Learning outcome

The student has also knowledge about basic concepts of the theory of point random fields, including K- and L-statistics, spatial Poisson, Cox and hard-core fields, and simulation of such point fields. (...) Lastly, the student has knowledge of parameter estimation in spatial random fields.

Extra text book: Statistical Analysis and Modelling of Spatial Point Patterns by Janine Illian, Antti Penttinen, Helga Stoyan, Dietrich Stoyan (2008)

TODAY:

Parameter estimation
- Maximum likelihood
- Bayesian
- Log-Gaussian Cox process

'Satatistics'
- Intensity, $\lambda(s)$
- 2nd order characteristics $K()$, $L()$, $\lambda_2()$ and $g(h)$

Simulation of Gibbs models

Spatial Point Processe (SPP)

From Cressie and Wikle:

4.3 SPATIAL POINT PROCESSES

A spatial point process is a stochastic process governing the location of events (equivalently, points) $\{\mathbf{s}_i\}$ in some set $D_s \subset \mathbb{R}^d$, where the number of such events in D_s is also random (e.g., Diggle, 2003). In the simplest case, the

and we only consider:

Only *simple* spatial point processes in $\mathbb{R}^d$ (i.e., almost surely, either no event or a single event occurs at any point) will be considered. We characterize the

Data

Number of points: $Z(D_s) = m$

Locations: $\{s_1, s_2, \ldots, s_m\}$

Need to find likelihood function

Intensity functions

1 order intensity:

$$\lambda_1(s) = \lim_{|ds|\to 0} \frac{E(Z(ds))}{|ds|}$$

2. order intensity:

$$\lambda_2(s_1, s_2) = \lim_{|ds_1|\to 0, |ds_2|\to 0} \frac{E(Z(ds_1)Z(ds_2))}{|ds_1||ds_2|}$$

...

mth order intersity

$$\lambda_m(s_1, s_2, \ldots, s_m) = \lim_{\cdots\to 0} \frac{E(Z(ds_1)Z(ds_2)\ldots Z(ds_m))}{|ds_1||ds_2|\ldots|ds_m|}$$

Ex. Random: Homogenous Poisson point process (HPPP)

An important, but simple, example is the *Poisson point process* Z, for which

$$Z(A)|\lambda^o \sim Poi(\lambda^o|A|), \qquad A \subset D_s, \tag{4.159}$$

where $\lambda^o > 0$ is a parameter of the Poisson point process, and recall that $|A|$ is the d-dimensional volume of A. More details on this spatial point process are given in Section 4.3.1.

Sampling from Poisson point process

Trees in $1km \times 1km$ domain, with intensity $\lambda^0 = 15 trees/km^2$

- Sample $m \sim Pos(15)$
- for $i = 1 : m$
  - sample location randomly in domain $\{s_i\}$
- end

Hierachical statistical models (HM), and Cox process

Data model: $[Z|Y,\theta]$

Process model: $[Y|\theta]$

Parameter model: $[\theta]$

Cox process, Inhomogenious Poisson Point Process (IPPP)

Data model: $[Z|\lambda] \sim Pos(\int_A \lambda(d)dx)$

Process model: $\lambda(s)$ (Cox process)

Pdf for Inhomogenious Poisson Process

(Cressie & Wikle page 215)

For example, because the (inhomogeneous) Poisson point process has independent counts in disjoint regions, for that model, (4.184) becomes

$$f(\mathbf{s}_1, \ldots, \mathbf{s}_m | m) = \prod_{i=1}^{m} \{\lambda(\mathbf{s}_i) / \int_{D_s} \lambda(\mathbf{x})\, d\mathbf{x}\},$$

and from (4.185),

$$p_m = \exp(-\int_{D_s} \lambda(\mathbf{x})\, d\mathbf{x}) \{\int_{D_s} \lambda(\mathbf{x}) d\mathbf{x}\}^m / m!, \qquad m = 0, 1, \ldots,$$

which is the probability mass function of a Poisson distribution. Conse-

Need to evaluate $\int_D \lambda(x) dx$, (same dimension as D, for us 2D)

Summery likelihood and simulation of Gibbs processes

- OK for models with onlt 1.order characteristics
- General for models with 2nd order characteristics:
  - Do not know $\int \dots \int_D \lambda_m(x_1, \dots x_m) dx_1 \dots dx_2$, and this is a function of the parameters (*intractable likelihoods*)
  - Solutions estimation: Integral(s) with MCMC, or composite likelihood.
- Simulations Gibbs processes (and general)
  - Can sample from $f(s_1, \dots, s_m|m)$ using Metropolise-Hasting (update one point at time)
  - To change number of points, use reversible jump (MH methods for varying dimensions).

Kernel estimation of intensity

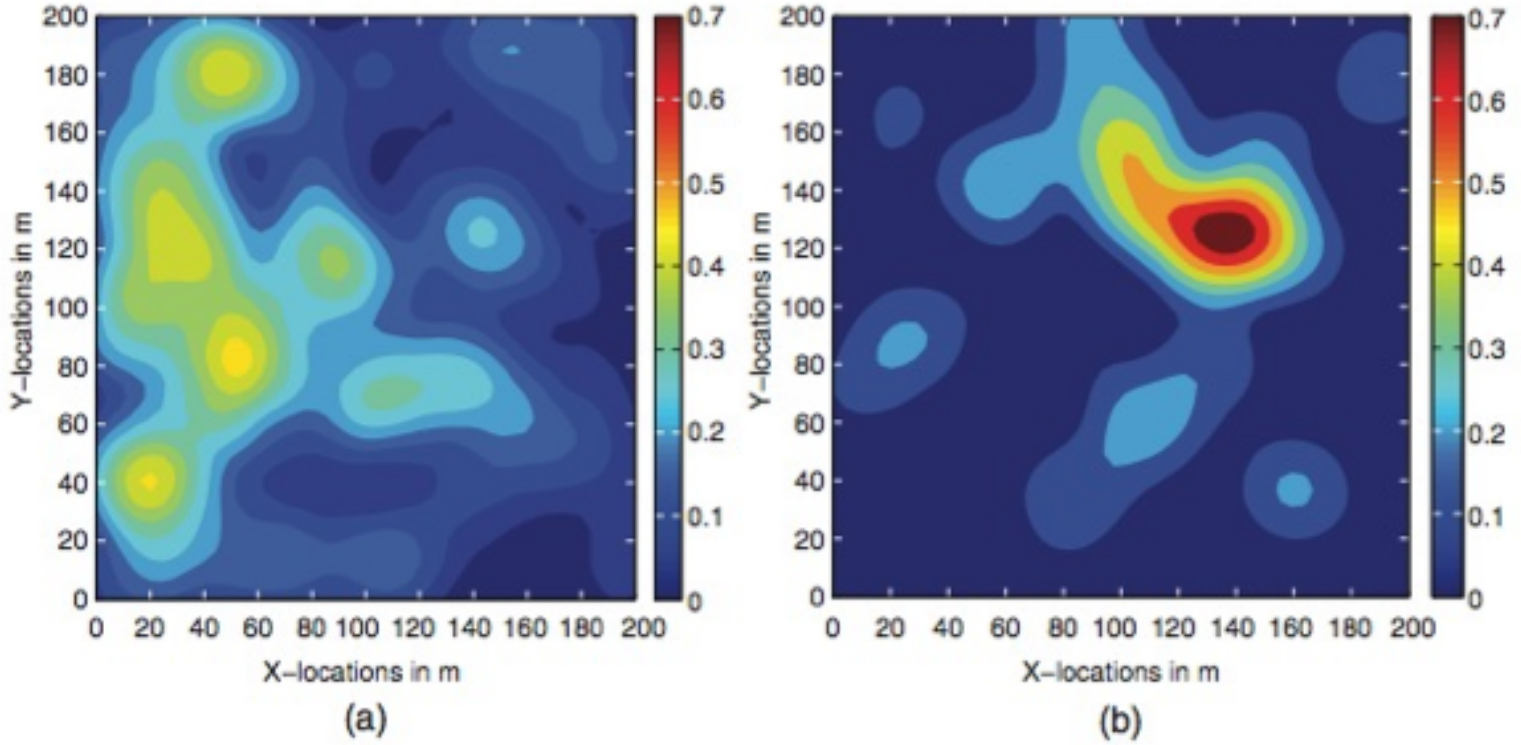

Figure 4.16 (a) Contour plot of estimated first-order intensity of $m_A = 271$ adult longleaf pine trees in the 4-ha study area in the Wade Tract; bandwidth $b = 30$ m. (b) Contour plot of estimated first-order intensity of $m_S = 159$ subadult longleaf pine trees in the same study area as (a); bandwidth $b = 30$ m.

Edge-effects fix

From Illian etat 2008, page 185

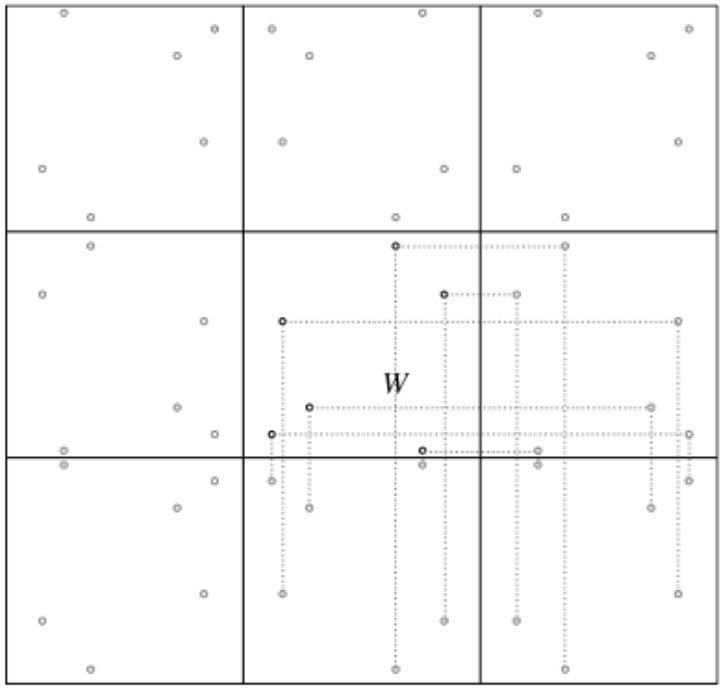

Figure 4.3 A point pattern in a rectangular window W and its continuation by reflection.

K-function (left) subadult longleaf pines

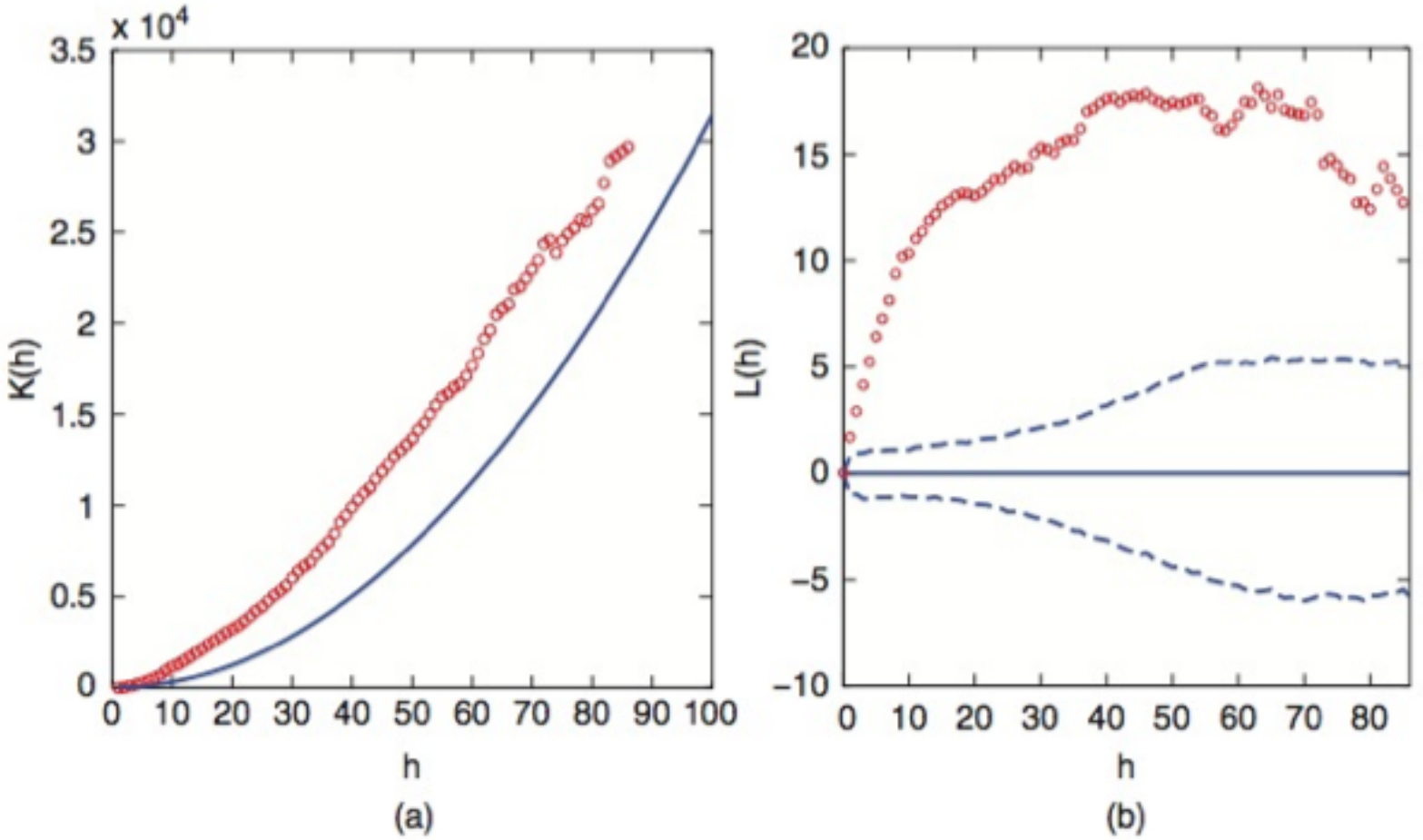

Figure 4.17 **(a)** Estimated K function for the $m_S = 159$ subadult longleaf pine trees (red circles) in the 4-ha study area in the Wade Tract; the theoretical K function for CSR is superimposed (solid blue line). **(b)** Estimated L function for the $m_S = 159$ subadult longleaf pine trees (red circles) obtained from **(a)**; the 95% pointwise confidence limits for L values based on 1000 CSR realizations (dashed blue lines) are superimposed.

K-function and L-function for isotropic point process

from Cressie & Wikle page 210 and 211

K-function

within distance h of an arbitary event. Finally, the definition of the *homogeneous* K function is (Ripley, 1976)

$$K(h) \equiv (\lambda^o)^{-1} E\begin{pmatrix}\text{number of extra events within}\\ \text{distance } h \text{ of an arbitrary event}\end{pmatrix}, \qquad h \geq 0. \quad (4.174)$$

L-function

$$L(h) \equiv \{K(h)\Gamma(1+(d/2))/\pi^{d/2}\}^{1/d} - h, \qquad h \geq 0. \quad (4.179)$$

(often without '$-h$')

2nd order intensity and Pair-correlation function

2nd order intensity

$$\lambda_2(s,x) = \lim_{|ds|\to 0, |dx|\to 0} \frac{E(Z(ds)Z(dx))}{|ds||dx|}$$

Pair correlation function

$$g(s,x) = \frac{\lambda_2(x,s)}{\lambda(s)\lambda(x)}$$

For HPPP: $\lambda_2(s,x) = \lambda(s)\lambda(x)$ and $g(s,x) = 1$

Example: Gold particles

From Illian et al (2008)

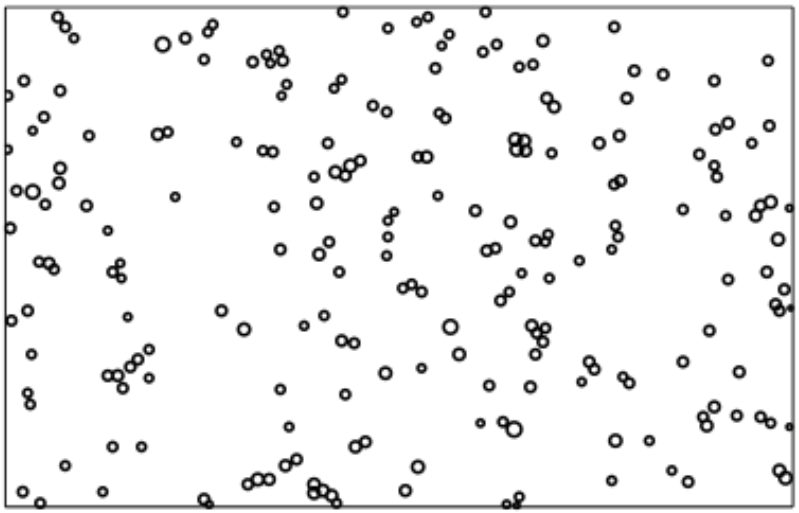

Figure 1.3 Ultrathin section of a pellet of purified tobacco rattle virus after immunogold labelling with a goat antirabbit gold (size 15 nm) probe in a rectangular window of size 1064.7×676 nm. The 218 gold particles are identifiable as dark spots in the electron-microscopic image. The diameters of the small circles are proportional to the gold particle diameters. Data courtesy of C. Glasbey.

Pair-correlation function for gold particles

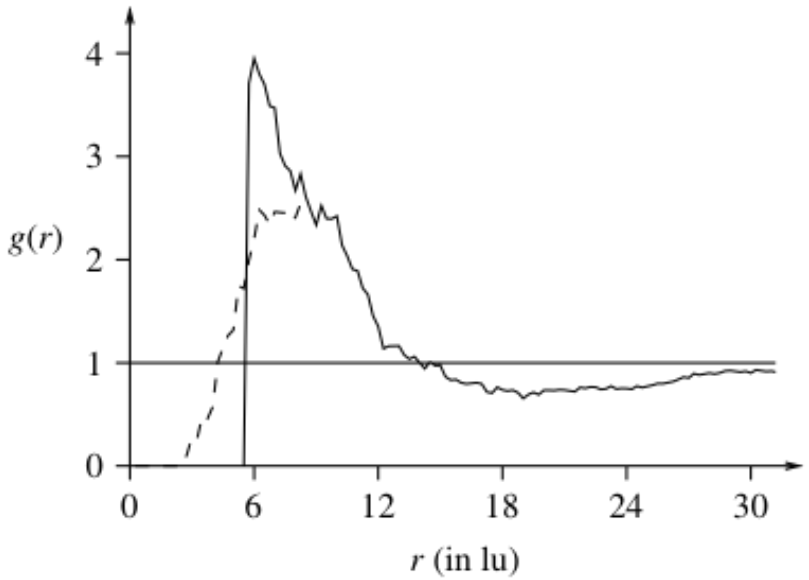

Figure 4.20 The empirical pair correlation function of the pattern of gold particles, obtained with the estimator (4.3.38) and bandwidths $h = 3$ lu for $r \leq 20$ lu and 6 lu for $r > 20$ lu and improved with the reflection method. The dashed line shows the result without this correction. A comparison with Figure 4.18 reveals the advantages of using $g(r)$ as opposed to $L(r)$ as an instructive summary characteristic.

Matern cluster process

From Illian et al 2008 page 379

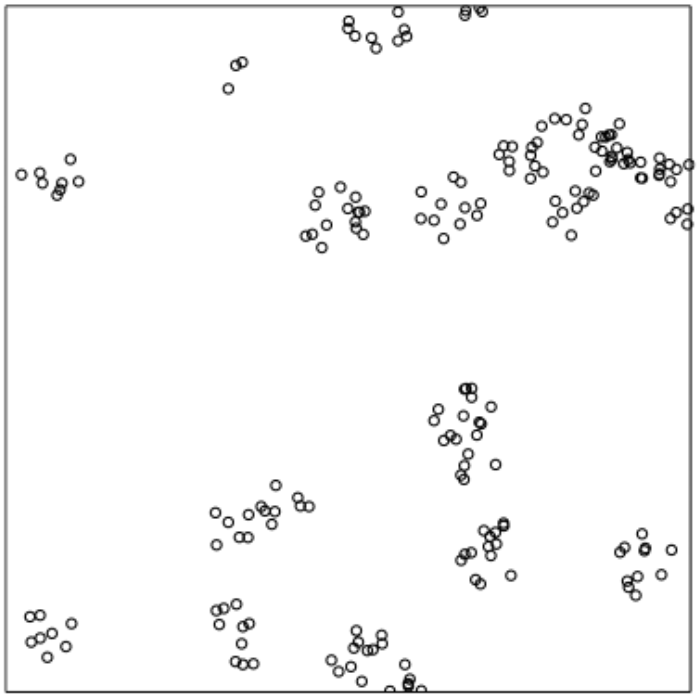

Figure 6.3 Simulated planar Matérn cluster process in [1]. The intensity is $\lambda = 200$, the cluster radius is $R = 0.05$ and the mean number of points per cluster is $\bar{c} = 10$.

Ex Bandwidth paire correlation

From Illians et al 2008 page 237

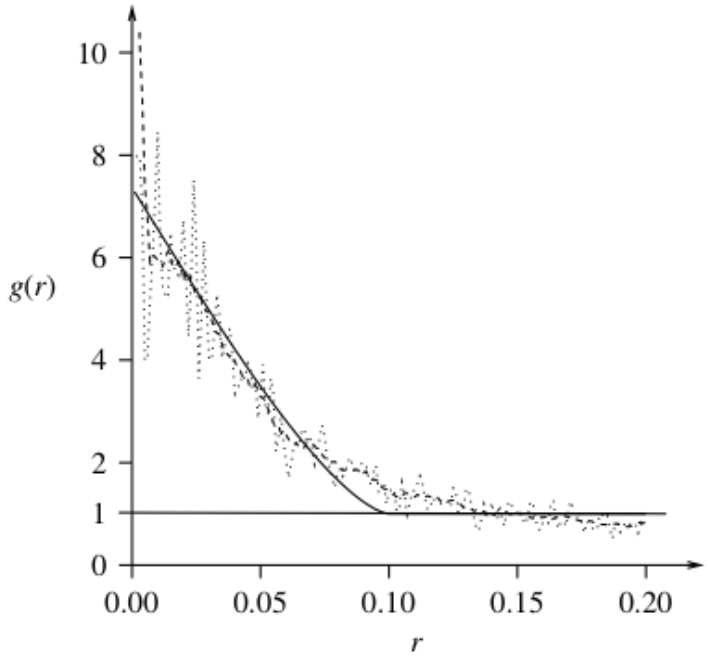

Figure 4.23 The theoretical pair correlation function $g(r)$ of a Matérn cluster process as described in the text (solid line) and its estimates $\hat{g}(r)$ with $h = 0.007$ (dashed line) and 0.001 (dotted line). The irregular fluctuations of the estimates for r larger than 0.1 do not provide any interpretable information.

Phlebocarya filifolia plants

From Illians et al 2008 page 8

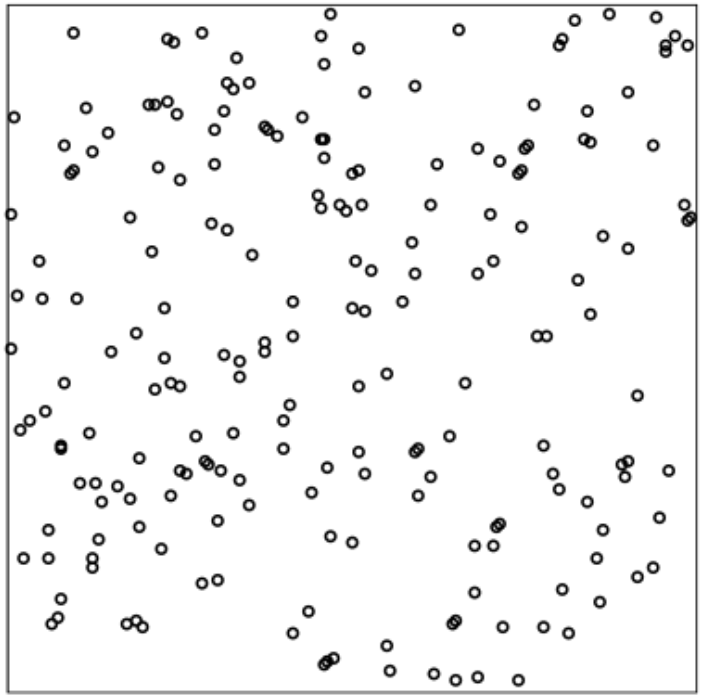

Figure 1.4 Positions of 207 *Phlebocarya filifolia* plants in a 22×22 m square at Cooljarloo near Perth, West Australia. Data courtesy of P. Armstrong.

Pair correlation function Phlebocarya

From Illians et al 2008 page 221

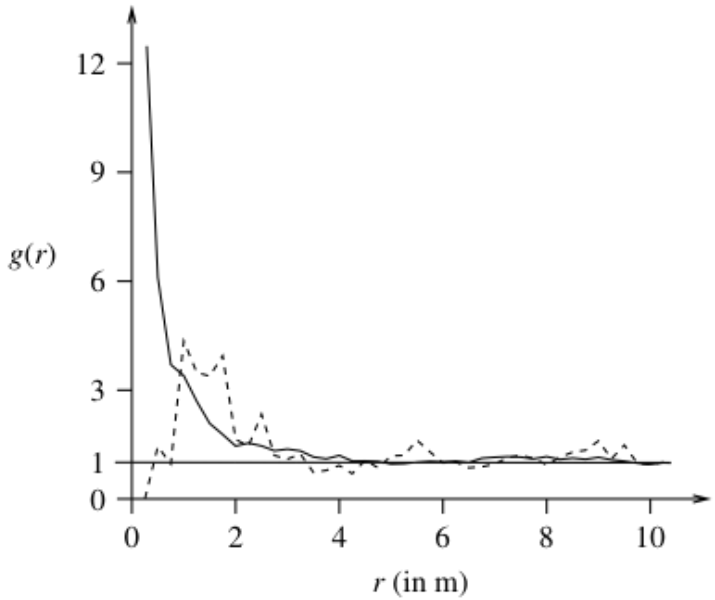

Figure 4.19 Empirical pair correlation functions for the pattern of *Phlebocarya* positions, obtained with bandwidths $h = 0.5$ (dashed line) and 2.0 m (solid line). The large values for small r obtained with $h = 2$ m indicate strong clustering, while the values larger than 1 for r around 8 m may indicate larger clusters. For the smaller h the lattice nature of the pattern becomes apparent. The use of adapted bandwidths makes sense here.