Cluster Analysis of Spatial Patterns in Malaysian Tree Species

Joshua B. Plotkin,[1,2,*] **Jérôme Chave,**[2] **and Peter S. Ashton**[3]

1. Institute for Advanced Study, Princeton, New Jersey 08540;
2. Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544;
3. Department of Organismal and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138

ABSTRACT: Tree species in tropical rain forests exhibit a rich panoply of spatial patterns that beg ecological explanation. The analysis of tropical census data typically relies on spatial statistics, which quantify the average aggregation tendency of a species. In this article we develop a cluster-based approach that complements traditional spatial statistics in the exploration and analysis of ecological hypotheses for spatial pattern. We apply this technique to six study species within a fully mapped 50-ha forest census in peninsular Malaysia. For each species we identify the scale(s) of spatial aggregation and the corresponding tree clusters. We study the correlation between cluster locations and abiotic variables such as topography. We find that the distribution of cluster sizes exhibits equilibrium and nonequilibrium behavior depending on species life history. The distribution of tree diameters within clusters also varies according to species life history. At different spatial scales, we find evidence for both niche-based and dispersal-limited processes producing spatial pattern. Our methodology for identifying scales of aggregation and clusters is general; we discuss the method's applicability to spatial problems outside of tropical plant ecology.

Subject heading: tropical forests, spatial statistics, spatial point processes, continuum percolation, dispersal limitation.

* Corresponding author; e-mail: plotkin@ias.edu.

It has been firmly established that woody plant species and lianas exhibit spatial aggregation in natural communities. This trend has been especially well documented in species-rich communities, such as tropical rain forests, through large-scale census experiments (Hubbell 1979; Hubbell and Foster 1983; He et al. 1997; Condit et al. 2000; Plotkin et al. 2000b). Although the primary cause of spatial patterns is hotly debated, the existence of spatial heterogeneity is fundamental to our theoretical and practical under-standing of complex ecosystems (Ashton 1976, 1998; Hubbell and Foster 1983; Hubbell 1997; Levin et al. 1997). The niche-based hypothesis attributes most species-level heterogeneity to biotic and abiotic environmental variability (Ashton 1969; Grubb 1977). However, even without habitat differentiation, neutral models of diversity predict that dispersal limitation can itself account for the emergence of spatial clustering (Wong and Whitmore 1970; Hubbell 1997, 2001).

Why should tree species be spatially clumped? Why do they not eventually occupy all suitable space evenly? Many ecologists (the niche theorists) argue that some sites are, at least at certain periods, unsuitable for certain species. This occurs because the species is not competitive in the local abiotic habitat, or pathogens or predators have eliminated it locally (Janzen 1970; Connell 1971), or it is due to some random vicissitude. Other ecologists (the neutral theorists) contend that dispersal limitation and differential mortality alone cause aggregated occupancy patterns.

Not until recently, however, with the advent of large botanical inventories and long-term surveys, have these fundamental issues been addressed with detailed empirical data. Methodologies for analysis of spatial data have become important for evaluating ecological theories. Such analyses usually employ the large body of spatial statistics that have been developed over the past century. Recent contributions (He et al. 1997; Batista and McGuire 1998; Condit et al. 2000; Plotkin et al. 2000b) have quantified the average clumping characteristics of tree species by utilizing a family of measures derived from the Ripley K statistic (Ripley 1976). The K statistic computes the number of conspecifics within a distance d from an individual, averaged over all individuals in the data set. To quantify clumping, the observed K values are compared with the null hypothesis of spatial randomness. Other common methodologies for analyzing spatial data include Fisher's quadrat-based variance-to-mean ratio (Fisher et al. 1922; David and Moore 1954; Lloyd 1967), nearest-neighbor statistics (Pielou 1959; Pollard 1971; Diggle 1983; Cressie 1991; He et al. 1997), and fitted spatial point processes (Diggle 1983; Cressie 1991; Batista and Maguire 1998; Plotkin et al. 2000b).

In this article we address a generic limitation of such

statistical studies. Despite their broad efficacy, spatial statistics essentially quantify the average aggregation pattern of a species. This limitation often has induced ecologists to assign a single scale of aggregation to each species (Condit et al. 2000; Plotkin et al. 2000b, He and Gaston 2000) even though a species can be aggregated at several scales simultaneously. Many ecological questions, such as the influence of habitat differentiation, can be better addressed by identifying the specific position, size, and shape of the spatial clusters in a data set. This task is tantamount to partitioning survey data into distinct clusters.

The problem of partitioning data has been investigated extensively in many disciplines (Duda et al. 1998). In this article we develop an approach to identifying clusters in spatial data, at several spatial scales, without making any a priori assumptions on the structure of clusters. We apply this technique to a 50-ha permanent sampling plot of tropical rain forest. Using select species from the 50-ha survey, we investigate the potential effects of habitat differentiation and dispersal limitation, that is, niche-based and dispersal-based hypotheses, on the spatial arrangement of individuals at various spatial scales. For each species we identify the scale(s) of spatial aggregation and the corresponding tree clusters. We study the correlation between cluster locations and abiotic variables such as topography, the distribution of cluster sizes, and the distribution of tree diameters within clusters. We conclude by discussing the relative merits of our cluster-based approach and spatial statistics in general.

Material and Methods

Study Site

The Pasoh Forest Reserve, Negiri Sembilan, Malaysia, is a 2,450-ha protected forest of the Keruing-Meranti type. Its flora has been studied since 1970 (Wong and Whitmore 1970; Ashton 1976; P. S. Ashton, unpublished manuscript, 1971). The canopy of the Pasoh forest is dominated by dipterocarp trees, mostly in the genera *Shorea*, *Dipterocarpus*, and *Neobalanocarpus*. Mean rainfall at the study site is relatively low, approximately 2,000 mm/yr, but without marked seasonality.

A 50-ha permanent sampling plot was established in 1987 by the Forest Research Institute of Malaysia to monitor long-term changes in primary forest (Manokaran and Kochummen 1987; Kochummen et al. 1990; Manokaran and LaFrankie 1990). The Pasoh plot, located at 102°18′W and 2°55′N, is a rectangular census 1 km long and 0.5 km wide. The plot's topography is fairly even, with a gentle slope rising 24 m toward the northwest. There are two primary streams within the plot. All 335,256 free-standing woody stems exceeding 1 cm diameter at breast height

(dbh) were identified to species and spatially mapped to <1 m accuracy (see Manokaran and LaFrankie 1990), yielding 816 species, 294 genera, and 74 families. The initial census of 1987 has been repeated in 1990 and 1996. The 1996 data are used throughout this article. The Pasoh plot is part of a long-term research program worldwide, coordinated by the Smithsonian's Center for Tropical Forest Science.

We have selected this plot because it has motivated several statistical studies in the past few years (He et al. 1997; Condit et al. 2000; Plotkin et al. 2000a, 2000b; He and Gaston 2000). Hence, our cluster analysis may be easily compared with prior results of a more statistical nature.

Study Species

Previous studies of spatial pattern in tropical forests (He et al. 1997; Condit et al. 2000; Plotkin et al. 2000b) have addressed the question of how many species are aggregated. These studies have clearly concluded that most species are clumped, and a few are randomly distributed. This conclusion was based on spatial statistics, which are often deficient for rare species.

Here, by contrast to previous studies, we focus on a handful of carefully chosen species. We have selected four abundant and two rare species in the Pasoh permanent plot. *Xerospermum noronhianum* (Blume) Blume Sapindaceae is a subcanopy species and the most common in the plot (8,820 stems, average dbh $\overline{D} = 3.19$ cm, maximal dbh $D_{max} = 37.3$ cm). It is recorded as a peat swamp forest specialist. It is abundant in mesic moist sites throughout western Malaysia but also is widespread in other lowland habitats in mixed dipterocarp forest. The seeds of *X. noronhianum* are dispersed by the monkey, the crab-eating or long-tailed macaque, and by the gibbon (Yap 1976). *Knema laurina* (Blume) Warb. Myristicaceae (4,088 stems, $\overline{D} = 2.53$ cm, $D_{max} = 35.4$ cm) is a subcanopy species often found on sand or clay.

Neobalanocarpus heimii (3,334 stems, $\overline{D} = 5.0$ cm, $D_{max} = 196.8$ cm) is an emergent dipterocarp and a commercial timber known as chengal. It is endemic to peninsular Malaysia and is widespread there on sandy alluvium and on ridges to about 1,000 m. It is sometimes gregarious. It is shade tolerant as a juvenile and has very hard durable wood (specific density 0.7–0.8 g/cm³). With large wingless fruit, the seed has no known means of dispersal. *Neobalanocarpus heimii* is the ultimate climax species, able to regenerate even in very small light gaps.

Shorea macroptera (1,508 stems, $\overline{D} = 6.1$ cm, $D_{max} = 102.5$ cm) is a main-canopy dipterocarp and a leading species in the so-called Red Meranti–Keruing type of mixed dipterocarp forest on suitable soils throughout western Malaysia. The fruit bears slightly twisted, aliform, ex-

panded sepals, so that it is dispersed by gyration and some-times by wind. It is also a light demander and regenerates from established seedlings after the formation of large gaps. *Shorea macroptera* follows life-historical strategies that are extremely different from those of *N. heimii.*

In addition, we have studied two less abundant species with peculiar spatial patterns. *Mallotus penangensis* Muell. Arg. Euphorbiaceae (1,404 stems, $\overline{D} = 4.56$ cm, $D_{max} = 24$ cm) is a subcanopy tree found on various habitats both in peninsular Malaysia and in Borneo (known as *Mallotus sarawacensis*). *Pternandra coerulescens* Jack Melastomataceae (438 stems, $\overline{D} = 5.31$ cm, $D_{max} = 25$ cm) is a subcanopy species found mainly along watercourses or wherever ground litter is minimal.

Clustering Spatial Data

Algorithm Specification

We present a technique of spatial analysis that identifies the location of intraspecific tree clusters in a fully censused forest plot. This technique will serve as the basis for analyzing the spatial patterns of our six study species. Tree locations are henceforth considered as punctual. Our technical problem is, therefore, to define and detect the spatial clusters for a set of points in the plane. In addition, we wish to detect "critical distances" or "critical scales" related to the observed data. When data contain more than one critical distance, classical statistics would yield an average of these distances and, thus, a fairly imprecise characterization. The most straightforward method for partitioning data into clusters is by visual inspection. However, this method is subjective and time consuming. We desire a rigorous technique of cluster detection that requires no a priori assumptions about the shape of the clusters. We also want to detect clusters even if they are not radially symmetric or approximately symmetric. We now present an efficient technique for addressing both questions.

We say that two trees are connected if their stems are separated by a distance d or less. Two trees are said to belong to the same cluster if they are related by a path of connected trees. For example, if tree A is connected to tree A, and tree B is connected to tree C, then all three trees will belong to same cluster, even if A and C are not directly connected. This definition of a cluster, illustrated graphically in figure 1, is extremely simple and natural. In addition, this definition allows us to interpret cluster analysis in the language of statistical mechanics (see "Percolation and Critical Distances").

Given a value of the clustering distance d, our algorithm partitions the data into distinct spatial clusters. The distance d is a free parameter of the algorithm. Once d is specified, there is a unique corresponding partition of the data into disjoint clusters. In the trivial case when $d = 0$, each tree forms its own distinct cluster, but when d is very large (larger than the maximal pairwise distance between two trees), there is only one cluster; all trees are pairwise connected. Interesting patterns emerge for intermediate values of the clustering parameter d. The appendix describes an efficient implementation of our algorithm as well as some related subtleties of its execution.

Percolation and Critical Distances

Spatial statistics typically are compared against the null model of randomly distributed points. In order to gain some intuition about our clustering algorithm, we temporarily restrict our focus to the case of random data points. An understanding of this scenario will improve our interpretation of the cluster results in the forthcoming applied setting. When data are random, the behavior of our algorithm is, in fact, equivalent to the "continuum percolation problem," a well-researched topic in statistical physics (Broadbent and Hammersley 1957; Stauffer and Aharony 1994). The partitions identified by our algorithm undergo a sharp transition as the parameter d increases. Before the transition, most data points are isolated; after the transition, most points belong to the same large cluster. This phenomenon is known as the "percolation transition." Percolation is usually studied when data points are
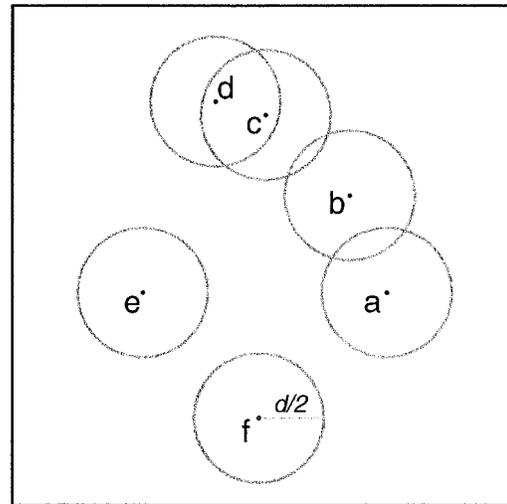

Figure 1: Schematic spatial arrangement of six data points in the plane. Points within a radius d are assigned to the same cluster; hence, points *a, b, c,* and *d* are in the same cluster. Points *e* and *f* are in separate clusters.
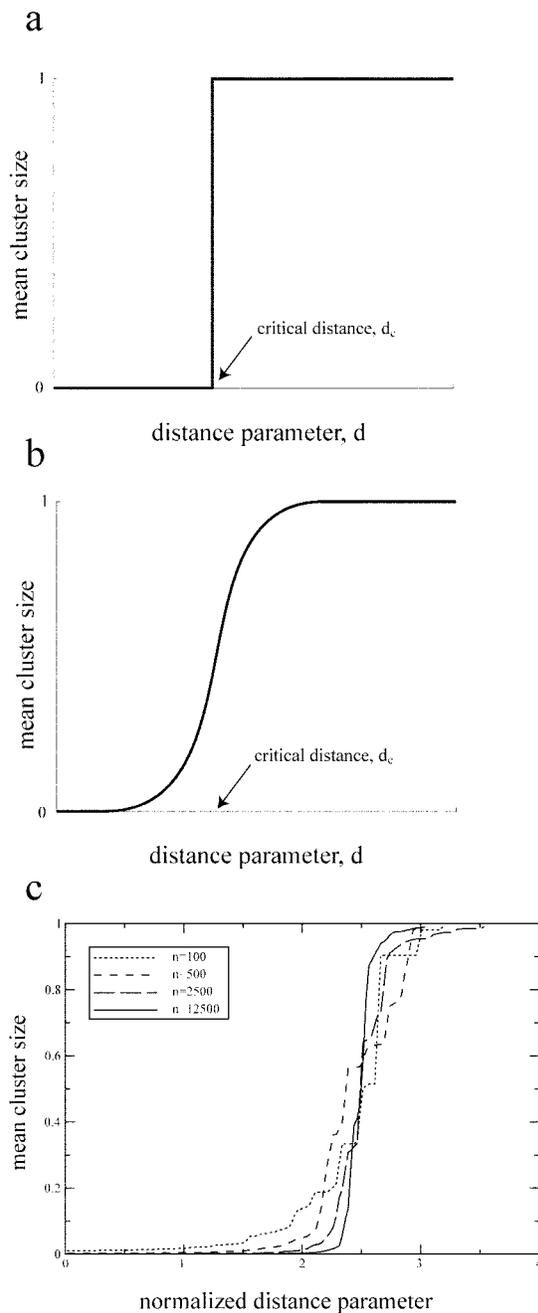
a

b

c

Figure 2: *a*, Idealized relationship between the distance parameter *d* and the mean cluster size. A sharp (critical) transition is predicted at a single distance d_c if the data points are randomly distributed in space and if the sample size is large. As *d* passes through the percolation threshold d_c, the data points become connected in space and the mean cluster size undergoes a phase transition from 0 to 1. *b*, Influence of finite sample size on the "continuum percolation" transition. The transition occurs at the same critical distance as in the large sample size limit (*a*), but the transition is not as sharp. *c*, Percolation behavior for simulated data. In each of the four data sets, points have been placed in space at random with density 0.1. The largest data set contains $n = 12,500$ points, and

constrained to remain on a discrete lattice (see, e.g., Keitt et al. 1997; and Keymer et al. 2000 for applications in ecology). In our case, however, tree locations are unconstrained, and the corresponding process is called "continuum percolation" (Hall 1985; Meester and Roy 1996).

As we vary the cluster parameter *d*, we obtain corresponding partitions of the data into disjoint clusters. For any given *d* value, we denote the number of resulting clusters by *m* and the size of the clusters by $c_1, c_2, \ldots, c_m$. Let $n = \sum c_i$ denote the total number of individuals in the sample. We can conveniently summarize a cluster arrangement by recording the mean cluster size, that is, the first moment of the cluster size distribution:

$$\langle c \rangle = \sum_{i=1}^{m} \frac{c_i}{n} \times c_i = \frac{1}{n} \sum_{i=1}^{m} c_i^2.$$

The mean cluster size equals the expected size of the cluster containing a randomly chosen individual. The normalized mean cluster size, $\langle c \rangle / n$, computes the probability that two randomly chosen data points lie in the same cluster.

Percolation theory predicts that below a certain critical distance, d_c, individuals will be distributed in many small clusters, while above this distance most individuals will belong to a single large cluster. The critical distance d_c at which the clusters undergo this transition is called the "percolation threshold." The transition will be increasingly sharp as number of individuals in the data sample increases. The percolation transition may also be stated in terms of the mean cluster size: the (normalized) mean cluster size of a very large random data set will equal 0 below the critical distance d_c and 1 above the critical distance. Figure 2 demonstrates the percolation transition by graphing the distance parameter *d* against the mean cluster size $\langle c \rangle$ for a simulated set of random data points.

In real data, as opposed to random data, we do not expect to find a single sharp transition at one critical distance. Instead, we may often discover several critical distances with varying degrees of sharpness. Such behavior is indicative of spatial clustering at several scales. As we vary the distance parameter *d*, producing different partitions of the data into clusters, we will detect critical distances by graphing *d* against the corresponding mean cluster size $\langle c \rangle$. Any range of distances *d* that corresponds to a plateau in the cluster size curve indicates a nonrandom scale of aggregation that is insensitive to perturbations in *d*. In general, the number of plateaus in the observed

the smallest data set, $n = 100$ points. As the sample size gets smaller, the transition becomes more gradual, as depicted schematically in *a*. See the appendix for a discussion of minimizing sample size artifacts.

cluster size curve will reflect the number of nonrandom scales of aggregation.

In order to compare cluster behavior across species with different abundances, we must normalize the distance parameter d. We will divide d by the quantity

$$\frac{1}{2}\sqrt{A/n},$$

which is the average nearest-neighbor distance for n randomly placed data points. The normalized parameter is denoted

$$\hat{d} = 2d\sqrt{n/A}.$$

Here A denotes the total sample area, and n denotes the abundance of the species. This normalization produces a nondimensional version of cluster parameter $\hat{d}$, suitable for interspecific comparison. In the limit of a large sample size and a random data set, the sharp transition will occur around $\hat{d} \approx 2.4$, regardless of the density of points.

Techniques for the Study of Habitat Associations

The association of species with habitat type was one of the first questions studied by tropical botanists (see, e.g., Richards 1936). This question was a primary motivation for many large-scale studies in Southeast Asia (Ashton 1964) and in the Neotropics (Gentry 1988). Such studies have stimulated the development of statistical methods for detecting "phytosociological units" or habitat associations.

In essence, statistical methods of habitat correlation query, with varying degrees of sophistication, whether or not a species is overrepresented in a particular habitat; χ^2 (Basnet 1992) and modified χ^2 (e.g., Harms 1996; Plotkin et al. 2000b) tests are the most common statistical analyses. Such techniques are undeniably useful in quantifying overall patterns of habitat associations. Yet their ability to characterize the specific effects of habitat on any particular species is modest: such tests simply inform us whether or not a species is associated (positively or negatively) with a given habitat. Moreover, such tests are only possible if complete and consistent abiotic measurements have been performed throughout the survey area. We will demonstrate how cluster analyses can complement existing statistical techniques. Cluster analysis can elucidate habitat associations for rare and common species alike, even in situations where more classical statistical analyses fail to detect or to characterize associations. The dual information of the location and the shape of a cluster is extremely useful in assessing potential habitat specializations.

Results

Identification of Spatial Clumps

As a preliminary example of our cluster algorithm, we first consider *Mallotus penangensis*, a species with a nonisotropic spatial distribution. Figure 3 shows the primary output of our algorithm. Cluster analyses are shown for two distances, $d = 12.4$ m and $d = 20.8$ m (fig. 3a). Figure 3b shows the relationship between the normalized cluster parameter $\hat{d}$ and the corresponding mean cluster size. Notice that the cluster behavior summarized in figure 3b is very different from the random case discussed above (fig. 2c). Instead of a single percolation threshold at $\hat{d} \approx 2.4$, *M. penangensis* displays two transitions occurring at approximately 1.5 and 8.0. The first transition occurs at a smaller distance than for random data and the second, at a much larger distance. The distance required to connect all points into a single cluster is much larger for the *Mallotus* data ($\hat{d} > 8$) than for the random data ($\hat{d} = 3$).

The *M. penangensis* clusters undergo their first transition at $\hat{d} \approx 1.5$, and the mean cluster size remains fairly constant on a plateau for a range of distances thereafter (fig. 3b). This indicates that the *M. penangensis* clusters are well defined over a large range of scales ($1.5 < \hat{d} < 6.0$). The two d values chosen in figure 3a correspond to $\hat{d} = 1.3$ and $\hat{d} = 2.2$; these two values demonstrate the cluster structure before and after the first transition. Figure 3b indicates that the right panel of figure 3a is a better clustering of the *M. penangensis* species than the left panel, in the sense that the clusters are robust to perturbation.

The precise relationship between cluster structure and distance parameter is best summarized by the graph in figure 3b. Nevertheless, figure 4 provides an alternative, three-dimensional summary of this relationship. The Z-axis of figure 4 corresponds to the distance parameter d. For convenience, clusters are represented schematically by circles (in fact, clusters are not symmetric). The top layer, $z = 0$, shows the locations of every tree in green. On the bottom layer ($d = 134$ m) all the trees are within the same large cluster, represented by a single circle. As d increases down the Z-axis, the number of clusters decreases and the cluster sizes increase. The critical transitions in this spectrum are visible in figure 4, but they are perhaps discerned more accurately by graphing the mean cluster size (fig. 3b).

Our analysis of *M. penangensis* highlights two important regimes of spatial patterns: stable regions in which the clusters are well defined and sensitive regions in which the cluster structure undergoes dramatic transitions. Depending on the ecological question at hand, either one of these regimes may be of interest. In the case of *M. penangensis*, the stable region most likely indicates the clustering that corresponds to dispersal-driven clumps of related trees. As
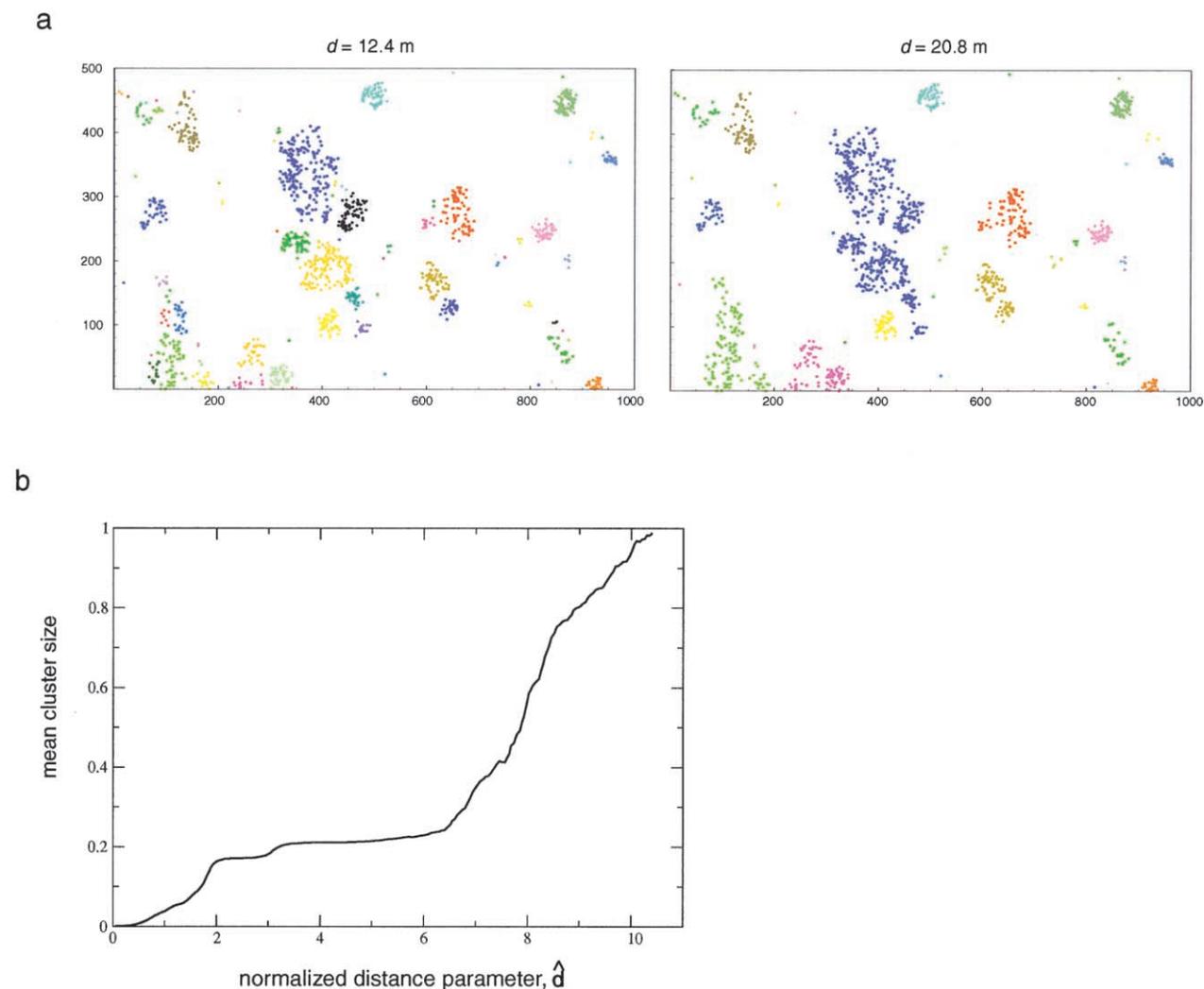
a

b

Figure 3: *a,* Cluster analysis of the understory species *Mallotus penangensis,* Euphorbiaceae. There are 1,404 stems of this species in the Pasoh 50-ha plot. When we set the clustering distance *d* equal to 12.4 m, the algorithm partitions the data into 75 distinct clusters, each plotted in a different color (*left panel*). Some colors may appear similar. Setting $d = 20.8$ m results in 34 clusters (*right panel*). *b,* Mean cluster size as a function of the normalized distance parameter for *M. penangensis.* The algorithm was run with a temperature $\chi = 0.08$ to minimize finite-size artifacts (see appendix for details).

we shall see, however, the critical regimes can also be useful for detecting the influences of habitat.

Habitat Associations

Pternandra coerulescens. The species *P. coerulescens* provides an instructive comparison with *M. penangensis.* Both species demonstrate essentially two critical distances (figs. 3*b,* 5*b*). For *P. coerulescens,* the transitions occur at $\hat{d} \approx$ 1.2 and $\hat{d} \approx 2.5$. The clusters of *P. coerulescens* shown in figure 5*a* occur in the stable region immediately after the

first critical distance. Unlike *M. penangensis,* the clusters of *P. coerulescens* are long, narrow, and linear in shape, with a few exceptions. Notice that both cluster geometries are simultaneously detected by our algorithm. *Pternandra coerulescens* also features a large number of singleton trees located far away from conspecifics. This feature, seen graphically in figure 5*a,* is reflected by the long tail of the cluster size curve in figure 5*b.* The tail is caused by the gradual assimilation of singleton trees into the main cluster at large distances *d.*

Pternandra coerulescens exhibits clusters with a very characteristic shape. The hypothesis that *P. coerulescens* is
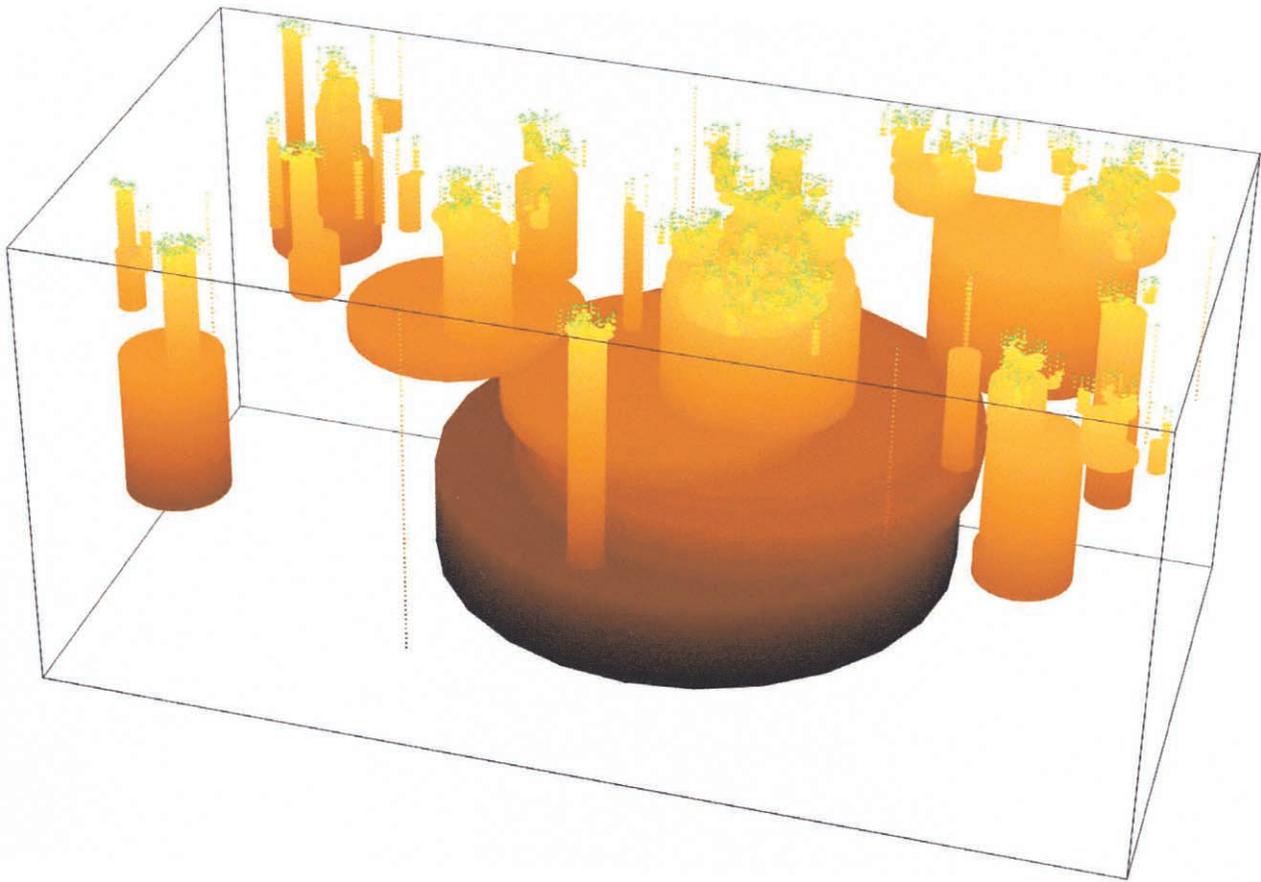
Figure 4: Dependence of clusters on the distance parameter *d* for *Mallotus penangensis*. The figure shows a three-dimensional graph of *M. penangensis* clusters. The *X*- and *Y*-axes of the figure correspond to the length and width of the 50-ha plot, as in figure 3*a*. The *Z*-axis corresponds to the value of *d*. The color also corresponds to the value of *d*. On the upper plane, $d = 0$, and each tree forms its own cluster, all plotted in green. As *d* increases down the *Z*-axis, larger clusters form. Each cluster is represented in the figure by a circle centered at the center of mass of the cluster. The radius of the circle equals the mean distance of the trees in the cluster to its center of mass.

associated with local streams, which stands out as the most natural in light of figure 5*b*, is confirmed by the natural history of the species (see "Study Species"). This suggests that the linear-shaped clusters are the direct result of *P. coerulescens*'s finicky habitat requirements. Indeed, *P. coerulescens* mainly grows wherever ground litter is minimal or absent. The latter observation was made during a tropical plant biomass experiment in 1971–1973 when all free-standing trees and litter were removed from a survey plot near Pasoh's 50-ha plot (Kira 1971). *Pternandra coerulescens*, normally a fairly rare species, colonized most of the logged region, presumably due to the complete absence of ground litter there (T. Kira, personal communication). In addition, *P. coerulescens* also clusters at the primary pig wallows in the plot (fig. 5*a*), another location with frequent litter disturbance. Despite the numerous, widespread,

small seeds produced by *P. coerulescens*, the species is severely limited by abiotic environmental factors. Since maps of Pasoh's local stream networks (and many other abiotic environmental variables) are not available, the present analysis would have been much more difficult, or impossible, via classical tests of habitat specificity.

Xerospermum noronhianum and *Knema laurina.* We turn now to the environmental influences on two of the most common Pasoh species: *Xerospermum noronhianum* and *Knema laurina*. In particular, we examine the effect of topography of their spatial distributions. A Ripley *K* analysis indicates that both of these species follow nonrandom spatial distributions (Plotkin et al. 2000*b*). The cause or nature of the nonrandomness is difficult to determine with the naked eye. Statistical tests indicate that neither of these
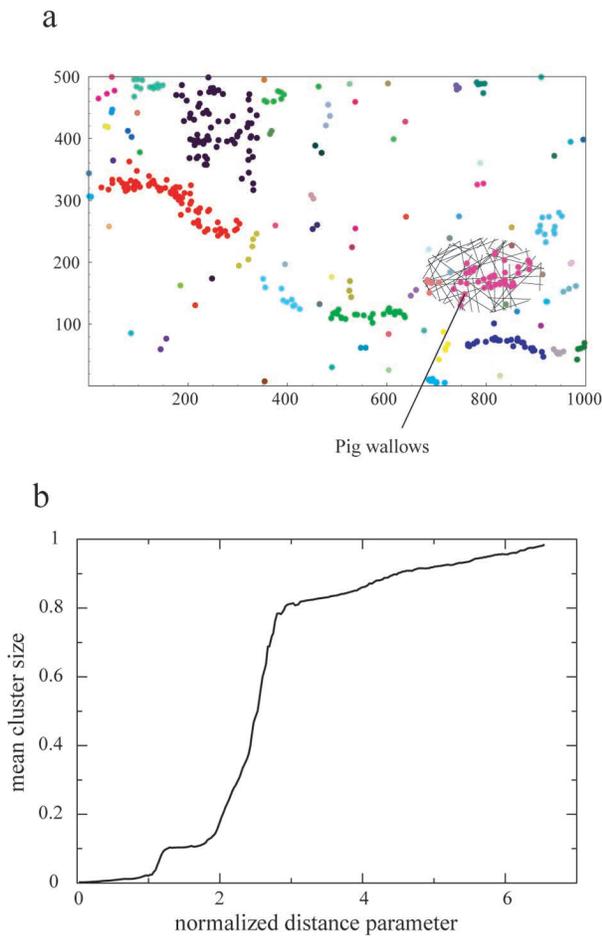
a

Pig wallows

b

Figure 5: *a*, Cluster analysis of *Pternandra coerulescens* that results from the distance parameter $d = 25.4$ m ($\hat{d} = 1.5$). There are 438 stems within the plot, which form 79 clusters. Most of the clusters of *P. coerulescens* are linear in shape. The arrow indicates the locations of pig wallows within the 50-ha plot. *b*, Mean cluster size curve for *P. coerulescens*. The algorithm was run with a positive temperature ($\chi = 0.08$) to minimize finite size effects. The resulting curve shows two principal transitions.

species is very strongly correlated with Pasoh's topography (Plotkin et al. 2000*b*). In particular, for each species, the total density of stems on the hill is roughly equal to the density of stems off the hill. Although topography does not limit the ranges of these species, it might, nevertheless, influence the spatial geometry of individual tree placement.

Figure 6*b* shows that topography does have an effect on the spatial distribution of these two species, despite the absence of a strong statistical correlation with topography. *Knema laurina* exhibits one large cluster whose boundaries clearly coincide with the detailed topography of the single
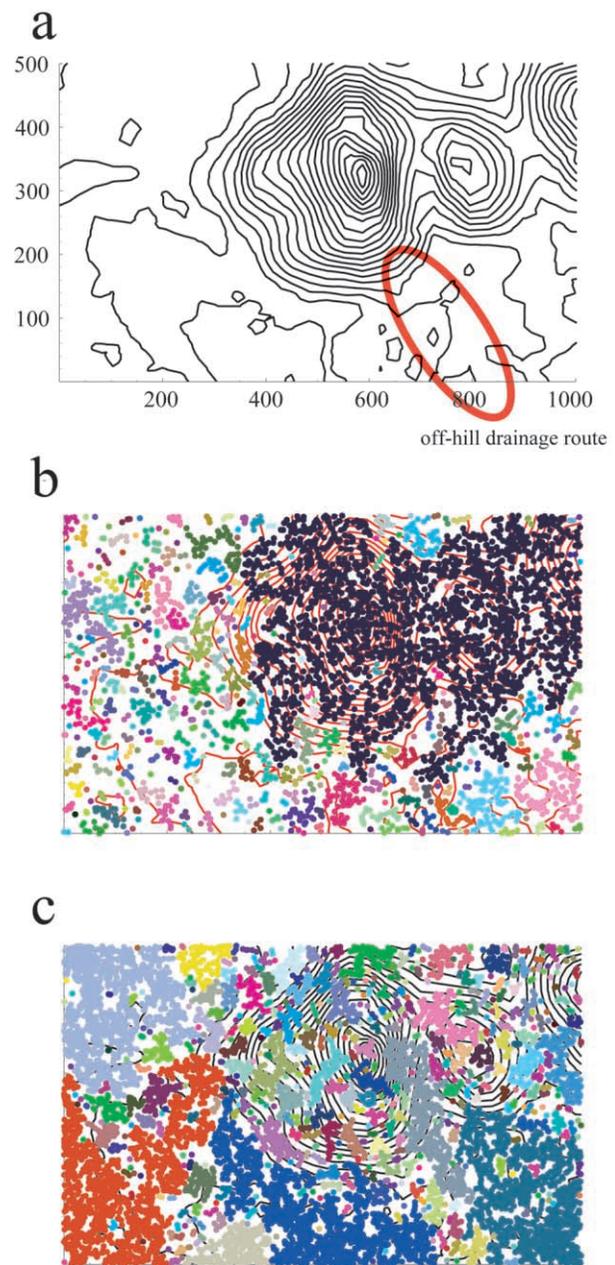
a

off-hill drainage route

b

c

Figure 6: *a*, Topographic contour map of the Pasoh 50-ha plot. The primary water drainage route running off the hill is indicated in red. *b*, Cluster analysis of *Knema laurina* for $d = 12.1$ m (330 disjoint clusters). Notice that *K. laurina* forms a single, stable cluster on the hill. *c*, Cluster analysis of *Xerospermum noronhianum* for $d = 9.0$ m (494 disjoint clusters). Notice that *X. noronhianum* tends to form larger, stable clusters off of the hill. The water drainage route apparently disrupts the *X. noronhianum* clusters. See figure 7 for the relationship between distance parameter *d* and mean cluster size.
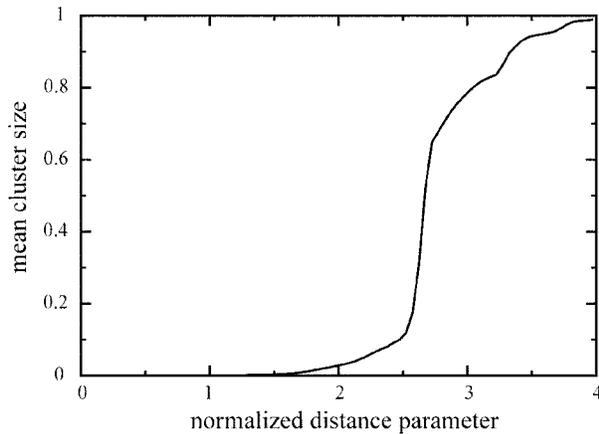
The ability to partition spatial data into clusters allows us to reexamine many classical questions about plant community dynamics, such as habitat associations. We can also now analyze community structure on a different scale than the census scale addressed by classical statistics. In particular, having identified specific tree clusters, it is natural to consider the range of observed cluster sizes as well as the within-cluster demographics.

Cluster Sizes. Null models of disturbance-driven communities would predict uniform cluster sizes within each species, reflecting the same history of disturbance. As is fairly clear from figures 3–7, there is a very wide range of cluster sizes within each species, both in terms of the minimum convex area containing a cluster and the number of trees belonging to each cluster. Figure 8 summarizes the distribution of cluster sizes (in terms of number of individuals) for *M. penangensis*, *P. coerulescens*, and *K. laurina*. Each of these species, as well as the vast majority of Pasoh species, demonstrates a wide range of cluster sizes, even if we discard extraneous clusters containing ≤ 5 individuals.

Distribution of Within-Cluster Tree Diameters. We now focus on the within-cluster demographics of our study species. Figure 9 shows the range of tree diameters within each of the four largest clusters of *M. penangensis*. For comparison, we also plot the dbh distribution of all *M. penangensis* stems in the entire 50-ha plot. Aside from the four or five largest trees, the clusters exhibit the same full
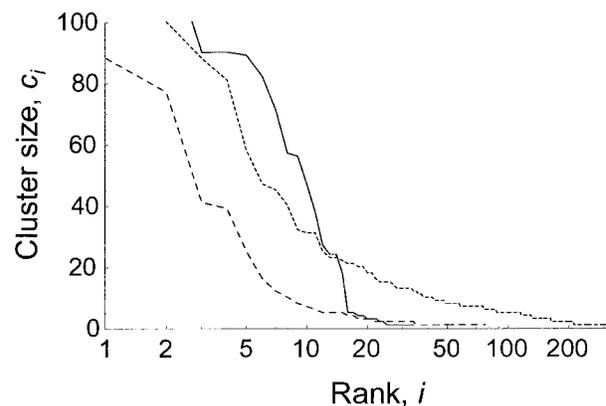
Figure 7: Mean cluster size curve for *Xerospermum noronhianum*. The algorithm was run with a positive temperature, $\chi = 0.05$ (see appendix for details). The critical value $\hat{d} = 2.4$ yields the cluster arrangement displayed in figure 6c. The mean cluster size curve for *Knema laurina* (not shown) is very similar.

hill within the Pasoh plot. *Knema laurina*'s cluster geometry is radically different off the hill as compared to on the hill. Off the hill, *K. laurina* forms tighter clumps that are more separated from each other than those on the hill. Since the total densities of trees on and off the hill are roughly equal, the within-clump density of stems is higher for clusters off the hill, an intriguing behavioral pattern whose underlying mechanism we cannot yet fully elucidate.

Although the cluster behavior of *X. noronhianum* is less dramatic than that of *K. laurina*, it also evidences a clear response to the topographic gradient. In the case of *X. noronhianum*, the clusters off the hill are large, stable, and evenly spaced. On the hill, *X. noronhianum* exhibits tighter aggregation patterns. In fact, the cluster geometry of *X. noronhianum* is also correlated with the main drainage streams running off the hill (as indicated in fig. 6c). Figure 7 shows the cluster size curve for *X. noronhianum*. The effect of topography on cluster geometry has been revealed by inspecting a critical value of $\hat{d} = 2.4$ ($d = 9.0$ m).

As figures 6 and 7 reveal, the spatial cluster geometries of *X. noronhianum* and *K. laurina* are clearly influenced by topography. It is unlikely that these species suffer from a dramatic difference in dispersal limitation depending on their location on or off the hill. Thus, habitat is probably the primary influence on the particular spatial arrangement of these two common species, despite their constitutive presence throughout the plot.

Figure 8: Distribution of cluster sizes for *Mallotus penangensis* (*dashes*), *Pternandra coerulescens* (*solid line*), and *Knema laurina* (*dots*). The figure shows the rank cluster size distribution on semilogarithmic axes. All three species demonstrate a full range of cluster sizes. (The largest cluster of *K. laurina*, containing 2,223 trees, is not visible within these axes.)
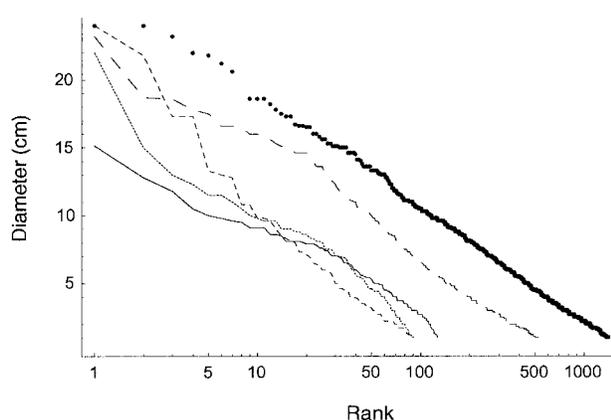
Figure 9: Distribution of tree diameters within the four largest *Mallotus penangensis* clusters, compared with the distribution of *M. penangensis* diameters in the entire 50-ha census (*large dots*). The figure shows a rank dbh plot on a semilogarithmic axis. The individual clusters of *M. penangensis* exhibit nearly the same full range of dbh classes as the 50-ha plot as a whole.

range of diameters as the 50-ha census taken altogether. Moreover, the dbh distribution is approximately exponential at both the cluster scale and the full 50-ha scale. These results indicate that the identified clusters of *M. penangensis* represent full-fledged microcommunities of trees, containing juvenile and mature stems alike. We do not find any examples of purely juvenile or purely adult trees, which would be obvious signatures of a system away from equilibrium. Instead, the clusters of *M. penangensis* are in equilibrium, in that we do not expect their diameter distribution to change significantly in time.

A full range of intracluster diameters is common among those Pasoh species that demonstrate obvious aggregation patterns. The trend demonstrated by *M. penangensis* also holds for our other study species. Species *Neobalanocarpus heimii* is a rare counterexample: its large clusters demonstrate an elevated proportion of juveniles, compared with the dbh distribution in the plot as a whole. The unusual nonequilibrium distribution of *N. heimii* reflects its characteristic ability to regenerate quickly in small gaps.

Cluster Structure and Tree Diameters. In addition to the range of within-cluster tree diameters, we are also interested, naturally, in the relative locations of large and small trees within a cluster. In particular, are clusters characterized by one or several large "mother trees" located near the cluster center? Or are large trees more often found on the periphery of clusters? Such questions are relevant to an understanding of those processes behind cluster formation.

We have found examples of both trends within the Pasoh plot. Many species exhibit random within-cluster tree placement, with respect to tree diameter. Many other species exhibit an elevated proportion of large peripheral trees. And only a few species exhibit large central trees. This finding, which we document in two particular instances below, may be somewhat surprising compared with the simplistic view of a dispersal-driven cluster with central large "mother trees." *Shorea macroptera* exhibits an elevated proportion of peripherally located large trees. Figure 10*a* shows a cluster analysis of *S. macroptera* in which the size of each circle corresponds to the tree's diameter. It is not immediately apparent to the eye whether large trees are central or peripheral. With the explicit knowledge of the cluster partition, however, we can test for statistical correlations between tree diameter and within-cluster placement.

We define the center of a cluster to be the average X and Y coordinate of all those trees in the cluster. This definition assumes some degree of isotropy. For each cluster, we inspect the relationship between a tree's dbh and its distance to the center of its cluster. We normalize this distance by dividing by the mean distance to the center among all trees in the cluster. Finally, we calculate the Pearson's correlation coefficient of the relationship between dbh and distance to center for each cluster.

The correlation between tree diameter and distance to center is shown for each cluster of *S. macroptera* in figure 10*b*. Notice that most of the large clusters demonstrate a positive correlation. Pooling the dbh distance data from the 10 largest clusters of *S. macroptera* together, the positive correlation between dbh and distance is statistically significant at the 95% confidence level: larger trees tend to lie on the periphery of their cluster. The same finding holds for many other Pasoh species, including, for example, *Dipterocarpus cornutus*, *Shorea guiso*, *Schoutenia accrescens*, and *Eugenia claviflora*.

Figure 11 demonstrates a parallel analysis for *N. heimii*. In this case, however, the large clusters demonstrate a negative correlation between dbh and distance to cluster center. Pooling the data from the 10 largest clusters together, there is again a statistically significant result at the 95% level: larger trees tend to be found closer to the center of the cluster. Among the Pasoh species, *N. heimii* is a somewhat rare example of this central large tree phenomenon. The tendency for large trees to be located in the center of clusters, as opposed to the periphery, is indicative of strong intraspecific competition. Despite an abundant seed rain around many large, reproductive, locally dispersed trees, seedlings have difficulty establishing themselves in the nearby vicinity of large conspecifics and find themselves marginalized on the periphery. Most of this intraspecific competition occurs before stems reach 1 cm diameter
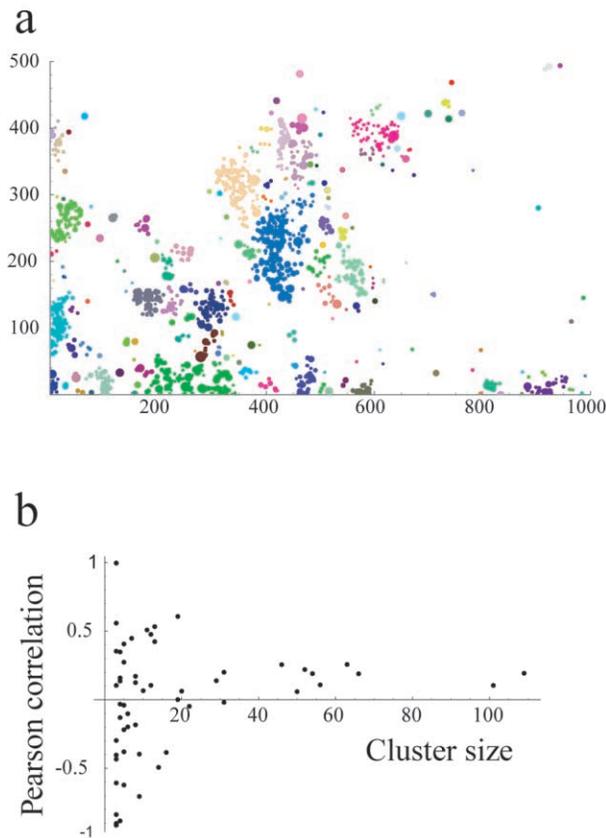
reproductively for centuries. Because of their large wingless fruit, most juveniles do not survive the stifling dark parental subcanopy light climate. As a result, there are several large central stems of *N. heimii* surrounded by a range of aspiring juveniles on the perimeter, most of which will not survive. This interpretation agrees with our earlier observation of an elevated proportion of juveniles found in clusters of *N. heimii*.

The seeds of *S. macroptera*, by contrast, are well dispersed, and the growth rates are also higher. Moreover, adult stems of *S. macroptera* do not persist as long as *N. heimii*. As a result, old clumps of *S. macroptera* may have new clumps growing in gaps and around their perimeters; we therefore find a positive correlation between dbh and distance to center.

It is also interesting to compare the cluster demograph-

Figure 10: *a*, Cluster analysis of *Shorea macroptera* corresponding to $\hat{d} = 1.4$, resulting in 171 clusters. The radius of each circle represents the diameter of the corresponding tree, although not to scale. *b*, Pearson correlation coefficient for the within-cluster relationship between dbh and distance to cluster center for all *S. macroptera* clusters. The *X*-axis denotes the size of the cluster, and the *Y*-axis denotes the correlation coefficient for the trees within the cluster. Notice that most of the large clusters exhibit a positive correlation between dbh and distance to center; large *S. macroptera* trees are found toward the periphery of their cluster.

(Wills and Condit 1999). Thus, within-cluster demographics may be a strong indicator of intraspecific forces and even Janzen-Connell effects (Janzen 1970; Connell 1971).

It is interesting to note that *N. heimii* and *S. macroptera* have radically different dispersal mechanisms (see "Study Species"). The dispersal differences between these two dipterocarps are clearly correlated with the different spatial population structures of their clusters (fig. 10 vs. fig. 11). Such an observation would perhaps be predicted from an understanding of life-history traits and from visual inspection but can only be verified rigorously by a complete cluster analysis. In the case of *N. heimii*, trees that succeed to maturity then escape competition and continue to live
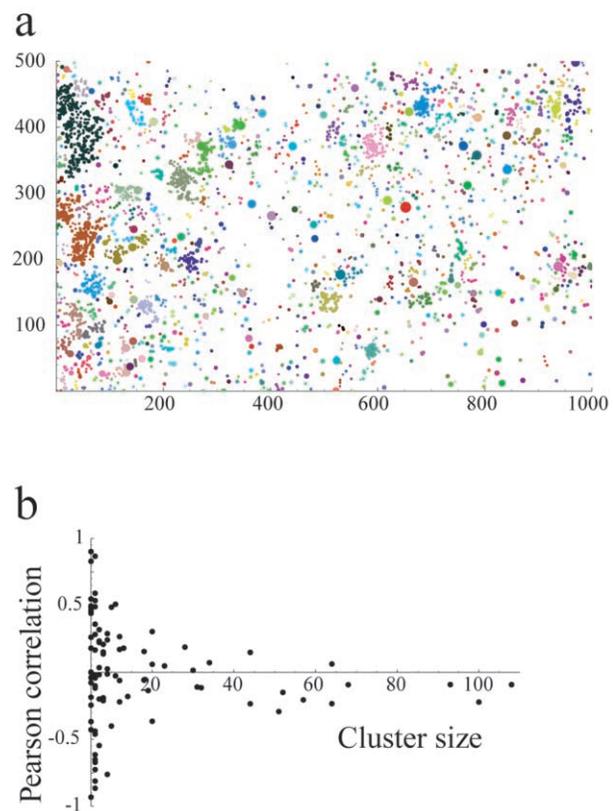

Figure 11: *a*, Analysis of *Neobalanocarpus heimii* parallel to figure 10. The cluster algorithm was run at $\hat{d} = 1.1$, resulting in 1,200 clusters. *b*, Pearson correlation coefficient for the within-cluster relationship between dbh and distance to cluster center for all *N. heimii* clusters. Notice that most of the large clusters exhibit a negative correlation between dbh and distance to center. In other words, large trees of *N. heimii* are found near the center of their cluster.

ics of *N. heimii* with those of the ballistically dispersed *M. penangensis*. On the face of it, these two species might be expected to demonstrate similar within-cluster spatial structure. However, *M. penangensis* does not, in fact, exhibit any significant correlation (positive or negative) between distance and size, in contrast to *N. heimii*'s negative correlation. This observation likely reflects the fact that *N. heimii* does not reproduce until its crown has emerged above the canopy, perhaps after at least 50 yr. *Mallotus penangensis*, which also is a slow grower, nevertheless starts reproducing when quite small, so that a range of sizes are dispersing seeds within the cluster, resulting in no consistent space-diameter correlations.

In summary, we have seen that there is a large range of cluster sizes within each species and that clusters generally exhibit a full range of tree sizes, whose distribution is similar to the dbh distribution found in the 50-ha plot as a whole. In addition, many species exhibit large trees on the periphery of their clusters, few species exhibit clusters centered around large trees, and many species exhibit no consistent spatial size class structure. And finally, many of these trends reflect the different dispersal mechanisms and growth patterns between species.

Discussion

Ecological Implications

Ecologists address questions of community dynamics on many spatial scales: individuals, censuses, regions, and landscapes. Once spatial data have been partitioned into clusters, however, we can inspect community dynamics on two new scales: the intracluster and intercluster scales. In our case, clustering has allowed us to examine several issues of importance in tropical plant ecology: the correlation between cluster locations and abiotic environmental variables, the distribution of cluster sizes, the distribution and spatial placement of within-cluster tree diameters.

This technique has allowed us to unveil hidden patterns (for *Xerospermum noronhianum* and *Knema laurina*) whose discovery might otherwise have been impossible. Such discoveries suggest that habitat may have a stronger effect on the spatial distribution of common species—even species that are present constitutively on all habitat types—than standard statistical analyses would suggest. In addition, cluster analysis of species such as *Pternandra coerulescens* demonstrate that spatial distribution can sometimes be driven almost entirely by abiotic habitat specificity, regardless of dispersal capability.

At the same time, our investigation of cluster demographics has revealed that dispersal mechanisms and growth patterns can play a large role in determining intracluster spatial population structure. Species-level variation in regeneration speed is reflected by variation in the range of tree diameters found within clusters. Gap colonizers (e.g., *Neobalanocarpus heimii*) may be identified, from static data alone, by a signature elevated proportion of purely juvenile clusters. Differences in dispersal mechanism (e.g., *Shorea macroptera* vs. *N. heimii*) are associated with characteristic differences in the placement of large trees within a cluster. Similarly, differences in growth patterns (e.g., *Mallotus penangensis* vs. *N. heimii*) are also reflected by within-cluster tree placement; species that reproduce throughout their lifetime may lack any correlation between diameter and intracluster placement.

In short, we have found that the diversity of tropical tree life-historical traits and of the diversity of responses to abiotic influences are reflected by the diversity of their spatial arrangements. Thus our study species evidence both niche-based and dispersal-based processes affecting spatial distributions, at different spatial scales.

We emphasize that, despite the analysis of our study species, our cluster method does not allow us to infer process unambiguously from pattern. We cannot soundly argue to have teased apart the influences of niche- and dispersal-based processes in general. In the end, we view the techniques developed here more as tools for exploratory data analysis and for generating hypotheses. Such hypotheses (e.g., the disequilibrium dynamics of *N. heimii*) must eventually be tested by field and lab experiments. Nevertheless, we have seen that careful direct observation of species' maps can, when combined with a clustering algorithm, provide a powerful method for ecological exploration.

Cluster Analysis and Spatial Statistics

We have developed a fairly general method for analyzing spatial distributions that is conceptually different from the classical paradigm of spatial statistics. Spatial statistics have been used to assess mean clumping behavior across the entire range of species in order to document universal trends of aggregation in forests. The cluster techniques developed here complement such gross statistics.

One might argue that a formal clustering technique provides little more information than an automated mapping device. The human eye is, in fact, a remarkably efficient and complex tool for detecting spatial patterns. The methodology developed here, however, permits us to reframe specific questions of tree dynamics in terms of spatial scales and to identify the relevant critical scale(s) for a given species. Several critical scales indicate multiple factors that influence a species' distribution. There is no single "correct

answer" to partitioning spatial data but, rather, a spectrum of answers among which our method guides us toward the critical and stable solutions.

Unlike most spatial statistics, our method of identifying clusters does not make a priori assumptions about underlying spatial structure. The edge-correction factors used to compute spatial statistics, by contrast, typically assume that the underlying stochastic point process is isotropic or, at least, stationary (Ripley 1976). We have seen, however, that abiotic influences often cause large departures from stationarity or isotropy in tropical tree distributions. Departures from these assumptions are also common in other spatial settings (Diggle 1983).

The primary distinction between cluster analysis and spatial statistics is that the former explicitly separates data points into disjoint subsets. Formal clustering is better used as a tool for exploratory data analysis of the specific spatial geometries in a census; spatial statistics, by contrast, are more limited and require more assumptions but allow for rigorous statistical tests of departure from spatial randomness (He et al. 1997; Condit et al. 2000; Plotkin et al. 2000*b*).

Unlike gross statistics, the identification of explicit clusters allows for further inquiry into the spatial geometry of a data set. In our case, clustering has allowed us to examine several questions of ecological importance: the correlation between cluster locations and abiotic environmental variables, the distribution of cluster sizes, and the distribution and spatial placement of within-cluster tree diameters. Such analysis of specific ecological and demographic questions in light of the known life-historical strategies would be impossible without partitioning the data set into clusters.

Our method for identifying scales of aggregation and clusters is general, and it can be applied in contexts other than tropical plant ecology. For example, spatiogenetic analysis of plants in general may benefit from a formal clustering methodology to complement statistical techniques that are currently employed (Ouborg et al. 1999). In that setting, such a methodology would allow researchers to contrast the genetic variation within and between clusters. Our method of identifying scales of aggregation can also be applied outside of a strictly spatial setting, for example, for the identification of quasi species within a data set of RNA virus sequences (Plotkin et al. 2002).

APPENDIX

Algorithmic Details

In this appendix we detail the algorithmic implementation of our data clustering method.

Algorithmic Implementation

Our basic clustering algorithm is extremely simple. We first fix the clustering parameter d, and we connect two trees if their pairwise distance is less than d. Next, we identify the resulting connected components (clusters). The latter stage is critical from a computational viewpoint. A naive implementation of this step would involve a double loop over all the data points. For n trees, the algorithm would require on the order of n^2 operations. An efficient alternative for identifying connected components has been developed by Hoshen and Kopelman (1976), and it requires $n \ln(n)$ operations.

Minimizing Sample Size Effects

Our basic algorithm can be modified to reduce the noise associated with finite sample sizes (as seen, e.g., in fig. 2*c*). In the original algorithm, we first fix the clustering parameter d, and we connect two trees if their pairwise distance is less than d. In the modified algorithm, we assume a small amount of noise in the parameter d, and we average over many replicas of the basic algorithm. The noise is somewhat analogous to temperature in statistical mechanics. For a fixed mean value of $\langle d \rangle$, we compare pairwise interevent distance to independent draws of d from a normal distribution with mean $\langle d \rangle$ and standard deviation $\chi \times \langle d \rangle$ (the normal distribution is naturally truncated at $d = 0$). Our simplest algorithm corresponds to the deterministic case $\chi = 0$. When χ is positive, however, the algorithm is no longer deterministic. For each replica, the radius around each data point varies slightly. We first

fix $\langle d \rangle$, and we average the resulting mean cluster size over multiple runs of the stochastic algorithm. This averaging procedure reduces the finite-size artifacts. Empirical investigations suggest that $\chi \approx .05 \pm .03$ is a good range of values for minimizing noise in ecological data.

As an aside, we mention one important technicality regarding the stochastic version of our algorithm. The critical distance d_c studied extensively in physics and mathematics depends on the value of χ. In fact, simulations suggest that as χ ranges from 0 to 0.25, the normalized critical distance of the Poisson process varies from to 2.44 to 2.18. Fortunately, a result of Meester and Roy (1996) guarantees that d_c depends continuously on χ. (Meester and Roy [1996] prove that if two series of distributions for the radii converge weakly, then their corresponding critical distances converge as well.) In most theoretical studies of percolation, χ is assumed to be 0.

Other Clustering Algorithms

There exists a large variety of clustering algorithms (k means, fuzzy k means [Duda et al. 1998], neural k means [Kohonen 1989], nearest neighbor, furthest neighbor, centroid, Ward, hypervolume, minimal spanning tree, Mojena's upper tail, Wolfe's test) designed to partition data into clumps (Everitt 1993). Many of these techniques require explicit or implicit a priori assumptions about cluster shapes or the total number of clusters. In an extensive comparison of clustering techniques, Hardy (1996) concluded that the implicit assumptions of most algorithms often lead to erroneous cluster classifications. Our method of clustering is closely related to the so-called single-linkage hierarchical method. Our method is distinguished, however, by the introduction of a temperature, or statistical averaging, into the cluster calculation. Moreover, the connection between single-linkage clustering and continuum percolation provides an important tool for analyzing the spectrum of possible cluster arrangements and for identifying either the critical or the stable solutions.

A new generation of clustering algorithms, also inspired by percolation theory, has recently appeared in the physics literature (Rose et al. 1990; Blatt et al. 1997; Angelini et al. 2000). The unifying concept of these methods is to optimize the temperature variable χ for each value of the average distance $\langle d \rangle$. Statistical mechanics can be used to solve this problem in a broad class of dynamical systems (coupled map lattices and superparamagnetic systems), and the resulting algorithm is, therefore, powerful and, in a sense, optimal. We have verified that one of these algorithms (SPC, provided by E. Domany) gives very similar results to our more simple algorithm. We believe that the relative simplicity of our algorithm is valuable in the present context.

Literature Cited

Angelini, L., F. DeCarlo, C. Marangi, M. Pellicoro, and S. Stramaglia. 2000. Clustering data by inhomogenous chaotic map lattices. Physical Review Letters 85: 554–557.

Ashton, P. S. 1964. Ecological studies in the mixed dipterocarp forests of Brunei state. Oxford Forestry Memoirs 25. Clarendon, Oxford.

———. 1969. Speciation among tropical forest trees: some deductions in the light of recent evidence. Biological Journal of the Linnean Society 1:155–196.

———. 1976. Mixed dipterocarp forest and its variation with habitat in the Malayan lowlands: a re-evaluation at Pasoh. Malaysian Forester 39:56–72.

———. 1998. Niche specificity among tropical trees: a question of scales. Pages 491–514 *in* D. M. Newbery, H. H. T. Prins, and N. D. Brown, eds. Dynamics of tropical communities. Blackwell Science, Oxford.

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., and D. Maguire. 1998. Modeling the spatial structure of tropical forests. Forest Ecology and Management 100:293–314.

Blatt, M., S. Wiseman, and E. Domany. 1997. Data clustering using a model granular magnet. Neural Computation 9:1805–1842.

Broadbent, S. R., and J. M. Hammersley. 1957. Percolation processes. Proceedings of the Cambridge Philosophical Society 53:629–641.

Condit, R., P. S. Ashton, P. Baker, S. Bunyavejchewin, S. Gunatilleke, N. Gunatilleke, S. P. Hubbell, et al. 2000. Spatial patterns in the distribution of tropical tree species. Science (Washington, D.C.) 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. Pages 298–312 *in* P. den Boer and G. Gradwell, eds. Dynamics of populations. Center for Agricultural Publishing and Documentation, Wageningen, Netherlands.

Cressie, N. 1991. Statistics for spatial data. Wiley, New York.

David, F. N., and P. G. Moore. 1954. Notes on contagious distributions in plant populations. Annals of Botany 18: 47–53.

Diggle, P. 1983. Statistical analysis of spatial point patterns. Academic Press, London.

Duda, R., P. Hart, and D. Stork. 1998. Pattern classification. 2d ed. Wiley, New York.

Everitt, B. S. 1993. Cluster analysis. Edward Arnold, London.

Fisher, R. A., H. G. Thornton, and W. A. Mackenzie. 1922. The accuracy of the plating method of estimating the density of bacterial populations. Annals of Applied Biology 9:325–359.

Gentry, A. H. 1988. Changes in plant community diversity and floristic composition on environmental and geographical gradients. Annals of the Missouri Botanical Garden 75:1–34.

Grubb, P. 1977. The maintenance of species richness in plant communities: the importance of the regeneration niche. Biological Reviews 53:107–145.

Hall, P. 1985. On continuum percolation. Annals of Probability 13:1250–1266.

Hardy, A. 1996. On the number of clusters. Computational Statistics and Data Analysis 23:83–96.

Harms, K. 1996. Habitat-specialization and the seed dispersal–limitation in a neotropical forest. Ph.D. diss. Princeton University, Princeton, N.J.

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

He, F., P. Legendre, and J. V. LaFrankie. 1997. Distribution patterns of tree species in a Malaysia tropical rain forest. Journal of Vegetation Science (Washington, D.C.) 8:105–114.

Hoshen, J., and R. Kopelman. 1976. Percolation and cluster distribution. I. Cluster multiple labeling technique and critical concentration algorithm. Physical Review B 1:3438–3445.

Hubbell, S. P. 1979. Tree dispersion, abundance and diversity in a dry tropical forest. Science 203:1299–1309.

———. 1997. A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. Coral Reefs 16:S9–S21.

———. 2001. The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, N.J.

Hubbell, S., and R. Foster. 1983. Diversity of canopy trees in neotropical forest and implications for conservation. Pages 25–41 *in* S. Sutton, T. Whitmore, and A. Chadwick, eds. Tropical rain forest: ecology and management. Blackwell Scientific, London.

Janzen, D. H. 1970. Herbivores and the numbers of tree species in tropical forests. American Naturalist 104:501–528.

Keitt, T. H., D. L. Urban, and B. T. Milne. 1997. Detecting critical scales in fragmented landscapes. Conservation Ecology 1(1). http://www.consecol.org/Journal/vol1/iss1/art4/index.html.

Keymer, J. E., P. A. Marquet, J. X. Velasco-Hernandez, and S. A. Levin. 2000. Extinction thresholds and metapopulation persistence in dynamic landscapes. American Naturalist 156:478–494.

Kira, T. 1971. Biomass and NPP for Pasoh research station, Malaysia. Oak Ridge National Laboratory server: http://www-eosdis.ornl.gov/DAAC.

Kochummen, K. M., J. V. LaFrankie, and N. Manokaran. 1990. Floristic composition of Pasoh Forest Reserve, a lowland rain forest in Peninsular Malaysia. Journal of Tropical Forest Science 3:1–13.

Kohonen, T. 1989. Self-organization and associative memory. 3d ed. Springer Information Sciences Series. Springer, New York.

Levin, S. A., B. T. Grenfell, A. Hastings, and A. S. Perelson. 1997. Mathematical and computational challenges in population biology and ecosystem science. Science (Washington, D.C.) 275:334–343.

Lloyd, M. 1967. Mean crowding. Journal of Animal Ecology 36:1–30.

Manokaran, N., and K. M. Kochummen. 1987. Recruitment, growth and mortality of tree species in a lowland dipterocarp forest in Peninsular Malaysia. Journal of Tropical Ecology 3:315–330.

Manokaran, N., and J. V. LaFrankie. 1990. Stand structure of Pasoh Forest Reserve, a lowland rain forest in Peninsular Malaysia. Journal of Tropical Forest Science 3:14–24.

Meester, R., and R. Roy. 1996. Continuum percolation. Cambridge University Press, Cambridge.

Ouborg, N. J., Y. Piquot, and V. Groenendael. 1999. Population genetics, molecular markers and the study of dispersal in plants. Journal of Ecology 87:551–568.

Pielou, E. C. 1959. The use of point-to-plant distances in the pattern of plant populations. Journal of Ecology 28:575–584.

Plotkin, J. B., M. D. Potts, D. W. Yu, S. Bunyavejchewin, R. Condit, R. Foster, S. Hubbell, et al. 2000*a*. Predicting species diversity in tropical forests. Proceedings of the National Academy of Sciences of the USA 97:10850–10854.

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

Plotkin, J. B., J. Dushoff, and S. A. Levin. 2002. Hemagglutinin sequences clusters and the antigenic evolution of influenza A virus. Proceedings of the National Academy of Sciences of the USA 99:6263–6268.

Pollard, J. H. 1971. On distance-estimators of density in randomly distributed forests. Biometrics 27:991–1002.

Richards, P.W. 1936. Ecological observations on the rain

forest of Mount Dulit, Sarawak. I, II. Journal of Ecology 24:1–37, 24:233–250.

Ripley, B. D. 1976. The second-order analysis of stationary point processes. Journal of Applied Probability 13: 255–266.

Rose, K., E. Gurewitz, and G. C. Fox. 1990. Statistical mechanics and phase transitions in clustering. Physical Review Letters 65:945–948.

Stauffer, D., and A. Aharony. 1994. Introduction to percolation theory. 2d ed. Taylor & Francis, London.

Wills, C., and R. Condit. 1999. Similar non-random processes maintain diversity in two tropical rainforests. Proceedings of the Royal Society of London B, Biological Sciences 266:1445–1452.

Wong, Y. K., and T. C. Whitmore. 1970. On the influence of soil properties on species distribution in a Malayan lowland dipterocarp forest. Malayan Forester 33:42–54.

Yap, S. K. 1976. Ph.D. thesis. University of Malaya.