

**Thorsten Wiegand, Savitri Gunatilleke, Nimal Gunatilleke, and Toshinori Okuda. 2007. Analyzing the spatial structure of a Sri Lankan tree species with multiple scales of clustering. *Ecology* 88:3088–3102.**

Appendix A. Analytical formulas of the point processes.

Essential ingredient of the Thomas process is that the distance  $r$  between offspring and parents follows a radial symmetric bivariate normal distribution  $h(r, \sigma)$  where  $\sigma^2$  is the variance of the distribution. For the derivation of analytical formulas we define the probability  $H(r, \sigma_1, \sigma_2)$  that two offspring of type 1 and 2 of the same parent (but with different variances  $\sigma_1^2$  and  $\sigma_2^2$ ) are a distance  $r$  away:

$$H(r, \sigma_1, \sigma_2) = \int h(x, \sigma_1) h(x - r, \sigma_2) dx \quad (\text{A.1})$$

The normal distribution has the convenient property that  $H(r, \sigma_1, \sigma_2)$  is also a normal distribution with variance  $\sigma_1^2 + \sigma_2^2$ . Thus, Eq. A.1 yields

$$H(r, \sigma_1, \sigma_2) = h(r, \sqrt{\sigma_1^2 + \sigma_2^2}) \quad (\text{A.2})$$

*Bivariate double-cluster process under antecedent condition*

The bivariate double-cluster process under antecedent condition assumes that pattern 1 follows a Thomas process and that points of pattern 2 are offspring of the points of pattern 1. We calculate the probability  $\lambda_{12}(r)$  to find a type 1 point at locations  $I_1$  and a type 2 point at a location  $I_2$  which is a distance  $r = |I_1 - I_2|$  away from location  $I_1$ . We need to separate the two cases where the type 2 point is offspring of the type 1 point and where the type 2 point is offspring of another type 1 point (Fig. A1). In the first case the calculation is simple: we multiply the intensity of type 1 points ( $= \lambda_1$ ) with the expected number of offspring ( $\mu_2 = \lambda_2/\lambda_1$ ) with the probability that an offspring is distance  $r$  away from the type 1 point [ $= h(r, \sigma_2)$ ]. To derive the contribution of the second case we need to multiply the probability to find a second type 1 point at a location  $I_2$  which is a vector distance  $\mathbf{R}$  from a first type 1 point [ $= \lambda_1^2 g(\mathbf{R}, \sigma_1)$ ] with the expected number of offspring ( $\mu_2 = \lambda_2/\lambda_1$ ) with the probability that the type 2 offspring is a distance  $|\mathbf{R} - \mathbf{r}|$  away from the second type 1 point [ $= h(\mathbf{R} - \mathbf{r}, \sigma_2)$ ] and integrate over all possible vector distances  $\mathbf{R}$ :

$$\begin{aligned}
& \lambda_{12}(r) \\
&= \lambda_1 \mu_2 h(r, \sigma_2) + \int \lambda_1 \lambda_1 g(R, \sigma_1, \rho_1) \mu_2 h(R-r, \sigma_2) dR \\
&= \lambda_1 \left[ \frac{\lambda_2}{\lambda_1} h(r, \sigma_2) + \lambda_2 \int \left[ 1 + \frac{1}{\rho_1} h(R, \sqrt{2}\sigma_1) \right] h(R-r, \sigma_2) dR \right] \\
&= \lambda_1 \lambda_2 \left[ \frac{1}{\lambda_1} h(r, \sigma_2) + \int h(R-r, \sigma_2) dR + \int \frac{1}{\rho_1} h(R, \sqrt{2}\sigma_1) h(R-r, \sigma_2) dR \right] \\
&= \lambda_1 \lambda_2 \left[ \frac{1}{\lambda_1} h(r, \sigma_2) + 1 + \frac{1}{\rho_1} h(r, \sqrt{2\sigma_1^2 + \sigma_2^2}) \right]
\end{aligned} \tag{A.3}$$

Note that  $H(r, \sigma, \sigma) = h(r, \sqrt{2}\sigma)$  (equation A2),  $\mu_2 = \lambda_2/\lambda_1$ , and  $\lambda_1 = \rho_2$ . Thus, the  $g_{12}(r)$  for the bivariate double-cluster process under antecedent condition yields:

$$g_{12}(r) = 1 + \frac{1}{\rho_2} h(r, \sigma_2) + \frac{1}{\rho_1} h(r, \sqrt{2\sigma_1^2 + \sigma_2^2}) \tag{A.4}$$

#### *Univariate double-cluster process*

Figure A1 shows that the second-order intensity function  $\lambda_{22}(r)$  of the univariate double-cluster follows directly from the formula for the bivariate double-cluster process under antecedent condition (Eq. A.4):

$$\lambda_{22}(r) = \int \lambda_2 g_{12}(R-r) h(R, \sigma_2) dR. \tag{A.5}$$

Integrating Eq. A.5 using Eq. A.4 yields:

$$g_{22}(r) = 1 + \frac{1}{\lambda_1} h(r, \sqrt{2}\sigma_2) + \frac{1}{\rho_1} h(r, \sqrt{2\sigma_1^2 + 2\sigma_2^2}) \tag{A.6}$$

#### *Superposition of independent component processes*

Diggle (2003) provides a formula for the  $K$ -function of the superposition of two independent component processes with intensities  $\lambda_1$  and  $\lambda_2$  where  $\lambda = \lambda_1 + \lambda_2$ :

$$K(r) = \lambda^{-2} (\lambda_1^2 K_1(r) + \lambda_2^2 K_2(r) + 2\lambda_1 \lambda_2 \pi r^2) \tag{A.7}$$

With  $g(r) = K'(r)/(2\pi r)$  we find

$$g(r) = \lambda^{-2}(\lambda_1^2 g_1(r) + \lambda_2^2 g_2(r) + 2\lambda_1 \lambda_2). \quad (\text{A.8})$$

Application of Eq. A.8 for superposition of two independent Thomas processes

$$g(r, \sigma_p, \rho_p) = 1 + \frac{1}{\rho_p} \frac{\exp(-r^2/4\sigma_p^2)}{4\pi\sigma_p^2}, \quad (\text{A.9})$$

where  $p = 1, 2$  for component process 1 and 2, respectively, yields the pair-correlation function of the superposition process: where  $p = 1, 2$  for component process 1 and 2, respectively, yields the pair-correlation function of the superposition process:

$$g(r, \sigma_1, \rho_1, \sigma_2, \rho_2) = 1 + \underbrace{p_2^2 \frac{1}{\rho_2} \frac{\exp(-r^2/4\sigma_2^2)}{4\pi\sigma_2^2}}_{\text{contribution of Thomas process 2}} + \underbrace{(1-p_2)^2 \frac{1}{\rho_1} \frac{\exp(-r^2/4\sigma_1^2)}{4\pi\sigma_1^2}}_{\text{contribution of Thomas process 1}} \quad (\text{A.10})$$

where  $p_2 = \lambda_2/\lambda = (1 - p_1)$ .

Application of Eq. A.8 for the superposition of a homogeneous Poisson process [ $g_H(r) = 1$ ] and the nested double-cluster process Eq. A.6 [abbreviated as  $g_D(r) = 1 + c(r)$ ] yields

$$g(r, \sigma_1, \rho_1, \sigma_2, \rho_2) = 1 + p_C^2 c(r) \quad (\text{A.11})$$

where  $p_C$  is the proportion of the points of the pattern belonging to the double-cluster component process.

Note that the functional form of the nested double-cluster process (Eq. A.6) and the superposed double-cluster processes (Eq. A.10 and Eq. A.11) are the same. It is thus not possible to distinguish between these processes based on second-order characteristics alone. Therefore, other characteristics of the pattern, such as the distribution of the nearest neighbor distances, need to be analyzed additionally (Stoyan and Stoyan 1994; Diggle 2003).

Comparison of the nested double-cluster process (Eq. A.6) with the superposed double-cluster processes (Eq. A.10 and Eq. A.11) shows that the parameters  $\rho_1$  and  $\rho_2$  of Eq. A.6, respectively, which give the intensity of the first- and second-generation parents, may not give the true parents intensities if the process is a superposed process. If a pattern was created by the superposed process (Eq. 10), but fitted with the pair-correlation function of the nested process (Eq. A.6), the fit will yield the parameters  $\rho_1^*$  and  $\rho_2^*$ . However, the true parents intensities are smaller and yield  $\rho_2 = \rho_2^2 \rho_2^*$  and  $\rho_1 = (p_1 - 1)^2 \rho_1^*$ . Similarly, if a pattern was created by the superposed process (Eq. 11), the true parents intensities are smaller and yield  $\rho_2 = p_C^2 \rho_2^*$  and  $\rho_1 = p_C^2 \rho_1^*$ .

The fitted intensities  $\rho_1^*$  and  $\rho_2^*$  of the first- and second-generation parents can be used for finding evidence that a given double-cluster process may have a superposition component. A nested process requires that there are more offspring than parents (i.e.,  $\rho_1 < \rho_2 = \lambda_1 < \lambda_2$ ).

Thus, if the fit with Eq. A.6 yields e.g.  $\rho_1 > \rho_2$  this could be taken as evidence for a superposition with  $p_2 \gg p_1$ . Similarly,  $\rho_2 > \lambda_2$  would indicate that  $p_1 \approx p_2$  since in this case the estimated  $\rho_2$  would be four times the “true” one.

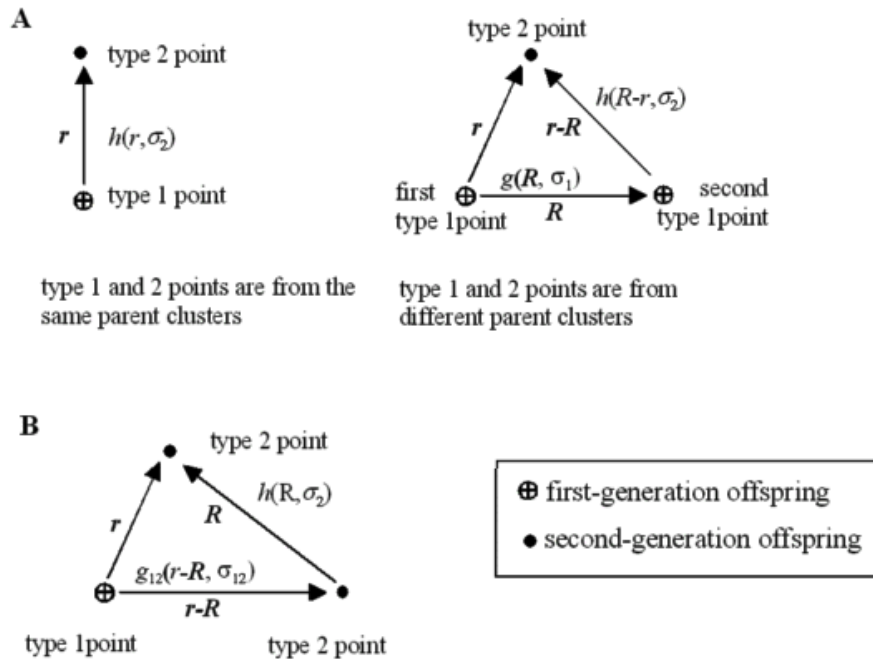


FIG. A1. Scheme for calculating the pair-correlation function of the double-cluster processes. (A) Bivariate double-cluster process under antecedent condition (Eq. A 4). (B) Univariate double-cluster process (Eq. A 6).

#### LITERATURE CITED

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

Stoyan, D., and H. Stoyan. 1994. Fractals, random shapes and point fields. Methods of geometrical statistics. John Wiley and Sons, New York, New York, USA.

**Thorsten Wiegand, Savitri Gunatilleke, Nimal Gunatilleke, and Toshinori Okuda. 2007. Analyzing the spatial structure of a Sri Lankan tree species with multiple scales of clustering. *Ecology* 88:3088–3102.**

Appendix B. Parameter fitting.

We first calculate the discrepancy between model and data separately for the  $g$ -function (=  $error\_g$ ) and the  $L$ -function (=  $error\_L$ ) and by using the minimal-contrast method (e.g., Stoyan and Stoyan 1994; Diggle 2003) we minimize their geometric mean (=  $error\_Lg$ ):

$$\begin{aligned}
 error\_L &= \frac{\sum_{r=r_0}^{r_{\max}} [\hat{L}(r)^c - L(r, \sigma, \rho)^c]^2}{\sum_{r=r_0}^{r_{\max}} [\hat{L}(r)^c]^2} \\
 error\_g &= \frac{\sum_{r=r_0}^{r_{\max}} [\hat{g}(r)^c - g(r, \sigma, \rho)^c]^2}{\sum_{r=r_0}^{r_{\max}} [\hat{g}(r)^c]^2} \\
 error\_Lg &= \sqrt{error\_g * error\_L}
 \end{aligned}
 \tag{B.1}$$

with tuning constants  $r_0$ ,  $r_{\max}$ , and  $c$ . The constant  $r_0$  is the minimal scale of the fit,  $r_{\max}$  the maximal scale of fit, and  $c$  a power transformation. The error functions  $error\_g$  and  $error\_L$  gives the fraction of the total sum of squares of the transformed empirical  $g$ -function and  $L$ -function, respectively, which is not explained by the model. To fit the parameters we minimize the average contrast  $error\_Lg$  of the  $g$ - and the  $L$ -function (Eq. B1). The usual approach is to minimize the contrast between the  $K$ -function and the data (e.g., Stoyan and Stoyan 1994; Diggle 2003). Theoretically, the  $g$ - and the  $L$ -function contain the same information and fitting using  $g$  or  $L$  should therefore yield the same parameter estimates. However, in practice we found improved results by fitting both simultaneously. The explanation for this is that the  $g$ -function is sensitive at smaller scales  $r$ , but approaches the asymptote  $g = 1$  relatively quick at larger scales (compared to the way  $L$  approaches the asymptote 0). In turn, the accumulative  $L$ -function is not very sensitive at small scales but sensitive at larger scales. Therefore we balance between a good fit at small scales and a good fit at larger scales when optimizing the fit for both.

An immediate question is how to choose appropriate values for the tuning constants  $r_0$ ,  $r_{\max}$ , and  $c$ . If we aim to fit the large-scale component of a double-cluster process (i.e., the parameters  $\sigma_1$  and  $\rho_1$  of Eq. 2) using a Thomas process, comparison of the fitted Thomas process with the empirical pair-correlation function provides in general clear inductions about selection of  $r_0$  if the critical scales of clustering are well separated. The best approach, however, is to try several values for  $r_0$  and visualize the  $g$ - and  $L$ - function of the data and the fit. In our experience it was not necessary to use minimal-contrast methods for determination of  $r_0$ . Because of the memory of the  $K$ -function, it was observed in the literature that the fit may depend sensitively on the tuning constant  $r_{\max}$  (e.g., Batista and Maguire 1998; Plotkin et al. 2000). Using the pair-correlation function and the transformed  $K$ -function (Eq. 6) this sensitivity largely disappears. However,  $r_{\max}$  should not be much larger than the range of the clustering. A power transformation with  $c > 1$  weights larger values of  $L(r)$  or  $g(r)$  more than a transformation with  $c = 1$ , whereas a transformation with  $c < 1$  weights larger differences

less. We use in all analyses a power transformation with  $c = 1$  for the  $L$ -function and a power transformation with  $c = 0.5$  for the  $g$ -function to reduce the high sensitivity of the  $g$ -function to smaller scales. The error functions *error\_g*, *error\_L*, and *error\_Lg* give the fraction of the sum of squares of the transformed  $L$ - or  $g$ -functions of the data which are not explained by the fit. Therefore, an assessment of the uncertainty in the estimates of the parameter  $\sigma$  and  $\rho$  can be done by determining the area in the  $\sigma$ - $\rho$  parameter space for which the error is smaller than a certain level of say, 0.025 or 0.01.

#### LITERATURE CITED

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

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

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.

Stoyan, D., and H. Stoyan. 1994. *Fractals, random shapes and point fields. Methods of geometrical statistics*. John Wiley and Sons, New York, New York, USA.

---

Thorsten Wiegand, Savitri Gunatilleke, Nimal Gunatilleke, and Toshinori Okuda. 2007. Analyzing the spatial structure of a Sri Lankan tree species with multiple scales of clustering. *Ecology* 88:3088–3102.

Appendix C. Qualitative reconstruction of the 34 large-scale clusters of the double-cluster component pattern and cluster occupancy for recruits and adults.

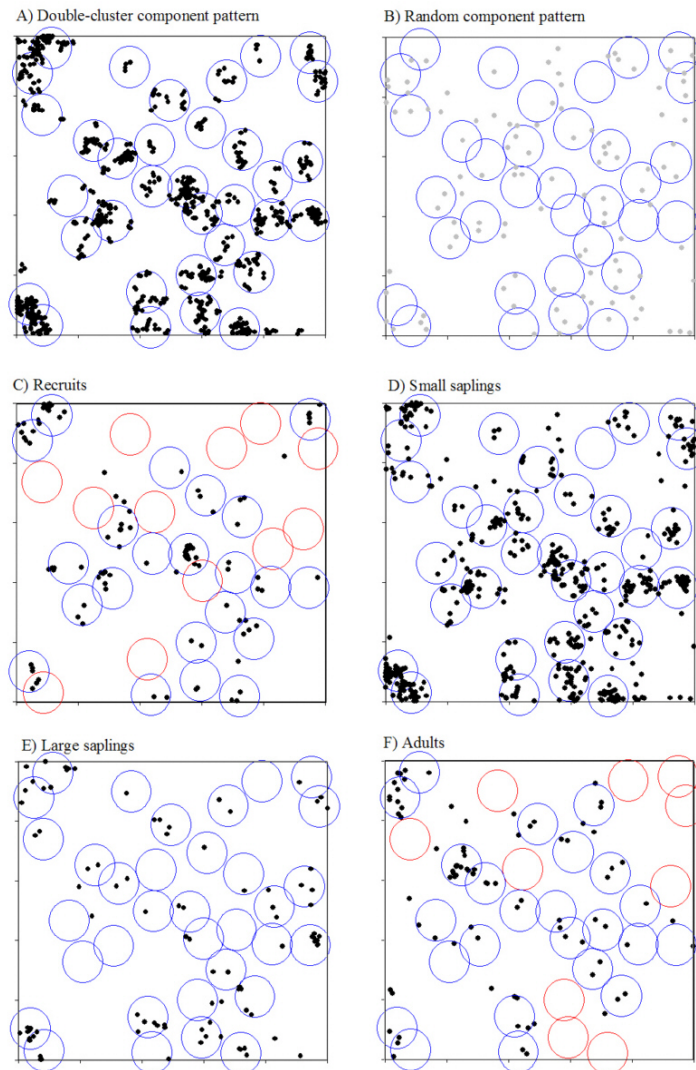


FIG. C1. Qualitative reconstruction of the 34 large-scale clusters of the double-cluster component pattern and cluster occupancy for recruits and adults. (A) The reconstruction was done visually by allocating 34 circles of the expected size of the clusters over the points of the pattern. The size of the clusters represented by the open circles corresponds a radius of  $2.5\sigma_1 = 32\text{m}$ , which is the radius where about 95% of all points of a cluster should be within the corresponding circle. (B) The random component pattern compared to the reconstructed clusters. (C) – (F) The different size classes compared to the reconstructed patterns. Empty clusters for recruits and adults are indicated by red color. Recruits were absent in 12 clusters and adults in 10 clusters.