



Species-area Curves, Spatial Aggregation, and Habitat Specialization in Tropical Forests

JOSHUA B. PLOTKIN*[†], MATTHEW D. POTTS[‡], NANDI LESLIE[§], N. MANOKARAN^{||},
JAMES LAFRANKIE[¶] AND PETER S. ASHTON**

**Institute for Advanced Study and Princeton University, Olden Lane, Princeton, NJ 08540, U.S.A.*

[‡]*Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, U.S.A.,*

[§]*Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08540, U.S.A.*

^{||}*Forest Research Institute of Malaysia, Kepong, Malaysia,* [¶]*Center for Tropical Forest Science, National Institute of Education, 1025 Singapore and* ^{**}*Department of Organismal and Evolutionary Biology, Harvard University, Cambridge, MA 02138, U.S.A.*

(Received on 26 April 2000, Accepted in revised form on 26 July 2000)

The relationship between species diversity and sampled area is fundamental to ecology. Traditionally, theories of the species–area relationship have been dominated by random-placement models. Such models were used to formulate the canonical theory of species–area curves and species abundances. In this paper, however, armed with a detailed data set from a moist tropical forest, we investigate the validity of random placement and suggest improved models based upon spatial aggregation. By accounting for intraspecific, small-scale aggregation, we develop a cluster model which reproduces empirical species–area curves with high fidelity. We find that inter-specific aggregation patterns, on the other hand, do not affect the species–area curves significantly. We demonstrate that the tendency for a tree species to aggregate, as well as its average clump size, is not significantly correlated with the species' abundance. In addition, we investigate hierarchical clumping and the extent to which aggregation is driven by topography. We conclude that small-scale phenomena such as dispersal and gap recruitment determine individual tree placement more than adaptation to larger-scale topography.

© 2000 Academic Press

Introduction

Faced with an ecological assemblage, a natural question is “How many species are found in a given area?” This question has puzzled naturalists since the early 19th century (Connor & McCoy, 1979). Aside from an intriguing question, the species–area relationship is a conceptual cornerstone for almost all theories of community ecology (McGuinness, 1984; Rosenzweig, 1995). Referring to their seminal work on biogeography,

MacArthur and Wilson write, “Theories, like islands, are often reached by stepping stones. The ‘species–area’ curves are such stepping stones” (MacArthur & Wilson, 1967). Beyond their theoretical significance, species–area curves also form the basis for almost all estimates of extinction due to habitat loss (May, 1995; Pimm & Raven, 2000). More generally, the species–area relationship helps ecologists to assess the relative importance of those factors—competition, dispersion, adaptation to environment, chance, etc.— which determine species' geographic ranges.

[†]Author to whom correspondence should be addressed.
E-mail: plotkin@ias.edu

The most common characterization of the species–area relationship (SAR) posits a power law: the number of species is proportional to a constant power of sampled area. This relationship, $S = cA^z$, was first postulated by Arrhenius (1921). Nevertheless, this model did not enjoy universal popularity until the pioneering work of Preston (1962) followed by MacArthur & Wilson (1963). MacArthur and Wilson considered the power law in light of a dynamic equilibrium of species exchanges between islands.

In 1975, May endowed the power-law model of the SAR with a firm theoretical underpinning by relating it to the relative abundances of species (May, 1975). May argued that in a complex community the relative abundances of species should follow the lognormal distribution: the proportion of species with n individuals is a Gaussian function of $\log n$. Assuming this abundance distribution, May showed that the power-law SAR must hold over a wide range of spatial scales. In order to perform this derivation, May assumed that individuals are drawn *independently* from the species abundance distribution, yielding a species–individual curve. In order to translate the species–individual curve into the species–area curve, May assumed that individuals are placed in space with a constant density. The entire process is equivalent to drawing individuals from the abundance distribution and placing them in space (Poisson-) randomly.

Following May's approach, Coleman developed a random-placement null model of the SAR (Coleman, 1981; Coleman *et al.*, 1982). Coleman provided explicit equations for the species–area curve which results from given species abundances. Specifically, given N species with abundances $n_1, n_2, n_3, \dots, n_N$, Coleman derived equations for the species–area curve assuming that individuals are placed in space randomly and independently. As Coleman states, the randomness assumption “presupposes a lack of correlation in the locations of individuals. [As such,] it can be considered a zeroth-order hypothesis”. In other words, Coleman's model ignores the possibility of inter- and intra-specific spatial aggregation. Although the random-placement hypothesis is violated in many ecological systems, few data sets were available in 1981 by which to form or judge more explicit models. (Coleman

did employ avifaunal data from Pymatuning Lake, Pennsylvania for comparison to his theory. Nevertheless, the region contained fewer than 40 identified species.)

Since the work of Coleman and May, there have been relatively few explicit efforts to extend the “zeroth-order hypothesis” of a random placement SAR. In fact, current theories of biogeography and diversity often still utilize Coleman's model (e.g. Hubbell, 1997). The prevailing wisdom generally holds that the species–area curve is driven by random sampling from the abundance distribution—although this belief is not held without exceptions (Hubbell & Foster, 1983). The explicit, theoretical efforts towards a non-random theory include work by Leitner & Rosenzweig (1997), Buckley (1982), McGuinness (1984), and Gotelli & Graves (1996). Recently, Kunin (1998) and Ney-Nifle & Mangel (1999) have developed spatial models based upon the geographic range of each species. In general, these spatial-refinements upon Coleman's model have seldom been compared to extensive, empirical data.

With less of an emphasis on a spatial model for the species–area curve, previous research has analysed spatial patterning in tropical forests. He *et al.* (1987) analyse conspecific patterning by using nearest-neighbor statistics and the Donnelly clumping index. They also investigate the interplay of aggregation and topography. A comparison of our results with theirs will prove interesting; their analysis does not quantify topography as a hierarchical aggregation effect. Batista & Maguire (1998) provide a comparative overview of stochastic point processes and their application to tropical forests. They focus on the effect of the forest canopy on its understory. Condit *et al.* (2000) offer an extensive, comparative study of aggregation across tropical forest plots. In a seminal, qualitative investigation, Hubbell & Foster (1983) provide a largely biological discussion of spatial patterning in the canopy of a tropical forest, focusing on the maintenance of diversity. Most of their analysis relies on visual characterization of spatial patterns. Their discussion of the biological and ecological factors driving aggregation provides a uniquely well-informed interpretation of the more quantitative results in this paper.

This paper aims to develop a spatially explicit, theoretical framework for the species–area curve in a tropical forest. By doing so, we will assess the relative importance of small-scale (e.g. dispersal, gap-recruitment) vs. large-scale (e.g. topographic) phenomena as determinants of tree-placement. In the spirit of Coleman and May, we will assume throughout that the abundance of each species is known. Given these abundances, we will first investigate the impact of local aggregation on the species–area curve. Then, we will investigate hierarchical clumping and larger-scale effects of environment. We develop our models of the SAR hand in hand with comparisons to extensive data.

Tropical Forest Data

There could hardly be an ecological system more well-suited for species–area and spatial-aggregation research than tropical forests. Unlike avifauna, trees have the obvious advantage of a sedentary life history—which makes them relatively easy to locate and identify. Moreover, tropical forests boast an astounding diversity of tree species (up to 1000 within 50 h), providing excellent resolution for our analyses. Finally, due to the impressive efforts of the Smithsonian Center for Tropical Forest Sciences, we have currently identified the location and species for a sum-total of over two million individual trees.

In this paper, we investigate spatial aggregation and the SAR in a 50-h, fully censused tree plot from the Pasoh forest on peninsular Malaysia. Compared with several other 50-ha censuses, Pasoh has the advantage of a fairly homogenous environment and a relatively rich species diversity. These qualities make Pasoh an excellent choice as the focus of our study. In addition, throughout the paper we also refer to two other 50-ha plots as verification of the generality of our methods. Each of the 50-ha Forest Dynamics Plots is part of a long-term research program coordinated by the Center for Tropical Forest Science. The plots are located in the following forests: Pasoh Forest Reserve, Peninsular Malaysia, 1996 census; Huai Kha Khaeng Wildlife Sanctuary (HKK), Thailand, 1995 census; Lambir Hills National Park, Sarawak, Malaysia, 1999 census. In every plot, each woody stem > 1 cm diameter has been identified to species,

TABLE 1
Tree density and diversity at each of three 50-h plots of tropical forest.

Plot name	Location	Stems	Species
Pasoh	Malaysia	320 902	817
Lambir	Malaysia	325 335	1171
HKK	Thailand	96 072	251

We develop and test our spatial model at Pasoh, and then we verify the results at Lambir and HKK. The table indicates the total number of woody stems > 1 cm in diameter and the total number of species found in each plot. For a complete list of references, consult the CTFS web site at <http://www.si.edu/ctfs>.

measured for girth, and spatially mapped to at least 1 m. The number of such stems, and the number of species among them, varies greatly from plot to plot (Table 1).

We will include all free-standing stems > 1 cm diameter throughout our analyses. In general, aggregation patterns benefit from a separate large-tree/small-tree treatment—especially for comparison with the theories of Janzen (1970) and Connell (1971). Nevertheless, for the purpose of investigating the species–area curve, it is best to include every individual in the data set. Almost all of our results remain true if we analyse, instead, stems > 5 cm in diameter. When summarizing results, we will often classify species as either rare or abundant. Following Hubbell & Foster (1986), we define a species as rare if it has one or fewer stems per hectare, on average.

The Random-placement Model

We begin with a brief review of Coleman’s “zeroth-order”, random-placement theory of species–area curves. Consider a region of total area A_0 within which individuals of various species are located. Assume that there are N species and that the i -th species is represented by n_i individuals. Consider any sub-region of area $A < A_0$. Under the assumption of independent, random placement of individuals, the probability that a given member of the i -th species does not reside in a sub-region of size A is simply $(1 - A/A_0)$. Similarly, the probability that all members of species i lie outside of A is given by $(1 - A/A_0)^{n_i}$.

Thus, the probability that at least one member of the i th species resides in the sub-region A is $1 - (1 - A/A_0)^{n_i}$. This in turn yields an expression for the mean number of species in A , denoted $S(A)$, and the variance, denoted $\sigma^2(A)$ (Coleman, 1981):

$$S(A) = N - \sum_{i=1}^N (1 - A/A_0)^{n_i}, \quad (1)$$

$$\sigma^2(A) = \sum_{i=1}^N (1 - A/A_0)^{n_i} - \sum_{i=1}^N (1 - A/A_0)^{2n_i}. \quad (2)$$

It is often assumed that, given the abundances of the species, eqns (1) and (2) provide an adequate model of the SAR in tropical forests (e.g. Hubbell, 1997).

Fortunately, it is easy to test the adequacy of eqns (1) and (2) for tropical forests. Early tests by Hubbell at Barro Colorado Island (BCI) indicate a failure of the random model for 186 species in the canopy > 20 cm in diameter (Hubbell & Foster, 1983). Our independent test of the random-model will include all trees > 1 cm in diameter. In the interest of precision we must define the species–area curve carefully. Following Harte *et al.* (1999) and Plotkin *et al.* (2000), we define the species–area curve for a rectangular region of total area A_0 . Let $A_i = A_0/2^i$ denote the area of a rectangular patch obtained after i bisections of A_0 (bisections chosen perpendicular to the longer dimension). There are 2^i disjoint patches of area A_i which naturally partition A_0 . Loosely speaking, the empirical species–area function $S(A)$ is defined as the average number of species found in an area of size A . By longstanding convention, we evaluate $S(A)$ by averaging over disjoint patches of area A (Connor & McCoy, 1979). In this manner, we let $S(A_i)$ denote the average number of species found in a patch of area A_i . In other words, we define

$$S(A_i) = \frac{1}{2^i} \sum_{j=1}^{j=2^i} (\# \text{ species in the } j\text{-th patch of area } A_i). \quad (3)$$

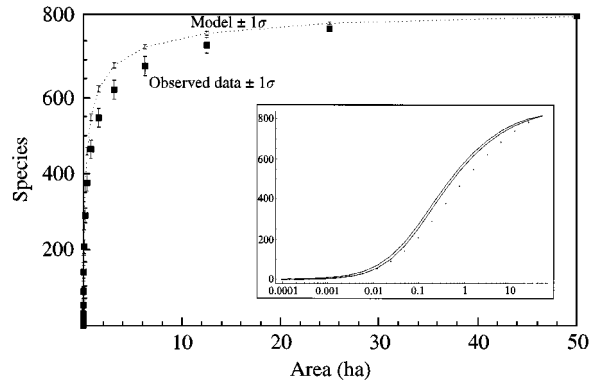


FIG. 1. The random placement model significantly overestimates diversity for areas in the 0.04- to 45-ha range. The figure shows graphs of the actual SAR measured at Pasoh (± 1 S.D.) compared to the SAR predicted by the random-placement model (± 1 S.D.). The inset repeats the graphs on log-linear axes; the solid lines demarcate the ± 1 S.D.-confidence interval predicted by Coleman's model. As is clear from the graphs, the observed SAR is far outside of the random-placement prediction. The same result holds at HKK and Lambir (not shown).

In particular, $S(A_0)$ denotes the total number of species found in the entire plot. We define M as the total number of individuals in the plot, i.e. $M = \sum_{i=1}^N n_i$.

The SAR predicted by random placement significantly overestimates diversity at Pasoh (Fig. 1). In fact, for all three forests the measured SAR is well outside of the two-standard-deviation confidence interval of the random model. This confidence interval is given by 2σ from Eq. (2). The discrepancy between the random model and the empirical data is even larger than the discrepancy originally measured for large trees alone at BCI (Hubbell & Foster, 1983). A cursory glance at Fig. 1 may suggest that Coleman's model is inaccurate for mid-sized areas, but very accurate near 0 and 50-ha. This is an artifact. By construction, Coleman's model must agree with the actual data at 0 and 50 ha.

The extent to which random placement fails to capture the SAR is, in fact, quite severe. Coleman's model significantly overestimates diversity for all areas within the range 0.04–45 ha (see Fig. 1, inset). For example, at Pasoh the measured value of $S(A_4) = S(3.1 \text{ ha})$ is more than nine standard deviations (S.D.) less than the predicted value. On average over areas ranging from 1.5 to 25 ha,

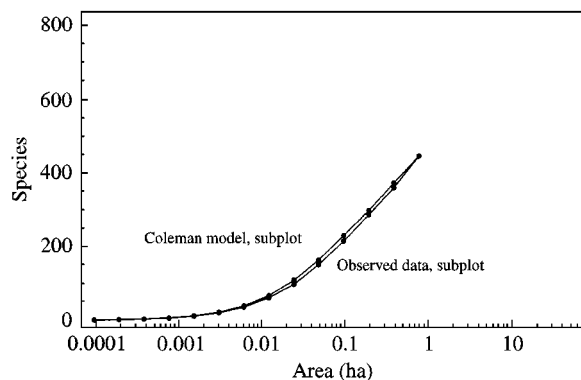


FIG. 2. The observed SAR of Pasoh compared to the SAR predicted by Coleman's random-placement model, on a log-linear graph. Unlike 50-ha curves in Fig. 1, the curves displayed here are generated from a single subplot of area $A_6 \approx 0.78$ ha. At this small-scale, the random-placement prediction is much closer to the observed SAR than at the 50-ha scale.

the random-placement model overestimates diversity 7.2 S.D.s at Pasoh, 4.0 S.D.s at HKK, and 17.2 S.D.s at Lambir. As these results indicate, beyond a doubt the random-placement model fails to account for the observed SAR. The actual data must therefore follow some non-random, spatial patterning.

Figure 1 also reveals another signature of spatial aggregation. For each i , the diversity found in patches of area A_i shows larger variance in reality than in Coleman's model. This large variance suggests spatial aggregation: some patches have large diversity, while others are dominated by a few, clumped species. Observed variance is much larger than Coleman's model in all three plots.

The departure from random placement is apparent to us because of the large scale at which the plots have been censused. We test this by temporarily restricting our attention to a *single* subplot of small area. In particular, we choose a single subplot from Pasoh of area $A_6 \approx 0.78$ ha, measure the abundances within the subplot, and again compare Coleman's model to the observed SAR. When restricting our attention to this subplot, the random-placement model provides an adequate characterization of the SAR (Fig. 2). At small spatial scales, the entire subplot is smaller than the correlation length of spatial clumps; hence random placement describes the spatial

pattern fairly well. This trend—that the random-placement model can be rejected for large plots, but not for small plots—suggests that random placement will provide an increasingly poor model for plots even larger than 50 ha.

Taken together, the results in Figs 1 and 2 underscore the need for a decidedly non-random model of tropical forests, especially for large areas. Throughout the sequel, we aim (i) to analyse the spatial aggregation in the forest plots, (ii) to develop a model which characterizes this aggregation, producing the observed species-area curves, and (iii) to investigate the relative strength of topographic effects on aggregation.

A Preliminary Measure of Aggregation: proportion of Neighbors which are Conspecific

We begin our analysis of the aggregation patterns with a somewhat non-standard approach. A more traditional analysis—which eventually yields a spatial model—will be delayed until the next section. Nevertheless, the statistic which we investigate in this section will later provide important information for determining the parameters of our spatial model.

Most univariate measures of aggregation essentially ask the question “How far apart are two conspecific trees?” Nevertheless, in this preliminary section we will ask the inverse question: “Given two trees a distance d apart, how often are they conspecific?” This question yields an interesting measure of aggregation. We will calculate this aggregation statistic for each plot and compare it to the value predicted under a random-placement model. Specifically, for each distance d we define P_d as the proportion of trees in the plot, distance d apart, which are the same species. We calculate this proportion by considering each pair of trees in turn. Notice that P_d is inherently inter-specific: it requires knowledge of tree locations for all species.

In a random-placement model, P_d clearly does not depend on the distance d . P_d is determined by the relative abundances of the species alone. If individuals are located randomly and independently, then P_d is simply the chance that two randomly-chosen individuals are of the same species. This probability does not depend upon distance, and it is easily expressed as a quantity

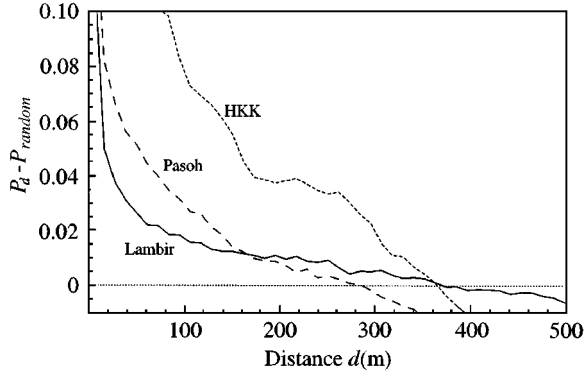


FIG. 3. Graph of the aggregation statistic $P_d - P_{random}$ for each of the forests Pasoh, HKK, and Lambir. All three forests possess an elevated proportion of conspecifics at distances $d < 200$ m. The x -intercept of the graph $P_d - P_{random}$ yields an upper bound on mean dispersal distance and gap-size. HKK demonstrates the most aggregation at large distances ($d < 385$ m).

closely related to the Simpson index:

$$P_{random} = \sum_{i=1}^N \frac{n_i(n_i - 1)}{M(M - 1)}. \quad (4)$$

As indicated by Monte-Carlo simulations of random placement, the S.D. of P_{random} is small (< 0.005) in all three plots. A measured value of P_d greater than P_{random} indicates a higher proportion of conspecifics distance d apart relative to the random model. Trees generally propagate locally, and hence we expect that $P_d > P_{random}$ for small distances, but $P_d \leq P_{random}$ for large distances. This trend is verified in all three forests by graphing $P_d - P_{random}$ (Fig. 3). (Note that the value of P_{random} is different in each forest because the relative abundances differ.)

Figure 3 illustrates the aggregation metric $P_d - P_{random}$ in the three forest plots. For example, Fig. 3 reveals that on average, when standing at a given tree in Pasoh, one finds an elevated proportion of conspecific trees at distances $d < 250$ m away, but thereafter one finds no more conspecifics than expected at random, given the relative abundances of the species in Pasoh. This observation provides an upper-bound on those factors which determine correlation length for small-scale cluster formation—e.g. dispersal distance, gap size, etc. In particular, at Pasoh we do not expect significant clumping to occur at dis-

tances greater than 250 m. On the other hand, clumping seems to be more widespread at HKK and Lambir. These preliminary observations will prove useful in later sections.

5 The Ripley-K Measure of Aggregation

Unlike the statistic $P_d - P_{random}$, which requires information about the location of all individuals in the plot, we henceforth focus on strictly intra-specific aggregation measures. In particular, we will employ the well-understood, second-moment measure called Ripley's K (Ripley, 1976). As with the statistic $P_d - P_{random}$, we will compare the observed values of K to those predicted under the assumption of random placement. Eventually, we aim to parameterize our spatial model using the information gleaned from Ripley's K —which is the primary reason why we choose to measure K . In the meantime, we will use Ripley's K to assess whether each species is significantly clumped or not.

COMPUTING RIPLEY'S K

If the individuals of a given species are placed randomly in the plot—i.e. via a Poisson-process with intensity λ —then the expected number of stems within a circle of radius d is simply $\lambda\pi d^2$. Ripley's K quantifies the departure from the randomized situation. Clustering increases K , while regularity decreases it. Specifically, given a point-process on the plane, Ripley's K -function is defined as

$$K(d) = \lambda^{-1} \mathbb{E}(\text{number of extra events within a distance } d \text{ from an arbitrary event}). \quad (5)$$

Given a particular map of n events s_1, s_2, \dots, s_n within a region of area A_0 (in our case, the locations of stems of a fixed species), the canonical edge-corrected estimator of K is given by Ripley (1976):

$$\hat{K}(d) = \hat{\lambda}^{-1} \sum_{i=1}^n \sum_{j=1, j \neq i}^n w(s_i, s_j)^{-1} \mathbb{I}(\|s_i - s_j\| \leq d) / n. \quad (6)$$

In eqn (6), $\hat{\lambda} = n/A_0$ estimates the intensity, Π is the indicator function of an event, and the weight $w(s_i, s_j)$ is the proportion of the circumference of the circle centered at s_i , passing through s_j , which lies in A_0 . Given λ , if the underlying point-process is stationary and isotropic, then \hat{K} is an unbiased estimator of K (Cressie, 1991). $\hat{\lambda}$ is a biased estimate of λ , but only slightly so. Assuming random placement, $K(d) = \pi d^2$.

We have computed Ripley's K using eqn (6) for every species in Pasoh, HKK, and Lambir. More specifically, we calculate $\hat{K}(d)$ for 20 values of d equally spaced between 0 and d_{max} . The quantity d_{max} signifies the largest distance at which we measure clumping in the plot. We use Fig. 3 as a guideline for choosing d_{max} in each plot: we use $d_{max} = 250$ m at Pasoh, $d_{max} = 385$ m at HKK and Lambir. As an example, at Pasoh we have calculated $\hat{K}(d)$ for $d = 12.5, 25, 37.5, \dots, 250$ m. In order to test spatial randomness, we compile the information about $\hat{K}(d)$ into a single Cramer-von Mises-type statistic

$$k = \int_0^{\infty} (\sqrt{\hat{K}(h)} - h\sqrt{\pi})^2 dh. \quad (7)$$

In practice, we evaluate this statistic using a Riemann sum from 0 to d_{max} . In order to test random placement, for each species we compare the measured k -statistic to the maximum k -statistic generated from 19 Monte-Carlo simulations of a Poisson process (i.e. of random placement). If the measured k is larger than the maximum simulated k , denoted k_{max} , then we conclude at the 5% confidence level that the species is not distributed at random (Diggle, 1983).

In principle, this procedure allows us to classify each species as either clumped or not clumped. Of course, we only consider species with ≥ 2 stems. "Not-clumped" is the null hypothesis; failure to reject the hypothesis should not compel us to accept it. With this warning in mind, we classify a species as not clumped meaning, simply, that it was not possible to reject this hypothesis on the basis of the k -statistic. When a species is rare ($n_i \leq 50$), then the power of the k -statistic—or any other statistic—is dramatically diminished. We will take special care when interpreting the results of our statistics on rare species.

RESULTS OF RIPLEY-K ANALYSIS

Of the 798 species at Pasoh with at least two stems, 661 of them (83%) are classified as aggregated via k . Of these aggregated species, 513 of them (78%) have more than 50 stems. On the other hand, out of the 137 non-aggregated species, most of them (85%) have 50 stems or less. In other words, fewer of the rare species are classified as clumped, relative to the abundant species. This trend, which He *et al.* (1987) noticed as well, is sometimes overemphasized in the literature. The trend certainly results as an artifact of the statistics on small sample sizes. In general, when $n_i < 50$ at least one Monte-Carlo simulation receives a very large k -statistic, making it difficult to reject the null-hypothesis of random-placement. In other words, when a species is rare, for statistical reasons alone we often cannot justify labeling it as clumped.

Ignoring the rare species for the moment, of the 534 species with > 50 stems, 513 of them (96%) are aggregated according to k . Species at HKK and Lambir demonstrate a comparable tendency to aggregate. In short, almost all species which are not rare are clumped. These results contrast sharply with analyses of a temperate forest (Szwagrzyk & Ptak, 1981; Bodziarczyk *et al.*, 1999; Bodziarczyk & Szwagrzyk, 1996), but agree with recent, comprehensive, inter-plot analyses of tropical forests by Condit *et al.* (2000).

We may also use the k -value to rank species by their tendency to aggregate (cf. Condit *et al.*, 2000). Both a larger number of clumps and a tighter average clump-size increase the k -statistic. Hence, the k -statistic provides a first-cut, agglomerate measure of overall aggregation. To be precise, the value of k itself means little without comparison to the distribution of k -statistics generated by many simulations of random placement. Given an observed k -value, $k_{observed}$, if a very large number of random simulations were feasible then the quantile in which $k_{observed}$ lies would yield a good index of aggregation. As a surrogate to this computationally intensive index, we use 19 simulations and the value of $k - k_{max}$ in order to rank the species by aggregation. For each species, the value $k - k_{max}$ indicates the extent to which the species is more (or less) aggregated than the most aggregated-looking random-placement simulation.

Figure 4 illustrates species at Pasoh which range from the most-clumped towards the least-clumped according to their $k - k_{max}$ ranking. The rankings seem to agree with an intuitive assessment of aggregation. For example, the species with the largest $k - k_{max}$, *Pentace strychnoidea*, is visibly more aggregated than any other species. As Fig. 4 indicates, this ranking allows us to stratify all the species by their aggregation tendencies.

Poisson Cluster Model

We have seen that the k -statistic, compiled from the entire Ripley $K(d)$ -curve, yields a convenient, single measure of aggregation. Nevertheless, the $K(d)$ -curve contains far more information than the value of k alone. In particular, we desire a more specific characterization of aggregation than a single index. We desire a model of the aggregation patterns. In this spirit, we choose to characterize local aggregation patterns by (i) a measure of the number of clumps and (ii) a measure of the mean clump size. Furthermore, for each species, we wish to distill these two parameters, denoted by ρ and σ , from the observed $K(d)$ -curve.

SPECIFICATION OF THE CLUSTER MODEL

We model the spatial pattern of each species by an independent, Poisson cluster process. This point-process is well understood theoretically (Cressie, 1991; Neyman & Scott, 1958), and it has the advantage of simplicity. Despite its simplicity, as we shall see, this model of aggregation captures enough information about spatial-patterning to reproduce the observed species-area curve with great accuracy. Most important, we choose the Poisson cluster process because its probabilistic properties (in particular, its expected Ripley K -curve) are well-understood in the literature; this fact will allow us to estimate parameters without much difficulty.

We use the following axiomatic definition of the Poisson cluster process.

1. “Parents” form a Poisson process in the plane with intensity ρ .
2. Each “parent” produces a random number of “offspring”, drawn independently from a fixed distribution.
3. The positions of the “offspring” relative to their parent are drawn independently from a fixed bivariate probability density function h .
4. The final pattern consists only of the “offspring” events.

More specifically, we stipulate that the offspring of a parent follows a radially symmetric Gaussian distribution with distribution function

$$h(x, y) = (2\pi\sigma^2)^{-1} \exp\left(\frac{-(x^2 + y^2)}{2\sigma^2}\right). \quad (8)$$

In particular, the mean squared distance from an offspring to its parent is $2\sigma^2$ and the mean distance is $\sigma\sqrt{\pi/2}$. As desired, ρ measures the density of clumps and σ measures clump size. We choose to describe ρ in units of clumps-per-square-meter and σ in units of meters.

In practice, once the parameters (ρ_i, σ_i) have been estimated for species i , we simulate the cluster process by placing $\lfloor \rho_i 5 \times 10^5 (\text{m}^2) + 1/2 \rfloor$ “parents” in the plot according to a uniform distribution. Next, we assign each of n_i stems to a randomly chosen parent, and position the stem according to eqn (8), relative to the parent. In particular, the expected number of stems per clump is given by $n_i/(\rho_i 5 \times 10^5)$. We impose toroidal boundary conditions in the event that a stem is placed outside of the 50-ha plot. Finally, we erase the “parents”—which were only used in order to position the individual clumps. For each species i , the cluster process is completely determined by the parameters (n_i, ρ_i, σ_i) .

By choosing this model, we certainly imply that clusters arise from local propagation. In this sense, the Poisson cluster process models aggregation caused by local seed dispersal or gap recruitment. The observed clump size for each species is surely determined by a number of biotic and abiotic factors other than seed dispersal and gap recruitment. Nevertheless, dispersal and recruitment are prominent among these factors. We have used the words “parent” and “offspring” as an analogy to these processes. We emphasize, however, that our notation does not imply, however, that every observed cluster arises from a single “parent” tree. Nor do we believe that our

stochastic process represents the actual mechanics which caused the formation of clusters in the forest. Instead, we are using the cluster process because it offers a convenient, phenomenological model to describe the pattern which results from the myriad of mechanistic forces which determine local aggregation patterns. More global patterns of aggregation (e.g. those guided by altitude specificity) and hierarchical aggregation will be considered later.

ESTIMATING CLUSTER PARAMETERS

Having specified the cluster model, all that remains is to estimate the parameters (ρ, σ) for each species. We estimate the best-fit parameters by using the measured $\hat{K}(d)$ curve. To this end, we rely on the well-established fact that a Poisson cluster process (ρ, σ) results in the following $K(d)$ curve (Cressie, 1991):

$$K(d)^{PCP} = \pi d^2 + \rho^{-1} \left(1 - \exp\left(\frac{-d^2}{4\sigma^2}\right) \right). \quad (9)$$

Given the empirical values $\hat{K}(0), \dots, \hat{K}(d_{max})$, we choose (ρ, σ) such that $K(d)^{PCP}$ most closely fits the observed values. Choice of the upper limit d_{max} has a significant effect on the resulting parameter fit. For instance, using $d_{max} < 385$ m at the HKK plot does not provide near as good a fit as $d_{max} = 385$ m. In this sense, the statistic $P_d - P_{random}$ illustrated in Fig. 3 provides crucial information for each plot: it indicates the proper range of distances over which to parameterize our spatial model.

Diggle (1983) suggests fitting (ρ, σ) by minimizing the integral

$$\int_0^{d_{max}} (K(h)^c - (K(h)^{PCP})^c)^2 dh, \quad (10)$$

for some tuning constant $c \approx 1/2$. (We found that $c = 1/4$ was most effective.) In order to minimize eqn (10) we must specify an initial parameter guess for (ρ, σ) . This choice has a dramatic effect on the eventual parameter fit. Let $(\bar{d}, \hat{K}(\bar{d}))$ be the maximum point on the observed $\hat{K}(d)$ curve (in a tie, choose the largest such \bar{d}). Following Diggle, we specify our initial parameter guess as $(\rho, \sigma) = (1/\hat{K}(\bar{d}), \bar{d}/4)$. We have estimated para-

eters by fitting eqn (9) to the observed values via minimization [eqn (10)] and via a more traditional curve-fitting technique (minimizing the χ^2 merit function of the sum squared residuals). Both methods provide comparable parameters.

Having computed the parameters (ρ_i, σ_i) for each species, we may compare the resulting Poisson cluster model to the actual data. Visually, the cluster model of each species is strikingly similar to the actual data (Fig. 4). Of course, the simplistic, radially symmetric model does not reproduce the fine details of all tree placement, but our method has the ability to separate a species into a best-fit number of clusters and best-fit cluster size. Careful inspection of Fig. 4 reveals that the parameter estimation, although largely accurate, is not perfect. For example, the method appears to overestimate slightly the number of *Mallotus leucodermis* clusters, and underestimate the average *Phaeanthus ophthalmicus* cluster size. The method also fails to capture the sharpness of the cluster margins in some cases—e.g. *Pentace strychnoidea* and *Cleistanthus sumatranus*. To a first approximation, however, the model matches our complex, visual intuition of aggregation (Fig. 4).

A rigorous goodness-of-fit between model and data, for each species at each plot, is certainly possible. Nevertheless, we refrain from this statistical exercise. Given the topic of our investigation, the important criterion by which to judge the cluster model should be its ability to reproduce the observed species-area curve. This criterion provides a practical, ecologically motivated metric. Moreover, whether or not the model produces the correct SAR, we will deduce information about the extent and manner in which individual-level aggregation effects species-level patterning.

Before examining the SAR, we pause briefly to inspect the distribution of clump sizes estimated by our best-fit parameters. The distribution of σ values is right-skewed normal at all three plots. Given σ , recall that $\sigma\sqrt{\pi/2}$ yields the mean distance from an individual to the center of the clump. This corresponds to mean patch size, and is determined primarily by dispersal distance and gap-sizes. Figure 5 shows a histogram of mean clump radius ($\sigma\sqrt{\pi/2}$) for those species at Pasoh which are classified as clumped. From Fig. 5 we

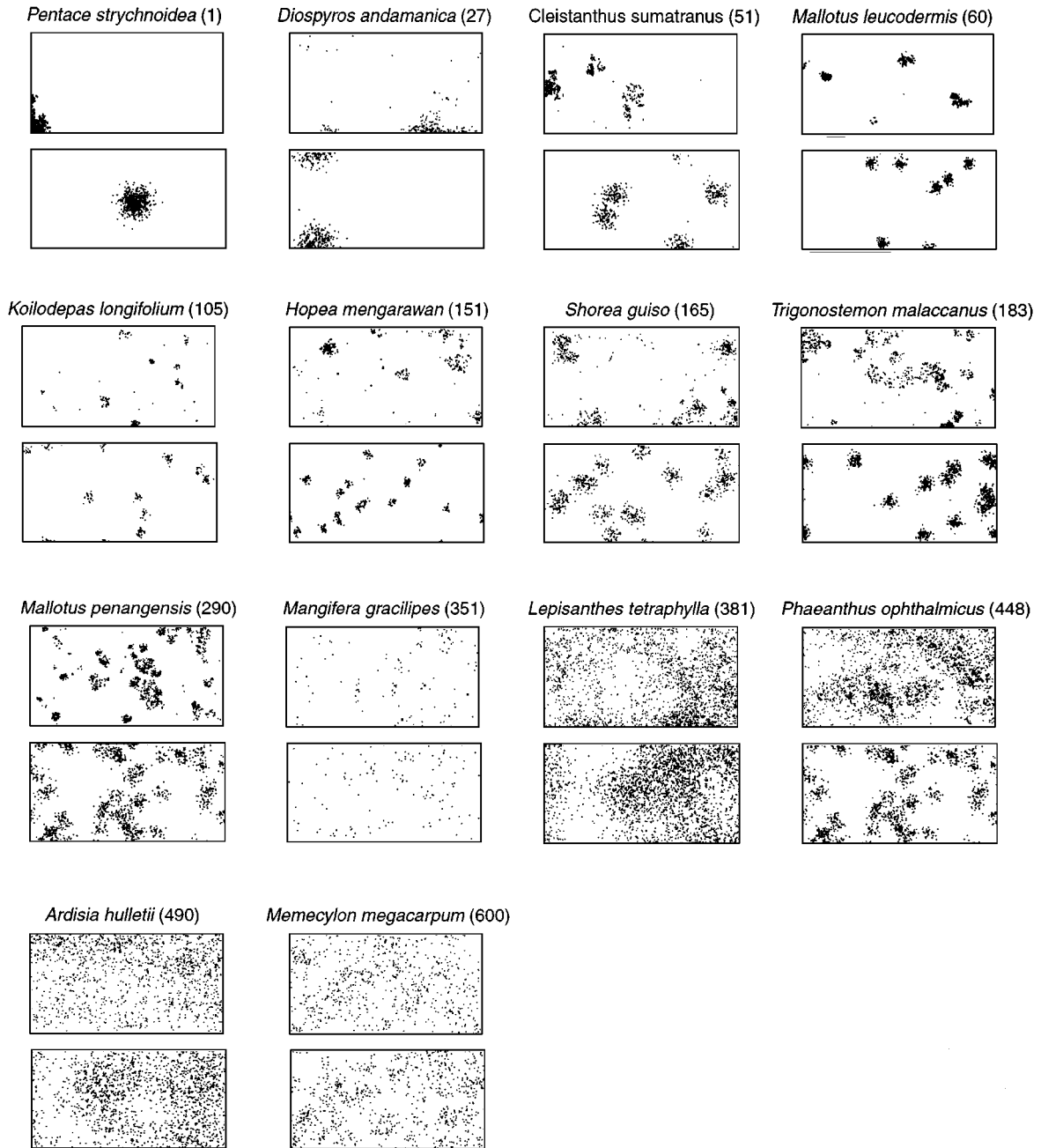


FIG. 4. Examples from Pasoh of the observed spatial pattern of a species (upper rectangle of each pair) compared with the pattern simulated by the Poisson cluster model (lower rectangle). Each rectangle encompasses 50 ha. The species are ranked according to their values of $k - k_{max}$; their ranked position is denoted in parentheses. This aggregation statistic appears to agree with an intuitive measure of clumping. Despite minor discrepancies, the Poisson cluster model provides a visually acceptable reproduction of the spatial pattern for each species.

see that 250 m is a decent upper-bound on mean patch size at Pasoh—which agrees with our earlier estimate in Section 4 (cf. Fig. 3).

THE SAR GENERATED BY THE CLUSTER MODEL
 For Pasoh we have simulated the entire 50-ha plot by overlaying each independently simulated

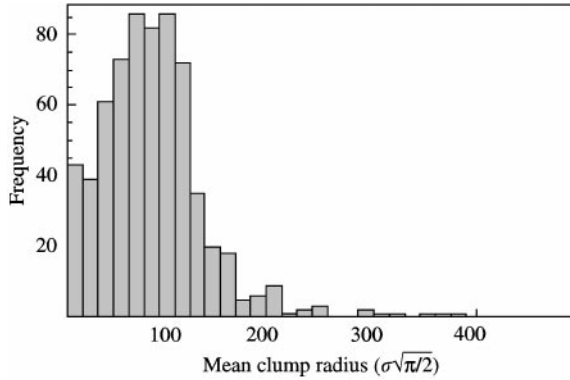


FIG. 5. A frequency chart of mean clump radius (estimated by $\sigma\sqrt{\pi/2}$) for aggregated species at Pasoh. Notice that 250 m provides a decent upper bound on dispersion distance, as seen independently in Fig. 3.

species. The resulting model SAR is almost indistinguishable from the actual SAR. The model very slightly overestimates diversity for areas < 0.2 ha, but accurately characterizes the SAR from 0.2 to 50 ha. For this large range of areas, the observed SAR falls within the confidence intervals for the model (which are extremely tight, smaller than the rectangular dots used Fig. 6). We have repeated the entire method (fitting parameters and simulating the point process) at HKK and Lambir; in all three plots we find an excellent agreement between model and data (Fig. 6). In short, we may conclude that the cluster model accurately reproduces enough information about spatial patterning to generate the correct species–area curve, given the abundances of species. In this sense, the Poisson cluster model completes the “zeroth-order analysis” pioneered by Coleman.

Given the simplicity of the cluster model, the fidelity with which it reproduces the SAR in each forest is somewhat surprising. We have accounted for clustering only on a relatively local scale. We have not accounted for larger-scale, environmentally driven patterns. We have not used information about tree-diameter. Most importantly, we have ignored all inter-specific spatial patterns. The Poisson cluster model, the Ripley-K measurements, and the (ρ_i, σ_i) parameter estimation are all intrinsically univariate. Despite these simplifications, the resulting SAR agrees

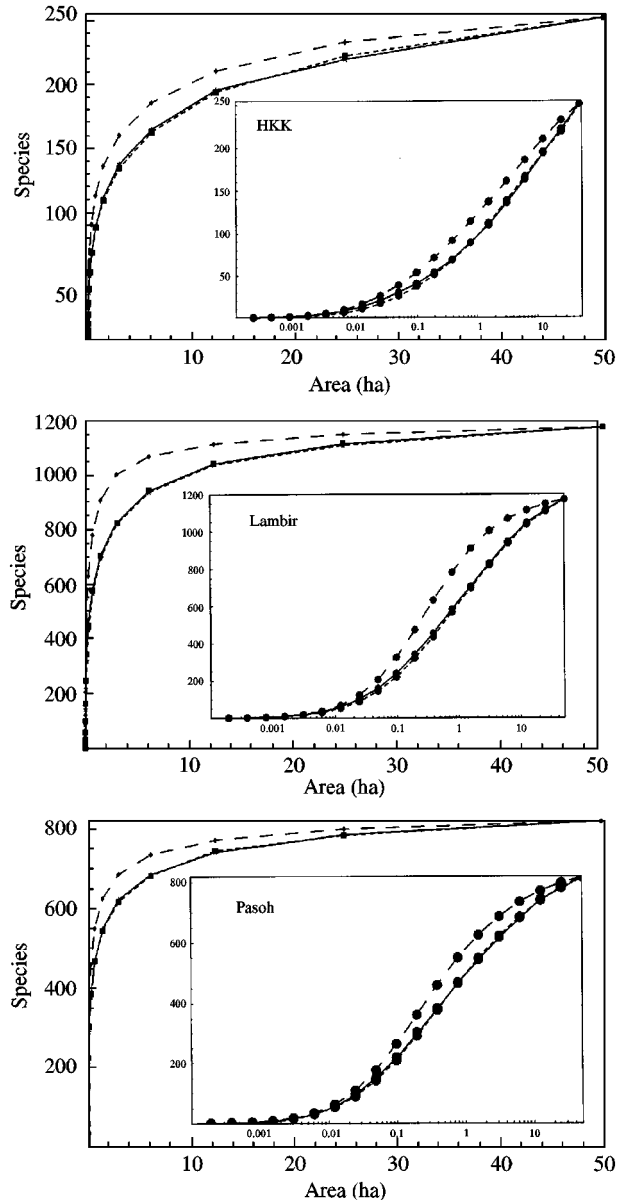


FIG. 6. The observed species–area curve at Pasoh, Lambir, and HKK along with the predictions via a random-placement model and the Poisson cluster model. In contrast to the random-placement model which overestimates diversity, the cluster model reproduces the SAR accurately—especially for areas > 0.2 ha. In each graph, the inset displays the same information on log-linear axes: Random placement model, ---; Cluster model, —; Observed data,

with the empirical SAR. In other words, whatever non-random, inter-specific patterns or environmental patterns exist, they do not influence the species–area curve. For each species, the SAR depends only on the location of conspecific trees;

the placement of other species is relatively unimportant. In some sense, this result suggests that interspecific competition may have only a limited or indirect effect on the species–area curve at this scale. Our results also suggest that, despite its prominence in ecology at large, the species–area curve may be a somewhat insensitive indicator of community structure.

Abundance and Aggregation

The methods developed thus far are sufficient to rank species by aggregation (Fig. 4), to reproduce intra-specific patterns in all three plots (Fig. 4), and to characterize accurately the species–area curve (Fig. 6). Thus, we may have some confidence in the practical and biological relevance of our model and its estimated parameters. We now investigate which biological features are correlated with the parameters of our model. We will address the relationship between a species' abundance and its tendency to aggregate.

Figure 7 shows a graph of species abundance n_i versus the k -statistic for Pasoh (compare with Fig. 3 in He *et al.*, 1987). As mentioned before, the statistic k alone, without reference to its quantile position in Monte-Carlo trials, may be a poor choice of aggregation index due to autocorrelation problems. Alternatively, we could use $k - k_{max}$ over 19 simulations. These provisos aside, we will investigate the relationship between n_i and k .

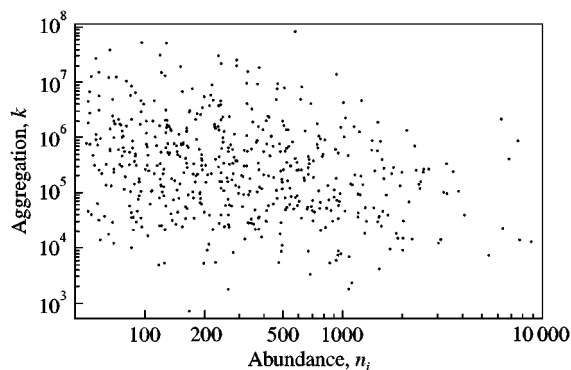


FIG. 7. Graph of abundance vs. the clumping index k for Pasoh. We only graph those species with abundance greater than 50 stems. Spearman rank correlation reveals a statistically significant, but very small, negative correlation between n_i and k for these species.

We find that there is a statistically significant, but extremely slight negative correlation between abundance and aggregation tendency at Pasoh, indicated by k (cf. Fig. 7). This result agrees with the parallel result of He *et al.* (1987) based upon the Donnelly index of aggregation. Similar results hold at HKK and Mudumalai. Our correlation statistics are based only on those species with $n_i > 50$. We cannot place much confidence in the clumping index for less abundant species where k is likely prone to autocorrelation errors. This is unfortunate because roughly one-third of the species in Pasoh have fewer than 50 stems. Although the standard Pearson correlation coefficient of (n_i, k_i) is insignificant ($r = -0.0884$), the Pearson correlation coefficient on the log–log transform is statistically significant ($r = -0.2688$). The latter statistic should be trusted more; each variable follows a nearly lognormal distribution. As further evidence, the Spearman rank coefficient of correlation between n_i and k is also statistically significant ($r = -0.2644$). We have tested significance using the fact that $|r|\sqrt{(n-2)/(1-r^2)}$ follows a Student's t distribution.

We emphasize, however, that statistical significance here by no means implies biological or practical importance. On the contrary, the negative correlation between abundance and aggregation is extremely slight. The statistical significance arises only out of the large number of data points. In fact, the correlation coefficient reveals that only r^2 , or less than 7% of the variation in k may be predicted from the abundance n_i .

If, however, we include all species with $n_i > 5$, we find a somewhat stronger negative correlation between abundance and k (Pearson $r = -0.4792$), suggesting that rarer species are more aggregated. This trend is very likely a statistical artifact, and we cannot be confident that it has much biological significance. Even when trees are placed at random, the k statistic becomes large when n_i is small. In other words, the k statistic suffers from autocorrelation problems when n_i is small (Fig. 8). When considering all species with $n_i > 5$, the alternative statistic $k - k_{max}$ yields an insignificant Spearman coefficient ($r = -0.0677$), suggesting that aggregation is not correlated with abundance after all. The experiment in Fig. 9

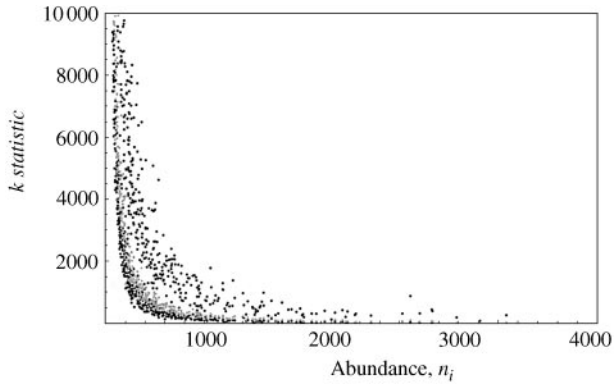


FIG. 8. Graph of abundance versus the maximum (upper $\cdots\cdots$), mean ($\cdots\cdots$), and median (lower $\cdots\cdots$) k -statistic obtained over 19 Monte-Carlo trials of spatial randomness. Even when trees are located randomly, k automatically becomes large for small n_i . Hence, the k statistic alone is a poor choice for comparing aggregation with abundance.

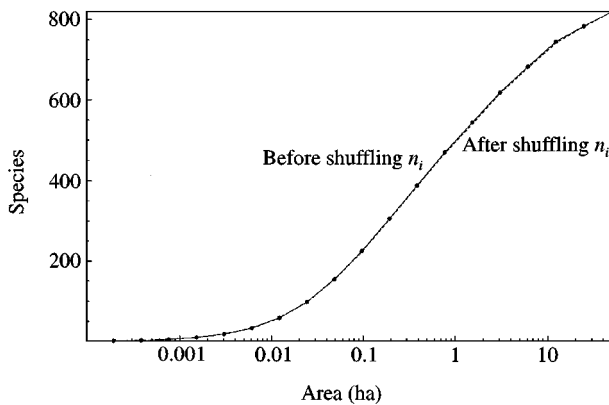


FIG. 9. The results of an experiment which tests the biological impact of abundance on aggregation. We graph the SAR generated by the Poisson cluster model of Pasoh (as in Fig. 6), as well as the SAR generated by the cluster model whose aggregation parameters (ρ_i, σ_i) have each been assigned to a randomly chosen abundance n_j . The resulting species–area curves are almost identical, and they both agree with the actual SAR at Pasoh. In other words, randomly shuffling the abundances of the species does not effect the SAR. The same phenomenon also occurs at HKK and Lambir (not shown). Hence, abundance and aggregation are conclusively uncorrelated insofar as the species–area curve is concerned. Before shuffling (—), after shuffling (---).

should help to resolve this statistical quandary in a practical way.

Aside from overall aggregation (k), we may investigate the more specific issues of abundance correlation with clump density (ρ) or with clump size (σ). In fact, given autocorrelation difficulties

with the k -statistic, a comparison between n_i and ρ_i may be more useful. Among the 534 species at Pasoh classified as clumped ($k > k_{max}$) with $n_i > 50$, there is no statistically significant correlation between abundance and ρ (Spearman rank correlation $r = -0.0057$). This gives further evidence that the negative correlation between abundance and aggregation tendency (k) is probably spurious.

We also find a slight negative correlation between abundance and σ (Spearman rank correlation $r = -0.1152$)—suggesting, if anything, that rare species are less tightly clumped than common species.

Given our results about ρ , σ , and $k - k_{max}$, we believe that the observed correlation between k and abundance reflects a statistical artifact more than a biological reality. This claim may conflict somewhat with analyses of Hubbell (1979) or recent, extensive analyses of Condit *et al.* (2000). The biological importance of the correlation may be tested—insofar as the species–area curve is concerned—in a revealing, practical manner. Recall that three parameters, n_i , ρ_i , and σ_i , determine the distribution pattern of a species in the Poisson cluster process. As an experiment, we may assign the previously measured parameters (ρ_i, σ_i) to the abundance of a randomly chosen species, n_j . In other words, we randomly shuffle the abundances of the species, and re-run the cluster simulation. This has the effect of removing whatever correlation may have existed between abundance and cluster parameters. For all three plots, upon randomly shuffling abundances the resulting SAR of the cluster model is almost identical to the SAR with non-shuffled abundances (Fig. 9). This somewhat surprising result reflects the fact that abundance was not strongly correlated with aggregation to start with.

Although we originally fit parameters (ρ_i, σ_i) to each species i , our model produces the correct SAR even if we randomly draw each species' parameters from the distribution of the fitted parameters (Fig. 9). This result should allow us to predict an SAR fairly well without detailed knowledge of best-fit, species-by-species parameters. For example, if we use the abundances of Pasoh and draw aggregation parameters randomly from the distribution of parameters fit for Lambir, we nevertheless obtain a fairly accurate

model of the Pasoh species–area curve. In other words, only the *distribution* of aggregation parameters matters, and the distributions at Lambir and Pasoh are fairly similar.

Environment and Aggregation: Hierarchical Clumping

Thus far we have modeled clumping on a local scale. The cluster model loosely mimics local aggregation patterns. Although aggregation at this scale alone has been demonstrated to determine the SAR (insofar as 50-ha data sets may verify), larger-scale aggregation driven by habitat certainly occurs as well. In this section, we investigate the extent and importance of environmentally driven aggregation. We are particularly interested in the possibility of hierarchical aggregation in which the local patches are themselves clustered following a more global pattern. We will employ our Poisson cluster model in order to factor out local clumping, and thereby properly examine larger-scale patterns.

Topography, soil differentiation, water stress, etc. are all examples of environmental factors

which we suspect influence spatial patterns at scales larger than mean dispersal distance and gap size (cf. Hubbell & Foster, 1983). In addition, there is certainly an undersurge of abiotic influences on all spatial scales. We demonstrate an analysis of hierarchical clumping driven by topography at the Pasoh forest. As with local aggregation, we follow a simple, first-cut approach. In particular, we focus on the main topographic gradient found at Pasoh: the single hill within the plot.

In particular, we divide the Pasoh 50-ha plot into two habitats called simply “on the hill” and “off the hill”. Our methods easily generalize to multiple environments—such as in a valley, on the slope, on a plateau, on humult soils, in old-growth forest, etc. Subdividing the plot into 800 squares each 25×25 m large, we define each square as either on the hill or off the hill. This subjective process relies only on the topographic contour lines (Fig. 10). According to our (subjective) definition, the “hill” at Pasoh accounts for 37% of plot’s total area; 36% of all stems in the plot are on the hill.

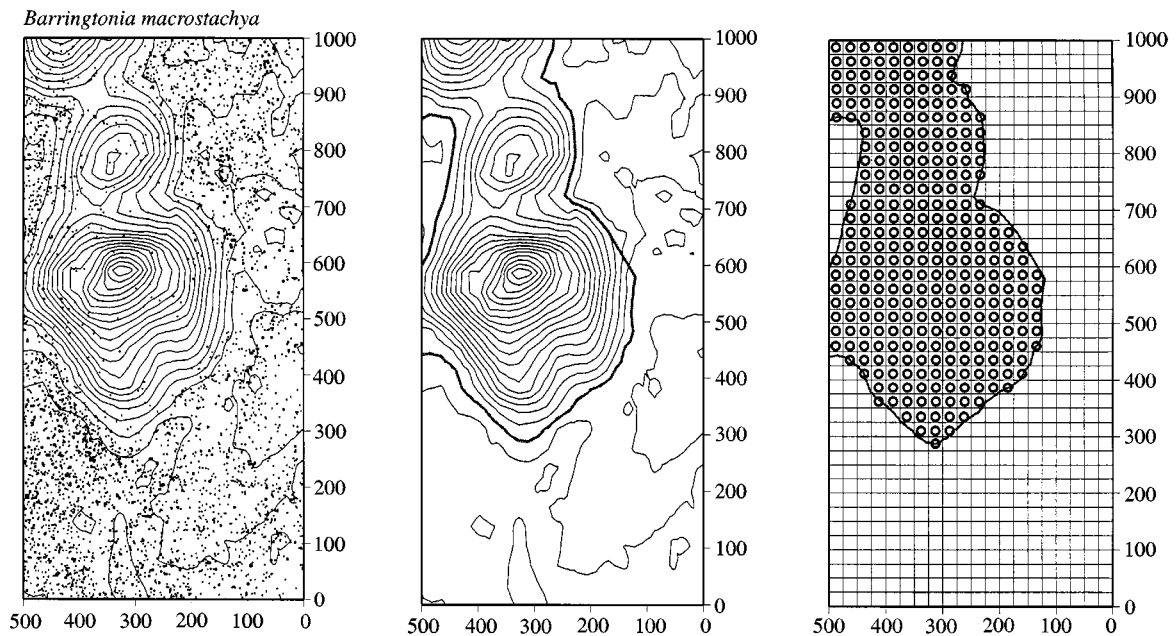


FIG. 10. A schematic diagram depicting our subjective definition of the two environments at Pasoh: “on the hill” and “off the hill”. The example species *Barringtonia macrostachya* in (a) is used as a general guide by which we select a cut-off topographic contour (b). Altitudes above the cut-off are defined as on the hill. (c) illustrates the 25×25 m boxes into which we subdivide the plot. Boxes with a circle are defined to be on the hill, and the other boxes off. As defined, the hill occupies about 37% of the 50-ha plot. *Barringtonia macrostachya* has 88% of its stems off the hill.

RANDOM-PLACEMENT TESTS OF TOPOGRAPHIC
CORRELATION

Given a specification of the two habitats, a straightforward analysis of aggregation driven by topography would proceed via a χ^2 test (cf. Basnet, 1992). For each species, the χ^2 test determines if an extraordinary number of trees are on (or off) the hill, assuming that tree locations are independent of one another. As we have demonstrated already, the independence assumption is certainly violated by most species. Therefore, we will design a slightly more sophisticated method which accounts for local aggregation. Our method offers an alternative to the elegant, torus-randomization method of Harms (1997); Harms randomizes the locations of the habitats, while we randomize the locations of the trees. In preparation for our more sophisticated test, we begin by re-formulating the χ^2 test in a Monte-Carlo setting. For a given arrangement of trees (of a particular species i) in the plot, we define

$$\Gamma = \frac{\text{proportion of trees on the hill}}{\text{number of trees on the hill}} = \frac{1}{n_i} \quad (11)$$

In order to test if the observed arrangement of species i is significantly hill-correlated (positively or negatively), we perform 1000 Monte-Carlo simulations. For each simulation, we place n_i trees of species i in the plot randomly, and we measure Γ . If the true, observed value of Γ falls in either the top 2.5% or bottom 2.5% tail of the simulations, we can reject the null hypothesis of no correlation at the 5% confidence level.

The Γ test with random-placement Monte-Carlo simulations is absolutely equivalent to the standard χ^2 test. As verification, 1000 simulations reveal that the two methods differ on $< 1\%$ of the species. The χ^2 test indicates that roughly two-thirds of all species at Pasoh are hill-correlated (Table 2). Figure 11 illustrates the six most hill-correlated species (positively or negatively) as ranked by their χ^2 value.

Figure 11 illustrates that the topography at Pasoh can have a strong effect on spatial patterning. Nevertheless, one of the species shown in Fig. 11, *Pentace strychnoidea*, despite its extremely high χ^2 value, would likely not meet our intuitive notion of topographic specificity. *Pentace*

TABLE 2
The proportion of all 817 species which are classified as hill-correlated (leftmost column)

	All species	Species with $n_i > 50$	Species with $n_i > 50$, $n_{clumps} > 3$
Γ – test			
random-model	511 (62.5%)	423 (79.2%)	361 (79.2%)
Γ – test			
cluster-model	241 (29.5%)	212 (39.7%)	183 (40.1%)

Results are given for the Γ test with random-placement Monte-Carlo simulations (equivalent to a χ^2 test) and for the Γ test with cluster-placement. Both tests were performed at the 5% significance level. Species by species, the latter test indicates whether or not the clusters of trees are themselves clustered on (or off) the hill. The cluster test reveals that, by taking account of local aggregation, topography influences the geographic range of nearly half as many species as suggested by a naive χ^2 -analysis. The same result holds among the 534 species with > 50 stems (middle column) or among the 456 species with > 50 stems and > 3 clumps (rightmost column).

strychnoidea is an extremely clumped species (in fact, it has the highest aggregation index in Fig. 4), and it is unclear whether its association with the hill is driven by environment *per se*. Perhaps, instead, this species is found only on the hill because its single, tight, dispersal-driven clump happens by chance to lie on the hill. The χ^2 test—or, equivalently, the Γ test with random-placement simulations—cannot distinguish between these two possibilities. Many species at Pasoh follow patterns similar to *P. strychnoidea*. We desire a new test to disentangle aggregation on different scales.

 A CLUSTER-BASED TEST OF TOPOGRAPHIC
CORRELATION

Fortunately, we can adjust our Γ test so as to take advantage of our knowledge of local aggregation. Instead of using random-placement for our Monte-Carlo simulations, we may use the Poisson cluster process with the best-fit parameters for each species. In essence, the Γ test with cluster-based Monte-Carlo simulations will determine whether or not an extraordinary number of *clumps* of each species are on (or off) the hill. For each species, the improved Γ test provides a viewpoint which is more coarse than the indi-

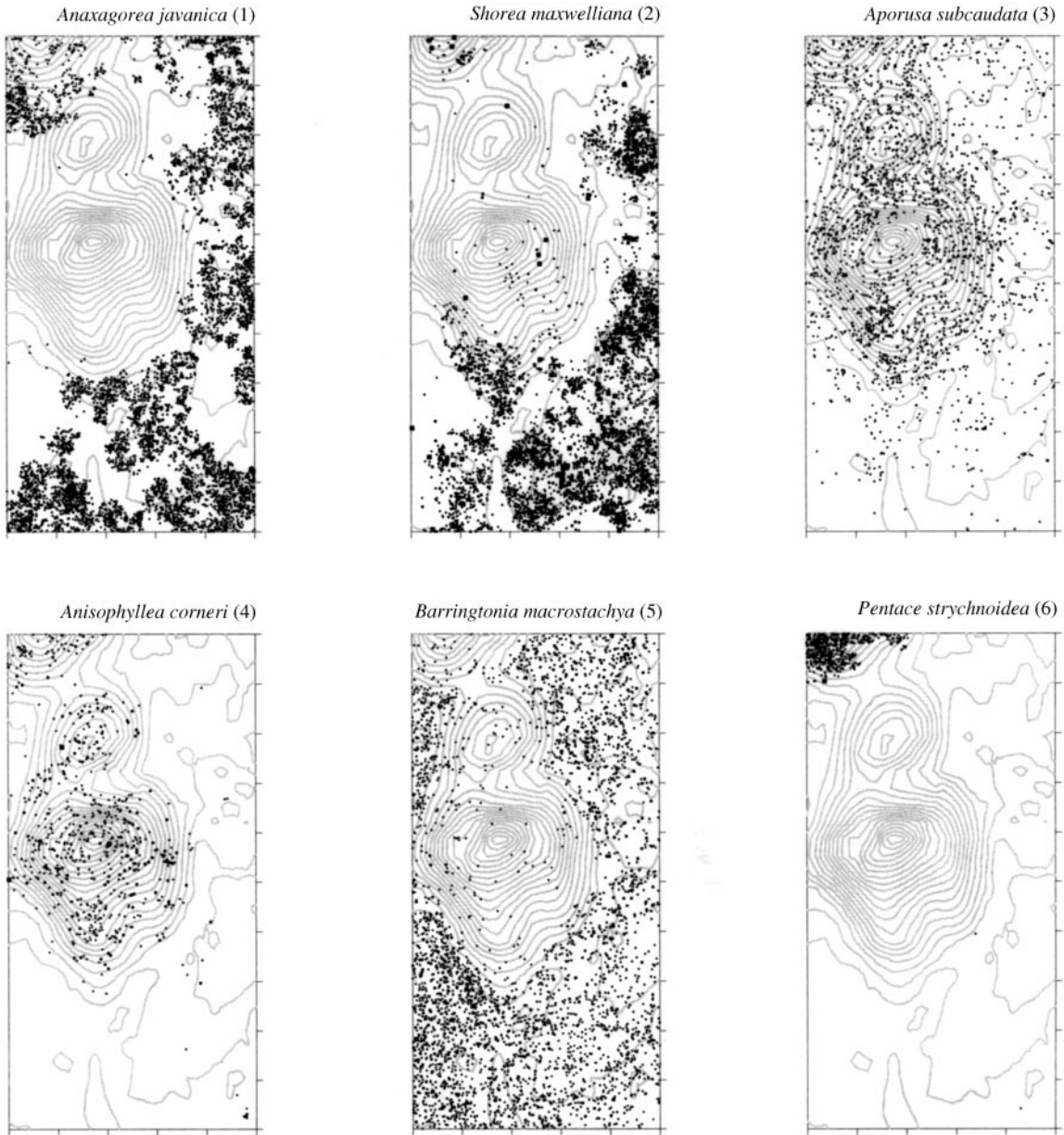


FIG. 11. Spatial maps of the six most topographically driven species at Pasoh, as indicated by their χ^2 scores. Three of the species are positively correlated with the hill, and three negatively. Even though the χ^2 test ranks *Pentace strychnoidea* as strongly correlated with the hill, it seems plausible that the single, tight, clump of this species lies on the hill by chance, as opposed to genuine topographic specificity. This example highlights the need for a more sophisticated test than the χ^2 analysis.

vidual level—and the level of coarseness is controlled by the cluster parameters (ρ_i, σ_i). (For those species for which we estimated very many clumps or very large clump radius, the cluster- Γ test still operates on an individual level.) For significantly clumped species, the Γ test with cluster-based simulations will detect the extent to

which the clumps are themselves aggregated on the hill, i.e. a test of hierarchical clumping (Table 2).

The cluster-based Γ test reveals that about one-half of the species classified as on or off the hill by the χ^2 test are not, in fact, significantly hill-correlated. In other words, by taking account of their local-clumping properties, only half as

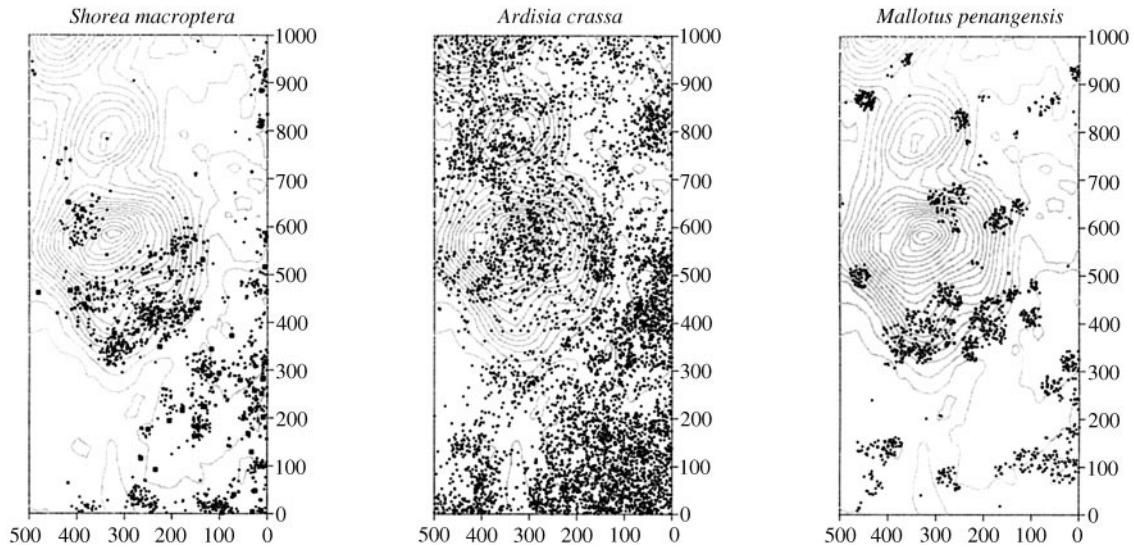


FIG. 12. Three examples of species classified as hill-correlated by the χ^2 test, but uncorrelated by the cluster-based Γ test. These species do not show an extraordinary number of clumps either on or off the hill (although they do have an extraordinary number of individuals on or off the hill). In other words, insofar as may be inferred from our 50-ha data set, the cluster-based Γ test indicates that small-scale dispersion/recruitment constricts the geographic ranges of these three species, but topography does not.

many species are actually guided by topography per se as compared to the naive estimate. This would suggest that adaptation to topography limits the placement of individuals less severely than dispersal, gap recruitment, and other local factors. For instance, at the 5% level, we have seen that 83% of species show dispersion-scale aggregation, while only 30% show true topographic aggregation.

We may have confidence that the cluster-model Γ test performs its stated task well. Figure 12 illustrates three example species for which the naive χ^2 analysis concludes hill-correlation, but for which the modified Γ test indicates no correlation. In these examples, by broadening our focus to the cluster level (instead of the individual level), we do not find a disproportionate amount of hill-correlation. In particular, as desired, the cluster-based Γ test indicates that *Pentace strychnoidea* is not significantly correlated with the hill. There are no species which the χ^2 test finds uncorrelated but which the Γ test indicates are correlated; the cluster-based test is uniformly more conservative. Finally, among those species which the cluster-based Γ test classifies as hill-correlated, there is no significant disparity between the number positively (116) and negatively

(125) correlated. The Pasoh species have apparently adapted to fill both ecological niches equally well.

We conclude this section with some general observations and provisos about the cluster-based Γ test. Although we have used it to test topography, the Γ test can be used to disentangle local clumping from—and therefore query the strength of—any environmental factor which operates on scales larger than 50–100 m. Nevertheless, one should be aware that the Γ test is inherently conservative. As a null hypothesis we assume that habitat has no effect on the spatial arrangement; hence, if a species fails the Γ test, we cannot soundly conclude that habitat has no effect, but rather that our particular test did not discover an effect. In this sense, the results reported in Table 2 are conservative lower bounds. For example, when properly interpreted, Table 2 indicates that *at least* 30% of the species in Pasoh are driven by topography, but possibly more. Furthermore, the Poisson cluster model used in the Γ test assumes that the abundance of each species is known. These abundances—and, in particular, the total species richness of the plot—are certainly determined to a large extent by the diversity of habitats in the plot. Hence, the

results of the I test should not be interpreted as evidence against the importance of environmental determinants overall, but rather as evidence against the specific, relative strength of topography versus local factors as determinants of spatial patterning within our particular 50-ha plot.

Discussion and Conclusions

We have modeled local clumping via a Poisson cluster process. Fitting the proper parameters of this model to each of three tropical forests has yielded significant results on a few fronts.

First, the cluster model characterizes the species–area curve with extremely high fidelity. The model can be viewed as a completion of Coleman's, zeroth-order random-placement model. This result highlights those biological factors which are sufficient—and which are unnecessary—to determine the species–area curve. The fact that the cluster parameters can be shuffled between the species (or transplanted from one forest to another) without affecting the SAR suggests that the species–area curve may be a somewhat insensitive indicator of community structure.

Second, the best-fit parameters of each species allow us to address possible correlates between biological factors (such as abundance) and clump density or clump size. In the future, one may and should use these parameters to look for spatial patterns correlated with genus, functional groups, dispersal syndrome, and other biological factors.

Finally, the cluster model—which characterizes clumping at the dispersion distance/gap size scale—may be used to investigate the hierarchical effects of habitat on aggregation. This method applies to those environmental factors which operate on scales larger than dispersion. In particular, we have used the cluster model to ask if, within each species, the clusters of trees are *themselves* aggregated following a topographic gradient. Once parameterized the cluster model allows us to factor out small-scale clumping, and to investigate accurately aggregation driven by large-scale topography. We conclude quantitatively that, within the 50-ha plot at Pasoh, topography determines the geographic range of a species less often than small-scale factors such as dispersion and gap recruitment.

Despite the merits of the Poisson cluster model, we must mention some of its severe drawbacks. Foremost, the model is phenomenological. It does not provide a dynamic understanding of the processes which form the spatial patterns. Instead, the model is simply a static, retroactive characterization of spatial patterns. Indeed, estimating the parameters of the model required detailed knowledge of all tree locations. Thus, the cluster model does not provide a predictive theory. Instead, the cluster model is useful as a tool for assessing the effects and causes of aggregation, for comparing aggregation parameters across plots, and for investigating biological correlates to aggregation.

We conclude by recalling one of the early investigations into spatial aggregation in a large tropical forest plot. In 1983, before the first census at BCI was even complete, Hubbell and Foster categorized the significant spatial patterns into three disjoint groups (direct quote, Hubbell & Foster, 1983):

- (i) Species which appear to be randomly or near-randomly distributed over the plot.
- (ii) Species which are clumped and whose patches follow easily recognized topographic features of the plot.
- (iii) Species which are clumped but whose patches are spatially uncorrelated with topography.

At the time, a limited number of species in the BCI canopy were identified according to these categories, relying mainly upon intuition and visual inspection. We believe that the three categories above, even though they were identified in the BCI canopy, remain the most descriptive, qualitative characterization of the major spatial patterns in tropical forests. The rich theory of stochastic point processes, however, now allows us to approach these categories more quantitatively. The Poisson cluster model and the I test provide a rigorous, systematic method for categorizing each species according to Hubbell's original delineations.

The authors sincerely thank Helene Muller-Landau, Simon Levin, Bert Leigh, Bill Bossert, Ran Nathan, Liz Losos, Richard Condit, and Burt Singer for their helpful suggestions. The authors are especially indebted to Lee Hua Seng

for the use of Lambir data, and to Sarayudh Bunyavejchewin for the HKK data. J.B.P is supported by a fellowship from the National Science Foundation.

REFERENCES

- ARRHENIUS, O. (1921). Species and area. *J. Ecol.* **9**, 95–99.
- BASNET, K. (1992). Effects of topography on the pattern of trees in Tabonuco (*Dacryodes excelsa*) dominated rain forest of Puerto Rico. *Biotropica* **24**, 31–42.
- BATISTA, J. & MAGUIRE, D. (1998). Modeling the spatial structure of tropical forests. *Forest Ecol. Manage.* **110**, 293–314.
- BODZIARCZYK, J. & SZWAGRZYK, J. (1996). Species composition and structure of forest stands in Phyllitido-Aceretum community. *Ekologia-Polska* **43**, 153–173.
- BODZIARCZYK, J., MICHALCEWICZ, J. & SZWAGRZYK, J. (1999). Secondary forest succession in abandoned glades of the Pieniny National Park. *Polish J. Ecol.* **47**, 175–189.
- BUCKELY, R. (1982). The habitat-unit model of island biogeography. *J. Biogeography* **9**, 339.
- COLEMAN, B. (1981). Random placement and species–area relations. *Math. Biosci.* **54**, 191–215.
- COLEMAN, B., MARES, M., WILLIG, M. & HSIEY, Y. (1982). Randomness, area, and species richness. *Ecology* **63**, 1121–1133.
- CONDIT, R., ASHTON, P. S., BAKER, P., BUNYAVEJCHEWIN, S., GUNATILLEKE, S., GUNATILLEKE, N., HUBBELL, S. P., FOSTER, R. B., ITOH, A., LAFRANKIE, J. V., LEE, H. S., LOSOS, E., MANOKARAN, N., SUKUMAR, R., YAMAKURA, T. (2000). Spatial patterns in the distribution of tropical tree species. *Science* **288**, 1414–1418.
- CONNELL, J. H. (1971). On the roles of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. In: *Dynamics of populations, proceedings of the Advanced Study Institute on Dynamics of Numbers in Populations* (den Boer, P. & Gradwell, G., eds). Wageningen, Netherlands: Cent. Agric. Publ. Doc.
- CONNOR, E. & MCCOY, E. (1979). The statistics and biology of the species–area relationship. *Am. Nat.* **113**, 791–833.
- CRESSIE, N. (1991). *Statistics for Spatial Data*. New York: Wiley & Sons.
- DIGGLE, P. (1983). *Statistical Analysis of Spatial Point Patterns*. London: Academic Press.
- GOTELLI, N. J. & GRAVES, G. (1996). *Null Models in Ecology*. Washington, DC: Smithsonian Institution Press.
- HARMS, K. (1997). Habitat-specialization and the seed dispersal-limitation in a neotropical forest. Ph.D. Thesis, Princeton University.
- HARTE, J., KINZIG, A. & GREEN, J. (1999). Self-similarity in the distribution and abundance of species. *Science* **284**, 334–336.
- HE, F., LEGENDRE, P. & LAFRANKIE, (1987). Distribution patterns of tree species in a Malaysian tropical rain forest. *J. Vegetation Sci.* **8**, 105–114.
- HUBBELL, S. (1979). Tree dispersion, abundance, and diversity in a tropical dry forest. *Science* **203**, 1299–1309.
- HUBBELL, S. (1997). A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. *Coral-Reefs* **16**, 29–31.
- HUBBELL, S. & FOSTER, R. (1983). Diversity of canopy trees in neotropical forest and implications for conservation. *Tropical Rain Forest: Ecology and Management* (Sutton, S., Whitmore, T. & Chadwick, A., eds). London: Blackwell Scientific.
- HUBBELL, S. & FOSTER, R. (1986). Commonness and rarity in a neotropical forest: implications for tropical tree conservation. In: *Conservation Biology: The Science of Scarcity and Diversity* (Soule, M., ed.). Sunderland, MA: Sinauer Assoc. Inc.
- JANZEN, D. H. (1970). Herbivores and the number of tree species in tropical forests. *Am. Nat.* **104**, 501–528.
- KUNIN, W. (1998). Extrapolating species abundance across spatial scales. *Science* **281**, 1513–1515.
- LEITNER, W. A. & ROSENZWEIG, M. (1997). Nested species–area curves and stochastic sampling: a new theory. *Oikos* **79**, 503–512.
- MACARTHUR, R. & WILSON, E. O. (1963). An equilibrium theory of insular zoogeography. *Evolution* **17**, 373–387.
- MACARTHUR, R. & WILSON, E. O. (1967). *Island Biogeography*. Princeton, NJ: Princeton University Press.
- MCGUINNESS, K. A. (1984). Equations and explanations in the study of species–area curves. *Bio. Rev.* **59**, 423–440.
- MAY, R., (1975) *Ecology and Evolution of Communities* (Cody, M. & Diamond, J., eds), pp. 81–120. (Cambridge, MA: Belknap Press.
- MAY, R., LAWTON, J. & STORK, N. (1995). *Extinction Rates* (Lawton, J. & May, R., eds), pp. 1–24. Oxford: Oxford University Press.
- NEYMAN, J. & SCOTT, E. (1958). Statistical approach to the problems of cosmology. *J. R. Statist. Soc. B* **20**, 1–43.
- NEY-NIFLE, M. & MANGEL, M. (1999). Species area curves based on geographic range and occupancy. *J. Theor. Biol.* **196**, 327–342.
- PIMM, S. & RAVEN, P. (2000). Extinction by numbers. *Nature* **403**, 843–845.
- PLOTKIN, J. B., POTTS, M. YU, D., BUNYAVEJCHEWIN, S., CONDIT, R., FOSTER, R., HUBBELL, S., LAFRANKIE, J., MANOKARAN, N., SEUY, L., SUKUMAR, R., NAUAK, M., ASHTON, P. (2000). Predicting species diversity in tropical forests. *Proc. Natl Acad. Sci. U.S.A.* **97**, in press.
- PRESTON, F. (1962). The canonical distribution of commonness and rarity: Part I. *Ecology* **43**, 185–215.
- RIPLEY, B. (1976). The second-order analysis of stationary point processes. *J. Appl. Probab.* **13**, 255–266.
- ROSENZWEIG, M. (1995). *Species Diversity in Space and Time*. Cambridge, MA: Cambridge University Press.
- SZWAGRZYK, J. & PTAK, J. (1981). Analyses of spatial structure of populations and communities based on mapped point patterns of individuals. *Wiadomosci Ekol.* **37**, 107–124.