Title

A theory of spatial structure in ecological communities at multiple spatial scales

Permalink

https://escholarship.org/uc/item/0v7177zt

Journal

Ecological Monographs, 75(2)

ISSN

0012-9615

Authors

Harte, J
Conlisk, E
Ostling, A
et al.

Publication Date

2005-05-01

Peer reviewed

Ecological Monographs, 75(2), 2005, pp. 179–197
© 2005 by the Ecological Society of America

A THEORY OF SPATIAL STRUCTURE IN ECOLOGICAL COMMUNITIES AT MULTIPLE SPATIAL SCALES

JOHN HARTE,[1,3] ERIN CONLISK,[1] ANNETTE OSTLING,[1] JESSICA L. GREEN,[2] AND ADAM B. SMITH[1]

[1]*Energy and Resources Group, University of California, Berkeley, California 94720 USA*
[2]*School of Natural Sciences, University of California, Merced, California 95344 USA*

Abstract. A theory of spatial structure in ecological communities is presented and tested. At the core of the theory is a simple allocation rule for the assembly of species in space. The theory leads, with no adjustable parameters, to nonrandom statistical predictions for the spatial distribution of species at multiple spatial scales. The distributions are such that the abundance of a species at the largest measured scale uniquely determines the spatial-abundance distribution of the individuals of that species at smaller spatial scales. The shape of the species–area relationship, the endemics–area relationship, a scale-dependent community-level spatial-abundance distribution, the species-abundance distribution at small spatial scales, an index of intraspecific aggregation, the range–area relationship, and the dependence of species turnover on interpatch distance and on patch size are also uniquely predicted as a function solely of the list of abundances of the species at the largest spatial scale. We show that the spatial structure of three spatially explicit vegetation census data sets (i.e., a 64-m^2 serpentine grassland plot, a 50-ha moist tropical forest plot, and a 9.68-ha dry tropical forest plot) are generally consistent with the predictions of the theory, despite the very simple statistical assumption upon which the theory is based, and the absence of adjustable parameters. However, deviations between predicted and observed distributions do arise for the species with the highest abundances; the pattern of those deviations indicates that the theory, which currently contains no explicit description of interaction mechanisms among individuals within species, could be improved with the incorporation of intraspecific density dependence.

Key words: *abundance distribution; Barro Colorado Island (BCI), Panama; endemics–area relationship; range–area relationship; San Emilio, Costa Rica; scale dependence; serpentine grassland; spatial distribution; species–area relationship; tropical forest.*

INTRODUCTION

Understanding the abundance and spatial distribution of species at multiple spatial scales is a central concern of ecology (Fisher et al. 1943, Preston 1948, Krebs 1994, Rosenzweig 1995, Gaston and Blackburn 2000, He and Legendre 2002). Patterns in the distributions of individuals and species across space provide information critical to our ability to decipher the forces that structure and maintain ecological diversity (Pielou 1969, Brown et al. 1995). Within biomes, knowledge of the scale-dependent frequency of patch occupancy can lead to improved estimates of extinction rates under perturbations, more effective land protection policies, the design of more efficient and accurate censusing strategies, and improved estimates of species richness from sparse census data (Whitmore and Sayer 1992, May et al. 1995, Rosenzweig 1995).

Spatial models that explicitly assume knowledge of processes such as birth, death, dispersal, speciation, migration, extinction, and niche differentiation have significantly advanced our understanding of patterns in the abundance and distribution of species across space and time (see reviews by Hubbell [2001], Chave [2004], and Leibold et al. [2004]). Statistically based models, while not derived from explicit biological mechanisms, also have provided a tractable theoretical framework for quantifying the spatial structure of ecological communities. For example, in an early investigation of the distribution and abundance of species, Preston (1962) and May (1975) argued that a particular species-abundance distribution (SAD), the canonical lognormal, was related to a power-law form of the species–area relationship (SAR). Their approach was based on randomly sampling individuals from an ecological community with a lognormal SAD. Coleman (1981) extended the reach of the random model, developing a general framework for deriving the shape of the species–area relationship under arbitrary species-abundance distributions, while Green and Ostling (2003) derived the form of an endemics–area relationship under the random model. Although the random distribution assumption does provide a general model of spatial pattern in ecology, numerous studies have shown that this assumption inadequately describes spatial patterns of aggregation of individuals within

Manuscript received 7 September 2004; revised 3 November 2004; accepted 4 November 2004. Corresponding Editor: W. S. C. Gurney.

[3] E-mail: jharte@socrates.berkeley.edu

species and of species across landscapes (Condit et al. 1996*b*, 2002, Plotkin et al. 2000, Green et al. 2003).

To model aggregated species distributions, and to investigate the effect of aggregation on macroecological properties such as the SAR or species turnover, a number of authors (Wright 1991, He and Gaston 2000, 2003) have examined the following statistical model:

$$p = 1 - [(k + \mu)/k]^{-k} \qquad k = \mu^2/(\sigma^2 - \mu) \quad (1)$$

where p is the probability of a species occurrence in a grid cell as a function of the mean occupancy μ, the variance σ^2, and a parameter k, that can take on values $k < -\mu$ or $k > 0$ and that can be estimated from census data (Bliss and Fisher 1953). In general, k will be species and scale dependent (Plotkin and Muller-Landau 2002). For positive k values, the model derives from the negative binomial distribution (NBD), which has been used in the following ways: to describe SARs and abundance–aggregation patterns (He and Gaston 2000, 2003, He and Hubbell 2003), to derive a formula for the fraction of species in common to two separated cells of a specified area (Plotkin and Muller-Landau 2002), and to derive an ‘‘endemics–area relationship’’ (EAR) between the number of species in a cell that are unique to that cell and the area of the cell (Green and Ostling 2003). In applications of the NBD, values of the parameter k at every scale and for every species are determined from data, rather than from first principles, and thus the model contains a sizeable number of adjustable parameters when applied to a community of species. Plotkin et al. (2000) assumed and explored another spatial model based on the Poisson cluster distribution; with even more adjustable parameters than the NBD, this model can generate a wide variety of patterns that resemble those observed in nature.

Another statistically based approach to describing species-level and community-level spatial patterns across spatial scales has focused on the explicit scaling properties of the SAR and, for individual species, the range–area relationship (RAR). The RAR describes the dependence of the range size of a species (defined by a box-counting procedure as the total area of all occupied grid cells at a given scale) on cell area (Kunin 1998, Gaston and Blackburn 2000, Harte et al. 2001). The scaling approach starts with the observation that the shape of the SAR and the RAR can be expressed in terms of certain fundamental probability parameters. At the species level, these probabilities describe how the number of occupied census cells depends on the spatial scale of the cells (Harte et al. 2001), while at community level, the probabilities describe how species richness depends on spatial scale (Harte et al. 1999, 2001). The assumption of self-similarity, or fractality, is equivalent to the assumption that the probability parameters are independent of scale (Ostling et al. 2003). When the community-level probabilities are scale independent, a power-law SAR results; when the parameters for a particular species are scale independent, a power-law RAR results for that species. Although the theory presented here is substantially different from this previous work, and we make no a priori assumptions about the scaling properties of these probability parameters, it is convenient to express the theory with those same parameters, and so formal definitions are given (see *Theoretical background; The probability parameters*).

The statistical theory presented here departs in varying ways from previous approaches. In one respect, our approach is similar to that of random placement models in that all the macroecological predictions follow from a single statistical assumption, along with knowledge of the total abundances of the species (Coleman 1981). In particular, our results derive from what we denote as the ‘‘hypothesis of equal allocation probabilities’’ (HEAP). This fundamental statistical assumption is akin to, but significantly different in detail from, the ‘‘hypothesis of equal a priori probabilities’’ that underpins statistical mechanics (Ruhla 1992). The latter, for classical molecules, results in a binomial abundance distribution across space, as does the random placement model in ecology. Instead, HEAP results in a nonrandom, more aggregated distribution of individuals.

In contrast to models based on the negative binomial or the Poisson cluster model, which contain adjustable parameters for each species, we introduce no adjustable parameters nor, in contrast to McGill and Collins (2003), do we require as input to our theory any empirical information about the shape of species-abundance distributions across the ranges of the species. However, our work shares with those approaches the objective of rigorously deriving testable predictions for the relationships among different macroecological patterns, such as the relationship between species-level aggregation or RAR curves and community properties such as the SAR or species turnover in space.

In contrast to our previous work on fractal scaling properties of species distributions, we impose here a top-down boundary condition on the recursion relations that generate the probability distribution for each species; the boundary condition at the ‘‘top’’ is simply the total abundance of a species in some large area. Whereas the bottom-up approach requires prior knowledge of the species-level probability parameters describing the RAR at all spatial scales being considered, our top-down approach uniquely predicts the values of all those parameters, along with the community-level probability parameters describing the SAR at all scales as a function solely of the list of total abundances of the species (the SAD) within the large area.

Assuming a SAD at some largest scale, our theory predicts scale-dependent probability distributions describing a wide range of ecological patterns at both species and community levels at multiple spatial scales. For each species, it predicts the distribution of patch occupancy frequencies across multiple spatial scales, the RAR, an ''O-ring'' index of spatial clustering (Condit et al. 2000), and the scaling behavior of the mean–variance relationship for the species-level spatial distributions. At community level, it predicts the scale-dependent patch occupancy distribution for total abundances summed over all species, as well as the SAR, the EAR, the SAD at all smaller spatial scales, and species turnover as a function of census patch area and interpatch distance. Knowledge of the SAD at some largest spatial scale is all that is needed by the theory to uniquely generate all these spatial properties of ecosystems at smaller scales.

Although power-law SARs are not assumed here, and in fact under HEAP can never hold over all scales, HEAP can predict the SADs for which a power-law SAR will arise over a limited scale range. On the other hand, the theory predicts that the species-level probability parameters can never be scale independent, and thus the RARs cannot have exact power-law behavior, over any scale range. More specifically, it predicts that on log–log plots, the RAR for each species will always exhibit negative curvature, the strength of which at any scale depends on the abundance of the species.

Our approach here is to create a theory of spatial structure in macroecology that can be tested against a wide array of empirical spatial patterns in ecosystems. Because the theory contains no adjustable parameters and no explicit biological mechanisms such as density dependence, dispersal, birth, death, and interspecific interactions, we expect the theory to perform poorly under at least some circumstances. From knowledge of the patterns of success and failure when we compare the theoretical predictions with data from three sites for which spatially explicit plant census data are available, we show that we can gain insight into which of the many neglected biological mechanisms are likely to most significantly influence spatial patterns in ecology. Because our fundamental statistical assumption leads to predictions that greatly outperform random placement predictions, we argue that HEAP is a useful ''null assumption'' for macroecology. Although we currently lack a mechanistic understanding of the origin of HEAP, we suggest that an alternative formulation of the theory in terms of a spatially explicit ''assembly function'' may provide the basis for such an understanding.

Theoretical Background

Types of abundance distributions

Consider a plot or landscape of area A_0 populated by S_0 species, each containing a total of $n_0^{(j)}$ individuals, where (j) is a species label that ranges from $j = 1 \ldots S_0$. To introduce a scale parameter, let A_0 be repeatedly bisected into similar-shaped patches. Then we can define a scale parameter i such that patches of area A_i are formed from the ith bisection (Harte et al. 1999). Note that larger spatial scales correspond to smaller values of i. A formal procedure for this bisection process exists (Harte et al. 1999, Plotkin et al. 2000, Ostling et al. 2003) and a consistent procedure for avoiding potentially unrealistic, artifactual consequences has been described (Ostling et al. 2004).

We distinguish here two related types of, potentially scale-dependent abundance distributions: the species-abundance distribution and spatial-abundance distributions. We denote the former by $\Phi_i(n)$; it is the distribution of abundances across all the species in a community at scale i. The meaning of $\Phi_i(n)$ is based on the idea of a species occurrence or nonoccurrence at scale i. At scale $i = 0$, all S_0 species occur and $\Phi_0(n)$ is the customarily defined species-abundance distribution (Fisher et al. 1943, Preston 1948, Gaston and Blackburn 2000): the fraction of all the S_0 species in A_0 with n individuals. At finer scales, $\Phi_i(n)$ is a straightforward generalization of this. For example, consider the two A_1 cells formed by the first bisection. There are $2S_0$ species occurrences or nonoccurrences in those two cells. Suppose there are ϕ_3 occurrences in which a species has 3 individuals in either one of the A_1 cells; then $\Phi_1(3) = \phi_3/(2S_0)$. Note that $\Phi_1(0)$ will be the fraction of nonoccurrences in the A_i cells and is thus nonzero if at least one of the species is absent from one of the two A_1 cells. In general, $\Phi_i(n)$ is the probability that an A_i cell has a species occurrence with n individuals (or a species nonoccurrence if $n = 0$); equivalently it is estimated by the fraction of all species occurrences at scale i with n individuals (or nonoccurrences for $n = 0$). Fig. 1 illustrates how $\Phi_i(n)$ and the other distributions are calculated from a data set.

Spatial-abundance distributions can be defined at the species level and the community level. Anticipating that each species-level spatial distribution in our theory will be entirely determined at all scales by the abundance of that species, n_0, we hereafter use the species label (n_0) to label the distributions. Thus we define a spatial distribution, $P_i^{(n_0)}(n)$, to be the probability that a census cell A_i contains n individuals of a species with abundance n_0 in A_0, where n can be any integer from 0 to n_0. Equivalently, $P_i^{(n_0)}(n)$ is estimated by the fraction of all the $2^i A_i$ cells that contain n individuals. We note that

$$P_0^{(n_0)}(n) = \delta(n, n_0) \qquad (2)$$

where $\delta(n, n_0)$ is the Kronecker delta function, equal to 1 if $n = n_0$ and 0 otherwise. This is our top-down boundary condition.

At the community level, a spatial-abundance distribution $F_i(N)$ is defined here to be the probability that an A_i cell contains a total of N individuals of all species combined; equivalently, it is estimated by the fraction of all the A_i cells that contain a total of N individuals. The calculations of Φ_i, P_i, and F_i from census data are illustrated in Fig. 1.

AAAA BB C D	A B C E
A D	AAA BBB

FIG. 1. Illustration of how Φ, P, and F are computed from a simple ''data'' set in which $S_0 = 5$. The five species (A–E) have total abundances of 9, 6, 2, 2, and 1 individuals, respectively. Hence the species-abundance distribution at scale $i = 0$ is $\Phi_0(9) = 1/5$, $\Phi_0(6) = 1/5$, $\Phi_0(2) = 2/5$, and $\Phi_0(1) = 1/5$. At scale $i = 2$ (since the data are resolved to quadrants) there are potentially $2^2S_0 = 20$ species occurrences or nonoccurrences in the four quadrants. In fact, there are eight nonoccurrences (0 individuals), eight occurrences with 1 individual, one with 2 individuals, two with 3 individuals, and one with 4 individuals. Hence $\Phi_2(0) = 8/20$, $\Phi_2(1) = 8/20$, $\Phi_2(2) = 1/20$, $\Phi_2(3) = 2/20$, $\Phi_2(4) = 1/20$. Consider next the species-level spatial distribution function for species A at scale $i = 2$ (since the data are resolved to quadrants). In the four quadrants, there are 4, 3, 1, and 1 individuals; $n_0 = 9$ for species A. Hence $P_2^{(9)}(4) = 1/4$, $P_2^{(9)}(3) = 1/4$, and $P_2^{(9)}(1) = 1/2$. The total numbers of individuals in the community that lie in the four quadrants are 8, 6, 4, and 2, so the community-level spatial distribution is given by $F_2(8) = F_2(6) = F_2(4) = F_2(2) = 1/4$.

The probability parameters

We define a community-level parameter, a_i, as follows. Consider a cell randomly selected from the set of A_{i-1} cells, and a species occurrence (at $i - 1$ scale) randomly chosen from the set of species occurrences in that cell. We define a_i to be the probability that this species occurrence is also present in at least a prespecified (say, the left-hand) one of the two cells of area A_i that comprise the selected A_{i-1} cell. Equivalently, we have shown (Ostling et al. 2003) that a_i can be reexpressed as

$$a_i = S_i/S_{i-1} \tag{3}$$

where S_i is the mean number of species in the A_i cells. Eq. 3 implies that

$$S_i = S_0 \prod_{j=1}^{i} a_j. \tag{4}$$

If a_i is scale independent, so that $a_i \equiv a$ for all i, then the power-law form of the SAR,

$$S_i = cA_i^z \tag{5}$$

follows, with $z = -\log_2(a)$ (Harte et al. 1999, Ostling et al. 2003).

In parallel with this, at the species level we can define a set of species-level parameters, $\alpha_i^{(n_0)}$, where i is a scale label, and we have anticipated that the α_i could depend on n_0 and spatial scale. Each $\alpha_i^{(n_0)}$ is the conditional probability that if a species with abundance n_0 in A_0 is found in an A_{i-1} cell, then it is present in at least a prespecified one of the two A_i cells that comprise the A_{i-1} cell. Equivalently, the α's can be reexpressed as

$$\alpha_i^{(n_0)} = R_i^{(n_0)}/R_{i-1}^{(n_0)} \tag{6}$$

where $R_i^{(n_0)}$, the range size of the species at scale i, is equal to $A_iW_i^{(n_0)}$, with $W_i^{(n_0)}$ equal to the number of grid cells of area A_i in which the species is found (Harte et al. 2001).

Recalling the conditional nature of the definition of the probability $\alpha_i^{(n_0)}$, and the fact that $1 - P_i^{(n_0)}(0)$ is the probability that a specified cell A_i is occupied by that species, then it is straightforward to show that

$$\alpha_i^{(n_0)} = [1 - P_i^{(n_0)}(0)]/[1 - P_{i-1}^{(n_0)}(0)]. \tag{7}$$

If an α for a particular species is scale independent, then in analogy with Eq. 5 another kind of power-law relation, the range–area relationship, is obtained. In particular, for each species with scale-independent $\alpha^{(n_0)}$,

$$R_i^{(n_0)} = c'A_i^{y'(n_0)} \tag{8}$$

where, in analogy with the relationship between z and a, $y'(n_0)$ for each species is related to the value of α for that species by $y'(n_0) = -\log_2\alpha^{(n_0)}$ (Harte et al. 2001).

Because presence/absence data are easier to obtain than complete abundance counts, Eq. 7 provides a potential means of estimating the total abundance of a species in some large area from more accessible data (Kunin 1998). In particular, if for each species a theoretical relationship exists between n_0 for that species and the value of α_i for that species, then estimation of the right-hand side of Eq. 7 (from presence/absence data at two successive scales) allows estimation of n_0. We shall see that HEAP does indeed lead to a unique value of n_0 for each specified value of α_i and thus allows us to predict the value of n_0 from the measured value of the right-hand side of Eq. 7 at arbitrary value of the scale parameter i.

We also note the following exact relationship (Harte et al. 2001) among the set of probabilities, $\alpha_i^{(n_0)}$, and the a_i

$$\prod_{j=1}^{i} a_j = \left\langle \prod_{j=1}^{i} \alpha_j^{(n_0)} \right\rangle_{\text{species}} \tag{9}$$

where $\langle \cdot \rangle_{\text{species}}$ refers to the average over all species. Eq. 9 implies the following important result: if the a_i are independent of scale over any scale interval (k, $k + 1$), then either α_k does not equal α_{k+1} for at least one of the species (that is, α is not scale independent for that species over that scale interval), or all the $\alpha_i^{(n_0)}$ are equal to each other for both $k = i$ and $i + 1$. In other words, unless the α's are the same for all species, both the a's and all the α's cannot be simultaneously scale independent. The implication of this is that Eqs. 5 and 8 cannot hold simultaneously; either the SAR, or the RAR for at least one species, must be scale dependent over every scale interval, unless all species have equal α_i values. Empirical evidence suggests that there is generally scale dependence in the RARs for most species, but that over at least some limited scale range the SAR sometimes deviates insignificantly from scale independence (Kunin et al. 2000, Green et al. 2003).

Using Eq. 9, we can rewrite Eq. 4 in a form that will be more convenient later:

$$S_i = \langle \lambda_i^{(n_0)} \rangle_{\text{species}} S_0 \tag{10}$$

where

$$\lambda_i^{(n_0)} = \prod_{k=1}^{i} \alpha_k^{(n_0)} = 1 - P_i^{(n_0)}(0). \tag{11}$$

The RARs can also be expressed in terms of the $\lambda_i^{(n_0)}$. Using Eq. 6, and the fact that $W_0 = 1$, we obtain, in analogy with Eq. 10,

$$R_i^{(n_0)} = \lambda_i^{(n_0)} A_0. \tag{12}$$

The HEAP assumption and the fundamental recursion relationship

To proceed we make an assumption that will uniquely determine the functional dependence of the $P_i^{(n_0)}(n)$ on n_0, n, and i. We call it the ''hypothesis of equal allocation probabilities'' (HEAP). To illustrate HEAP, consider a species with $n_0 = 3$. What are the probabilities that govern how those three individuals in A_0 are distributed among the two A_1 cells that comprise it? Under HEAP, we assume that the four options (0, 3), (1, 2), (2, 1), and (3, 0) are equally likely. The implication of this is clearly that

$$P_1^{(3)}(0) = P_1^{(3)}(1) = P_1^{(3)}(2) = P_1^{(3)}(3) = 1/4. \tag{13}$$

Hence, from Eqs. 2, 7, and 13,

$$\alpha_1^{(3)} = 1 - P_1^{(3)}(0) = 3/4. \tag{14}$$

In general, for any value of n_0, and for all n, HEAP implies

$$P_1^{(n_0)}(n) = (n_0 + 1)^{-1} \tag{15}$$

$$\alpha_1^{(n_0)} = n_0(n_0 + 1)^{-1}. \tag{16}$$

We assume that HEAP holds at smaller scales, as well. Thus, for example, suppose it is known that of the three individuals from a particular species in A_0, there are two individuals of that species in a particular A_1 cell. Then, regardless of the value of n_0, the following distributions of those two individuals are equally likely in the two A_2 cells that comprise the A_1 cell: (0, 2), (1, 1), (2, 0). So now, by multiplying conditional probabilities, we can calculate $\alpha_2^{(3)}$. From Eq. 14, the probability that the left-hand A_1 cell is unoccupied is 1/4, and if it is unoccupied, then the probability that the A_2 cell that constitutes the upper half of that cell has no individuals is 1. The probability that the left-hand A_1 cell has one individual in it is 1/4 and if so, the probability is 1/2 that there are no individuals in the upper A_2 cell. The probability that the left-hand A_1 cell has two individuals in it is 1/4 and if so, the probability is 1/3 that there are no individuals in the upper A_2 cell. The probability that the left-hand A_1 cell has three individuals in it is 1/4 and if so, the probability is 1/4 that there are no individuals in the upper A_2 cell. Hence the probability that A_2 is unpopulated is given by

$$P_2^{(3)}(0) = (1/4)(1) + (1/4)(1/2) + (1/4)(1/3) + (1/4)(1/4) = 25/48. \tag{17}$$

It now follows from Eqs. 7 and 13 that

$$\alpha_2^{(3)} = (1 - 25/48)/(1 - 1/4) = 23/36. \tag{18}$$

Note that $\alpha_2^{(3)} \neq \alpha_1^{(3)} = 27/36$. We shall see that in general, $\alpha_i^{(n_0)}$ is a decreasing function of i and thus the theory predicts a specific scale dependence of the $\alpha_i^{(n_0)}$.

Consider next the calculation of $P_2^{(3)}(1)$ under HEAP. The probability that there is 1 individual in the left-hand A_1 cell is 1/4 and if that is the case then the probability that the upper A_2 cell has one individual is 1/2. The probability that there are two individuals in the A_1 cell is 1/4 and if so the probability that there will be just one individual in the A_2 cell is 1/3. The probability that there are three individuals in the A_1 cell is 1/4 and if so the probability that there will be just one individual in the A_2 cell is 1/4. If the A_1 cell contains no individuals, then the probability that the A_2 cell contains one individual is 0. Hence,

$$P_2^{(3)}(1) = (1/4)(1/2) + (1/4)(1/3) + (1/4)(1/4) = 13/48. \tag{19}$$

The same reasoning leads to

$$P_2^{(3)}(2) = (1/4)(1/3) + (1/4)(1/4) = 7/48 \tag{20}$$

$$P_2^{(3)}(3) = (1/4)(1/4) = 3/48. \tag{21}$$

The result of these calculations can be summarized by writing

$$P_2^{(3)}(n) = \sum_{q=n}^{3} \frac{P_1^{(3)}(q)}{(q + 1)}. \tag{22}$$

Eq. 22 readily generalizes to all values of n_0 and i. Thus, more generally, HEAP results in the following recursion relationship for the species-level spatial-abundance distributions:

$$P_i^{(n_0)}(n) = \sum_{q=n}^{n_0} \frac{P_{i-1}^{(n_0)}(q)}{(q + 1)}. \tag{23}$$

At the species level, Eq. 23 is the fundamental result of our theory; for a species with n_0 individuals in A_0, it yields the probability distribution, over grid cells at smaller scales, of the numbers of individuals per cell.

The solutions to Eq. 23 at all scales are entirely determined by the value of n_0 and the boundary condition, Eq. 2.

An alternative derivation of the fundamental recursion relationship

Eq. 23 can also be derived starting from an alternative assumption to HEAP. To see this, we introduce an assembly (or colonization) function, $\beta_i^{(n_0)}(1\,|\,p, q - p)$, that can be used to describe the sequential assembly of the n_0 individuals of a species onto A_0. Suppose that q individuals of the species have been allocated to the two A_i cells that make up an A_{i-1} cell. Of those q allocated individuals, p are known to be in the left-hand A_i cell and $q - p$ are in the right-hand one. Then we define

$$\begin{aligned}&\beta_i^{(n_0)}(1\,|\,p, q - p)\\ &\quad = \text{the conditional probability that the } q + 1\text{st}\\ &\qquad \text{individual is on the left.}\end{aligned} \tag{24}$$

$\beta_i^{(n_0)}(0\,|\,p, q - p)$ is analogous, except that it is the probability that the $q + 1$st individual is on the right. The functional form of the $\beta_i^{(n_0)}(1\,|\,p, q - p)$ could depend on the abundance of the species in A_0 (i.e., n_0, but we will assume it does not and hereafter leave the superscript off). From the definition of β, the following constraints hold:

$$\beta_i(1\,|\,p, q - p) = 1 - \beta_i(0\,|\,p, q - p) \tag{25a}$$

$$\beta_i(1\,|\,p, q - p) = \beta_i(0\,|\,q - p, p) \tag{25b}$$

$$\beta_i(1\,|\,p, p) = \beta_i(0\,|\,p, p) = 1/2. \tag{25c}$$

The functions $P_i^{(n_0)}(n)$ can be expressed in terms of the $\beta_i(1\,|\,p, q - p)$ functions and, as shown in Appendix A, the following form for β then results in Eq. 23:

$$\beta_i(1\,|\,p, q - p) = (p + 1)/(q + 2) \tag{26}$$

for all scales, i. Note that this function satisfies all the constraints in Eqs. 25a–25c and that it implies the scale independence of the β's. Eq. 26 also has implications for the degree of clustering in the distributions of individuals within a species. Recalling that p is the number of individuals on the left, $q - p$ is the number on the right, and $\beta_i(1\,|\,p, q - p)$ is the probability that the $p + 1$st individual is allocated to the left, Eq. 26 implies that ''the richly populated half cell gets more richly populated''; in other words, it produces a level of aggregation greater than expected under a random distribution.

Embedding HEAP within an infinite family of distributions

Every assembly function $\beta_i(1\,|\,p, q - p)$ that obeys Eqs. 25a–25c will produce a set of spatial probability distributions $P_i^{(n_0)}(n)$. The particular choice of Eq. 26 produces species-level spatial-abundance distributions that are identical to those resulting from the assumption of HEAP. But other forms for β will yield other spatial distributions. A very broad class of scale-independent assembly functions is of the following form:

$$\beta_i(1\,|\,p, q - p) = (p\theta + 1)/(q\theta + 2). \tag{27}$$

The parameter θ in Eq. 27 is an aggregation index. The choice $\theta = 0$ yields the random placement model; the choice $\theta = 1$ yields the HEAP case of Eq. 26; and the choice $\theta = \infty$ yields the maximally aggregated case, in which at every scale all individuals are confined to one A_i cell. Distributions that are more uniform than under random placement (that is, repulsive distributions) are obtained for certain negative values of θ. In particular, if n_0 is even, then any integer value of θ^{-1} that satisfies $-n_0/2 \geq \theta^{-1} \geq -n_0$ or any value of $\theta^{-1} < -n_0$ yields a repulsive distribution and $\theta^{-1} = -n_0/2$ is the limit of perfect uniformity. If n_0 is odd, then $\theta^{-1} = -(n_0 + 1)/2$ results in the maximally uniform distribution and n_0 is replaced by $n_0 + 1$ in the inequalities above. Using likelihood tests, we can evaluate the comparative success of the predictions of the entire family of distributions defined by Eq. 27, of which HEAP is a special case.

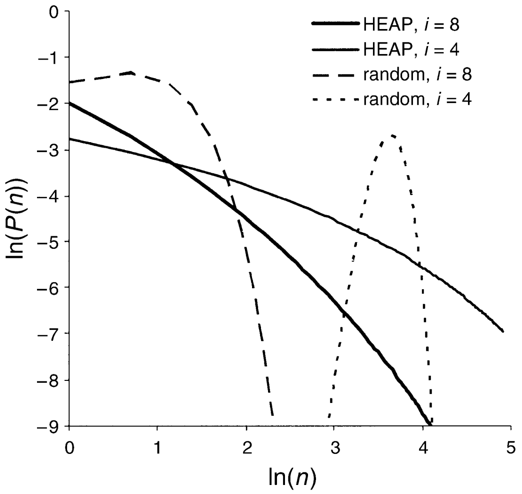

FIG. 2. Shape of the predicted probability function $P_i^{(617)}(n)$ plotted on ln–ln axes at two scales, $i = 8$ and $i = 4$. For $n_0 = 617$, the binomial distribution that results from the random placement model (Coleman 1981) at the same two scales is also shown for comparison.

SPECIES-LEVEL PREDICTIONS

The species-level spatial-abundance distribution

Fig. 2 shows the typical shape of the $P_i^{(n_0)}(n)$ at two scales, $i = 4, 8$. The selected value of $n_0 = 617$ happens to characterize one of the species in the serpentine plot that we used to test the theory, but the monotonic decrease of $P_i^{(n_0)}$ shown in Fig. 2 is a feature of the solution to Eq. 23 for all $i > 1$ and all values of n_0. In comparison, the random placement probability distributions predictions for $n_0 = 617$, $i = 4, 8$, are hump shaped and narrower than the HEAP prediction (Fig. 2). To

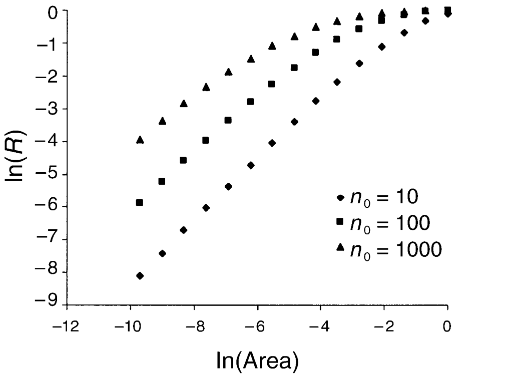

FIG. 3. The predicted range–area relationships for three abundances: $n_0 = 1000$, $n_0 = 100$, and $n_0 = 10$. The range, R, is the number of occupied grid cells of area A multiplied by area A. Units of area are such that $A_0 = 1$.

our knowledge, despite its simplicity, the recursion relation Eq. 23 describes a probability distribution that has not been described in the mathematics literature nor utilized in probability analysis.

Using Eqs. 7 and 23, a recursion relation that determines the $\alpha_i^{(n_0)}$ can be derived. It is most conveniently expressed in terms of products of the λ parameters, using the definition of $\lambda_i^{(n_0)}$ in Eq. 11:

$$\lambda_i^{(n_0)} - \lambda_{i+1}^{(n_0)} = [\lambda_{i-1}^{(n_0)} - \lambda_i^{(n_0)}][1 - \lambda_1^{(n_0)}] + \lambda_1^{(n_0)}[\lambda_i^{(n_0-1)} - \lambda_{i+1}^{(n_0-1)}]. \qquad (28)$$

This recursion relation for the products of α's allows straightforward numerical calculation of all the α's at any scale, i.

The range–area relationships

The theoretically predicted $\alpha_i^{(n_0)}$ values determine the shapes of the range–area relationship (RAR) curves (Eq. 8) for each species. Using Eqs. 11, 12, and 23, the RARs can be readily computed. The predicted RAR curves for species with $n_0 = 10$, 100, and 1000 are shown in Fig. 3. We note the tendency for the RAR curves, plotted on ln–ln axes, to be slightly curved at small scales and then to curve over and level off at larger scale. Because the slope of the RAR is equal to $-\log_2(\alpha)$, the predicted curvature in the RARs is equivalent to the prediction, for $n_0 > 1$, that the α's decrease with increasing i (decreasing cell size).

The transition to greater curvature occurs at smaller scale as n_0 increases. Although the $\alpha_i^{(n_0)}$ are strictly scale dependent for all $n_0 > 1$, we note that at small scales (large i) the predicted RAR behaves sufficiently like a power law (straight line on log–log plot) that it could readily be mistaken for such.

An aggregation index

A useful index of clustering or aggregation in the distribution of the individuals within a species was used by Condit et al. (2000) to explore clustering of individual trees within tropical forest species. The index, $\Omega_{x_1,x_2}^{(n_0)}$, measures the average density of conspecific individuals of a species (with n_0 individuals in A_0) in neighborhoods around an average individual, relative to the density expected for a random distribution. The ring-shaped neighborhoods have inner radius x_1 and outer radius x_2. We have shown elsewhere (Ostling et al. 2000) that the $\alpha_i^{(n_0)}$ uniquely determine $\Omega_{x_1,x_2}^{(n_0)}$. In particular, letting x_1 refer to the radius of a circle of area A_{i+1} and x_2 the radius of a circle of area A_i, and using the simpler notation of $\Omega_i^{(n_0)}$ in place of $\Omega_{x_1,x_2}^{(n_0)}$, then

$$\Omega_i^{(n_0)} = \frac{2}{\lambda_i^{(n_0)}} - \frac{1}{\lambda_{i+1}^{(n_0)}}. \qquad (29)$$

Plots of this index against abundance, n_0, at three spatial scales, using Eq. 28 to calculate the $\lambda_i^{(n_0)}$, are plotted in Fig. 4. The distinguishing features of these plots are that aggregation is greatest at small scales (large i) and that above a relatively small abundance threshold that is weakly scale dependent, aggregation decreases with increasing abundance. Both of these qualitative features are observed in the BCI tropical forest data (Condit et al. 2000).

The relationship between variance and mean

An exact solution to Eq. 23 takes the form

$$P_i^{(n_0)}(n) = \sum_{q=n}^{n_0} (-1)^{n+q}(q+1)^{-i}C(n_0, q)C(q, n) \qquad (30)$$

where $C(x, y)$ denotes $x!/[y!(x-y)!]$. From this result, an exact analytical expression for the variance of n, for a species with total abundance n_0, follows:

$$\sigma_i^2 = n_0^2(3^{-i} - 4^{-i}) + n_0(2^{-i} - 3^{-i}). \qquad (31)$$

Using the fact that the average occupancy across all cells of area A_i, $\langle n_i \rangle$, is given by

$$\langle n_i \rangle = n_0/2^i \qquad (32)$$

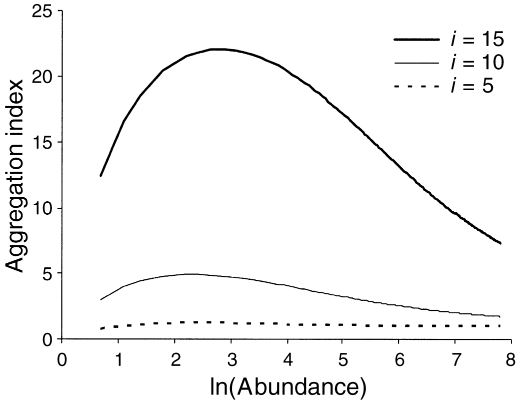

FIG. 4. Aggregation index (Eq. 29) plotted against ln(abundance) for $i = 15$, $i = 10$, and $i = 5$. For $A_0 = 50$ ha, $i = 15$, 10, and 5 correspond to outer radii of ~2.2 m, 12.5 m, and 70.5 m, respectively.

we can write

$$\sigma_i^2 = [(4/3)^i - 1]\langle n_i\rangle^2 + [1 - (2/3)^i]\langle n_i\rangle. \qquad (33)$$

Thus the variance contains terms that vary quadratically and linearly with the mean occupancy, indicating behavior that is intermediate between the prediction derived under the assumption of self-similarity (Banavar et al. 1999), for which $\sigma^2 \sim (\text{mean})^2$, and the continuous (Poisson) version of the random placement model, for which $\sigma^2 \sim$ mean. From Eqs. 32 and 33 we can also calculate the scale dependence of the specific combination of σ_i^2 and $\langle n_i\rangle$ that comprises the parameter k in the negative binomial distribution (NBD; Eq. 1):

$$\frac{\langle n_i\rangle^2}{\sigma_i^2 - \langle n_i\rangle} = \frac{1}{\left(\frac{4}{3}\right)^i\left(1 - \frac{1}{n_0}\right) - 1}. \qquad (34)$$

Using the fact that $A_i = A_0/2^i$, this can be rewritten as

$$\frac{\langle n_i\rangle^2}{\sigma_i^2 - \langle n_i\rangle} = \frac{1}{\left(\frac{A_0}{A_i}\right)^{\log_2(4/3)}\left(1 - \frac{1}{n_0}\right) - 1}. \qquad (35)$$

For sufficiently large A_0/A_i and n_0, this expression approaches $(A_i/A_0)^{0.42}$, where we have used $\log_2(4/3) \cong 0.42$. In practice, for $n_0 > 10$ and $i > 8$, this ought to be a good approximation. This implies that the effective value of the NBD parameter, k, in our theory scales as $A^{0.42}$ for large i.

Community Properties

To derive theoretical expressions for community-level patterns, including the species–area relationship (SAR), endemics–area relationship (EAR), and the scale-dependent species-abundance distribution, $\Phi_i(n)$, we need only assume that HEAP applies to all species; no assumptions are needed about spatial correlations between species because these community-level properties depend only on the collection of species-level spatial distributions and are independent of interspecific correlations. To derive the scale-dependent community-level spatial-abundance distribution, $F_i(N)$, we additionally assume that the spatial distribution of the individuals in each species is independent of the locations of individuals in the other species; in other words, HEAP applies independently to each of the species and thus interspecific spatial correlations are absent.

The species–area and the endemics–area relationships

The SAR can be expressed either in terms of the λ's, as shown in Eq. 10, or equivalently in terms of the species-level spatial distribution functions using the formula

$$S_i = \sum [1 - P_i^{(n_0)}(0)] \qquad (36)$$

where the sum is over all species. Eq. 36 states that the expected number of species in an A_i cell equals the sum over species of the probability of species occurrence in that cell, with the probability of occurrence equal to 1 minus the probability of absence. Because the P_i's and the λ's depend only on n_0, the shape of the SAR is predicted uniquely in terms of the list of abundances n_0. Although we cannot write down a simple analytical form for the SAR in terms of the list of abundances (because neither Eq. 23 nor Eq. 28 has, to our knowledge, a simple closed form analytical solution), we can solve for the SAR numerically from the list of abundances using Eqs. 10 and 28, or equivalently Eqs. 30 and 36.

A simple analytic expression for the EAR can be derived from the theory. We define a species (from the list in A_0) to be endemic to an A_i cell if all its individuals are found only in that A_i cell. Consider a single cell of area A_i. We seek the mean number of species that are found only in such a cell and not within the other $2^i - 1$ A_i cells that comprise A_0. The dependence of that number on area A_i is what we mean by an EAR. Denoting by E_i the expected number of endemic species in A_i, we have, from the definition of endemicity,

$$E_i = \sum P_i^{(n_0)}(n_0) \qquad (37)$$

where the sum is over all species.

But from Eq. 23 it is straightforward to show that

$$P_i^{(n_0)}(n_0) = \frac{1}{(n_0 + 1)^i} \qquad (38)$$

and hence the EAR takes the form

$$E_i = \sum \frac{1}{(n_0 + 1)^i} \qquad (39)$$

with the sum taken over all species. We note that Eq. 39 yields $E_0 = \Sigma\, 1 = S_0$, as it must because all the species are endemic to A_0 by our definition of endemicity.

Scale-dependent species-abundance distribution

As illustrated in Fig. 1, the scale-dependent species-abundance distribution $\Phi_i(n)$ is given by

$$\Phi_i^{(n_0)} = \frac{\sum P_i^{(n_0)}(n)}{S_0}. \qquad (40)$$

The argument of the function $\Phi_i(n)$ can take on values ranging from 0 to the largest value of n_0 in the community and the summation in Eq. 40 is taken over all species.

Community-level spatial-abundance distribution

The theory also yields expressions for the community-level spatial-abundance distribution, $F(N)$, provided that we assume HEAP applies independently to all species. In particular, the distribution of the total number of individuals (that is, summed over species) over the grid cells is given by the following:

$$F_i(N) = \sum_{\forall n_{(j)}} \prod P_i(n_j)\delta\left(N, \sum_j n_j\right) \tag{41}$$

where Π denotes a product over all species. Here $n_{(j)}$ is an abundance variable for the jth species and the equation simply states that the probability of a total of N individuals in an A_i cell is given by the sum over the joint probabilities for all the various combinations of individual species abundances that add up to the value N.

Comparisons with Data

The primary data set against which we test the theoretical predictions is census data from serpentine grassland habitat at Little Blue Ridge (Green et al. 2003) located at the University of California's Homestake Mine/Donald and Sylvia McLaughlin Natural Reserve (38°51′ N; 122°24′ W). We counted all individuals of all plant species in all 256 1/4-m^2 grid cells comprising an 8 m × 8 m plot. Our completed data set consists of the locations at a resolution of 1/4-m^2 cells of 37 182 individuals, divided among 24 species. From those data, aggregations at larger scale are readily computed.

We also tested the HEAP predictions using spatially explicit data on tree locations in the 500 m × 1000 m tropical forest plot at Barro Colorado Island (BCI) in Panama (Hubbell and Foster 1983, Condit et al. 1996*a*, Condit 1998) and in a 220 m × 440 m dry forest plot at San Emilio (SanEm) in Costa Rica (Enquist et al. 1999). The latter plot is the largest 1 × 2 shaped plot contained within a larger, but irregularly shaped, plot in which tree census data are available. For these plots, the coordinates of every individual in every tree species with dbh $\geq$1 cm is recorded. The BCI data set contains 235 308 individuals divided among 305 species and the dry forest data set contains 12 851 individuals divided among 138 species. For all three sites, the empirical species-level and community-level distributions are computed from data on all 2^i A_i cells, and thus the SAR and EAR are derived from a complete nested analysis.

Where appropriate, we calculate 95% confidence intervals on linear regression slopes of predicted vs. observed measures, and we use a likelihood-ratio test to compare the HEAP model predictions to those from distributions that fall along the continuum from random to maximally aggregated as defined by Eq. 27. To simply express the explanatory power of HEAP we use R^2 values to quantify the fraction of variance in the data that is explained by HEAP. In our comparisons to a random model, we use the random placement model of Coleman (1981), in which individuals from a specified species-abundance distribution are at every scale placed on a gridded landscape according to a binomial distribution. We emphasize that our goal is not the impossible task of establishing the superiority of HEAP relative to all other theories; rather it is to show the extent to which it captures the ecologically significant features of numerous spatial patterns and to identify the dominant trends in its failures to fit all the details of empirical patterns.

The species-level spatial distributions and the α parameters

A test of Eq. 23, the fundamental species-level prediction of HEAP, consists of comparing the observed and predicted spatial-abundance distributions. Qualitatively, the spatial-abundance distributions are monotonically decreasing for nearly all species at the three sites across the range of scales examined here; thus for $n_0 > 2^i$ they differ from the hump-shaped predictions expected from the random (binomial distribution) model. To see if the HEAP prediction captures the quantitative features of the data, we compare the actual shapes of the $P_i^{(n_0)}(n)$ to the data. These comparisons are carried out at scale $i = 6$ and 8 for the serpentine data and at $i = 8$ and 15 for the BCI and San Emilio data. The scale $i = 15$ corresponds to 15-m^2 quadrats, with a mean of 7 individuals in each, at BCI, and 3-m^2 quadrats, with a mean of 0.4 individuals in each, at San Emilio. The scale $i = 8$ corresponds to 0.25-m^2 quadrats, with a mean of 144 individuals in each, at the serpentine site; this is the smallest scale censused at that site.

Fig. 5 shows predicted and empirical values of the spatial-abundance distributions for four serpentine species with abundances ranging from $n_0 = 49$ to 3095. These species represent the range of abundances present in the data set, but were otherwise randomly selected. For the two most abundant species shown, the random model prediction at $i = 6$ is ~0 over the range of abundances plotted and does not even show up in the figure. We note three qualitative features of these comparisons. (1) The HEAP predictions match the general shape of the observed distributions and outperform the random model predictions at both scales and for all four species. (2) The differences between HEAP and random predictions diminish, but do not become negligible, with decreasing abundance. (3) As abundance increases, the HEAP prediction increasingly overpredicts the fraction, $P_i^{(n_0)}(0)$, of unoccupied cells. Equivalently, it tends to overpredict the abundances of occupied cells and thus the level of aggregation. Although not shown in the figure, the spatial distribution data are also well described by HEAP for the very lowest-abundance species ($n_0 < 30$).

Comparisons of theory and observation reveal the same three features for the BCI tropical forest data set (Fig. 6) and the San Emilio data set (Fig. 7) at scales $i = 8$ and 15, except that feature (3) is no longer true at $i = 15$. In particular, the HEAP prediction for $P_i^{(n_0)}(0)$ matches the data well for all abundances at that fine scale. Accurate prediction of the values of $P_i^{(n_0)}(0)$ at fine scales implies accurate prediction of the values of the α's at those scales. For both these tropical sites, the species tested in Figs. 6 and 7 were chosen to have

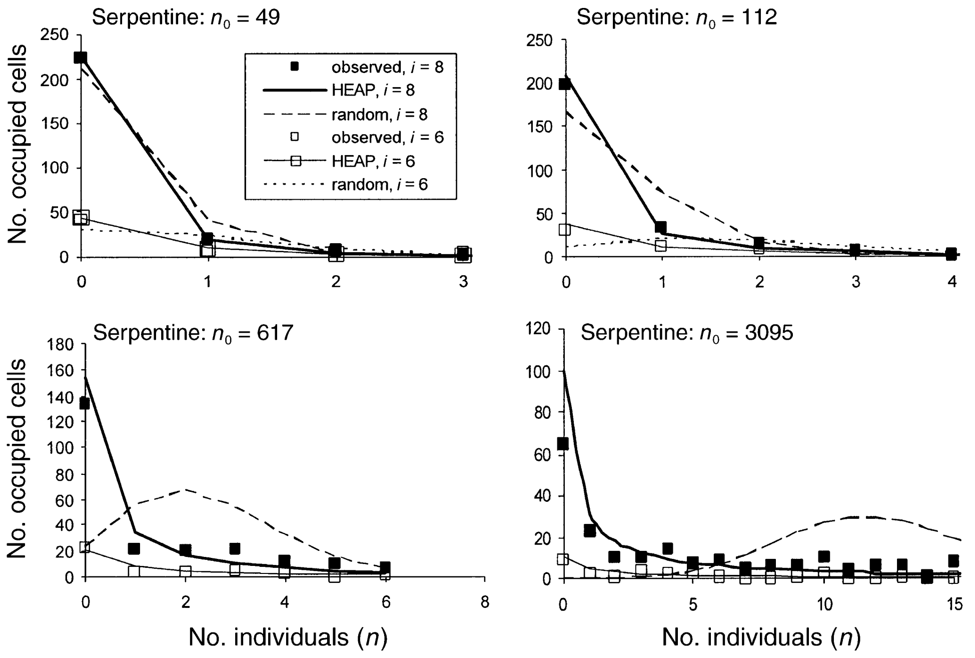

FIG. 5. Observed and predicted species-level distributions $P_i^{(n_0)}(n)$, multiplied by 2^i to yield the expected number of cells occupied with n individuals for the serpentine plot. The abundances of the species selected for comparison and the scales of analysis are shown in the figures; the four species represent the range of abundances but were otherwise randomly selected. Random model predictions are from the Coleman (1981) random placement model; random model predictions for $n_0 = 617$ and 3095 are not shown at scale $i = 6$ because they are ~0 over the range of n values plotted.

abundances closest to the four species selected from the serpentine site, although at the San Emilio site there are no species with $n_0 > 1000$.

By focusing on the $\alpha_i^{(n_0)}$ parameters, we can gain a more detailed picture of the tendency for Eq. 23 to overpredict the amount of aggregation in the distributions of the higher-abundance species. A graph of predicted vs. measured $\alpha_i^{(n_0)}$ values, averaged over scales $i = 4$, 6, and 8, for all the species in the serpentine plot, shows generally good agreement, but a general tendency for HEAP to underpredict the α's of the species with the highest α values (Fig. 8a). Because at any given scale, the predicted value of $\alpha_i^{(n_0)}$ is a monotonic increasing function of n_0, this implies a trend toward underprediction of α for the higher-abundance species. This is consistent with the general trend

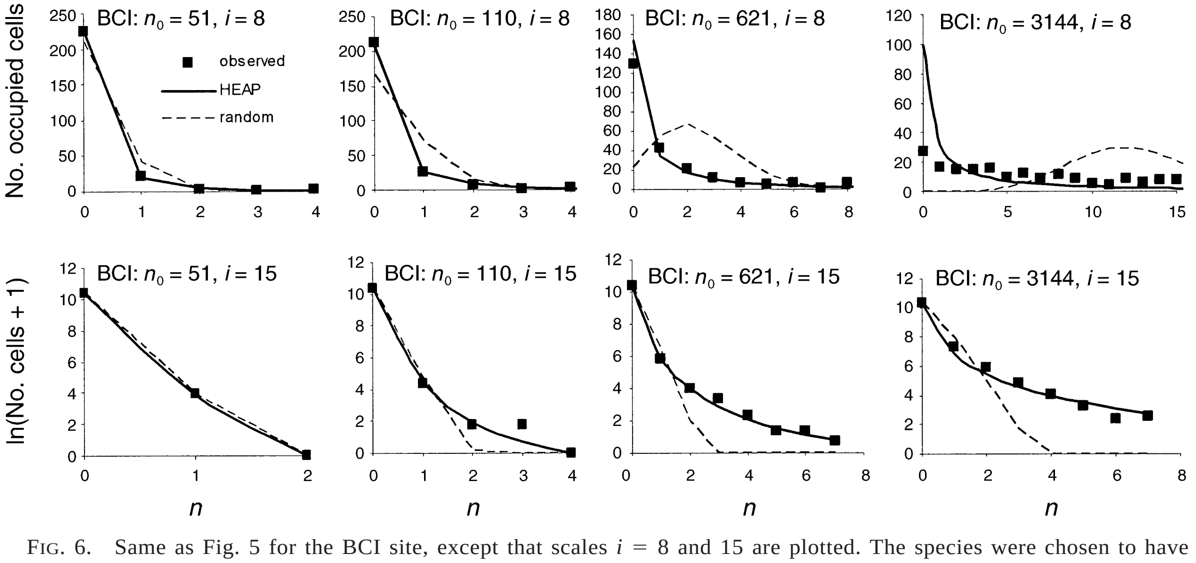

FIG. 6. Same as Fig. 5 for the BCI site, except that scales $i = 8$ and 15 are plotted. The species were chosen to have abundances closest to those in Fig. 5.

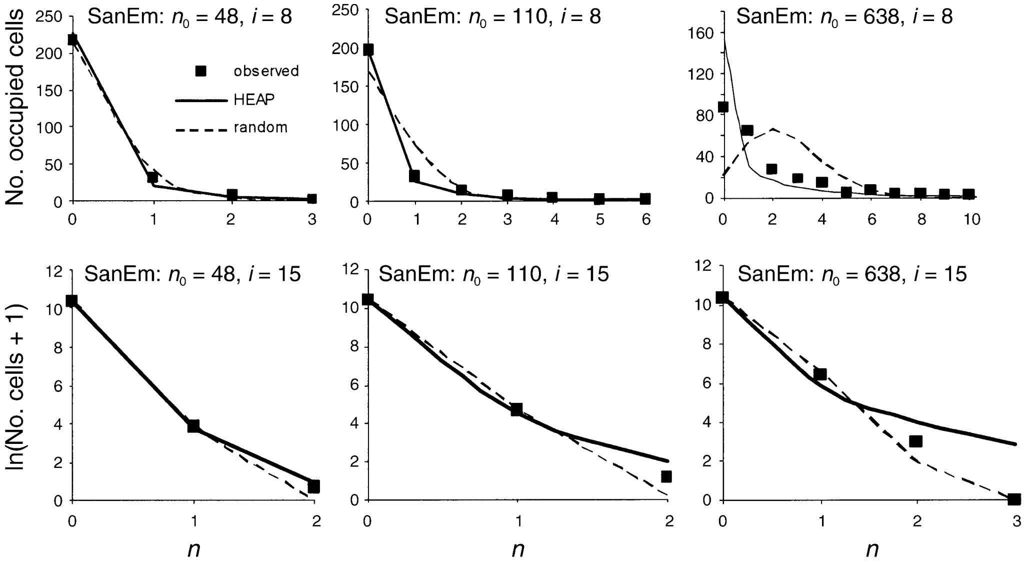

FIG. 7. Same as Fig. 6 for the San Emilio site, except that no species at the San Emilio site had $n_0 > 1000$.

that we saw in Figs. 5–7 for HEAP to overpredict $P_i^{(n_0)}(0)$ for the higher-abundance species. The slope of the linear regression (solid line in figure) is 0.91 with 95% confidence limits of (0.80, 1.02). For each species, Fig. 8a also shows the predicted value of α, averaged over the same scales, from the random model. We note that compared to HEAP and to the data, the random model underpredicts the α's at small abundance and overpredicts them at large abundance. To assess the relative performance of HEAP and the random model, we compare the ratio of sum of squared errors (RSSE) of each model using the ratio

$$\mathrm{RSSE} = \frac{\sum (Y - \hat{Y}_{\mathrm{random}})^2}{\sum (Y - \hat{Y}_{\mathrm{HEAP}})^2} \tag{42}$$

where the sums are over species, Y are the observed variables, and $\hat{Y}$ are the random or HEAP model predictions. RSSE > 1 implies that HEAP fits the observed data better than the random model. The calculated RSSE for the serpentine data in Fig. 8a is 2.9.

For each of the scales i = 1, 2, 4, 6, 8, the fractions of variance in the serpentine α data that are explained by HEAP (the squared correlations between the HEAP-predicted values and the measured values) are 0.70, 0.79, 0.74, 0.92, and 0.94, respectively. We note that even at scale i = 1, where the distribution of individuals between the right and left halves of the 64-m^2 plot are examined, HEAP explains 70% of the variance in the α's, but that the explained variance increases at finer scales.

The HEAP predictions for the α's at the BCI and San Emilio sites are not as accurate as at the serpentine site (Fig. 8b,c). The slopes of the linear regressions of predicted against measured α values are 0.75 and 0.73,

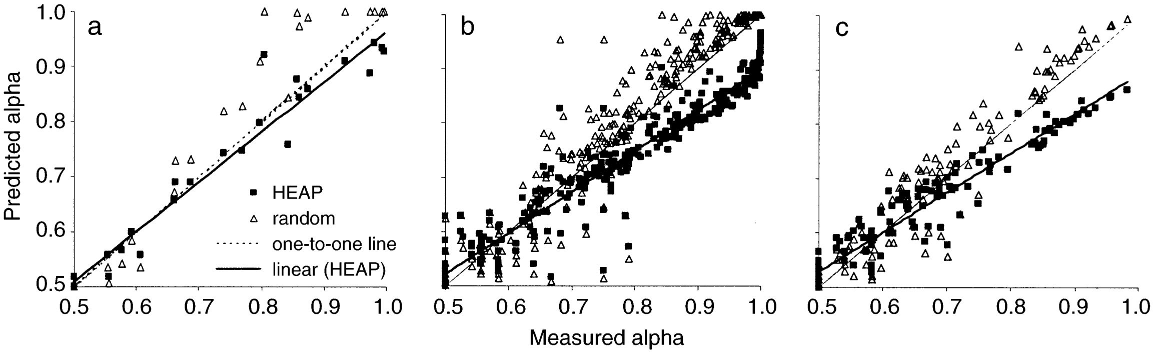

FIG. 8. Comparison of observed and predicted values of $\alpha_i^{(n_0)}$ for all plant species in (a) the serpentine plot, (b) the BCI plot, and (c) the San Emilio site. Comparisons are averaged over scales i = 4, 6, and 8. The dashed line is the one-to-one line, and the solid line is the linear regression of predicted on measured values. Solid squares are the HEAP predictions, and open triangles are predicted values of the $\alpha_i^{(n_0)}$ from the random placement model (Coleman 1981), calculated from the measured abundances, n_0.

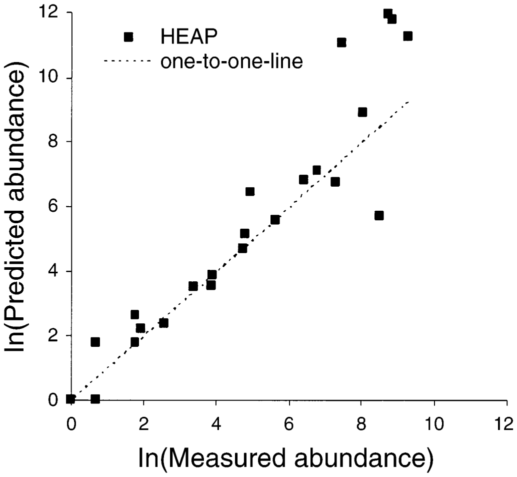

FIG. 9. Comparison of observed and predicted values of abundance for the serpentine species. The HEAP predictions for abundance are derived by calculating the value of n_0 that yields, from Eq. 7, a value of α_8 equal to the measured value of α_8 for each species. The HEAP prediction explains 88% of the variance in the data shown in the figure, and the slope of the linear regression is 1.16, with 95% confidence limits of (0.92, 1.40).

respectively, with 95% confidence intervals of ± 0.03 and ± 0.05 for BCI and SanEm, respectively. HEAP explains 93% of the variance in the α values, at each of the forest sites. As at the serpentine site, the random model underpredicts the α's at low abundance (small α) and overpredicts the α's at high abundance in the forest sites; the observed values of α for the high-abundance species fall between the HEAP and random model predictions, and with all species included, the RSSE (Eq. 42) values are ~1 at both sites.

Another way to portray the relationship between the predicted and measured α's is to plot actual species abundances, n_0, against the abundances predicted from the measured α's (see *Theoretical background: The probability parameters* for the discussion following Eq. 8). This comparison is of practical value because there is interest in being able to estimate overall abundance of a species within some large area based only on an empirical estimate of cell occupancy (i.e., presence/absence data) in small cells within that large area. Fig. 9 shows this comparison for the serpentine species. HEAP explains 88% of the variance in the measured values of $\ln(n_0)$ and the slope of the regression line in the figure is 1.16 ± 0.24 (95% CI). Fig. 9 shows that HEAP tends to overpredict abundance (equivalent to underpredicting α) for the higher-abundance species, consistent with Fig. 8a. We will return to this systematic pattern of discrepancy in the HEAP predictions later when we discuss future directions.

The mean–variance relationship

For $A_0 = 50$ ha, corresponding to the BCI data, and expressing A in units of m^2, HEAP predicts $k = 0.004 \times A^{0.42}$ at small scales and for all but the lowest-abundance species (Eq. 34). This predicted scale dependence for the quantity k is close to the empirical finding of J. Plotkin (*personal communication*) that if k is written as cA^w, then for the BCI species, fitted values of c and w cluster around values of ~0.003 and ~0.45, respectively.

The SAR and the EAR

Starting with the list of abundances of the species in A_0, the $\lambda_i^{(n_0)}$ can be calculated for all the species in a community from Eq. 28. Then, using Eq. 10, the species–area relationship (SAR) is determined. Equivalently, Eqs. 23 and 36 can be used to determine the predicted SAR. Because of our top-down boundary condition, in our tests of the SAR and the endemics–area relationship (EAR) the total number of predicted species is constrained empirically at the largest scale, A_0. The theoretical prediction captures the general features of the serpentine SAR (Fig. 10a,b) but tends to underpredict species richness at the forest sites. This discrepancy is consistent with the pattern in which the spatial distributions of the high-abundance species deviate from the theoretical predictions: the theory underpredicts species richness because it overpredicts the number of unoccupied cells for the high-abundance species. Equivalently, HEAP underpredicts the α's for those species and thus by Eq. 9 and the relationship $z = -\log_2(a)$, it overpredicts the slope of the SAR. We note that the gross differences across sites in the slopes of the empirical species richness vs. ln(area) plots ($\text{slope}_{\text{BCI}} > \text{slope}_{\text{SanEm}} > \text{slope}_{\text{SERP}}$) are captured by the theory (Fig. 10b) and that these gross differences in predicted slopes result solely from the different lists of species abundances at the three sites.

The theoretical EAR (Eq. 39) matches quite accurately the empirical EAR at all three sites (Fig. 11), again with no adjustable parameters. If the individuals of each species are distributed randomly, then we would expect the following (Green and Ostling 2003):

$$E_i = \sum 2^{-in_0} \tag{43}$$

where the sum is over all species. Compared to Eq. 39, this expression significantly underestimates the observed data at scales $i = 1$ and 2, but fits the data as well as our theory at smaller spatial scales. The reason is that at small spatial scales, only species with $n_0 = 1$ contribute significantly to endemism, and for those species $\alpha = 1/2$ in HEAP, which is the value that a random model would predict for those species.

The scale-dependent species-abundance distribution

We also tested Eq. 40, the predicted expression for the scale-dependent species-abundance distribution $\Phi_i(n)$ for the serpentine data at $i = 6$ and 8 scales (Fig. 12a). At both scales the empirical cumulative distribution is very well described by the theoretical pre-

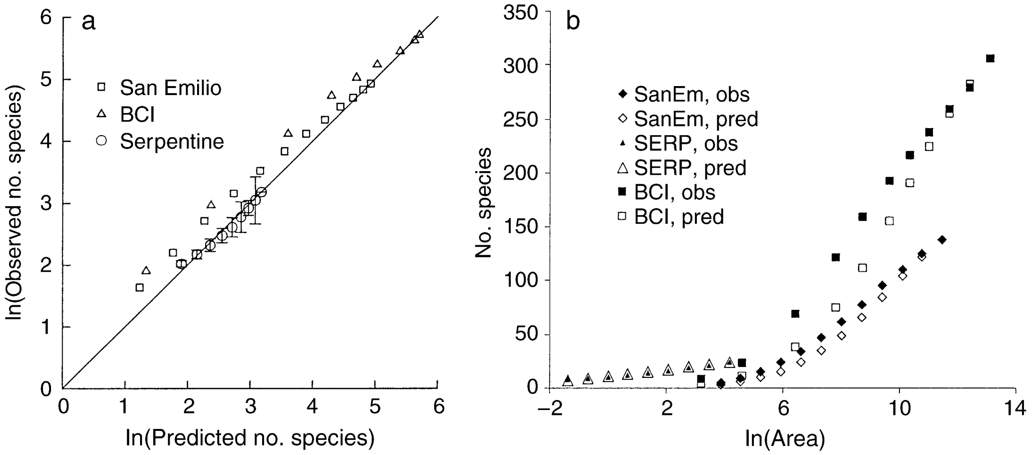

FIG. 10. Observed and predicted values of species richness as a function of scale for the three sites. (a) Predicted vs. observed species–area relationship (SAR); the straight line is a 1:1 line. (b) Predicted and observed species richness vs. ln(area); area is measured in square meters. Error bars on the serpentine data in panel (a) are ±1 SE; error bars on the BCI and San Emilio forest data points are generally no larger than the symbols and are not shown.

diction. The random placement model underpredicts $\Phi_i(n)$ at $n = 0$ and overpredicts the distribution at larger n, by approximately a factor of 2 relative to the data and to HEAP (Fig. 12b). As with all our other tests of theory here, the comparisons in Fig. 12 are not best fits, for there are no adjustable parameters.

The community-level spatial-abundance distribution

Calculation of the full expression for the community-level spatial-abundance distribution, $F_i(N)$, directly from Eq. 41 is a difficult numerical task for values of $N \geq 5$ because of the huge number of combinations of abundance values that can add up to a selected value of N, even for just the 24 species in the serpentine plot; it becomes enormously difficult for the more species rich forest plots. Hence we have simulated landscapes using HEAP and then calculated the predicted $F_i(N)$ directly from those landscapes. To verify the accuracy of the simulations, we compared simulated $F_i(N)$ with the exact theoretical $F_i(N)$ calculated from Eq. 41 for $i = 6$ and 8, and for $N = 0 \ldots 4$, using a species-abundance distribution identical to that of the serpentine site. The simulated $F_i(N)$ were in excellent agreement with the exact theoretical results.

We carried out two types of comparison of theory with observation; first with all the species at each site, and second using just those species with $n_0 < 1000$ at the serpentine and BCI plots and with $n_0 < 200$ at the SanEm site. The second case, with restricted abundance, was examined because we have seen that the theoretical species-level spatial-abundance distributions do not describe well the empirical distributions for those species with the very highest abundances; we chose the abundance cutoff at SanEm to be 1/5 that at the BCI plot because the area of the SanEm plot is ~1/5 that of the BCI plot. These restrictions leave us with 17 out of 24 species in the serpentine community, 259 out of 305 BCI species, and 116 out of 138 SanEm species. Although these restrictions exclude only a minority of the species, they exclude a majority of the individuals in the entire study areas.

For each of the three sites, the theoretical prediction for the cumulative community-level spatial-abundance distribution, $\sum_{k=0}^{N} F_8(k)$, agrees closely with the data when abundances are restricted; see the comparison of solid squares (data) with solid line (HEAP prediction) in Fig. 13. On the other hand, if all species are included, then the theoretical prediction rises faster at small N and reaches a plateau at lower N than the observed cumulative distribution. We note that the predicted community-level spatial-abundance distribution, $F_i(N)$, is a hump-shaped function that is right-skewed relative

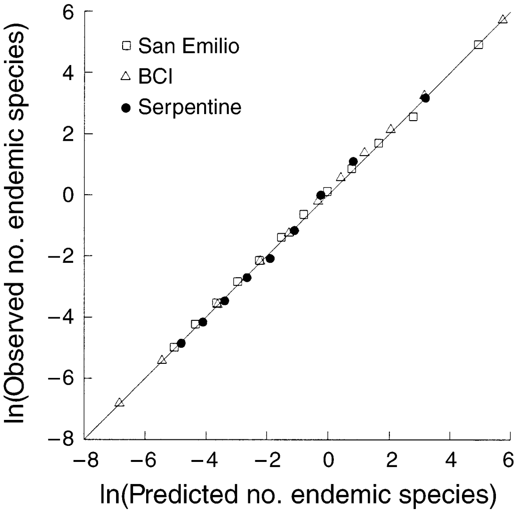

FIG. 11. Observed and predicted values of endemic species richness as a function of scale for the three sites. The straight line is a 1:1 line.

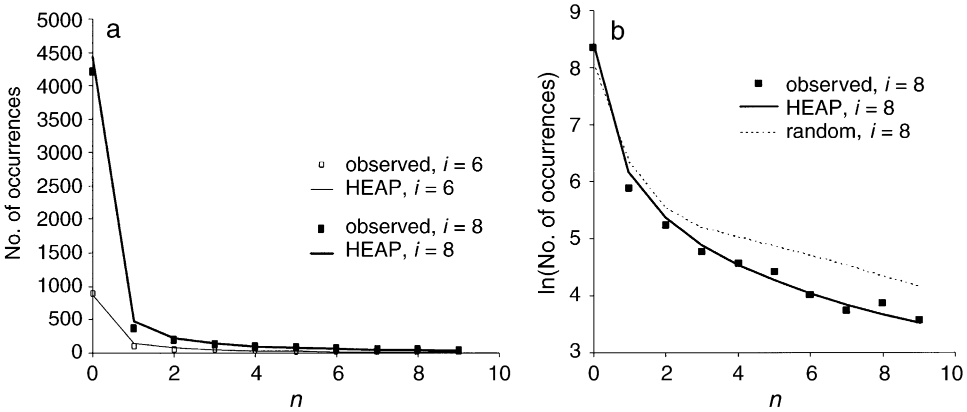

FIG. 12. Observed and predicted values of $\Phi_i(n)$ for the serpentine site. (a) Comparison of predicted and empirical species-abundance distribution $\Phi_i(n)$ for $i = 6$ (lower curve and data) and $i = 8$ (upper curve and data). (b) Comparison of HEAP and the random model predictions for the $i = 8$ data, with $\ln(\Phi_i(n))$ plotted to better highlight the differences at small $\Phi_i(n)$.

to a normal distribution and left-skewed relative to a lognormal distribution.

The prediction in Eq. 41 for the community-level spatial-abundance distribution from HEAP relies on two assumptions: that the species-level spatial-abundance distributions are well described by the solution to our recursion relation, Eq. 23, and that the species distributions are independent of one another. Given that the species-level spatial-abundance distributions of all but the highest-abundance species are well described by the solutions to the recursion relation, $P_i^{(n_0)}(n)$, Fig. 13 thus lends support to the assumption of interspecific independence for that subset of all but the highest-abundance species.

To examine this more directly, we also carried out two additional tests of the assumption of interspecific independence. First, we generated a pairwise correlation matrix of the abundances of species across grid cells, as before restricting the analysis to those species with $n_0 < 1000$ in the BCI and serpentine data sets, and $n_0 < 200$ in the SanEm data set. Only a small fraction of the pairwise comparisons are statistically significant, with Bonferroni correction, at each of the three sites (<8.2% at i = 4, 6, and 8 at the serpentine site, <2.4% at i = 5, 7, and 9 at the forest sites). A majority of pairwise correlations were nonsignificant and negative, but the small fraction of pairs of species correlations that were statistically significant were all positive at each site.

For a second test, we took the observed species-level spatial distributions at scale i = 8 and simulated a new landscape in which the species-level spatial-abundance distributions were maintained but the species were distributed independently of one another. For each cell of size A_8 on that landscape, the abundance of each species was drawn without replacement from the list of observed abundances for that species at the A_8 scale, independent of the abundances chosen for all other species in that cell. From that landscape, the cumulative community spatial-abundance distribution was computed, averaged over 1000 such simulations, and compared to the actual observed distribution.

Here, a small but systematic deviation between the observed (solid squares) and the ''independent'' dis-

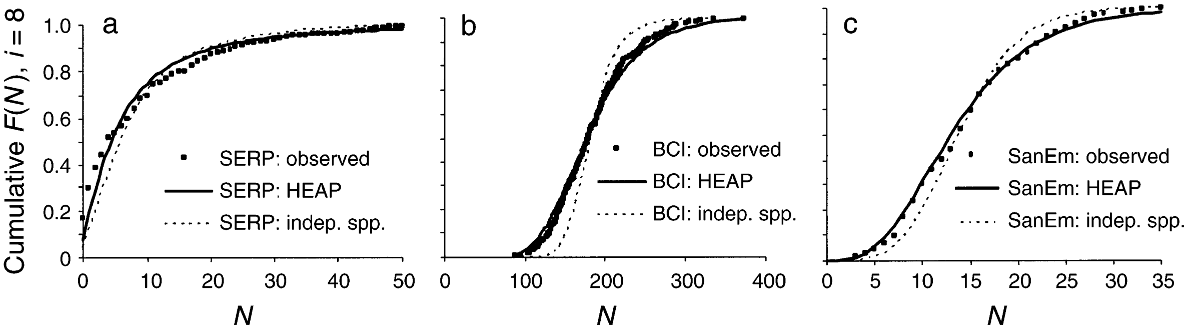

FIG. 13. Comparisons of the predicted (solid line) and observed (solid squares) cumulative values of the community spatial-abundance distribution $F_8(N)$; also shown are the results of a computed value for the cumulative distribution obtained by resampling from the observed distribution under the assumption that the species are distributed independently of each other (dashed line). (a) All serpentine species with $n_0 < 1000$; (b) all BCI species with $n_0 < 1000$; (c) all SanEm species with $n_0 < 200$.

tributions (dashed line) did show up at all three sites (Fig. 13). The observed cumulative $F_i(N)$ is a little less steep than the cumulative $F_i(N)$ calculated from the landscapes created assuming species are distributed independently of one another (the ''independence landscape''), indicating that the observed $F_i(N)$ is broader than would be expected under complete interspecific independence. This implies that some fraction of species exhibit positive interspecific correlations in abundance, which is qualitatively consistent with the results of analyzing the correlation matrix.

The HEAP-predicted $F_i(N)$ are virtually indistinguishable from observed (Fig. 13). Because the HEAP predictions ignore interspecific correlations, if HEAP predicted the observed species-level spatial-abundance distributions perfectly, then one would expect the HEAP predictions for the low-abundance subcommunities to fall on top of those from the ''independence landscapes.'' The fact that they agree closely with the observed distributions implies that the small divergences between observed and predicted species-level spatial distributions tend to correct for the discrepancies that arise from ignoring the small, positive interspecific correlations.

TABLE 1. Likelihood results for the serpentine data at scale $i = 8$.

n_0	$\theta_{\max}$	$R_{\text{HEAP,max}}$	$R_{\text{ran,max}}$	$R_{\text{HEAP,ran}}$
6	1.82	0.5	6.6*	6.2
6	3.33	2.5	17.1*	14.6
7	2.50	1.2	8.5*	7.4
13	2.00	1.7	25.5*	23.8
30	2.86	8.9*	124.5*	115.5
49	0.87	0.1	42.8*	42.7
50	0.83	0.2	69.1*	68.9
112	0.54	5.3	123.2*	117.9
120	0.59	5.0	136.2*	131.2
139	0.05	64.8*	5.5	−59.3
272	0.77	1.5	671.2*	669.6
617	0.29	45.3*	930.4*	885.2
885	0.65	9.0*	1917.0*	1908.0
1418	1.00	0.0	4517.9*	4517.9
1759	0.13	194.0*	817.5*	623.5
3095	0.33	74.7*	3877.3*	3802.6
4827	1.43	5.4	17 707.9*	17 702.5
5989	0.08	357.6*	917.4*	559.8
6990	0.22	153.9*	5341.1*	5187.2
10 792	0.31	117.2*	9481.2*	9363.9

Notes: Asterisks in the $R_{\text{HEAP,max}}$ column indicate species for which HEAP is rejected, and asterisks in the $R_{\text{ran,max}}$ column indicate species for which the random model is rejected, in each case by the chi-square test at the 1% level. The 1% cutoff is $R = 6.63$. The 5% cutoff is $R = 3.84$. The R's are log-likelihood ratios defined by Eqs. 44–46.

Comparison of HEAP to a continuum of distributions

To compare HEAP at each site, and at various spatial scales, to the continuum of species-level distributions that result from application of Eq. 27, we carried out likelihood-ratio tests. We first used Eq. 27 to derive the likelihood of observing a particular distribution of individuals over cells at scale i. We denote that likelihood function by $P(\mathbf{N}, \theta)$, where $\mathbf{N}$ is shorthand for a matrix of 2^i abundances that sum to n_0. We denote by $\theta_{\max}$ the value of θ that maximizes $P(\mathbf{N}, \theta)$. For convenience we leave the scale index off of $\theta_{\max}$ and $P(\mathbf{N}, \theta)$. We then evaluated the $P(\mathbf{N}, \theta)$ and $\theta_{\max}$ for each species in the forest and serpentine sites.

Table 1 reports values of $\theta_{\max}$ for each of the 20 serpentine species with $n_0 > 2$. The mean value of $\theta_{\max}$ is 0.99 (SE = 0.22), in good agreement with the HEAP assumption of $\theta = 1$. However, some of the individual values of $\theta_{\max}$ differ significantly from 1, particularly for the species with highest abundances. To test the HEAP hypothesis $\theta = 1$, and also to test the random hypothesis $\theta = 0$, we used the standard likelihood-ratio statistics:

$$R_{\text{HEAP,max}} = 2\log\left[\frac{P(\mathbf{N}, \theta_{\max})}{P(\mathbf{N}, 1)}\right] \qquad (44)$$

$$R_{\text{ran,max}} = 2\log\left[\frac{P(\mathbf{N}, \theta_{\max})}{P(\mathbf{N}, 0)}\right]. \qquad (45)$$

Both ratios are positive by definition. A sufficiently large value of $R_{\text{HEAP,max}}$ allows statistical rejection of the HEAP model, and a large value of $R_{\text{ran,max}}$ does the same for the random model. In particular, when $\theta = 1$, $R_{\text{HEAP,max}}$ has an approximate chi-square distribution with one degree of freedom, which can be used to test the null hypothesis $\theta = 1$ against the alternative $\theta \neq 1$ (Sokal and Rohlf 1995). Likewise, when $\theta = 0$, $R_{\text{ran,max}}$ has an approximate chi-square distribution with one degree of freedom, which can be used to test the null hypothesis $\theta = 0$ against the alternative $\theta > 0$. We verified by Monte Carlo simulation that the result works well here even though the standard technical conditions are not exactly satisfied in the $\theta = 1$ case. For a 1% level of significance, the critical chi-square is 6.63. For the serpentine species, at scale $i = 8$, HEAP is rejected for eight of the 20 species (mostly the higher-abundance species), and the random model is rejected for 19 of the 20 species (Table 1). For a less stringent test, the 5% critical value is 3.84.

The species for which HEAP is rejected are highly skewed toward those with the highest abundances. For those species, the value of $\theta_{\max}$ is generally between 0 and 1, indicating spatial distributions more aggregated than the random model predicts but less aggregated than HEAP predicts, consistent with our previous observation that HEAP generally overpredicts the number of empty cells for a high-abundance species.

To focus on comparisons between HEAP and the random model ($\theta = 1$ vs. $\theta = 0$), Table 1 also reports

$$R_{\text{HEAP,ran}} = R_{\text{HEAP,max}} - R_{\text{ran,max}} = 2\log\left[\frac{P_i(\mathbf{N}, 0)}{P_i(\mathbf{N}, 1)}\right]. \qquad (46)$$

This statistic can be positive or negative (and thus is not another chi-square), with positive values favoring HEAP over the random model and negative values favoring the opposite. Table 1 shows that $R_{\text{HEAP,ran}}$ is positive for 19 of the 20 serpentine species with $n_0 > 2$.

Qualitatively similar results hold at the BCI and San Emilio sites, with HEAP outperforming the random model for ~3/4 of the species at both sites by the likelihood-ratio criterion. For example, at scale $i = 13$ at BCI, HEAP outperforms the random model for 215 out of the 283 species with $n_0 > 1$. At these forest sites, however, HEAP can be statistically rejected for a larger fraction (~2/3) of species than at the serpentine site (2/5). As at the serpentine site, species for which HEAP and/or the random model are rejected by chi-square tests are generally the ones with highest abundance. And as at the serpentine site, those rejected species nearly all have a $\theta_{\max}$ between 0 and 1, indicating more aggregation than under the random model but less than under HEAP.

We note that statistically significant rejection of HEAP does not imply ecologically significant failure of specific HEAP predictions such as the spatial-abundance distributions that result from Eq. 23. For example, $R_{\text{HEAP, max}} = 45.3$ for the serpentine species with $n_0 = 617$, which implies a highly significant rejection of the exact HEAP value $\theta = 1$ at $i = 8$. Yet Fig. 5 indicates that HEAP captures well the essential features of the distribution $P_i^{(617)}(n)$. Moreover, for any given species a test of the HEAP prediction for $P(\mathbf{N}, \theta)$ is a more stringent test than that of Eq. 23 because it tests the entire landscape pattern, including correlations across cells, and not just the distribution of abundances within a single cell.

A more comprehensive test of HEAP, and insight into the implications of strict statistical rejection of the theory, are obtained by looking at the community likelihoods. Under the assumption that spatial distributions are independent across species, the $P(\mathbf{N}, \theta)$ for a community is the product of the $P(\mathbf{N}, \theta)$ for the species within the community and hence the community log likelihood is the sum of the log-likelihood values of the individual species. We compare the full community log likelihoods at each site, as well as the log likelihoods for the subset of each community that excludes the most abundant species, in Fig. 14. Because θ ranges from 0 to ∞, we compare log likelihoods against the transformed parameter

$$\phi(\theta) = \theta/(1 + \theta) \qquad (47)$$

which ranges from 0 to 1, with HEAP as the midpoint at $\phi(1) = 1/2$. And because the total number of individuals and the accessible scales differ markedly at each site, we have taken a second log to allow comparisons across extremely different log-likelihood values on a single graph. Thus we have plotted $\log[-\log\{P(\mathbf{N}, \theta)\}]$. The negative sign is necessary because $P(\mathbf{N}, \theta)$ is less than 1, and thus $\log\{P(\mathbf{N}, \theta)\}$ is negative. Because the negative reverses the orientation of the graph, the maximum likelihood now occurs at the minimum of the graph.

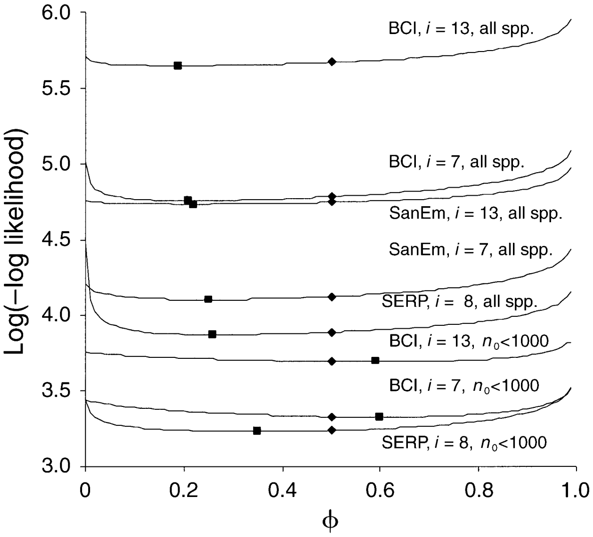

FIG. 14. Community likelihood estimates for a continuum of models (Eq. 43) as a function of the aggregation parameter ϕ (Eq. 47). The log of the negative of the log-likelihood function is plotted so that all comparisons can be shown on one graph. Sites, scales, and subsets of species are designated in the figure. Diamonds and squares mark the HEAP and optimum values of ϕ for each curve, respectively.

The plots in Fig. 14 indicate that between the lowest points on the curves (at a value of ϕ that yields maximum likelihood) and the HEAP point at $\phi = 1/2$, the function is quite flat, whereas it rises rapidly at the random extreme ($\phi = 0$) or the completely aggregated extreme ($\phi = 1$). We note that for the subcommunity of species in the serpentine and BCI communities with $n_0 < 1000$ (the majority of species), the optimum ϕ is closer to the value 1/2 than for the full community. For that subset of the BCI community, ϕ exceeds 1/2, indicating that the less abundant species are somewhat more aggregated than predicted by HEAP, whereas for the entire community the optimal value is less than 1/2, indicating a distribution somewhat more random than HEAP. Overall, the community likelihoods indicate that HEAP performs better than the random model and nearly as well as the model with the θ that maximizes the community likelihood at each site. The flatness of the curves in Fig. 14 between $\phi = 1/2$ and $\phi = \theta_{\max}/(1 + \theta_{\max})$ indicates that the small divergence from perfect prediction, while statistically significant, may not be ecologically significant.

DISCUSSION

Our theory is based on the hypothesis of equal allocation probabilities, which states that the allocation of the individuals that are found in an A_{i-1} cell to the two A_i cells that comprise it is governed by the simple rule that all distinct combinations (without internal permutations) of allocation are equally probable. Our "hypothesis of equal allocation probabilities" (HEAP) leads directly to Eq. 23, the master recursion relation that predicts the $P_i^{(n_0)}(n)$, and to predictions for the species–area relationship (SAR), the endemics–area relationship (EAR), and the community-level species-abundance distribution $\Phi_i(n)$; these predictions depend only on the species-abundance distribution at scale $i = 0$, $\Phi_0(n)$. The further assumption that HEAP applies independently to all the species in a community results in a prediction for the community-level spatial-abundance distribution functions $F_i(N)$.

There are no adjustable parameters in the theory and thus none of our comparisons between theory and observation involve fitting procedures. These comparisons, carried out over a wide range of scales with data from a serpentine meadow and two tropical forest sites, point to a general pattern of consistency between predictions and data, although significant deviations between the theoretical and empirical $P_i^{(n_0)}(n)$ are evident for species with relatively high n_0. The fact that for those high-abundance species the theory tends to overpredict the number of empty cells suggests that some degree of intraspecific density dependence needs to be included in the theory. Density-dependent intraspecific regulation would tend to spread species around on the landscape more uniformly, and thus if the theory contained a density correction to HEAP, the theory would better describe the spatial distributions of the high-abundance species.

Because the predicted species-level and community-level spatial-abundance distributions appear to fit the data best when the highest-abundance species are excluded, we also examined the predicted and observed SAR for the subset of species that excludes the same high-abundance species that were excluded in the comparisons in Fig. 13. For all three sites, the agreement between predicted and observed SAR was improved relative to the comparisons with all species included.

We emphasize that our theory is not based on any assumptions of self-similarity or fractality in nature. The species–area relationship predicted from HEAP may or may not exhibit power-law behavior, depending on the distribution of the species abundances. The predicted RARs *cannot* exhibit power-law behavior, however, because the derived α parameters *must* be scale dependent in our theory. The predicted scale dependence of the α's is such that they decrease with increasing i (decreasing spatial scale), which is both the way observed α's tend to behave and the way they must behave if Eq. 9 is to allow even the possibility of scale-independent values for the a_i's and thus a power-law SAR.

Although the probability parameters, $\alpha_i^{(n_0)}$, cannot be scale independent under HEAP, the conditional probability statement that defines the allocation rule is, in fact, scale independent. This can be most easily seen by observing that under HEAP the $\beta_i^{(n_0)}$ defined in Eq. 24 are independent of the scale index i (Eq. 26). We note that they are independent of n_0 as well. Numerical evaluation of the $\lambda_i^{(n_0)}$, and thus of the occupancy probabilities, for large i, large $n_{0,}$ and n_0 of order, or greater than, 2^i shows that occupancy probability exhibits a power-law dependence on patch area. In contrast to the assumption of species-level self-similarity in macroecology (Banavar et al. 1999, Harte et al. 2001), in which the RARs and thus the occupancy probabilities obey exact power-law behavior, HEAP is only asymptotically scale invariant.

Rationale for the assumption of HEAP

Because HEAP makes no explicit assumptions about biological mechanisms such as competition, dispersal, and density dependence, it can be considered as a kind of ‘‘null theory’’ of spatial structure. While such an assumption cannot possibly predict correctly all details of spatial pattern, to the extent that it leads to a wide range of testable and reliable predictions, understanding is advanced. Moreover, to the extent that there are patterns in the discrepancies between observation and prediction, a null theory provides an opportunity to evaluate the role of those behaviors that are assumed away at the outset.

We based our theory on HEAP in the same spirit that the ‘‘hypothesis of equal a priori probabilities’’ (HEAPP) is made in statistical mechanics and thermodynamics. HEAPP is the assumption that a many particle system such as a gas will ‘‘visit’’ all of the accessible states available to it in phase space with equal probability and, like HEAP, it fills the vacuum of evidence created by the absence of reasons to assume anything more complicated. Whereas HEAPP in classical statistical mechanics results in a binomial distribution for the numbers of distinguishable gas molecules in the two halves of a box, HEAP results in a flat distribution. This difference arises because under HEAP we effectively treat individuals as indistinguishable objects and therefore do not consider as independent allocation options the many permutations over individuals within a species.

The possibility of an eventual mechanistic explanation for the relative success of HEAP is suggested by the fact that the entire theory can be recast mathematically in terms of a simple functional form of an assembly (or colonization) function, $\beta_i(1|p, q - p)$. Direct tests of the derived functional form of the β's (Eq. 26) could help us understand the ecological interpretation of the β's and perhaps provide a more mechanistic foundation for HEAP. It is not clear to us whether the β's should be interpreted literally as descriptors of an assembly process, or whether they are simply mathematical devices that provide an alternative way to derive the consequences of HEAP, such as the form of the $P_i^{(n_0)}(n)$. This is an empirical question that can be answered with data from successional sites, where the sequence of individual appearances over time on a spatial grid can be determined.

Further tasks

Because the shape of the SAR depends only on the distribution of abundances, n_0, over the species pool, there could be one or more species-abundance distributions that actually result in an exact power-law SAR over a finite scale range. Determination of the analytic form of such species-abundance distributions (SADs) remains to be carried out.

The mismatch between our predicted $P_i^{(n_0)}(n)$ and the data is in the direction (see Figs. 5–7) that suggests that the theory would yield improved predictions if it incorporated some degree of density-dependent damping of high occupancy for the high-abundance species. This is consistent with our observation that the theory tends to underpredict species richness at intermediate scales. Stated differently, HEAP imposes slightly too much clustering, thus overpredicting the frequency of empty cells. Formulation of the theory in terms of the β functions (Appendix A) is a reasonable starting point for modifying HEAP for the high-abundance species by injecting intraspecific density dependence into the theory. Although many modifications of Eq. 26 that cause a dampening of the ‘‘rich get richer’’ effect at high occupancy are possible, a simple Monod-kinetics type approach would be a sensible first attempt.

The theory also leads to a recursion relationship for the dependence of species turnover on interpatch distance and on patch size given only the list of abundances of the species at the largest spatial scale (J. Harte, *unpublished manuscript*). As with our test of the community-level spatial distribution functions $F_i(N)$, in Fig. 13, tests of the predictions for species turnover will provide insight into how well the theory predicts entire landscape structure, as opposed to just single cell distributions. A further task is to solve these recursion relations and compare the HEAP prediction to observations.

Only plant data have been used for theory testing here. Further tests of HEAP for taxonomic groups of animals, such as census data on locations of breeding bird nest sites, or spatially explicit butterfly or mammal census data sets, would help us understand limitations on the scope of the theory. The availability of spatially explicit data sets in which all individuals are censused within a large contiguous area is unfortunately limited. Although we used three such data sets to test HEAP, most data sets consist of censuses within small scattered subplots or along sparsely located transects within a larger region. Another task, then, is to use HEAP to develop improved statistical techniques for extrapolating to landscape or biome scale the information about species abundance and species richness that derives from sparse censusing at local scale.

Conclusion

We have presented a theory of spatial structure in the distribution of individuals within species and species within ecological communities, and compared its predictions to spatially explicit vegetation census data from three sites. Assuming knowledge only of the species-abundance distribution at some largest scale, the theory predicts a diversity of measures of spatial structure in these data over a range of smaller spatial scales. These measures include, at species level, the spatial-abundance distributions and the range–area relationships, and at community level, the species–area and endemics–area relationships, a community-level spatial-abundance distribution, and the smaller scale species-abundance distributions. The spatial structure of three spatially explicit census data sets are generally consistent with the theory’s predictions for these measures, although a systematic pattern of failure for the high-abundance species is apparent.

Because this theory is based on hugely simplified assumptions and incorporates no explicit ecological processes, it will be of interest to develop a deeper understanding of why it makes successful predictions. At the same time, from the patterns in the failures of the predictions of this null theory, it appears possible to identify which of the many ecological processes that are neglected in its current formulation will need to be incorporated to improve the theory. Our analysis identifies the particular need to explicitly incorporate intraspecific carrying capacity constraints at high population density in an improved modification of the theory. A mathematically equivalent formulation of the theory in terms of an assembly (or colonization) function may provide the means with which to incorporate density dependence and develop a mechanistic foundation for the theory.

Acknowledgments

We thank the Class of 1935 Distinguished Professorship at UC Berkeley, the EPA STAR Fellowship Program, and the NSF Bioinformatics Postdoctoral Program for financial support, and Eric Berlow, Michael Bloxham, Richard Condit, Brian Enquist, Murray Gell-Mann, Nigel Goldenfeld, Ann Kinzig, William Kunin, and Joshua Plotkin for useful conversations. Two anonymous reviewers provided constructive suggestions for improving the manuscript. We are most grateful to Uthara Srinivasan for many insightful discussions.

Literature Cited

Banavar, J. R., J. L. Green, J. Harte, and A. Maritan. 1999. Finite size scaling in ecology. Physical Review Letters **83**: 4212–4214.

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

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

Chave, J. 2004. Neutral theory and community ecology. Ecology Letters **7**:241–253.

Coleman, B. 1981. On random placement and species–area relations. Journal of Mathematical Biosciences **54**:191–215.

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

Condit, R., et al. 2000. Spatial patterns in the distribution of tropical tree species. Science **288**:1414–1418.

Condit, R., S. P. Hubbell, and R. B. Foster. 1996*a*. Changes in tree species abundance in a neotropical forest: impact of climate change. Journal of Tropical Ecology **12**:231–256.

Condit, R., S. P. Hubbell, J. V. LaFrankie, R. Sukumar, N. Manokaran, R. Foster, and P. S. Ashton. 1996*b*. Species–area and species–individual relationships for tropical trees: a comparison of three 50-ha plots. Journal of Ecology **84**: 549–562.

Condit, R., et al. 2002. Beta-diversity in tropical forest trees. Science **295**:666–669.

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

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

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

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

Green, J., and A. Ostling. 2003. Endemics–area relationships: the influence of species dominance and spatial aggregation. Ecology **84**:3090–3097.

Harte, J., T. Blackburn, and A. Ostling. 2001. Self similarity and the relationship between abundance and range size. American Naturalist **157**:374–386.

Harte, J., A. Kinzig, and J. Green. 1999. Self similarity in the distribution and abundance of species. Science **284**: 334–336.

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

He, F., and K. Gaston. 2003. Occupancy, spatial variance, and the abundance of species. American Naturalist **162**: 366–375.

He, F., and S. Hubbell. 2003. Percolation theory for the distribution and abundance of species. Physical Review Letters **91**(19):198103.

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

Hubbell, S. 2001. The unified neutral theory of biodiversity and biogeography. Monographs in population biology 32. Princeton University Press, Princeton, New Jersey, USA.

Hubbell, S. P., and R. B. Foster. 1983. Diversity of canopy trees in a neotropical forest and implications for conservation. Pages 25–41 *in* S. L. Sutton, T. C. Whitmore, and A. C. Chadwick, editors. Tropical rain forest: ecology and management. Blackwell Scientific Publications, Oxford, UK.

Krebs, C. 1994. Ecology: the experimental analysis of distribution and abundance. Harper Collins, New York, New York, USA.

Kunin, W. E. 1998. Extrapolating species abundance across spatial scales. Science **281**:1513–1515.

Kunin, W. E., S. Hartley, and J. Lennon. 2000. Scaling down: on the challenge of estimating abundance from occurrence patterns. American Naturalist **156**:553–559.

Leibold, M. A., M. Holyoak, N. Mouquet, P. Amarasekare, J. M. Chase, M. F. Hoopes, R. D. Holt, J. B. Shurin, R. Law, D. Tilman, M. Loreau, and A. Gonzalez. 2004. The metacommunity concept: a framework for multi-scale community ecology. Ecology Letters **7**(7):601–613.

May, R. M. 1975. Patterns of species abundance and diversity. Pages 81–120 *in* M. L. Cody and J. M. Diamond, editors. Ecology and evolution of communities. Belknap Press, Cambridge, Massachusetts, USA.

May, M., J. Lawton, and N. Stork. 1995. Assessing extinction rates. Pages 1–24 *in* J. H. Lawton and R. M. May, editors. Extinction rates. Oxford University Press, Oxford, UK.

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

Ostling, A., J. Harte, and J. Green. 2000. Self-similarity and clustering in the spatial distribution of species. Technical comment. Science **290**:671a.

Ostling, A., J. Harte, J. Green, and A. Kinzig. 2003. A community-level fractal property produces power-law species–area relationships. Oikos **103**:218–224.

Ostling, A., J. Harte, J. Green, and A. Kinzig. 2004. Self similarity, the power-law form of the species–area relationship, and a probability rule: a reply to Maddux. American Naturalist **163**:627–633.

Pielou, E. C. 1969. An introduction to mathematical ecology. J. Wiley and Sons, New York, New York, USA.

Plotkin, J., and H. Muller-Landau. 2002. Sampling the species composition of a landscape. Ecology **83**:3344–3356.

Plotkin, J., M. Potts, N. Leslie, N. Manokaran, J. LaFrankie, and P. Ashton. 2000. Species-area curves, spatial aggregation, and habitat specialization in tropical forests. Journal of Theoretical Biology **207**:81–99.

Preston, F. 1948. The commonness, and rarity, of species. Ecology **84**:549–562.

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

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

Ruhla, C. 1992. The physics of chance. Oxford University Press, Oxford, UK.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry: the principles and practice of statistics in biological research. Third edition. W. H. Freeman, New York, New York, USA.

Whitmore, T., and J. Sayer. 1992. Tropical deforestation and species extinction. Chapman and Hall, London, UK.

Wright, D. 1991. Correlations between incidence and abundance are expected by chance. Journal of Biogeography **18**: 463–466.

APPENDIX

Derivation of the hypothesis of equal allocation probabilities (HEAP) from an assembly function is available in ESA's Electronic Data Archive: *Ecological Archives* M075-007-A1.