

Percolation Theory for the Distribution and Abundance of Species

Fangliang He^{1,*} and Stephen P. Hubbell²

¹Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia, V8Z 1M5, Canada

²Department of Plant Biology, University of Georgia, Athens, Georgia 30602, USA; Smithsonian Tropical Research Institute, Unit 0948, APO, AA 34002-0948, USA

(Received 5 June 2003; published 7 November 2003)

We develop and test new models that unify the mathematical relationships among the abundance of a species, the spatial dispersion of the species, the number of patches occupied by the species, the edge length of the occupied patches, and the scale on which the distribution of species is mapped. The models predict that species distributions will exhibit percolation critical thresholds, i.e., critical population abundances at which the fragmented patches (as measured by the number of patches and edge length) start to coalesce to form large patches.

DOI: 10.1103/PhysRevLett.91.198103

PACS numbers: 87.23.Cc, 89.75.Fb, 89.75.Kd

Understanding the relationship between the distribution and abundance of species is a central goal of ecology and biogeography [1,2]. To study this relationship, the distribution and abundance of a species are typically measured by gridding a study area into a lattice and then enumerating the species in each cell of the lattice (Fig. 1). The distribution of a species is commonly recorded by occupancy (the number of occupied cells), and the abundance by the total number of individuals in all occupied cells. Although occupancy is a useful measure of distribution, it fails to capture significant features of distribution [3]. Thus, two species having the same occupancy may nevertheless exhibit very different distribution patterns. In order adequately to describe the spatial structure of species distribution, it is also important to consider the number of patches that the occupied cells form as well as the total length of the perimeter (edge length) of all the patches. The subject of the relationship between occupancy and abundance is currently one of the most intensively investigated topics in ecology [3–5]. However, the relationships between perimeter and abundance and between the number of patches and abundance have not been studied. These latter relationships are of particular significance because they measure, to a great extent, the degree of fragmentation of landscapes. This study is to establish these two new relationships and to report their percolation phenomena.

The perimeter of a distribution is defined as the length of the joins between occupied and empty cells, and a patch is defined as a group of occupied cells which are connected side by side [Fig. 1(b)]. Here a join is a common edge between two cells. Given an occurrence map, the occupancy of a cell can be modeled as the result of a stochastic process according to the theory of percolation [6]. Let p denote the probability that a cell is occupied. One can create a statistical model of p as a function of three factors: the abundance of the species, its spatial distribution, and the cell size of the map. A function incorporating all three factors that has been in widespread use in ecology is [7,8]

$$p = 1 - \left(1 + \frac{aN}{Ak}\right)^{-k}, \quad (1)$$

where A is the total area of the map, a is the cell size, N is the total number of individuals of the species, and k is a clumping parameter describing the spatial dispersion pattern of the species in the area. Parameter k is defined in the domain of $(-\infty, -\mu)$ or $(0, +\infty)$, where $\mu = aN/A$ is the mean density of the species per cell. In Eq. (1), positive k describes aggregated dispersion patterns; negative k describes regular dispersion patterns of species [8]. It has been shown that the negative/positive binomial distribution on which Eq. (1) is based can be generated from a wide range of mechanisms including death and birth processes [9]. Note that the widely used occurrence probability, $p = 1 - (1 - \frac{a}{A})^N$, for a randomly placed species, is just a special case of Eq. (1), for $k = -N$ [10].

As a first approximation, we assume the occupancy of cells in a map are independent events. Under this assumption, the probability that, of two adjacent cells, one is occupied and one is not, is simply $p(1 - p)$. Then the

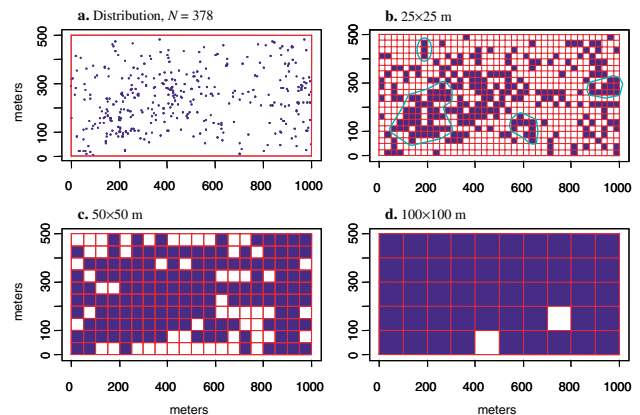


FIG. 1 (color online). Distribution of 378 plants of *Alibertia edulis* in a 50 ha (1000×500 m) plot on the Barro Colorado Island, Panama, and the occurrence maps at three mapping scales. The four circles shown in (b) illustrate the delineation of patches.

perimeter of a distribution, with the occurrence probability defined by Eq. (1), is given by

$$L = \sqrt{a}J \left[1 - \left(1 + \frac{aN}{Ak} \right)^{-k} \right] \left(1 + \frac{aN}{Ak} \right)^{-k}, \quad (2)$$

where J is the total number of neighboring joins for a map (regardless of the occupancy of the cells). For a regular lattice such as Fig. 1, $J = 2J_x J_y - J_x - J_y$, where J_x and J_y are the number of cells along the x and y axes, respectively. If cell occupancy is spatially correlated, Eq. (2) still holds, but the first term is no longer $\sqrt{a}J$. It is less than this value for aggregated cells and larger for regular cells. For random placement Eq. (2) becomes

$$L = \sqrt{a}J \left[1 - \left(1 - \frac{a}{A} \right)^N \right] \left(1 - \frac{a}{A} \right)^N.$$

Equation (2) shows that the perimeter L of a distribution depends on the abundance of the species N , the clumping parameter k , and the spatial scale a . The perimeter is a nonlinear function of the abundance. L first increases with N until reaching a peak at a critical abundance N_c . After that point, it decreases with N [Fig. 2(a)]. The critical abundance N_c is the so-called percolation

threshold, the point at which the coalescence of patches becomes sufficiently rapid to start causing a reduction in perimeter length. The value of N_c is easily obtained as

$$N_c = (2^{1/k} - 1) \frac{Ak}{a}. \quad (3)$$

The same percolation phenomenon is also observed for the relationship between L and a for fixed k and N [Fig. 2(b)]. In this case, the critical spatial scale a_c is the a value at which a phase transition in L occurs and a_c can be numerically solved for $(2a^2NkJ + aAN + A^2k) \times \{1 - [1 + (aN/Ak)]^{-k}\} - a^2NkJ = 0$.

Establishing a similar relationship for the number of patches (T) is a difficult task and no exact solution yet exists for the patch number. However, a good approximate solution given by [6] is

$$T \propto p^m (1-p)^j, \quad (4)$$

where m is the number of occupied cells for a patch, j is the number of occupied and empty joins for that patch.

Based on Eq. (4), we offer a very good approximation for the general scaling relationship between the number of patches and species abundance. Substituting Eq. (1) into (4), we have $T \propto \{1 - [1 + (aN/Ak)]^{-k}\}^m \times [1 + (aN/Ak)]^{-kj}$. If a or N is relatively small or A is large so that $aN/A < 1$, which is usually the case, the first bracketed term in this equation can be approximated by $(aN/A)^m$, whereas the second term can be approximated by $e^{-\beta N}$ where $\beta = ja/A$. These simplifications lead to a general scaling relationship

$$T \propto N^m e^{-\beta N}. \quad (5)$$

The number of patches T and abundance N , as given by Eq. (5), form a hump-shaped curve. At low abundance, the number of patches is low because of few individuals. At high abundance, the number of patches is also low because small patches coalesce into fewer, large patches. The number of patches reaches a maximum at an intermediate abundance that is the critical threshold N_c . From Eq. (5), the critical abundance is $N_c = m/\beta$, and the maximum number of patches at that threshold is $T_{\max} \propto (m/\beta e)^m$.

To test the theoretically predicted percolation relationships given by Eqs. (2) and (5), we analyzed distributions of populations of more than 300 species from a tropical rain forest on the Barro Colorado Island (BCI) of Panama. In 1981 a 50 ha (1000×500 m) forest plot on BCI was established. In the plot, all freestanding trees and shrubs ≥ 1 cm diameter at breast height have been enumerated, individually located on a reference map, and identified to species. The census was repeated in 1985, 1990, 1995, and 2000. The data from the 1990 census were used in this study where there are 229 048 plants belonging to 301 species with the most abundant species having 36 063 stems.

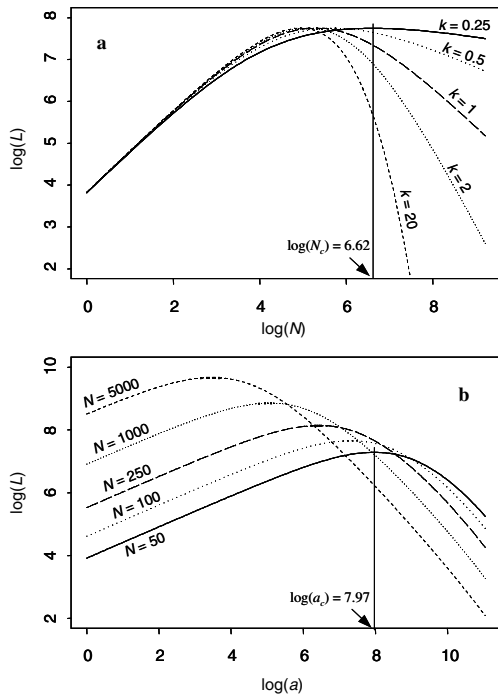


FIG. 2. (a) Perimeter (L)-abundance (N) relationships of Eq. (2) for spatial distribution patterns varying from highly aggregated to random patterns (k from 0.25 to 20) provided $A = 1000 \times 500$ m and $a = 50 \times 50$ m. Shown is also the critical abundance $\log(N_c) = 6.62$ for $k = 0.25$. (b) Perimeter (L)-scale (a) relationships for abundance $N = 50, 100, 250, 1000, 5000$ provided $A = 1000 \times 500$ m and $k = 0.75$. Also shown is the critical scale $\log(a_c) = 7.97$ for $N = 50$.

We converted the distribution of each of the 301 species into an occurrence map for a given scale a , for example, for *Alibertia edulis* (Fig. 1). For each map, we counted the number of trees in each cell and calculated the total length of edge (perimeter) and the number of patches. We then compared the observed perimeter and the number of patches for each species at each scale to the predictions of Eqs. (2) and (5). We made the computation on seven spatial scales: $a = 5 \times 5, 10 \times 10, 20 \times 20, 25 \times 25, 50 \times 50, 100 \times 100$, and 250×250 m.

We estimated the aggregation parameter k in Eq. (2) from the cell counting data according to $k = \mu^2 / (\sigma^2 - \mu)$, where μ is the average number of stems of a given species per cell, and σ^2 is the variance. Because k is scale dependent, it can be estimated either for each scale used or according to a scaling function across scales. The later approach is more appropriate in situations in which quadrat sampling across scales is prohibitively expensive to conduct. We find that an empirical scaling function, $k \propto a^{0.55}$, extrapolates k very precisely across scales in the BCI plot. Based on this scaling function, we can estimate k at any other scale a if k_0 at a base scale a_0 is known: $k = k_0(a/a_0)^{0.55}$. This scaling function works well as compared to the scale-specific estimates of k for

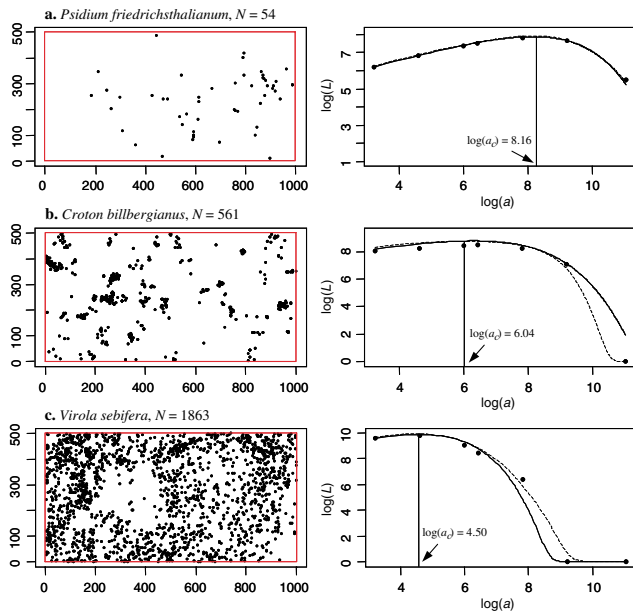


FIG. 3 (color online). Perimeter (L)-scale (a) relationships for three BCI species across scales from 5×5 to 250×250 m. The left-hand column shows the actual spatial distributions of the species. The observed perimeter data are shown on the right-hand column (dots) and the smooth solid curves are the predictions of Eq. (2) using a scaling estimation function for k across scale. For comparison, the curves using k estimated from each individual scale are also shown (the dashed curves). In this case k is estimated using the tree data counted at each scale and a curve is drawn by substituting these estimated k values into Eq. (2).

the BCI species, as shown in Fig. 3, in which $a_0 = 20 \times 20$ m was used as the base scale. However, other base scales give very similar results.

We illustrate the results for perimeter-scale relationships in Fig. 3 for three species chosen to represent the typical abundance range and distribution patterns in the BCI. For many species, the prediction of the theory is very good. However, for some highly aggregated species such as *Croton billbergianus*, theoretical predictions show a modest departure from observations at certain scales [Fig. 3(b)]. This departure is principally caused by violation of the independence assumption in cell occupancy for Eq. (2). Nevertheless, the fact that even for those aggregated species, Eq. (2) still predicts well the overall trajectory of the perimeter-scale relationship suggests that the independence assumption is quite robust for many species. Another reason for the good prediction is probably due to the fact that the BCI species indeed can be described well by the negative or the positive binomial distribution. The majority of the BCI species (259 out of 301) show no significant departure from the binomial distribution ($p > 0.05$, χ^2 test at the scale of 25×25 m).

Unlike the perimeter-scale relationship, conspecific data are not available to show the perimeter-abundance relationship for single tropical tree species. Such a relationship can be examined only by using interspecific data. As shown in Fig. 4, the observed perimeter-abundance relationships for the BCI species are well predicted by the percolation Eq. (2).

The relationship between the number of patches (T) and abundance (N) is shown in Fig. 5 for the BCI species. The critical abundance (N_c) decreases with an increase in scale a , forming a nearly perfect and identical power law

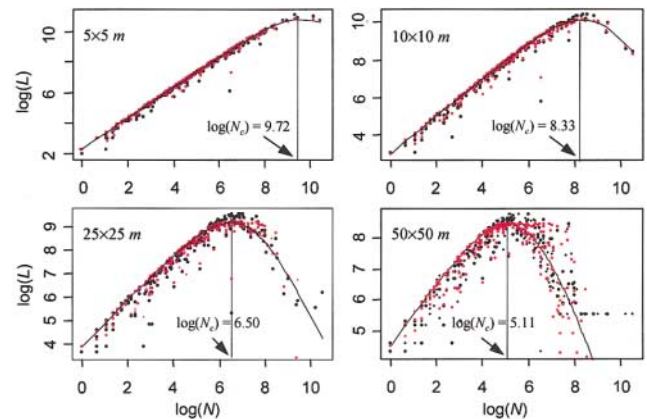


FIG. 4 (color). Perimeter (L)-abundance (N) relationships for the BCI species at four scales: 5×5 , 10×10 , 25×25 , and 50×50 m. Each black dot represents the observed perimeter for a species, the red dots are the predictions from Eq. (2) for each species using k estimated for the species at the given scale. The smooth curves are predicted from Eq. (2) using a universal $k = 2$ for all the species. The critical abundances N_c are calculated from Eq. (3) at $k = 2$ and $A = 500000$.

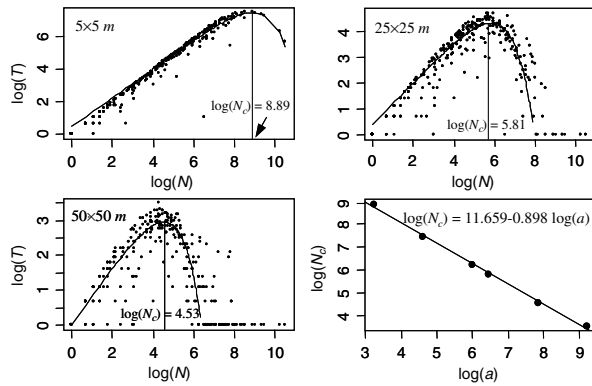


FIG. 5. Number of patches (T)-abundance (N) relationships for the BCI species at three scales: 5×5 , 25×25 , and 50×50 m. Each dot represents the observed number of patches for a species. The smooth curves are the scaling Eq. (5) fitted to the data. The last panel shows the power law ($r^2 = 0.997$ for the log-log linear regression) between the critical abundance N_c and spatial scale a for six scales from 5×5 to 100×100 m.

(Fig. 5, last panel). The nearly perfect identity of this power law suggests that this scaling relationship will be more generally true of tree communities, and perhaps in many other plant communities as well.

The ramifications of the discovery of the new theoretical relationships between distribution and abundance of species are potentially numerous and profound. For the first time, we have unified several fundamental ecological parameters including abundance (N), distribution pattern (k), scale (a), extent (A), edge length (L), and the number of patches (T) into a single mathematical framework for quantifying the spatial architecture of fragmented metapopulations. This unification is of important significance to ecology, biogeography, and conservation biology. First of all, this theoretical synthesis provides a far more rigorous and standardized mathematical foundation for analyzing and comparing the complex geometry of natural populations in ecology than the measurement of occupancy currently commonly used in ecology [3–5]. Second, this study connects fundamental landscape metrics (edge length and number of patches) to the spatial distribution of species abundance, thus building a bridge between biogeography and landscape ecology. Third, landscape connectivity is widely considered to be critical for the persistence of populations and maintenance of species diversity in ecosystems [11–16], the percolation theory developed here provides new tools for calculating fragmentation thresholds in population viability analysis.

In particular, the perimeter and the number of patches models would allow us to determine the level of landscape fragmentation (e.g., forest cutting [17]) so that to minimize the edge length and the number of patches in practice. It will be especially useful if these analyses are coupled with other information such as habitat quality, behavior, dispersal capability and life history properties of species.

We sincerely thank P. Ashton, R. Condit, R. Foster, E. Losos, and N. Manokaran for their contributions to the BCI and Pasoh forest dynamics plots. We also thank the NSF (U.S.), the John D. and Catherine T. MacArthur Foundation, the A. K. Mellon Foundation, the Pew Charitable Trusts, and The John Simon Guggenheim Foundation for support.

*Current address: Department of Renewable Resources, University of Alberta, Edmonton, Alberta, T6G 2H1, Canada.

- [1] R. H. MacArthur and E. O. Wilson, *The Theory of Island Biogeography* (Princeton University Press, Princeton, 1967).
- [2] S. P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton University Press, Princeton, 2001).
- [3] K. J. Gaston, *Rarity* (Chapman & Hall, London, 1994).
- [4] J. H. Brown, *Am. Nat.* **124**, 255 (1984).
- [5] J. Harte, T. Blackburn, and A. Ostling, *Am. Nat.* **157**, 374 (2001).
- [6] D. Stauffer and A. Aharon, *Introduction to Percolation Theory* (Taylor & Francis, London, 1994).
- [7] D. H. Wright, *J. Biogeogr.* **18**, 463 (1991).
- [8] F. He and K. J. Gaston, *Am. Nat.* **156**, 553 (2000).
- [9] M. T. Boswell and G. P. Patil, *Random Counts in Models and Structures*, edited by G. P. Patil (Pennsylvania State University Press, University Park, 1970), pp. 3–22.
- [10] D. B. Coleman, *Math. Biosci.* **54**, 191 (1981).
- [11] R. H. Gardner, B. T. Milne, M. G. Turner, and R. V. O'Neill, *Landscape Ecol.* **1**, 19 (1987).
- [12] K. A. With and T. O. Crist, *Ecology* **76**, 2446 (1995).
- [13] J. A. Wiens, R. L. Schooley, and R. D. Weeks, Jr., *Oikos* **78**, 257 (1997).
- [14] L. Tischendorf and L. Fahrig, *Oikos* **90**, 7 (2000).
- [15] F. K. A. Schmiegelow and M. Mönkkönen, *Ecol. Appl.* **12**, 375 (2002).
- [16] A. Gonzalez, J. H. Lawton, F. S. Gilbert, T. M. Blackburn, and L. Evans-Freke, *Science* **281**, 2045 (1998).
- [17] J. F. Franklin and R. T. T. Forman, *Landscape Ecol.* **1**, 5 (1987).