (22). Fig. 2b shows the dependence of the parameter \( a_i = S_i/S_{i-1} \) on area. The ratio \( a_i \) describes the average fraction of species that persist upon the ith bisection; therefore we call \( a_i \) the spatial persistence parameter. Self-similarity would require that this parameter be independent of area (18). As the figure

Abbreviations: SAR, species–area curve; ha, hectare; BCI, Barro Colorado Island; HKK, Huai Kha Khaeng.

\textsuperscript{1}To whom reprint requests should be addressed at: Institute For Advanced Study, Olden Lane, Princeton, NJ 08540. E-mail: plotkin@ias.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
shows, however, $a_i$ is not constant for any range of areas between 1 m² and 50 ha. Hence, tropical forests are conclusively not self-similar at these scales. The empirical form of the spatial persistence curve, and its departure from self-similarity, may result in part from aggregation of conspecifics—a possibility that we explore in detail elsewhere (29).

A Differential Equation Approach

Instead of self-similarity we find a consistent functional relationship between $a_i$ and the area, $A$, in all five forests (Fig. 2b). This observation is striking in light of the forests' disparate geographic locations, climates, and overall species diversities. We now introduce the spatial persistence function, $a(A)$, as a continuous extension of $a_i$. In the appendix, we use the persistence curve, and its departure from self-similarity, may result in part from aggregation of conspecifics—a possibility that we explore in detail elsewhere (29).

Using the diversity measured in a small area as the initial condition in Eq. 2, we may predict the diversity of a much larger area, if we know $a(A)$.

In the appendix, we derive Eq. 2 and find a general solution of the form $S = cA^e^{P(A)}$, where $P(A)$ is an infinite polynomial in $A$. We can truncate after the first $n$ terms to obtain an approximate solution. Truncating after the first term leads to the expression

$$S = cAe^{-kA}.$$  

Here $c$, $z$, and $k$ are constants determined by $a(A)$. This approximate solution is less accurate than the complete solution to Eq. 2, and it is only valid for a limited range of areas. Nevertheless, the approximation has the obvious advantage of simplicity. If we let $n \to \infty$, then we recover the full solution to Eq. 2; if we let $n \to 0$, then we reduce to the power law. Hence the power law is a zeroth-order, special case of our general model for the SAR.

Eq. 2 accurately predicts diversity given only a small amount of data. Because the persistence curves are similar across the five plots (Fig. 2b), we may use the canonical form of $a(A)$ fit at one plot to predict diversity for another plot. For example, using BCI's persistence curve to determine $a(A)$ and using the diversity in a single ha of Pasoh as the initial condition, we can predict the 50-ha diversity of Pasoh within 3% on average (Fig. 3). Conversely, Pasoh's persistence function predicts BCI's total diversity with 4% average error, and Lambir's diversity with 9% error, from a single ha of data.

Fig. 3 illustrates the extrapolative ability of Eq. 2 as compared with the classical models of the SAR. The precision of our method—namely, the ability to predict 50-ha diversity within 5% at Pasoh and BCI and 10% at Lambir—is an improvement over other previous methods. It is 1- to 7-fold more precise than Fisher's alpha (23), and 5- to 10-fold more precise than the power law. On the one hand, the increased precision of our method is not surprising; we have used two parameters to describe $a(A)$, as opposed to the classical models, which generally require one parameter. On the other hand, given the interplot similarity of the SAR, in practice we need only measure one parameter—the diversity of a single ha—to extrapolate diversity via Eq. 2.

Implications and Conclusions

We have analyzed the largest existing data set of location-mapped trees in tropical forests. We find that the SAR shows consistent deviations from the power law on all spatial scales that were studied, ranging from 1 m² to 50 ha. Hence, self-similarity does not hold over this range of areas. (There is the possibility that tropical forests are self-similar over scales greater than 50 ha, but in the absence of further data this remains speculative.)
These results might have some bearing on the longstanding controversy surrounding SARs. Previous research has focused on why the SAR has different slopes in different ecosystems (11), but in our extensive data the SAR does not possess a constant slope whatsoever.

Instead of self-similarity we propose a model of the SAR, based on the spatial persistence function, which generalizes the power law. This framework allows us to predict 50-ha diversity from small-scale samples with greater accuracy than ever before. Candidate logging protocols often are assayed at the 50- to 100-ha scale, and they are evaluated on the proportion of diversity that regenerates, as estimated from a small census (24). Hence, an accurate method to extrapolate 50-ha diversity from a small census will greatly benefit in the formulation of protocols for sustainable forestry and for biodiversity surveys (25). Furthermore, our methods may be extended to estimate landscape-scale diversity (see Appendix). These advances may induce ecologists to focus on the persistence curve itself as a unifying concept. The search for a biological mechanism that explains the observed persistence patterns offers an important challenge to ecology. In the meantime, our theory provides a valuable tool for conservation planning and a practical method for estimating diversity in the field.

Appendix

In this appendix we provide the details behind the derivation of Eq. 2, its solution, and its application to extrapolating diversity.

Derivation of Eq. 2. Given the definition of the spatial persistence parameter, \( a_i = S_i/S_{i-1} \), we start by deriving its relationship to the slope of the SAR. All logarithms are henceforth taken base two:

\[
-\log(a_i) = -\log\left(\frac{S_i}{S_{i-1}}\right) = \frac{\log(S_{i-1}) - \log(S_i)}{\log 2} = \frac{\log(A_i) - \log(A_{i-1})}{\log 2}.
\]

The last equality follows because \( A_{i-1}/A_i = 2 \). The final quantity in Eq. 4 measures the slope of a small chord on the log-log SAR to the right of \( \log(A_i) \). We conclude that

\[
-\log(a_i) = \frac{d \log S}{d \log A} \bigg|_{\log(A_i)},
\]

where \( i \) is now a continuous variable. To be more precise, the chord to the right of \( \log(A_i) \) is an estimate of the derivative at \( \log(A_i((i+1)/2)) \). Note that Eq. 5 clearly illustrates the equivalence between self-similarity (constant \( a \)) and the power law (constant \( \frac{d \log S}{d \log A} \)).

Eq. 2 follows easily from Eq. 5, using the fact that \( \frac{d \log S}{d \log A} = A/S \cdot dS/dA \). Note that Eq. 2 is a strict generalization of self-similarity: if \( a(A) \) is constant, then Eq. 2 reduces to the power law. If \( a(A) = \exp(-1/\log(A)) \), then Eq. 2 reduces to \( S(A) \propto \log A \). Hence Eq. 2 also generalizes the logarithmic law suggested by Gleason (26).

Diversity Extrapolation. Eq. 2 together with the interplot similarity of the persistence curves provide a method for extrapolating diversity over many spatial scales. For example, to extrapolate diversity from a subsample of Pasoh, we use BCI data to fit \( A(A) \), and then we solve Eq. 2 according to the small, initial condition measured at Pasoh. In effect, this process translates BCI’s log-log SAR so that it coincides with Pasoh’s initial condition; nevertheless, the universality between the forests is described more simply in terms of the persistence parameter, \( a_i \).

We choose to model the persistence values \( a_i \) with the simple, two-parameter family of curves

\[
\frac{1}{\alpha} \Phi\left(\frac{\alpha - i}{\beta}\right) + \frac{1}{4},
\]

(Other choices, such as a cubic model, are also possible and produce accurate predictions at these scales and beyond. See below.) Here \( \Phi(x) \) is the “error function” given by the cumulative distribution of the Gaussian: \( \Phi(x) = \left(2/\sqrt{\pi}\right) \int_0^x e^{-t^2} dt \). The parameter \( \alpha \) moves the inflection point of the persistence curve horizontally, and \( \beta \) determines the slope at the inflection point. Hence \( \beta \) measures the maximal “acceleration” of diversity with area, and \( \alpha \) measures the spatial scale at which acceleration is maximized. The best-fit at BCI is given by \( \alpha = 8.56, \beta = 8.08 \). For Pasoh, \( \alpha = 7.73, \beta = 7.41 \); for Mudumalai, \( \alpha = 7.06 \) and \( \beta = 7.76 \).

We may express all solutions to Eq. 2 in the form \( S = cA^\alpha \exp[P(A)] \), where \( P(A) \) is a polynomial in \( A \) of arbitrary degree \( n \), with no constant term. Once we specify \( \alpha \) and \( \beta \), we expand \(-\log(a(A))\) in a Taylor series of order \( n \) around the point \( A = 25 \) ha. The resulting separable equation can always be solved in closed form, yielding \( P \). For example, using Pasoh’s \( \alpha \) and \( \beta \) to determine \( a(A) \), the approximate solution of order \( n = 1 \) is
given by $S(A) = S(1\text{ha})A^z e^{kA}$, where $z = 0.125$ and $k = -5.66 \times 10^{-4}$. For Mudumalai, $z = 0.161$ and $k = -5.41 \times 10^{-4}$. This approximation is only valid for $A \leq 50$ ha, but its accuracy compares well with the complete numerical solution: it predicts 50-ha diversity with average error 4\% at Pasoh, 9\% at BCI, and 16\% at Lambir.

In practice, a numerical solution of Eq. 2 yields the most accurate SAR. We used the Fehlberg-Runge-Kutta method to generate the prediction shown in Fig. 3. The initial condition $S(1\text{ha})$ for Fig. 3 was determined by the diversity of a single, random, 1\text{ha} subplot of Pasoh. The confidence interval was constructed from 1,000, independently sampled, 1\text{ha} initial conditions.

We have divided our five plots into two categories: those that suffer regular disturbance and those that do not. The two tropical forests subject to regular fires, HKK and Mudumalai, generally should not be modeled via the persistence curve from a more stable, moist tropical forest. The values of $a_3$ to $a_6$ are generally smaller at HKK and Mudumalai than the other three forests. This reflects the fact that HKK and Mudumalai are subject to more disturbances, causing greater patchiness. The persistence curve at Mudumalai can predict HKK and conversely within 17\%, given 1 ha of data. Compared with BCI, Pasoh, and Lambir, the accuracy of Eq. 2 has been decreased by the disturbances at Mudumalai and HKK. Nevertheless, 17\% error is still preferable to the 28\% error-rate or worse given by Fisher’s alpha or a power law on these disturbed forests. In practice, when estimating diversity in a new forest, the ecologist should first determine the frequency of disturbances (e.g., fires, hurricanes, or roaming elephants) and choose a model forest, where $\alpha$ and $\beta$ are known, accordingly.

Extrapolation Beyond 50 ha. For 50-ha predictions, such as would be useful to assess logging protocols, $\Phi$ provides a simple, two-parameter model of the persistence curve. For larger areas, however, a cubic model (which works as well as $\Phi$ at 50 ha) is often more effective. For example, we can use a cubic persistence curve calibrated at Pasoh to extrapolate the diversity of the entire BCI, which occupies 1,500 ha, from a 1-ha sample. Using Eq. 2, the predicted diversity for all of BCI is $436 \pm 32$ species (1 SD). This estimate compares favorably with Croat’s floral count of 450 tree and shrub species on the island (27). For even larger areas, the persistence curve should be parameterized by using multiple, small censuses spread across the landscape (as in ref. 28), although such techniques require further development.

We sincerely thank Alice Chen, Jay Rosenheim, Stuart Davies, Egbert Leigh, Bill Bossert, Burt Singer, Steve Pacala, Jerome Chave, and Simon Levin for many helpful discussions. We are especially grateful to Helene Muller-Landau, Bob Holt, and Robert May for their conceptual and editorial input. J.B.P. is supported by a National Science Foundation graduate fellowship.
Corrections

BIOCHEMISTRY. For the article “Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation” by Ugo Mayor, Christopher M. Johnson, Valerie Daggett, and Alan R. Fersht, which appeared in number 25, December 5, 2000, of Proc. Natl. Acad. Sci. USA (97, 13518–13522), the authors note the following correction. In Results, in the last line of the last paragraph under the subheading “Kinetic Measurements,” \( \beta_{44} \) should be \( \beta_T \).

Correction published online before print: Proc. Natl. Acad. Sci. USA, 10.1073/pnas.031567698. Text and publication date are at www.pnas.org/cgi/doi/10.1073/pnas.031567698

BIOPHYSICS. For the article “Simultaneous two-photon excitation of distinct labels for dual-color fluorescence crosscorrelation analysis” by Katrin G. Heinze, Andre Koltermann, and Petra Schwille, which appeared in number 19, September 12, 2000, of Proc. Natl. Acad. Sci. USA (97, 10377–10382), the authors note the following correction. In Fig. 3 (wavelength dependence of emission), the legends for the dyes Texas red and Rhodamine green were accidentally permuted. The essence of the article is not altered.

ECOLOGY. For the article “Predicting species diversity in tropical forests” by Joshua B. Plotkin, Matthew D. Potts, Douglas W. Yu, Sarayuth Bunyavejchewin, Richard Condit, Robin Foster, Stephen Hubbell, James LaFrankie, N. Manokaran, Lee Hua Seng, Raman Sukumar, Martin A. Nowak, and Peter S. Ashton, which appeared in number 20, September 26, 2000, of Proc. Natl. Acad. Sci. USA (97, 10850–10854), the authors note the following correction. The name of the 10th author is printed incorrectly. The correct name is Hua-Seng Lee.

NEUROBIOLOGY. For the article “Neural stem cells display extensive tropism for pathology in adult brain: Evidence from intracranial gliomas” by Karen S. Aboody, Alice Brown, Nikolai G. Rainov, Kate A. Bower, Shaoxiong Liu, Wendy Yang, Juan E. Small, Ulrich Herrlinger, Vaclav Ourednik, Peter McL. Black, Xandra O. Breakefield, and Evan Y. Snyder, which appeared in number 23, November 7, 2000, of Proc. Natl. Acad. Sci. USA (97, 12846–12851), the authors note the following correction. The first two authors, Karen S. Aboody and Alice Brown, contributed equally to this work.

NEUROBIOLOGY. For the article “The neural representation of postural control in humans” by Hans-Otto Karnath, Susanne Ferber, and Johannes Dichgans, which appeared in number 25, December 5, 2000, of Proc. Natl. Acad. Sci. USA (97, 13931–13936), the authors note the following correction. In lines 6 and 16 of the third paragraph on page 13936, the word “zentrolateralis” should read “centrolateralis.” Moreover, a colon is missing in the legend of Table 2 on page 13935; the phrase should read “In the human nomenclature of Hassler (27): . . . .”

Correction published online before print: Proc. Natl. Acad. Sci. USA, 10.1073/pnas.031567398. Text and publication date are at www.pnas.org/cgi/doi/10.1073/pnas.031567398