Ecology, 88(12), 2007, pp. 3088–3102

ANALYZING THE SPATIAL STRUCTURE OF A SRI LANKAN TREE SPECIES WITH MULTIPLE SCALES OF CLUSTERING

THORSTEN WIEGAND,[1,4] SAVITRI GUNATILLEKE,[2] NIMAL GUNATILLEKE,[2] AND TOSHINORI OKUDA[3,5]

[1]*UFZ Helmholtz Centre for Environmental Research—Umweltforschungszentrum, Department of Ecological Modeling, PF 500136, D-04301 Leipzig, Germany*
[2]*Department of Botany, Faculty of Science, University of Peradeniya, Peradeniya 20400 Sri Lanka*
[3]*National Institute for Environmental Studies, Environmental Biology Division, Tsukuba, Ibaraki 305-8506 Japan*

Abstract. Clustering at multiple critical scales may be common for plants since many different factors and processes may cause clustering. This is especially true for tropical rain forests for which theories explaining species coexistence and community structure rest heavily on spatial patterns. We used point pattern analysis to analyze the spatial structure of *Shorea congestiflora*, a dominant species in a 25-ha forest dynamics plot in a rain forest at Sinharaja World Heritage Site (Sri Lanka), which apparently shows clustering at several scales. We developed cluster processes incorporating two critical scales of clustering for exploring the spatial structure of *S. congestiflora* and interpret it in relation to factors such as competition, dispersal limitation, recruitment limitation, and Janzen-Connell effects.

All size classes showed consistent large-scale clustering with a cluster radius of ~25 m. Inside the larger clusters, small-scale clusters with a radius of 8 m were evident for recruits and saplings, weak for intermediates, and disappeared for adults. The pattern of all trees could be divided into two independent patterns: a random pattern (nearest neighbor distance > 8 m) comprising ~12% of the trees and a nested double-cluster pattern. This finding suggests two independent recruitment and/or seed dispersal mechanisms. Saplings were several times as abundant as recruits and may accumulate several recruit generations. Recruits were only weakly associated with adults and occupied about half of the large-scale clusters, but saplings almost all. This is consistent with recruitment limitation. For ~70% (95%) of all juveniles the nearest adult was less than 26 m away (53 m), suggesting a dispersal limitation that may also be related to the critical large-scale clustering.

Our example illustrates the manner in which the use of a specific and complex null hypothesis of spatial structure in point pattern analysis can help us better understand the biology of a species and generate specific hypotheses to be further investigated in the field.

Key words: Janzen-Connell; multiple clustering; pair correlation function; point pattern analysis; Ripley's K *function;* Shorea congestiflora; *Sinharaja Forest Dynamics Plot, Sri Lanka; spatial point processes.*

INTRODUCTION

Patchiness, or the degree to which plant individuals are aggregated or dispersed, co-determines how a species uses resources, how it is used as a resource, and how it reproduces (Condit et al. 2000). In ecology there has been an increasing interest in the study of spatial patterns (e.g., Turner 1989, Levin 1992, Dale 1999, Liebhold and Gurevitch 2002). Spatial patterns have been a particularly important theme in tropical ecology and theories for explaining species coexistence and community structure rest heavily on spatial patterns. For example, niche assembly theories hypothesize that environmental heterogeneity and biological interactions may cause spatial clustering (Ashton 1969, Grubb 1977), dispersal–assembly theories predict that dispersal limitation can account itself for the emergence of spatial clustering (Wong and Whitmore 1970, Hubbell 1997, 2001), and the Janzen-Connell hypothesis (Janzen 1970, Connell 1971) predicts that wide dispersion and transportation of seeds away from parent plants is essential in avoiding the detrimental influence of pathogens, herbivores, seed predators, and seedling competition.

Prerequisite to the evaluation and testing of ecological theories regarding spatial patterns are methodologies for describing and analyzing spatial patterns. Methods for spatial pattern analysis have undergone a rapid development (Ripley 1981, Stoyan and Stoyan 1994, Dale 1999, Diggle 2003, Møller and Waagepetersen 2003). Point patterns, i.e., data sets consisting of mapped locations of plants, are especially important in plant ecology since plants can be approximated in many circumstances as points (but see Wiegand et al. 2006).

Manuscript received 10 August 2006; revised 16 January 2007; accepted 21 March 2007. Corresponding Editor: N. G. Yoccoz.

[4] E-mail: thorsten.wiegand@ufz.de

[5] Present address: Graduate School of Integrated Arts and Sciences, Hiroshima University, 1-7-1 Higashi-Hiroshima 739-8521 Japan.

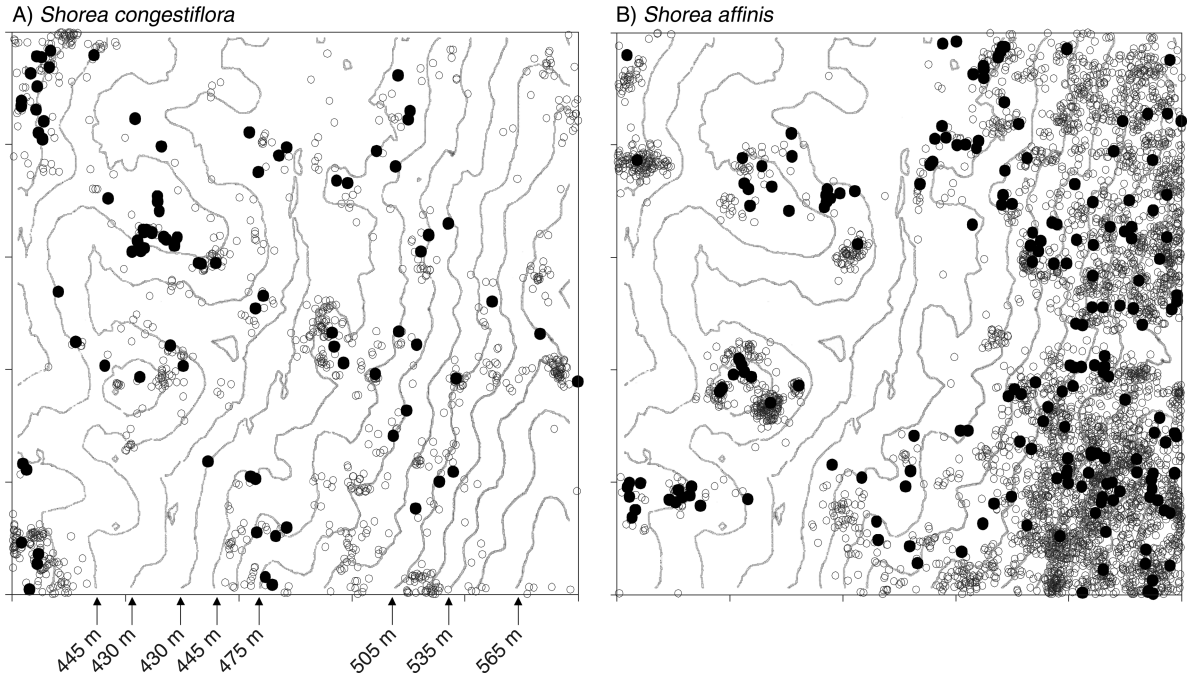

FIG. 1. (A) The spatial pattern of *Shorea congestiflora* trees in relation to the topography of the 500 × 500 m Sinharaja Forest Dynamics Plot, Sri Lanka. Adults are shown as solid circles; recruits, saplings, and intermediates are shown as open circles. (B) The same as (A) for the related species *Shorea affinis*, which shows a strong habitat association.

Second-order statistics such as the pair correlation function or Ripley's K, which are based on the distribution of distances of pairs of points (Ripley 1981), describe the characteristics of point patterns over a range of distances and can therefore potentially detect mixed patterns (e.g., occurrence of clustering at several critical spatial scales).

Detecting mixed patterns is especially important in ecological systems in which different processes may operate at different spatial scales (Levin 1992). For example, many factors and processes may cause clustered spatial patterns in tropical forests. However, since there is no a priori reason to assume that they all act at the same spatial scale, spatial pattern may show clustering at multiple scales, and the relative importance of the critical scales may change with size, age, or species. Thus, for interpreting spatial patterns of tropical tree species due to habitat niche, Janzen-Connell effects, dispersal constraints, and so on, it is critical to precisely determine the critical scale of clustering of the patterns. However, recent studies in tropical forests have assigned only a single scale of aggregation to each species (e.g., Condit et al. 2000, He and Gaston 2000, Plotkin et al. 2000), even though species in tropical forests are frequently aggregated at several scales simultaneously (Plotkin et al. 2000).

In this article, we analyzed the spatial structure of *Shorea congestiflora*, a dominant species at a 25-ha plot in a rain forest at Sinharaja World Heritage Site (Sri Lanka), which apparently shows clustering at several scales. To describe multiple scale of clustering we developed point processes that accommodate two critical scales of clustering. We used these processes and other null models to explore the spatial structure of *S. congestiflora* and interpret it in relation to competition, dispersal limitation, recruitment limitation, and Janzen-Connell effects. This example illustrates how the use of a specific and complex null hypothesis of spatial structure in point pattern analysis can help to better understand the biology of a species.

METHODS

Study site and study species

The area studied is the 25-ha Sinharaja Forest Dynamics Plot (FDP), a 500 × 500 m permanent study plot (Fig. 1). The Sinharaja FDP is located in the lowland rain forest of the Sinharaja UNESCO World Heritage Site at the center of the ever-wet southwestern region of Sri Lanka at 6°21–26′ N and 80°21–34′ E. The Sinharaja FDP is representative of the ridge–steep slope–valley landscape of the lowland and mid-elevational rain forests of southwestern Sri Lanka (see Plate 1). The forest has been classified as a *Mesua–Doona* community (de Rosayro 1942), and on a regional scale it represents a mixed dipterocarp forest (Ashton 1964, Whitmore 1984). The floristic ecology and forest structure within the plot as a whole have been documented in Gunatilleke et al. (2004). The elevation at the Sinharaja FDP ranges between 424 m and 575 m above sea level and includes a valley lying between two

slopes, a steeper higher slope facing the southwest, and a less steep slope facing the northeast (Fig. 1). Tree species at this topographically structured site show varying degree of associations to habitat types defined through elevation, slope, and convexity (Gunatilleke et al. 2006).

The study species, *Shorea congestiflora*, is a medium-large-sized tree up to 40 m tall and 2 m girth with low concave buttresses. In contrast to many other species at this site, *S. congestiflora* shows only minor habitat association, being slightly biased against the low-elevation habitats (<460 m; Gunatilleke et al. 2006). Fig. 1A shows the spatial distribution of *S. congestiflora* in relation to topography; there is no apparent strong habitat association as, e.g., observed for the related species *Shorea affinis* (Fig. 1B). Flowering occurs gregariously in August–September. *Shorea congestiflora* fruits have three wings and disperse from the tree by gyrating to the ground. Due to wind they may be carried just a short distance away from the crown; however, washing down the steep slopes with surface runoff may happen. Animal predation of *Shorea* seeds is minimal as they are highly resinous.

Vegetation sampling

The established methodology of Hubbell and Foster (1983) and Manokaran et al. (1990) was followed to maintain uniformity in the establishment and sampling of similar plots within the network of the Center for Tropical Forest Science (CTFS). The Sinharaja FDP was established in 1993 when it was demarcated on the horizontal plane into 625 plots of 20×20 m (400 m^2) each. The trees in the plot were censused over the period 1994–1996, when the diameter of all freestanding stems $\geq$1 cm diameter at breast height (dbh) was measured. Each stem was mapped and identified to species, using the National Herbarium of Sri Lanka and Dassanayake and Fosberg (1980–2000).

The trees were categorized by size into four classes: small saplings (1–5 cm dbh), large sapling (>5–10 cm dbh), intermediate (>10–20 cm dbh), and adult (>20 cm dbh, range up to 80 cm). Recruits and dead trees were determined in a second census approximately six years later. We classified as recruits all trees >1 cm dbh that appeared in the first census but were too small to be measured. All trees $\geq$ 1 cm dbh that were alive in the first census but dead (or alive but broken below 1.3 m on the trunk) were classified as dead.

Point pattern analysis

The pair correlation function and Ripley's K function, which are based on the distribution of distances of pairs of points, are powerful tools used to describe the second-order structure of a spatial point pattern, i.e., the small-scale spatial correlation structure of the point pattern. Ripley's K function can be defined using the quantity $\lambda K(r)$, which has the intuitive interpretation of the expected number of further points within distance r of an arbitrary point of the process that is not counted (Ripley 1976), where λ is the intensity of the pattern in the study area. The pair correlation function $g(r)$ is related to the derivative of the K function, i.e., $g(r) = K'(r)/(2\pi r)$ (Ripley 1977, Stoyan and Stoyan 1994). Bivariate extensions of $K(r)$ and $g(r)$ follow intuitively (e.g., Diggle 2003, Wiegand and Moloney 2004).

We followed the grid-based approach of Wiegand and Moloney (2004) and Condit et al. (2000) for implementation of Ripley's $K(r)$ and the pair correlation function $g(r)$. We used a grid size of 1 m^2 and a ring width of 3 m for estimation of the pair correlation functions. This is a sufficiently fine resolution compared to the 500×500 m size of the study plot (Fig. 1) and sufficient to respond to our objectives.

We used the distribution $n(y)$ of the distances y to the nearest neighbor and the corresponding accumulative distribution $G(y)$ (Diggle 2003) to describe the characteristic of the patterns not captured by the second-order statistics. The quantity $\lambda K(r)$ is the mean number of points located within a given distance r of each sampled point. However, the same mean [i.e., $\lambda K(r)$] may arise if many points have no neighbor but few points many neighbors or if all points have more or less the same number of neighbors. The distribution of nearest neighbor distances thus provides complementary information of how the number of points located within a given distance r of each point are distributed. We calculated $n(y)$ and $G(y)$ without edge correction (Diggle 2003).

We used a Monte-Carlo approach for construction of confidence limits of a given null model. Each of the n simulations of point process underlying the null model generates a g (or G) function, and approximate two-sided confidence limits with $\alpha = 0.02$ are calculated from the highest and lowest values of 99 simulations of the g (or G) function if the pattern had more than 300 points and from the tenth highest and tenth lowest values of 999 simulations otherwise (Stoyan and Stoyan 1994).

The univariate Thomas process

Examples for point processes that include an explicit clustering mechanism are Poisson cluster processes, Cox processes, or Gibbs processes (Tomppo 1986, Stoyan and Stoyan 1994, Diggle 2003); however, only a few have the advantage that the second-order statistics can be calculated analytically. Our primary interest was in constructing simple null models to be contrasted to our data, which nevertheless accommodate multiple scales of clustering. We therefore used the simplest family of cluster processes that can be solved analytically, so-called Thomas processes (Thomas 1949), as a basic module and combine them to yield point processes with multiple clustering. Univariate cluster processes have been used sporadically in ecological applications (e.g., Cressie 1991, Batista and Maguire 1998, Plotkin et al. 2000, Dixon 2002, Diggle 2003, Potts et al. 2004).

The univariate Thomas process (Fig. 2A, B; Thomas 1949) assumes that (1) the parents follow a homoge-

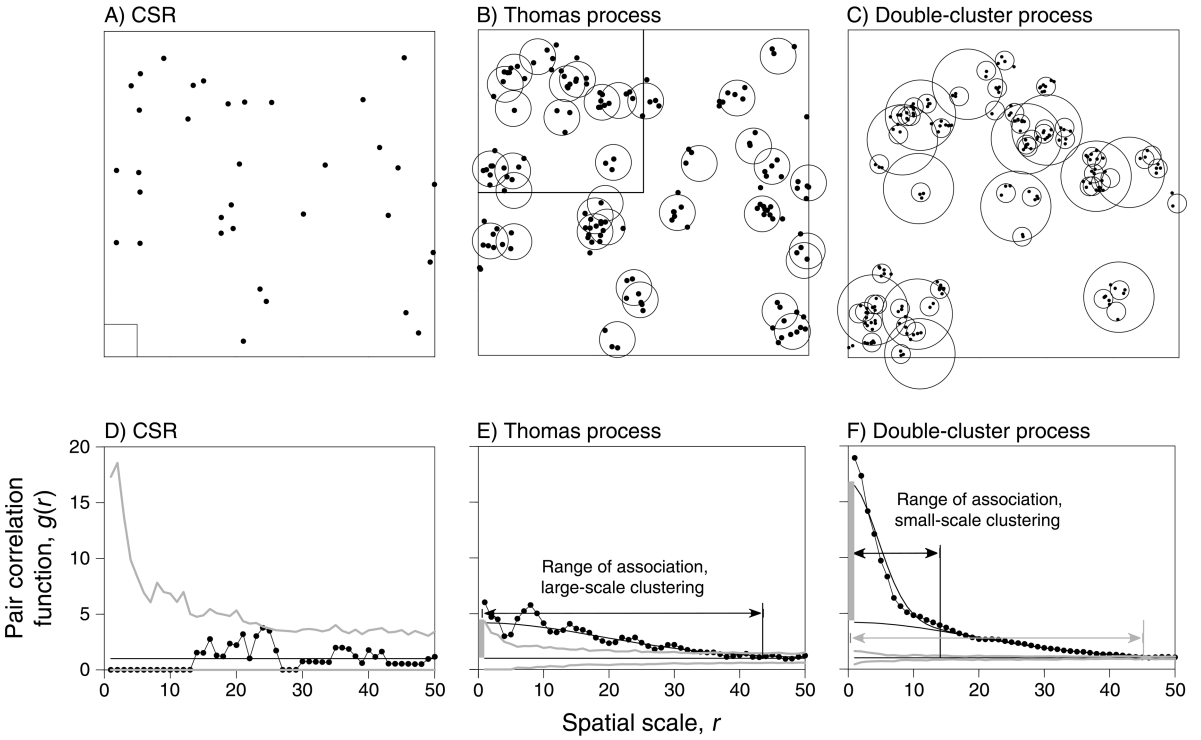

FIG. 2. Univariate nested double-cluster process and pair correlation functions. The process was simulated within a 500×500 m plot with parameters $\sigma_1 = 13.3$, $\sigma_2 = 3.18$, $A\rho_1 = 35$, and $A\rho_2 = 157$ (see Table 1 for an explanation of variable abbreviations). (A) First-generation parents ($n = 35$). The rectangle in the corner shows the maximal scale $r = 50$ for which the pair correlation functions are calculated. (B) Second-generation parents, constructed with the first-generation parents shown in (A). The larger scale clusters are represented by circles with radius $2\sigma_1 = 26.6$. (C) The final double-cluster process, constructed with the second-generation parents shown in the upper left rectangle of (B). The small-scale clusters are represented by circles with radius $2\sigma_2 = 6.4$, and the large-scale clusters are represented by circles with radius $2\sigma_1 = 26.6$. (D) Pair correlation function of the first-generation parents (data points) and confidence limits for complete spatial randomness (CSR) constructed from 999 simulations (gray lines). (E) Pair correlation functions (data points, simulated data; line, "real process") and range of association of the pattern shown in (B) together with confidence limits for CSR. The gray bold vertical line at $r = 0$ indicates g_{C1}, the overall degree of clumping of the process. (F) Same as (E) but for the final double-cluster pattern.

neous Poisson process with intensity ρ, (2) each parent produces a random number of offspring following a Poisson distribution with mean $\mu = \lambda/\rho$ (λ is the intensity of offspring), (3) the locations of the offspring, relative to the parents, have a bivariate Gaussian distribution $h(r, \sigma)$ with variance σ^2 (Stoyan and Stoyan 1994). The pair correlation function $g(r)$ of the Thomas process yields

$$g(r, \sigma, \rho) = 1 + \frac{1}{\rho} \frac{\exp(-r^2/4\sigma^2)}{4\pi\sigma^2}. \quad (1)$$

The unknown parameters ρ and σ are usually fitted by comparing the empirical $\hat{K}(r)$ with the theoretical K function using minimum contrast methods (Stoyan and Stoyan 1994, Diggle 2003). The Thomas process is a special case of the more general Neyman-Scott processes (sometimes also called Poisson cluster processes since the parents form a homogeneous Poisson process) in which the density function of the distances of the offspring from the parent and the distribution of points per cluster are not further specified.

Note that the Thomas process assumes a random distribution of clusters in the study area (i.e., a homogeneous pattern). In reality, however, this assumption may be violated for many species due to environmental heterogeneity and habitat association (e.g., Gunatilleke et al. 2006). If habitat association can be quantified by covariates such as altitude, Thomas processes could be applied in combination with inhomogeneous K functions (Baddeley et al. 2000).

The radius $r_C = 2\sigma$, in which 86% of all offspring are located away from the parent, can be used to describe the typical size of the clusters of the Thomas process (Fig. 2B). The approximate area covered by the cluster is thus $A_C = \pi r_C^2 = 4\pi\sigma^2$. Because formally distinct clusters may coalesce it is difficult to identify the sets of offspring with any confidence (Fig. 2B).

A useful characteristic of the Thomas process describing the overall degree of clustering is given by $g_C = (1/\rho)(1/A_C) = g(r = 0) - 1$ (Fig. 2E). This equation reflects the intuitive fact that the degree of clustering may increase if there are fewer clusters or if the area

covered by individual clusters is smaller. The range of association r_0 of the Thomas process is the distance for which $g(r) = 1$ for all $r > r_0$ (Stoyan and Stoyan 1994; Fig. 2E). Loosely speaking, this is the scale at which the clustering becomes small. It can be assessed approximately from a plot of the g function (e.g., Fig. 2E), although for empirical pair correlation functions irregular fluctuations of $g(r)$ around 1 may occur. Using Eq. 1 and the definitions of r_C and g_C, the range of association can be approximated as $r_0 = r_C\sqrt{\ln(g_C) - \ln(\delta)}$ where δ is a small "tolerance" with $g(r_0) = 1 + \delta$. However, the value of r_0 does not provide direct information regarding the significance of a departure from complete spatial randomness (CSR); this can be assessed by simulated confidence limits.

Fig. 2B shows a realization of the Thomas process with 35 parents and cluster size $r_C = 26.6$, an overall degree of clumping $g_C = 3.4$ and a range $r_0 = 45.6$ (using $\delta = 0.15$).

Nested double-cluster process

Double-cluster processes are rarely used, but see Stoyan and Stoyan (1996), Diggle (2003), and Watson et al. (2007). The Thomas process can be extended to a "multigeneration" process in which the offspring becomes the parent of the next generation. The offspring of the second generation forms the univariate point pattern. Indicating the parameters λ, ρ, and σ of the first generation by subscript 1 and those of the second generation with subscript 2, the pair correlation function of the double-cluster process yields

$$g_{22}(r, \sigma_1, \rho_1, \sigma_2, \rho_2) = 1 + \underbrace{\frac{1}{\rho_2}\frac{\exp(-r^2/4\sigma_2^2)}{2\pi\sigma_2^2}}_{\text{same second-generation parents}} + \underbrace{\frac{1}{\rho_1}\frac{\exp(-r^2/4\sigma_{\text{sum}}^2)}{4\pi\sigma_{\text{sum}}^2}}_{\text{same first but different second-generation parents}}$$

with

$$\sigma_{\text{sum}}^2 = \sigma_1^2 + \sigma_2^2 \quad (2)$$

(see Appendix A). The properties r_C, g_C, and r_0 defined for the Thomas process can be generalized for the single components of nested cluster processes that are based on the Thomas process (Fig. 2F).

Fig. 2C shows a realization of a nested double-cluster process with 35 first-generation parents (Fig. 2A) and 157 second-generation parents (Fig. 2B). This pattern has a small-scale cluster size $r_{C2} = 6.4$, an overall degree of small-scale clumping $g_{C2} = 12.5$, and a range $r_{02} = 13.4$ (using $\delta = 0.15$). Fig. 2F shows the pair correlation function of the original process (solid line) together with the pair correlation function of the simulated process (data points). Note that the simulated process does not reproduce the second-order characteristics of the original process perfectly; some smaller departures occur.

The two scales of clustering can only be separated if the second-generation clustering σ_2 is smaller than the first-generation clustering σ_1. In the other extreme if $\sigma_1 \ll \sigma_2$ we find $\sigma_{\text{sum}} \approx \sigma_2$, and Eq. 2 approximates the pair correlation function of the Thomas process (Eq. 1) with parents intensity $(1/\rho_2 + 1/\rho_1)$.

If the second-generation parents are known, the bivariate pair correlation function yields

$$g_{12}(r, \sigma_1, \rho_1, \sigma_2, \rho_2) = 1 + \underbrace{\frac{1}{\rho_2}\frac{\exp(-r^2/2\sigma_2^2)}{4\pi\sigma_2^2}}_{\text{same second-generation parents}} + \underbrace{\frac{1}{\rho_1}\frac{\exp(-r^2/4\sigma_{\text{sum}}^2)}{4\pi\sigma_{\text{sum}}^2}}_{\text{same first but different second-generation parents}}$$

with

$$\sigma_{\text{sum}}^2 = \sigma_1^2 + \frac{1}{2}\sigma_2^2 \quad (3)$$

(see Appendix A).

This process is useful for situations in which a hypothesis exists about the second-generation parents. For example, when studying the association of recruits to adult trees (which itself follow a Thomas process) an obvious hypothesis would be that the recruits are clustered in a shadow-like manner around the adults.

Superposition of cluster processes

The other extreme situation for a pattern showing two distinct critical scales of clustering is a situation in which the patterns are not nested as in Eq. 2, but results from the independent superposition of two Thomas processes with relative intensities p_1 and p_2 $(= 1 - p_1)$ (Stoyan and Stoyan 1996). Indicating the parameters ρ and σ for two Thomas processes with subscripts 1 and 2, the pair correlation function of the superposition process yields

$$g(r, \sigma_1, \rho_1, \sigma_2, \rho_2) = 1 + \underbrace{p_2^2\frac{1}{\rho_2}\frac{\exp(-r^2/4\sigma_2^2)}{4\pi\sigma_2^2}}_{\text{contribution of Thomas process 2}} + \underbrace{(1-p_2)^2\frac{1}{\rho_1}\frac{\exp(-r^2/4\sigma_1^2)}{4\pi\sigma_1^2}}_{\text{contribution of Thomas process 1}} \quad (4)$$

(see Appendix A).

Comparison of Eq. 4 with the pair correlation function of the nested double-cluster process (Eq. 2) shows that both have the same functional form. However, the intensities ρ_1 and ρ_2 of the superposed process can only be estimated if the relative intensity p_1 of process 1 is known. Additionally, the estimate of σ_1 will yield a slightly smaller value than σ_{sum}.

Another superposition process of interest is a process in which a random pattern is superposed to the nested double-cluster process (Eq. 2). Denoting p_C as the

proportion of the points belonging to the nested double-cluster component process, the pair correlation function of the superposition process yields the following:

$$g(r,\sigma_1,\rho_1,\sigma_2,\rho_2) = 1 + p_C^2 \frac{1}{\rho_2}\frac{\exp(-r^2/4\sigma_2^2)}{4\pi\sigma_2^2} + p_C^2 \frac{1}{\rho_1}\frac{\exp(-r^2/4\sigma_{\text{sum}}^2)}{4\pi\sigma_{\text{sum}}^2}$$

with

$$\sigma_{\text{sum}}^2 = \sigma_1^2 + \sigma_2^2 \quad (5)$$

(see Appendix A).

Again, the pair correlation function of the superposition (Eq. 5) yields the same functional form as the pair correlation function of the nested double-cluster process (Eq. 2). Thus, superposition with a random pattern does not affect the estimates of parameters σ_1 and σ_2, but the estimates of the intensities $\rho_1^* = \rho_1/p_C^2$ and $\rho_2^* = \rho_2/p_C^2$ of the first- and second-generation parents, respectively, are by factor $1/p_C^2$ larger than the true ones (ρ_1 and ρ_2).

Distinguishing between the nested and the superposed double-cluster processes is thus not possible based only on second-order characteristics. One possibility for the diagnosis of possible superposition is to analyze additional characteristics of the pattern, such as the distribution $G(y)$ of the nearest neighbor distances y (Stoyan and Stoyan 1994, Diggle 2003). If the pattern is a true nested double-cluster process, most points will have their nearest neighbor within the same cluster, thus yielding nearest neighbor distances $<2\sigma_2$. However, under superposition larger nearest neighbor distances will occur. A good indication may also be provided by visualization of the pattern. In the case of superposition small-scale clusters or isolated points would be scattered without forming characteristic larger clusters. Further evidence can be given by interpretation of the fitted parameters ρ_1 and ρ_2 (see Appendix A).

Parameter fitting

For parameter fit we followed the minimal contrast method, e.g., described in Stoyan and Stoyan (1994) and Diggle (2003). However, we fitted both the g function and the L function simultaneously because the g function is especially sensitive at smaller scales and the K function at larger scales. Details of the fitting procedure are provided in Appendix B.

A potential problem when using the K function for parameter estimation is that the K function has a memory (Wiegand and Moloney 2004). The problem arises here when fitting a Thomas process to a point pattern that shows an additional small-scale clustering with range r_{02}. In this case the observed values of the K function are influenced by this small-scale clustering, even for $r > r_{02}$. This may produce biased estimates of the parameters of the cluster process (Stoyan and Stoyan 1996) and leads to the observation that the parameter estimates depend sensitively on the upper limit $r_{\max}$ at which the K function is fitted (e.g., Batista and Maguire 1998, Plotkin et al. 2000).

To overcome this limitation we developed a transformation of the K function to remove the memory. We used the transformed K function,

$$K_t(r, r \geq r_0) = K_0 - K(r_0) + K(r) \quad (6)$$

for the fitting procedure instead of the common $K(r)$ function, which shows memory effects. K_0 is the observed value at scale r_0 (i.e., $\hat{K}(r_0) = K_0$), and $K(r_0)$ is the value of the K function of the theoretical point process at scale r_0.

Separation of the scales of clustering (i.e., $\sigma_1^2 \gg \sigma_2^2$ in Eq. 2) suggests a convenient approach to fit the four parameters of the double-cluster process. In a first step we fitted the parameters σ_{sum}^2 and ρ_1 of the overall larger-scale clustering using a Thomas process (Eq. 1), but we fitted only for scales r larger than the range r_{02} of the small-scale clustering. We assessed the range r_{02} of the small-scale clustering from comparing the plot of the estimated pair correlation function and the fitted pair correlation function (e.g., Fig. 2F). In the second step we used the estimates of σ_{sum}^2 and ρ_1 and fitted the two unknown parameters σ_2^2 and ρ_2 of the small-scale clustering using the full double-cluster model.

Biological Questions and Null Models

Double cluster structure (analysis 1)

Our working hypothesis was that *S. congestiflora* showed nested clustering at several critical scales. To test this hypothesis we first analyzed the spatial pattern of all trees, fitting the nested double-cluster process (Eq. 2) to the data. To find out whether this process describes the data well or a superposition process would be more likely, we performed Monte Carlo simulations of the fitted process and compared the resulting confidence limits with the pair correlation function and the distribution of nearest neighbor distances of the data. Next we analyzed the univariate patterns of each individual size class to find out whether the critical scales changed with life stage.

Smaller trees are more aggregated than larger trees (analysis 2)

A frequent observation in forests is that recruits are clustered at small scales, but lose this clustering with increasing size due to self-thinning. We used the results of the univariate analyses of the different size classes to find out whether clustering changes with size class.

Recruit–adult and juvenile–adult associations (analysis 3)

Shorea congestiflora seeds disperse from the tree by gyrating to the ground. Due to wind they may be carried only short distances away from the crown. This suggests a seed shadow around the stems of adult trees. Analyzing the distribution of the distances to the nearest

TABLE 1. Univariate analyses using the Thomas process (Eq. 1) and the univariate double-cluster model (Eq. 2).

	Patterns of compound larger-scale clustering								Patterns of small-scale clustering						
Pattern	n	Dead (%)	σ_{sum}	$A\rho_1$	μ_1	g_{C1}	r_{01}	er	g_{C2}/g_{C1}	σ_2	$A\rho_2$	μ_2	g_{C2}	r_{02}	er
All trees	986	14	13.3	44.3	22.3	2.8	43.5	0.003	2.3	3.8	223.9	4.4	6.2	14.6	0.0006
Component†	867	10	13.4	34.0	25.5	3.5	45.7	0.003	2.4	3.8	163.0	5.3	8.6	15.2	0.0029
Recruits	112		13.3	23.3	4.8	5.3	48.1	0.004	4.8	3.8	55.9	2.0	25.3	17.0	0.0046
Small saplings	626	17	13.3	35.2	17.8	3.4	45.6	0.001	3.7	3.2	157.8	4.0	12.5	13.4	0.0022
Large saplings	97	7	13.3	45.7	2.1	2.7	43.3	0.024	5.6	3.8	89.6	1.1	15.1	16.5	0.0122
Intermediates	64	13	14.6	44.3	1.4	2.7	43.9	0.024	4.1	5.3	64.0	1.0	11.1	22.0	0.0110
Adult‡	87	20	12.8	20.9	4.2	5.8	49.0	0.015	0.0	···	···	···	···	···	···
Dead‡	136	···	6.7	50.6	2.7	8.8	27.0	0.013	0.0	···	···	···	···	···	···

Notes: The variable abbreviations (where subscripts 1 and 2 refer to the small-scale and the large-scale component process, respectively) are: n, number of points of the pattern; dead, the percentage of dead trees in the size class; A, size of the study area; ρ_1, ρ_2, the intensity of the parents pattern; $A\rho$, the number of parents in the plot of size $A = (500 \times 500$ m); σ_2, parameter describing the cluster size (in meters); σ_{sum}: parameter describing the cluster size of a double-cluster process (in meters); $\mu_1 = n/(A\rho_1)$, $\mu_2 = n/(A\rho_2)$, the mean number of points in a cluster; g_{C1}, g_{C2}, the overall degree of clustering; r_{01}, r_{02}, range of clustering (in meters); er, fraction of the total sum of squares of the empirical g and L function not explained by the fit (combined as geometric mean; see Appendix B). We first analyzed the pattern of all trees and assumed in the analyses of individual size classes that they show the same larger-scale cluster size as the pattern of all trees together (i.e., we fixed $\sigma_1 = 12.8$). We fitted the remaining parameter ρ_1 of the larger-scale clustering using the Thomas process (Eq. 1) for scales $r = 15$–100. Next we used the double-cluster process (Eq. 2) to fit the parameters σ_2 and ρ_2 to the small-scale clustering.

† The component pattern of the pattern of all trees that have at least one nearest neighbor within 8 m.

‡ For this size class only a Thomas process (Eq. 1) was fitted to the data for scales $r = 1$–50, but no double cluster process.

adult from juvenile trees (i.e., trees with a dbh <10 cm) will provide indirect evidence for the shape of the seed shadow, although competition with adults and Janzen-Connell processes may reduce survival in the neighborhood of adults overproportionally. Our special interest was to find indications for possible dispersal limitation (i.e., only few juvenile are found further away than a certain distance from an adult) and to relate this to the scales of clustering. To avoid edge effects we used here only juveniles more than 50 m away from the border of the plot.

To explore whether other processes (e.g., limited availability of regeneration sites, competition, Janzen-Connell processes) substantially changed the presumed spatial pattern of the seed shadow, we contrasted the bivariate recruit–adult patterns to the null model of independence (Goreaud and Pelissier 2003). In case there was positive association, we further contrasted the bivariate patterns to the null model where the recruits are distributed as shadow around the adult trees (Eq. 3).

Association between size classes (analysis 4)

The spatial relationship between subsequent size classes should contain information about the spatial organization of the population. To find out whether trees of subsequent size classes tended to co-occur in the same clusters, thus producing a positive association, we contrasted the bivariate patterns to the null model of independence.

RESULTS

Double-cluster structure (analysis 1)

Univariate analysis of all trees.—Fit of the Thomas process (Eq. 1) for scales $r > 15$ m to the data of all trees yielded parameter estimates $\sigma_{sum} = (\sigma_1^2 + \sigma_2^2)^{0.5} = 13.3$ m and a total number of $A\rho_1 = 44.3$ larger scale clusters. By fitting the nested double-cluster process (Eq. 2) for scales $r = 1$–100 we estimated the parameters of small-scale clustering component as $\sigma_2 = 3.8$ m (indicating a cluster size of $r_{C2} = 2\sigma_2 = 7.6$ m) and $A\rho_2 = 224$ small-scale clusters. The range of small-scale clustering r_{02} yielded 14.6 m, confirming our selection of $r_0 = 15$ (Table 1). Using the estimate of σ_2 the larger scale cluster size yields $r_{C1} = 2\sigma_1 = 2(\sigma_1^2 - \sigma_2^2)^{0.5} = 25.5$ m. Thus, the estimated radius of the larger scale clusters was roughly four times that of the small clusters. The range of the larger scale clustering was $r_{01} = 44$ m (Table 1).

Fig. 3D shows that the empirical pair correlation function of all trees was well within the confidence limits of the fitted double-cluster process, indicating that the second-order properties of this process cannot be distinguished from that of our data. However, the empirical distribution of nearest neighbor distances did not agree well with that of the fitted process (small inset of Fig. 3D): some 10–20% of the trees had no nearest neighbors within 8 m, as expected by the fitted process, thus suggesting a superposition pattern.

To explore whether the pattern of all trees might be a superposition pattern we divided the pattern of all trees into two component patterns, one comprising trees for which the distance y to the nearest neighbor was smaller than 8 m (Fig. 3B) and a second component comprising the trees with at least one nearest neighbor within 8 m (Fig. 3C). We found that about $n = 119$ trees (=12%) had no nearest neighbor closer than 8 m. Fig. 3E shows that this pattern was a random pattern and the small inset of Fig. 3E shows that the two component patterns were independent (except small-scale repulsion caused by the way the patterns were constructed). Thus, the pattern of

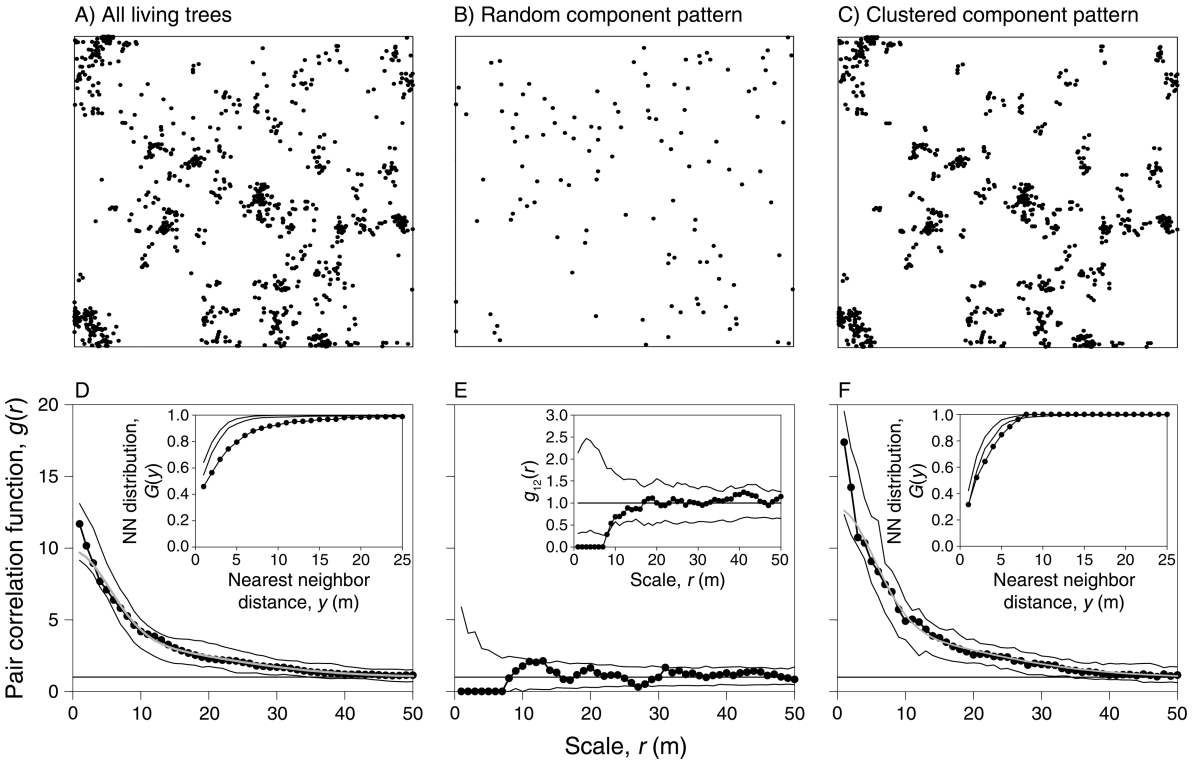

FIG. 3. Analyses of the pattern of all trees. (A) Spatial pattern of all *Shorea congestiflora* trees within the 500 × 500 m plot. (B) Component pattern of (A) comprising only trees that have no neighbor within 8 m. (C) Component pattern comprising only trees with at least one neighbor within 8 m. (D) Empirical pair correlation function (data points) of the pattern shown in (A) together with the fitted double-cluster process (gray line) and the confidence limits (solid line). The inset shows the analysis of the distribution $G(y)$ of the nearest neighbor (NN) distances y. (E) Empirical pair correlation function (data points) and confidence limits of the pattern shown in (C). The inset shows the bivariate pair correlation function $g_{12}(r)$ of the two component patterns and confidence limits constructed using 99 simulations of a toriodal shift null model testing for independence. (F) Same as (D), but for the pattern shown in (C). The confidence limits in (D) and (F) were constructed from 99 simulations of the fitted double-cluster process, and those in (E) from 999 simulations of complete spatial randomness (CSR). The ring width was 3 m in all analyses.

all trees fulfills the assumption for superposition of a nested double-cluster process with a random pattern, and the pair correlation function of the pattern of all trees should follow Eq. 5.

The second component pattern comprising only trees with at least one nearest neighbor within 8 m showed visually a clearer double-cluster structure than the pattern of all trees (cf. Fig. 3A, C). This allowed us to reconstruct the clusters qualitatively (see Appendix C). Repeating the fit with the double-cluster process yields parameters σ_{sum} =13.4, $A\rho_1 = 34$, $\sigma_2 = 3.8$, and $A\rho_1 = 163$. The proportion of points with at least one neighbor closer than 8 m was $p_C = 0.88$. Thus the inflation factor of the number of parents due to superposition with the random pattern yielded $1/p_C^2 = 1.29$, which was well confirmed by comparing the estimates of the number of first- and second-generation parents derived for the pattern of all trees and for the double-cluster component pattern (Table 1): 44.3/34.0 = 1.30 and 223.9/163.0 = 1.37. Comparison of the empirical distribution of nearest neighbor distances with that of the fitted process (inset of Fig. 3F) showed some smaller discrepancies that may stem from points that belong to the random component pattern but were accidentally close to a cluster and could therefore not be detected.

Univariate analysis of individual size classes.—The 119 trees that had no neighbor within 8 m distance were proportionally distributed among life stages, although they were slightly overrepresented in the intermediate and adult stages. They comprised 9% of the recruits (10 trees), 11% of the small saplings (72 trees), 8% of the large saplings (8 trees), 16% of the intermediates (10 trees), and 22% of the adults (19 trees). For all size classes, except small saplings for which the same results as for all trees hold (not shown), the sample sizes were too small to separate the patterns in the same way as done for the pattern of all trees. Because the superposition did not affect estimation of the cluster sizes and biased the estimates of the number of clusters in a predictable way, we analyzed the univariate patterns of individual size classes without dividing the pattern into two components, but we considered possible superposition in the interpretation of the parameter estimates.

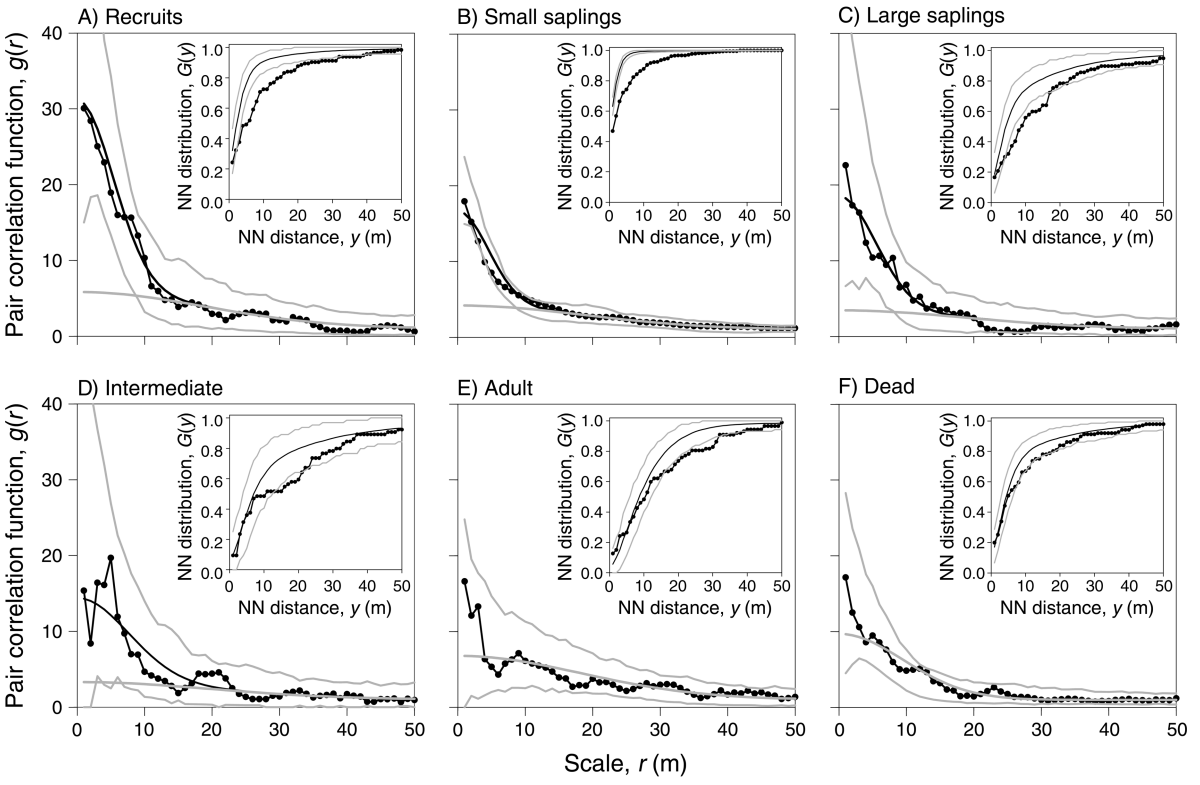

FIG. 4. Univariate point pattern analysis of the different size classes of *Shorea congestiflora*. The pair correlation functions estimated from the data (data points) are contrasted to a null model that assumes a common larger scale clustering with parameter $\sigma_1 = 12.8$ and additional small-scale clustering for recruits, saplings, and intermediates. The confidence limits (gray lines) of the null models were constructed using 999 Monte Carlo simulations of univariate double-cluster models (Eq. 2) and the Thomas process (Eq. 1) with parameters given in Table 1 [in (B) we used 99 simulations]. The fitted pair correlation function of the Thomas process is shown as gray solid lines, and that of the double-cluster model as a solid black line. The ring width was 3 m. The small inset figures show the empirical distribution $G(y)$ of the nearest neighbor (NN) distances (data points) together with confidence limits (lines).

For a given size class we estimated the intensity ρ_1 of first-generation parents and the parameters σ_2 and ρ_2 of the small-scale clustering mechanism under the assumption $\sigma_1 = 12.8$ (i.e., we assumed that the larger scale cluster size was the same for all size classes). For all size classes the fit with the Thomas process for scales $r > 15$ m showed an error coefficient er < 0.03, indicating that $<3\%$ of the total sum of squares of the empirical g and L functions was not explained by the fitted functions (Table 1). This can be considered a good fit that justifies a posteriori our choice of $\sigma_1 = 12.8$.

For recruits and saplings the fit with the double-cluster process showed for all classes an error coefficient er < 0.013, again indicating a good fit (see also Fig. 4). However, for adults and dead, the Thomas process yielded already a good fit (Table 1). Our analyses detected for all juvenile size classes clear indications for an additional small-scale clustering with cluster size of $\sim$8 m (Fig. 4, Table 1). Simulating the processes with the fitted parameters was then used to confirm that our data cannot be distinguished from the fitted processes. As expected by the low error coefficients, we found no significant departures of the pair correlation function from the confidence limits of the simulated processes (Fig. 4). Note that it is not obvious a priori that a fitted process describes the data well, especially in cases in which the fit is poor. However, dead trees did not show the signal of the common larger scale clustering; instead the Thomas process fitted the data well (er = 0.013; Table 1, Fig. 4F), yielded for scales $r = 1$–50 a cluster size of $r_{C1} = 2\sigma_1 = 13.4$ m and some $A\rho_1 = 51$ clusters.

The empirical distribution of nearest neighbor distances showed for all size classes the expected departure from the simulations of the fitted process due to a superposition (small insets in Fig. 4). For the recruits and saplings patterns with strong small-scale clustering the departure occurred at smaller scales (note that $G(y)$ is accumulative) and for intermediates and adults with weak or no small-scale clustering at larger scales.

Smaller trees are more aggregated than larger trees (analysis 2)

Our analyses clearly supported the hypothesis that smaller trees (i.e., recruits and saplings) were aggregated at two critical scales (Fig. 4A–C). Intermediates still showed a signal of the two scales of clustering (Fig. 4D)

but yielded a somewhat unstable fit, probably due to the low number of points ($n = 64$), and the pattern of adult trees did not show significant small-scale clustering (Fig. 4E). Interestingly, the small-scale cluster size was the same for recruits and the two sapling size classes and only slightly larger for intermediates (Table 1).

Given that the estimated large-scale cluster size was the same for all size classes, the monotonous decrease in the estimated overall degree g_{C1} of larger scale clustering from recruits ($g_{C1} = 5.3$) to intermediates ($g_{C1} = 2.7$) (Fig. 4, Table 1) was therefore probably caused by an increase in the number of occupied clusters (Table 1). A test with qualitatively reconstructed clusters confirmed this finding (Appendix C).

The estimated degree g_{C2} of small-scale clustering was largest for recruits and approximately half of that for saplings and intermediates (Table 1). The stronger clustering of recruits might therefore be caused by having fewer clusters rather than having a smaller cluster size. However, note that care is required with these interpretations because the estimates of g_{C1} and g_{C2} may be biased if the patterns would be superposition of a nested double-cluster process and a random pattern as suggested above (see *Double-cluster structure [analysis 1]: Univariate analysis of all trees*).

The pair correlation functions of the recruits (Fig. 4A) and saplings (Fig. 4B, C) show that recruits were more clustered than saplings. The much higher abundance of small saplings ($n = 626$) compared to recruits ($n = 112$) suggest that small saplings may accumulate several recruit generations. The observed differences in clustering between recruits and saplings could be caused by scarce and short-lived regeneration sites. Following this hypothesis, recruits should show fewer small-scale clusters than saplings.

Recruit–adult and juvenile–adult associations (analysis 3)

Shorea congestiflora seeds disperse from the tree by gyrating to the ground, but not very far from the crown. They should therefore accumulate under the canopy. However, we found that only 10% of all juveniles were located within a 5-m distance from the nearest adults (Fig. 5A). This might be due to competition from adults directly under the canopy or by Janzen-Connell effects. The most frequent nearest neighbor distances occurred between 4 and 25 m, and 95% of all juveniles were located within some 53 m from the nearest adult. Interestingly, the large-scale cluster size $\rho_1 = 25.5$ m is just the scale at which the juvenile–adult distances become less frequent (Fig. 5A). This scale should coincide with the maximum distance gyrating seeds disperse by wind away from the stem. However, ~30% of the juveniles, mostly small saplings, were located further than 26 m away from an adult tree (Fig. 5A). For some of these saplings the parent tree may have died before the census started, but also a secondary seed dispersal mechanism (e.g., washing down the slopes with surface runoff) could be involved.

The pair correlation function shows a pronounced peak in the intensity of recruits about 4–7 m away from the adult stems (Fig. 5B). Application of the null model of independence showed that the tendency of positive association between recruits and adults was not significant for scales $r = 10$–40, but significant for scales between 5 and 7 m (Fig. 5B). We therefore proceeded in testing the more specific hypothesis that the adults were the cluster centers of the recruits. However, the fit with the model Eq. 3, which assumes that adults are the parents of the recruits, failed: the parameter estimates yielded some 71 parents and $\sigma_2 = 21$, which was not consistent with the results of the univariate analysis. Thus, although there is a positive and significant small-scale association between recruits and adults, we found evidence that additional processes such as competition to adults or Janzen-Connell processes may have modified the seed shadow in a nonrandom way.

Association between subsequent size classes (analysis 4)

Small sapling and recruits.—Application of the null model of independence showed that there was a tendency to positive association between recruits and small saplings at scales $r < 40$ and a significant positive association for scales $r < 10$ (Fig. 5C). Interestingly, the bivariate pair correlation function describing the association of small saplings around recruits was for scales $r > r_{C2} = 6.4$, basically the same as the univariate pair correlation function describing the association of small saplings around small saplings (open discs in Fig. 5C). Thus, outside the range of the small-scale clustering, small saplings surrounded recruits in the same way as small saplings surrounded small saplings. However, at scales $r < r_{C2}$ saplings were more strongly associated to saplings (g_{22} in Fig. 5C) than saplings to recruit (g_{12} in Fig. 5C) or recruits to recruits (g_{11}, not shown). Thus, although recruits and small saplings were not randomly mixed in small clusters (in this case we would expect $g_{12} \approx g_{11} \approx g_{22}$), they co-occurred frequently enough in the same cluster to yield a clear positive small-scale association.

Small and large saplings.—We found a significant positive association between small and large saplings at scales $r < 25$ (Fig. 5D). The bivariate pair correlation function describing the association of small saplings around large saplings was for scales $r > 3$ basically the same as the univariate pair correlation function describing the association of small saplings around small saplings. Thus, small saplings occurred at small scales quite often around large saplings (cf. g_{12} and g_{22} in Fig. 5D), but not as frequently as small saplings occurred around small saplings.

Intermediates and large saplings.—For scales $r < 17$ there was a significant positive association between large saplings and intermediates (Fig. 5E). Again, g_{12} and g_{22} were quite similar outside the small clusters. Large saplings showed a significant positive association to intermediates, which, however, was clearly weaker than

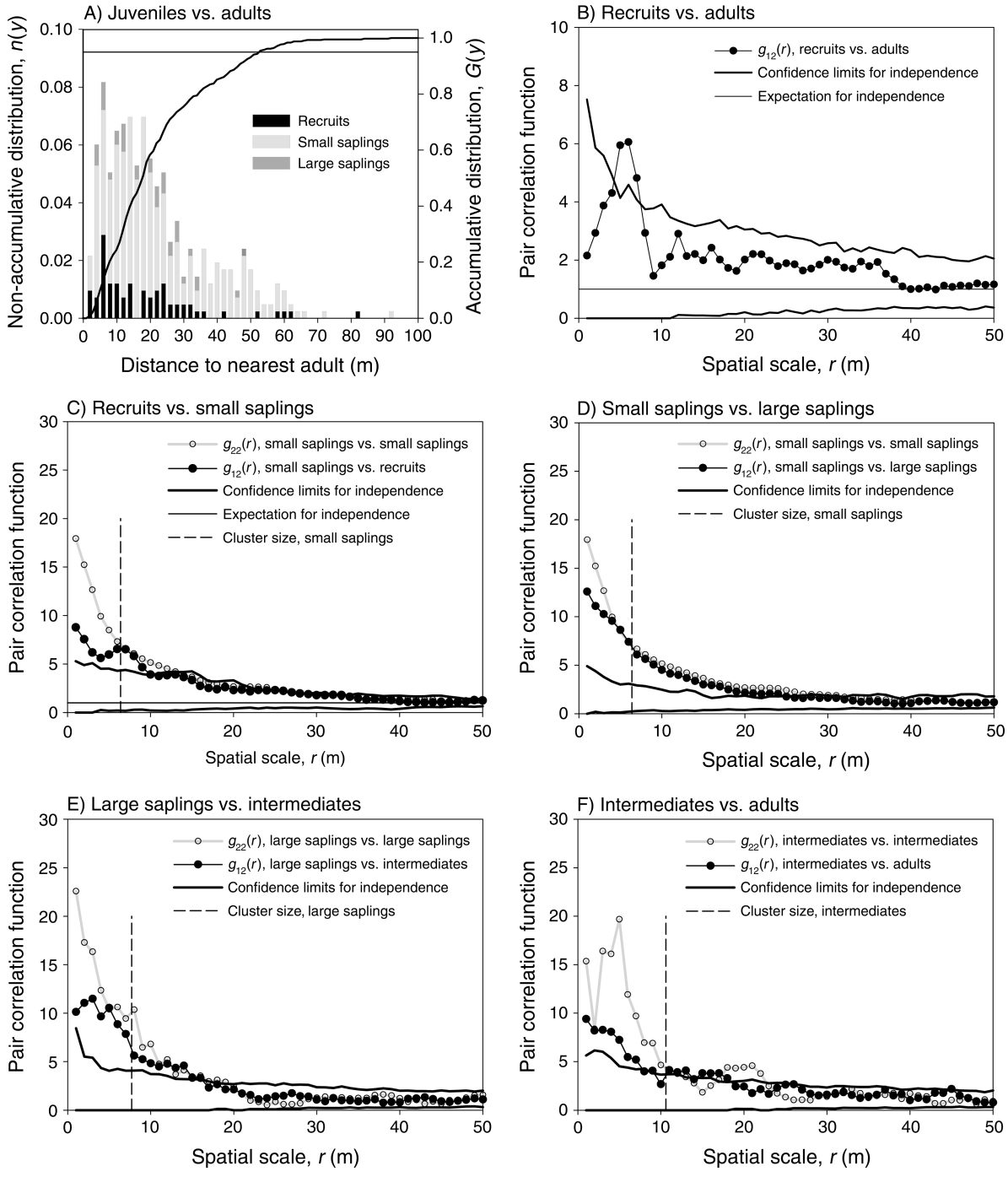

FIG. 5. Bivariate analyses. (A) Non-accumulative distribution $n(y)$ of the distances y from juveniles to the nearest adult neighbor (bars) and accumulative distribution $G(y)$ (line). (B–F) Bivariate pair correlation function $g_{12}(r)$ (solid lines with circles) and confidence limits (solid black lines) for independence were based on 999 simulations in (B), (E), and (F), but on 99 in (C) and (D). The pair correlation functions $g_{11}(r)$ and $g_{22}(r)$ for one univariate component pattern are shown as gray lines with open circles. The ring width was 3 m.

clustering of large saplings around large saplings (cf. g_{12} and g_{22} in Fig. 5E).

Adults and intermediates.—Application of the null model of independence revealed a significant positive association between adults and intermediates for most scales $r < 18$ (Fig. 5F). Comparing g_{12} and g_{22} showed that intermediates were, within their small-scale clusters, more associated to intermediates than to adults.

PLATE 1. View from a lowland hilltop into the canopy of the mixed dipterocarp forests at middle altitudes (400–700 m) of southwestern Sri Lanka which form, together with Western Ghats, a global biodiversity hotspot. Over 60% of the tree species in these forests are endemic to Sri Lanka and, despite habitat reduction and degradation, still retain some relict signatures of Gondwana ancestry. Photo credit: N. Gunatilleke.

However, the relatively low sample size for both life stages prevented a too literal interpretation of the relative shapes of the two pair correlation functions.

DISCUSSION

Our analyses demonstrated how the use of a specific and more complex null hypothesis of spatial structure in point pattern analysis can lead to a better understanding of the biology of a species. Our working hypothesis was that our study species *S. congestiflora* would show nested clustering at multiple critical scales. Double-cluster processes, which allowed for more realistic spatial structures, were critical ingredients of our approach and allowed us to escape the limitations in the current use of point pattern analysis in ecology, e.g., outlined by Plotkin et al. (2002). Standard null models such as complete spatial randomness, or even the Thomas process which incorporates one scale of clustering, often do not allow for a meaningful analysis in explorative point pattern analysis because they cannot address the complexity of real world data sets (Stoyan and Stoyan 1996, Plotkin et al. 2002).

Spatial structure of Shorea congestiflora

Our analyses provided a clear picture of the spatial structure of our study species. For most size classes we found strong evidence for two nested scales of clustering; a larger scale clustering with a cluster size of some 26 m for all live stages and a small-scale clustering with a consistent cluster size of some 8 m that persisted from recruits up to the intermediate stage. Interestingly, we found indications that the spatial pattern of *S. congestiflora* trees could be a superposition of two independent patterns. One component pattern (trees without a nearest neighbor within 8 m), comprising about 12% of the trees, was a random pattern. The second component could be approximated well by a nested double-cluster process. A hypothesis to explain this finding is that *S. congestiflora* has two dispersal mechanisms: primary dispersal, in which seeds gyrate to the ground, and secondary dispersal, in which seeds are occasionally washed down the steep slopes with surface runoff and are entrapped in the process. This hypothesis could be tested in the field. Secondary dispersal by animals is less likely because *Shorea* seeds are highly resinous.

Although there was a consistent scale of larger scale clustering among all size classes, our analyses suggested that not all larger scale clusters were occupied by all size classes. This caused differences in the overall degree of clustering. Recruits and adults may occupy about half of the larger scale clusters, small saplings 80%, and large saplings and intermediates all (Table 1). These results were consistent with a qualitative reconstruction of the large-scale clusters. We found that a substantial proportion of the clusters was not occupied by recruits and adults, whereas saplings occupied most clusters (Fig. 1C). Our results suggest that the large-scale cluster

size of 26 m might be related to a dispersal limitation of *S. congestiflora*. The distribution of distances from juveniles to the next adult (Fig. 5A) showed that for 70% of the juveniles the next adult was closer than $2\sigma_1 =$ 26 m. This distance should be the maximal primary seed dispersal distance by gyrating. The juveniles that had nearest neighbors further away than 26 m could be additionally dispersed by runoff or their parent trees were already dead at the time of the first census.

The finding that recruits do not occupy all clusters but saplings occupy most clusters is consistent with patchy recruitment due to limited regeneration sites that may not be present at any time (Hubbell et al. 1999). A high occupancy of larger scale clusters by saplings would arise if this size class accumulated several recruitment generations. Clusters without adults may appear because the "founder" adult already died, or it may indicate interspecific competition with large trees of other species (Condit et al. 2000) or it might be explained by patch mortality of cohorts, a large-scale Janzen-Connell effect rarely examined and, admittedly, difficult to "prove." However, it would be necessary to look at the precise locations of each patch to give more precise biological interpretations of this finding.

The larger scale clustering was overlaid by smaller scale clustering with a radius of ~8 m. Recruits showed a strong small-scale clustering that persisted up to the intermediate stage, but disappeared for adults. Thus, a classical self-thinning was evident for our study species; larger trees were less aggregated than smaller trees. The consistent 8-m small-scale cluster size may correspond to the typical size of gaps produced by dead canopy trees (Hubbell et al. 1999).

Subsequent size classes occurred frequently enough together in the same small-scale clusters to produce a significant positive small-scale association (Fig. 4C–E). This result is consistent with recruitment limitation in which the locations of the regeneration sites change for each recruit generation and where the temporal window of a recruitment site was large enough to allow for a certain overlap of subsequent size classes causing the overall positive smaller scale association.

Because *S. congestiflora* seeds are dispersed by gyrating there should be many recruits under the canopy of adult trees. However, recruits were only weakly associated with adults and few recruits occurred close to the stem of adult trees, but the peak density of recruits around adults occurred at ~5–7 m distance from the stem (Fig. 5B). Potential processes to explain this finding are competition from adults or Janzen-Connell effects in which pathogens, herbivores, and seed predators eliminate seedlings in the immediate neighborhood of adult trees. The latter would be consistent with work elsewhere (Wills and Condit 1999, Condit et al. 2000, Harms et al. 2000) that has shown that the greater part of Janzen-Connell mortality occurs below that size of 1 cm dbh. The lack of positive association at scales >8 m is consistent with our hypothesis of limited and short-lived regeneration sites.

Assumptions of our approach

Our approach of using simple and mathematically tractable point processes is a reasonable parsimonious approach since we used them as null models and did not intend to fit all idiosyncrasies of the real world patterns. The double-cluster processes based on the Thomas process capture the essence of multiple clustering in a simple way and should therefore be suitable null models for most practical application in ecology. Clearly, if data are scarce (in the order of a few hundred points) one may not be able to distinguish statistically among structurally similar candidate point processes. However, if strong biological evidence suggests violation of critical assumptions or if there is a substantial departure from the null model, other null models are required.

A critical assumption of our analyses that may frequently be violated in real data sets is large-scale homogeneity of the pattern. If the data stem from a single realization of the underlying process there is a fundamental ambiguity between clustering and heterogeneity; both cannot be distinguished statistically without additional biological information (Bartlett 1964, Diggle 2003). This is intuitively clear since both clustering and environmental heterogeneity generate patterns with locally elevated point densities. In general, however, large-scale aggregation is attributed to environmental heterogeneity, whereas small-scale clustering is attributed to point–point interactions. Thus, double-cluster processes might be used, within certain limits, to describe clustering due to environmental heterogeneity. This can be done if the broadest level of heterogeneity shows no clear large-scale trend in the study region, but can be considered as determined by randomly distributed intermediately sized patches displaying a clear mode in their size distribution. Our approach cannot be applied to the related species *S. affinis*, which showed a strong association to elevation and a large-scale trend in environmental heterogeneity (Fig. 1B).

Although we found evidence that the larger scale clustering was related to a dispersal limitation, a weak habitat association of *S. congestiflora* (it occurred less frequently at a habitat called "low less-steep gullies" in an elevation range between 424 and 460 m and a moderate slope; Gunatilleke et al. 2006) might be present. However, the spatial structure of the habitat was virtually not related to the size of the clusters (Fig. 1A). Large-scale heterogeneity of the pattern was therefore not a critical issue in our analysis. A considerable challenge for further development, however, is to expand our methods for heterogeneous cluster processes.

Point pattern analysis, hypotheses, and null models

Point-pattern analysis is most commonly used as a tool to assess departure from the simplest null models

(CSR for univariate null models and independence or random labeling for bivariate null models). However, an alternative, much richer approach is setting up explicit hypotheses prior to pattern analyses (e.g., Schurr et al. 2004) and to develop specific null models to test these hypotheses. This allows for a precise description of the properties of patterns, can provide deeper insight into the biology of the species, and can generate specific hypothesis to be tested in the field. We followed this approach and derived specific hypotheses on the spatial organization of our size-structured tree population. Monte Carlo simulations of the null model provided a rigorous test for detecting departure from these hypotheses if the null models had no unknown parameters. However, our working hypothesis that the patterns of different size classes showed two distinct scales of clustering required use of null models with four unknown parameters to be estimated by fitting the model to the data. Rejection of this hypothesis for adults and dead trees was unambiguous since the null model with only one critical scale of clustering yielded a good fit of the data. Similarly, rejection of the bivariate hypothesis that the adults were the cluster centers of the small-scale clusters of recruits was unambiguous since the formal fit yielded results inconsistent with the univariate analyses. Our general observation was that attempts to fit a double-cluster process to data that do not have a double-cluster structure failed either because the fitting algorithm did not find a solution or because the fitted parameters indicated only a single-cluster process (e.g., $\sigma_2 \to \infty$ or $\rho_2 \to \infty$). Thus, the first step of evidence against or in favor of our working hypotheses was provided by the fitting procedure itself. The next step, in case that the fitting seemed successful (i.e., for smaller size classes), was to perform Monte Carlo simulations of the null model with the fitted parameters to confirm that the fit was indeed satisfying (i.e., that we cannot distinguish several features of the data from the realizations of the fitted process). Stoyan and Stoyan (1994:300–302) discuss goodness-of-fit tests for fitted point processes in more detail; the probability of an error of type I is certainly greater than the chosen α.

One general problem with this approach is that one has usually only one realization of the process on hand, but different realizations of point processes may not always show exactly the same properties as the underlying process (e.g., Fig. 2F). The best we can do in this situation is to use the cluster processes to describe the spatial structure of our data, but be aware that the fitted parameters may only approximate the real parameters of the overall process.

We emphasize that our approach does not allow inferring process unambiguously from observed pattern. We cannot exclude the possibility that there may be other point processes that fit the data equally well, but that may suggest a different biological interpretation. This is, e.g., illustrated by the different double-cluster processes (Eqs. 2, 4, 5) that may arise by nested clustering or superposition processes. Therefore it is important to formulate the null hypothesis with care and make it as specific as possible and to use complementary information as, e.g., provided by the distribution of nearest neighbor distances. Point pattern analysis techniques are descriptive and inductive, i.e., they can test whether an observed pattern is well described by a given null model and suggest causal relationships that, however, must be proven experimentally (Levin 1992, Silvertown and Wilson 1994, Crawley 1997). We therefore view the techniques developed here more as tools for exploratory data analysis and for generating new hypotheses that can be tested in the field.

Acknowledgments

The authors gratefully acknowledge the permission given to work in Sinharaja World Heritage Site and the accommodation facilities provided by the Forest Department of Sri Lanka, as well as the generous financial assistance given to set up the plot and computerize the database by the John D. and Catherine T. MacArthur Foundation, the Smithsonian Tropical Research Institute, the U.S. National Science Foundation, the Centre for International Development at Harvard University, and the National Institute of Environmental Science of Japan. We also acknowledge the Japanese Society for Promotion of Science for providing a fellowship to N. Gunatilleke during which the initial part of this paper was done. We thank P. S. Ashton, two anonymous reviewers, and especially F. Goreaud for constructive criticism on earlier drafts of the manuscript.

Literature Cited

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

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

Baddeley, A., J. Møller, and R. Waagepetersen. 2000. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. Statistica Neerlandica 54:329–350.

Bartlett, M. S. 1964. Spectral analysis of two-dimensional point processes. Biometrika 51:299–311.

Batista, J. L. F., and D. A. Maguire. 1998. Modeling the spatial structure of topical forests. Forest Ecology and Management 110:293–314.

Condit, R., et al. 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. Pages 298–312 *in* P. den Boer and G. Gradwell, editors. Dynamics of populations. Center for Agricultural Publishing and Documentation, Wageningen, The Netherlands.

Crawley, M. J. 1997. The structure of plant communities. Pages 475–531 *in* M. J. Crawley, editor. Plant ecology. Blackwell, Oxford, UK.

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

Dale, M. R. T. 1999. Spatial pattern analysis in plant ecology. Cambridge University Press, Cambridge, UK.

Dassanayake, M. D., and F. R. Fosberg. 1980–2000. A revised handbook to the flora of Ceylon. Volumes 1–12. Amarind Publishing, New Delhi, India.

de Rosayro, R. A. 1942. The soils and ecology of the wet evergreen forests of Ceylon. Tropical Agriculturist 98:70–80, 153–175.

Diggle, P. J. 2003. Statistical analysis of point patterns. Second edition. Arnold, London, UK.

Dixon, P. M. 2002. Ripley's *K* function. Encyclopedia of Environmetrics 3:1796–1803.
Goreaud, F., and R. Pelissier. 2003. Avoiding misinterpretation of biotic interactions with the intertype K-12-function: population independence vs. random labelling hypotheses. Journal of Vegetation Science 14:681–692.
Grubb, P. 1977. The maintenance of species richness in plant communities: the importance of the regeneration niche. Biological Reviews 53:107–145.
Gunatilleke, C. V. S., I. A. U. N. Gunatilleke, S. Esufali, K. E. Harms, P. M. S. Ashton, D. F. R. P. Burslem, and P. S. Ashton. 2006. Species–habitat associations in a Sri Lankan dipterocarp forest. Journal of Tropical Ecology 22:371–384.
Gunatilleke, C. V. S., I. A. U. N. Gunatilleke, A. U. K. Ethugala, and S. Esufali. 2004. Ecology in Sinharaja rain forest and the forest dynamic plot in Sri Lanka's world heritage site. WHT Publications, Colombo, Sri Lanka.
Harms, K. E., S. J. Wright, O. Calderón, A. Hernández, and E. A. Herre. 2000. Pervasive density-dependent recruitment enhances seedling diversity in a tropical forest. Nature 404: 493–495.
He, F., and K. J. Gaston. 2000. Estimating species abundance from occurrence. American Naturalist 156:553–559.
Hubbell, S. P. 1997. A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. Coral Reefs 16(Supplement):S9–S21.
Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, New Jersey, USA.
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, editors. Tropical rain forest: ecology and management. Blackwell Scientific, London, UK.
Hubbell, S. P., R. B. Foster, S. T. O'Brien, K. E. Harms, R. Condit, B. Wechsler, S. J. Wright, and S. L. de Lao. 1999. Light-gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. Science 283:554–557.
Janzen, D. H. 1970. Herbivores and the numbers of tree species in tropical forests. American Naturalist 104:501–528.
Levin, S. A. 1992. The problem of pattern and scale in ecology. Ecology 73:1943–1967.
Liebhold, A. M., and J. Gurevitch. 2002. Integrating the statistical analysis of spatial data in ecology. Ecography 25: 553–557.
Manokaran, N., J. V. LaFrankie, K. M. Kochuman, E. S. Quah, J. E. Klahn, P. S. Ashton, and S. P. Hubbell. 1990. Methodology for the 50 ha research plot at Pasoh forest reserve. Forest Research Institute Malaysia Research Pamphlet number 104. Forest Research Institute, Kepong, Malaysia.
Møller, J., and R. Waagepetersen. 2003. Statistical inference and simulation for spatial point processes. Chapman and Hall/CRC, Boca Raton, Florida, USA.
Plotkin, J. B., J. Chave, and P. S. Ashton. 2002. Cluster analysis of spatial patterns in Malaysian tree species. American Naturalist 160:629–644.
Plotkin, J. B., M. D. Potts, N. Leslie, N. Manokaran, J. LaFrankie, and P. S. Ashton. 2000. Species-area curves, spatial aggregation, and habitat specialization in tropical forests. Journal of Theoretical Biology 207:81–99.
Potts, M. D., S. J. Davies, W. H. Bossert, S. Tan, and M. N. Nur Supardi. 2004. Habitat heterogeneity and niche structure of trees in two tropical rain forests. Oecologia 139:446–453.
Ripley, B. D. 1976. The second-order analysis of stationary point processes. Journal of Applied Probability 13:255–266.
Ripley, B. D. 1977. Modelling spatial patterns. Journal of the Royal Statistical Society, Series B 39:172–192.
Ripley, B. D. 1981. Spatial statistics. John Wiley, New York, New York, USA.
Schurr, F. M., O. Bossdorf, S. J. Milton, and J. Schumacher. 2004. Spatial pattern formation in semi-arid shrubland: a priori predicted versus observed pattern characteristics. Plant Ecology 173:271–282.
Silvertown, J., and J. B. Wilson. 1994. Community structure in a desert perennial community. Ecology 75:409–417.
Stoyan, D., and H. Stoyan. 1994. Fractals, random shapes and point fields. Methods of geometrical statistics. John Wiley & Sons, New York, New York, USA.
Stoyan, D., and H. Stoyan. 1996. Estimating pair correlation functions of planar cluster processes. Biometrical Journal 38: 259–271.
Thomas, M. 1949. A generalization of Poisson's binomial limit for use in ecology. Biometrika 36:18–25.
Tomppo, E. 1986. Models and methods for analysing spatial patterns of trees. Communicationes Instituti Forestalis Fenniae 138:1–65.
Turner, M. G. 1989. Landscape ecology: the effect of pattern on process. Annual Review of Ecology and Systematics 20: 171–197.
Watson, D. M., D. A. Roshier, and T. Wiegand. 2007. Spatial ecology of a parasitic shrub: patterns and predictions. Austral Ecology 32:359–369.
Whitmore, T. C. 1984. Tropical rain forests of the Far East. Clarendon Press, Oxford, UK.
Wiegand, T., W. D. Kissling, P. A. Cipriotti, and M. R. Aguiar. 2006. Extending point pattern analysis to objects of finite size and irregular shape. Journal of Ecology: 94:825–837.
Wiegand, T., and K. A. Moloney. 2004. Rings, circles, and null-models for point pattern analysis in ecology. Oikos 104: 209–229.
Wills, C., and R. Condit. 1999. Similar non-random processes maintain diversity in two tropical rainforests. Proceedings of the Royal Society of London B 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.

APPENDIX A

Analytical formulas of the point processes (*Ecological Archives* E088-191-A1).

APPENDIX B

Parameter fitting (*Ecological Archives* E088-191-A2).

APPENDIX C

Qualitative reconstruction of the 34 large-scale clusters of the double-cluster component pattern and cluster occupancy for recruits and adults (*Ecological Archives* E088-191-A3).