Hong Kong Baptist University
HKBU Institutional Repository

HKBU Staff Publication

2003

Spatial point pattern analysis by using Voronoi diagrams and Delaunay tessellations – A comparative study

S.N. Chiu
Hong Kong Baptist University, snchiu@hkbu.edu.hk

This document is the authors' final version of the published article.
Link to published article: http://dx.doi.org/10.1002/bimj.200390018

APA Citation
Chiu, S. (2003). Spatial point pattern analysis by using Voronoi diagrams and Delaunay tessellations – A comparative study. *Biometrical Journal, 45* (3), 367-376. https://doi.org/10.1002/bimj.200390018

This Journal Article is brought to you for free and open access by HKBU Institutional Repository. It has been accepted for inclusion in HKBU Staff Publication by an authorized administrator of HKBU Institutional Repository. For more information, please contact repository@hkbu.edu.hk.

Spatial Point Pattern Analysis by using Voronoi Diagrams and Delaunay Tessellations — A Comparative Study

S. N. Chiu

Hong Kong Baptist University

SUMMARY. Given a spatial point pattern, we use various characteristics of its Voronoi diagram and Delaunay tessellation to extract information of the dependence between points. In particular, we use the characteristics to construct statistics for testing complete spatial randomness. It is shown that the minimum angle of a typical Delaunay triangle is sensitive to both regularity and clustering alternatives, whilst the triangle's area or perimeter is more sensitive to clustering than regularity. These statistics are also sensitive to the Baddeley-Silverman cell process.

KEY WORDS: Complete spatial randomness, Delaunay tessellation, goodness of fit, spatial point pattern, Voronoi diagram.

RUNNING TITLE: Spatial Point Pattern Analysis

1. Introduction

An important task of the analysis of spatial point patterns is to examine the dependence between points. Many summary functions, for example, the nearest neighbour distance distribution, Ripley's K-function, the empty space function and the periodogram, have been proposed to extract such information. One way to assess how good a summary function can capture the dependence structure of a given point pattern is to see how sensitive it will be when it is used to test the complete spatial randomness (CSR) hypothesis against various alternatives.

However, the distribution theory for most summary functions has not yet been well developed and so simulation-based inference (Møller and Waagepetersen, 2002) has usually been adopted. Comparative studies can be found in the literature (Diggle, 1979; Grabarnik and Chiu, 2002; Mugglestone and Renshaw, 2001; Ripley, 1979; Thönnes and van Lieshout, 1999), and the conclusion from these studies is that there is no uniformly best test.

This paper discusses test statistics constructed from the Voronoi diagram and the Delaunay tessellation generated by the observed point pattern. Since the Voronoi diagram and the Delaunay tessellation are uniquely determined by the generating points and vice versa, they can be used to construct summary functions to test the CSR hypothesis.

The idea of course is not new. Geographers have used Voronoi diagrams and Delaunay tessellations to study point patterns for a long time, e.g. Brown (1965) and Evans (1967), but the idea was not widely developed until later (see Okabe *et al.*, 2000, Chapter 8, for a review). A typical difficulty in this approach is that the distributions of most characteristics of the Voronoi diagram and the Delaunay tessellation generated by a stationary Poisson process are unknown and can only be approximated by simulation. Using approximated distributions may lead to lower power tests.

Since the distributions of the interior angles and edge lengths of Poisson Delaunay triangles are known, Vincent *et al.* (1977) assessed goodness of fit of an empirical pattern to a Poisson process by using Pearson's χ^2 statistics for the interior angles and the edge lengths of Delaunay triangles, as well as the numbers of edges of Voronoi polygons; the distribution of the latter was approximated by ten realisations with one thousand Poisson points; Boots (1975, 1986) considered the Kolmogorov-Smirnov statistic, the Cramér-von Mises statistic, the Watson statistic and the Anderson-Darling statistic for the interior angles, the minimum angles and the maximum angles of Delaunay triangles. However, the correlation between Delaunay triangles and between Voronoi polygons in a realisation had not been taken into consideration in these studies.

Since numerically tractable expressions for the distributions of some characteristics were recently derived (see Okabe *et al.*, 2000, Sections 5.5.4 and 5.11), this short paper on one hand re-does the goodness-of-fit analyses by Boots (1975, 1986) and Vincent *et al.* (1977) in a more appropriate way and on the other hand compares the powers of tests based on more different characteristics. The Kolmogorov-Smirnov statistic $\sup_x |\sqrt{n}\{F(x) - F_n(x)\}|$ and the Cramér-von Mises statistic $\int_{-\infty}^{\infty} n\{F(x) - F_n(x)\}^2 \mathrm{d}F(x)$, where F and F_n respectively denote the distribution function under the null hypothesis and the empirical distribution function of a given sample of size n, will be used but their asymptotic critical values will not because we have neither independent nor a large number of observations; Monte Carlo tests will be used instead.

Since the statistics based on the characteristics of Voronoi diagrams and Delaunay tessellations are space-domain techniques, we compared the powers of these statistics with the statistic based on the most popular space-domain summary function, namely Ripley's K-function. We restrict ourselves to stationary alternatives. For a comparison between various space-domain tests and spectral tests against regular, clustered and non-stationary alternatives, see Mugglestone and Renshaw (2001).

2. Voronoi diagram and Delaunay tessellation

Given a realisation $\{x_i\}$ of a point process on the plane, we associate all locations on the plane with the closest member(s) of the realisation, i.e. x_i is associated with the polygon $V_i = \{x : \|x - x_i\| \leq \|x - x_j\| \text{ for } j \neq i\}$. The collection $\{V_i\}$ forms the Voronoi diagram of the given realisation and the points of the process are called generators. The polygon V_i is called the Voronoi polygon of x_i. The Delaunay tessellation of $\{x_i\}$ is obtained by joining all pairs of generators whose Voronoi polygons share a common edge. If the point process is simple and stationary, then almost surely the Delaunay tessellation is a collection of triangles, which are called Delaunay triangles.

The following nine characteristics will be examined:

- the length L_V of an edge in a Voronoi diagram,
- the distance R between a generating point and a vertex of the Voronoi polygon of the generating point (the radius of the circle circumscribing a Delaunay triangle)
- the area A_D of a Delaunay triangle,
- the perimeter P_D of a Delaunay triangle,

- the length L_D of an edge in a Delaunay tessellation,
- an interior angle α of a Delaunay triangle,
- the minimum angle $\alpha_{\min}$ of a Delaunay triangle,
- the middle angle α_{mid} of a Delaunay triangle,
- the maximum angle $\alpha_{\max}$ of a Delaunay triangle.

Numerically tractable expressions for their distributions under the CSR hypothesis are known (Okabe *et al.*, 2000, Sections 5.5.4 and 5.11).

Other characteristics and their combinations, such as the roundness factor $4\pi A_V/P_V^2$ (Marcelpoil and Usson, 1992), where A_V and P_V are the area and the perimeter of a Voronoi polygon, may be equally or even more capable of detecting deviations from CSR; however, their distribution functions can only be approximated by simulation. The powers of such combinations in testing CSR will be left for future endeavours. This paper focuses on characteristics with known distributions.

3. Simulation

In reality a point pattern is observed via a bounded observation window. Thus, when we simulate Voronoi polygons, no matter how many points there are, allowance must be made for edge effects introduced by the polygons close to the boundary of the observation window. One practical approach is to exclude from consideration any polygons for which a circle, centred at any vertex of the polygon and passing through the three points in patterns which are equidistant from the vertex, intersects the boundary of the observation window. If the window is a rectangle, another approach is to convert the window into a torus, which is usually referred to as a periodic boundary condition. The latter approach is adopted in this paper.

The alternative models used here generate regular or clustered point patterns with 100 points in a unit square with periodic boundary condition. The simple sequential inhibition process, which puts sequentially 100 non-overlapping discs of radius r uniformly and independently on the torus and then places points at the centres of these discs, was used to generate regular patterns. To get clustered patterns, we simulate a version of the Matérn cluster process by putting a cluster centre uniformly distributed in the torus and then placing a Poisson number, with mean ρ, of points one by one uniformly and independently within the disc of radius r centred at the cluster centre; this procedure is repeated until 100 points are generated.

For each simulated pattern, 999 realisations of a Poisson process with the same number of points were generated. If under the CSR hypothesis the value of the Kolmogorov-Smirnov statistic or the Cramér-von Mises statistic for the distribution of a characteristic from the pattern, when pooled together with the values from the simulated realisations of the Poisson process, is one of the largest fifty, then the CSR hypothesis is rejected. Tables 1 and 2, respectively, give estimated powers against regular alternatives with hard core radius $r = 0.05$, 0.2 and 0.3, and clustered alternatives with cluster radius $r = 0.1$ and 0.2 and mean cluster size $\rho = 5$ and 10. The powers were estimated from 100 realisations of the alternative models.

Table 1 about here

$\boxed{\text{Table 2 about here}}$

We can see from the estimated powers that neither the Kolmogorov–Smirnov statistics nor the Cramér–von Mises statistics are always superior to the other. Both $\alpha_{\max}$ and $\alpha_{\min}$ perform better than α, and $\alpha_{\min}$ has a higher power than $\alpha_{\max}$. This agrees with the conclusion in Boots (1986), even though asymptotic critical values were inappropriately used there. For alternatives that exhibit moderate regularity, $\alpha_{\min}$ yields the most powerful statistic among all these characteristics; A_D the second, and the rest performed poorly. However, for alternatives that exhibit moderate clustering, say cluster radius $r = 0.2$ and mean cluster size $\rho = 5$, $\alpha_{\min}$, although still better than $\alpha_{\max}$, α_{mid} and α, performs worse than L_V, R, A_D, P_D and L_D. Thus, we may conclude that these five characteristics are more able to detect clustering, while $\alpha_{\min}$ is more able to detect regularity.

Consider the probability density functions f and g, respectively, of L_D and $\alpha_{\min}$ in the Delaunay tessellation generated by a stationary Poisson process with intensity λ (Muche, 1996; Mardia *et al.*, 1977):

$$\begin{aligned} f(l) &= \frac{\lambda \pi l}{3} \left\{ \lambda^{1/2} l \exp\left(\frac{-\lambda \pi l^2}{4} \right) + \frac{2}{\sqrt{\pi}} \int_{(\lambda\pi)^{1/2} l/2}^{\infty} \exp(-x^2) \mathrm{d}x \right\}, \quad l \geq 0, \\ g(\alpha) &= \frac{2}{\pi} \left\{ (\pi - 3\alpha) \sin 2\alpha + \cos 2\alpha - \cos 4\alpha \right\}, \qquad 0 \leq \alpha \leq \pi/3. \end{aligned}$$

The reason for the difference in their sensitivities can be seen from Figures 1 and 2. For clustered patterns, the distributions of L_D are somewhat bimodal, including short edges within clusters and long edges between clusters. The distributions of L_D in regular patterns are unimodal. As a result L_D, as well as L_V, R, A_D and P_D, are much more sensitive to clustered alternatives than to regular alternatives. However, the distribution of $\alpha_{\min}$

is unimodal except in cases with strong clustering. Nevertheless, it does not mean that the empirical frequency distributions resemble the function g. Actually, as can be seen from Figure 2, the empirical frequency distributions of $\alpha_{\min}$ from either clustered or regular patterns show deviations from g, meaning that $\alpha_{\min}$ is sensitive to both alternatives.

From the distributions given in Figures 1 and 2, we can see that the shape of the empirical distributions of these characteristics may be used as a guideline for choosing alternative models if the CSR hypothesis is rejected. If the shape of the empirical distribution of a characteristic in a point pattern resembles a mixture of two distributions, one of which has a small mean and small dispersion and the other has a much larger mean and greater dispersion, then likely the point pattern is a realisation of a process that produces clustered patterns. The two peaks of the distributions give information on the cluster radii and inter-cluster distances. On the other hand, if the empirical distribution has a very narrow range or concentrates only on a limited number of values, then likely the point pattern exhibits regularity.

Figure 1 about here

Figure 2 about here

As a comparison, the powers of the statistic $\sup_t |\sqrt{\hat{K}(t)/\pi} - r|$, where $\hat{K}$ is the empirical Ripley's K-function with periodic boundary condition, are also given in Tables 1 and 2. We can see that the K-function was superior to the above characteristics. It is not surprising because it is well-known that the K-function is rather successful in this aspect, and so this superiority reinforces its practical value. The K-function, however, will fail if the given point pattern is a realisation of the cell process (Baddeley and Silverman, 1984). The cell process simultaneously exhibits regularity and clustering but has the same K-function as the stationary Poisson process, meaning that the K-function could not capture the dependence structure of such a point process.

Table 3 gives the estimated power in testing CSR by using these characteristics against the cell process. The simulation results here show that A_D and $\alpha_{\min}$ are rather able to detect the departure of the cell process from CSR. In order to see if the clustering or regularity nature can be revealed, we compare the empirical frequency distribution of L_D and $\alpha_{\min}$ of the cell process given in Figure 3 with those given in Figures 1 and 2. The empirical distribution of $\alpha_{\min}$ of the cell process resembles those of regular patterns but that of L_D does not particularly indicate clustering.

Table 3 about here

Figure 3 about here

4. Examples

The empirical frequency distributions of L_D and $\alpha_{\min}$ under periodic boundary condition of the three standard examples given in Diggle (1983, pp. 1-2) are given in Figure 4. The p-values for the null model being CSR are given in Table 4. The locations of Japanese black pine saplings do not show any significant clustering or regularity. The bimodal nature of L_D and $\alpha_{\min}$ of redwood seedlings shows significant clustering, whereas the empirical distributions of L_D and $\alpha_{\min}$ of the locations of biological cells clearly exhibit regularity.

Figure 4 about here

Table 4 about here

Acknowledgements

Research supported by the Research Grants Council of the Hong Kong Special Administrative Region, China (Project numbers HKBU 2075/98P and HKBU 2048/02P) and an FRG grant of the Hong Kong Baptist University. I thank Maggie Yu for writing simulation programmes, Fred Hickernell and Regina Hong for their advices on numerical integration of the densities and the referees for their helpful comments and suggestions.

References

Baddeley, A. J. and Silverman, B. W., 1984: A cautionary example on the use of second-order methods for analyzing point patterns. *Biometrics* **40**, 1089-1093.

Boots, B. N., 1975: Patterns of urban settlements revisited. *The Professional Geographer* **27**, 426-431.

Boots, B. N., 1986: Using angular properties of Delaunay triangles to evaluate point patterns. *Geographical Analysis* **18**, 250-260.

Brown, G. S., 1965: Point density in stems per acre. *New Zealand Forestry Service Research Notes* **38**, 1-11.

Diggle, P. J., 1979: On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrics* **35**, 87-101.

Diggle, P. J., 1983: *Statistical Analysis of Spatial Point Patterns*. Academic Press, New York.

Evans, I. S., 1967: The properties of patterns of points, measured by space filling and angular relationships. *Geographical Articles (Cambridge)* **8**, 63-77.

Grabarnik, P. and Chiu, S. N., 2002: Goodness-of-fit test for complete spatial randomness against mixtures of regular and clustered spatial point processes. *Biometrika* **89**, 411-421.

Marcelpoil, R. and Usson, Y., 1992: Methods for the study of cellular sociology: Voronoi diagrams and parametrization of the spatial relationships. *Journal of Theoretical Biology* **154**, 359-369.

Mardia, K. V., Edwards, R. and Puri, M. L., 1977: Analysis of central place theory. *Bulletin of the International Statistical Institute* **47**, 93-110.

Møller, J. and Waagepetersen, R. P., 2002: Simulation based inference for spatial point processes. In: M. B. Hansen and J. Møller (eds): *Spatial Statistics and Computational Methods*, Springer, New York, to appear.

Muche, L., 1996: Distributional properties of the three-dimensional Poisson Delaunay cell. *Journal of Statistical Physics* **84**, 147-167.

Mugglestone, M. A. and Renshaw, E., 2001: Spectral tests of randomness for spatial point patterns. *Environmental and Ecological Statistics* **8**, 237-251.

Okabe, A., Boots, B., Sugihara, K. and Chiu, S. N., 2000: *Spatial Tessellations. Concepts and Applications of Voronoi Diagrams*, 2nd edition. Wiley, Chichester.

Ripley, B. D., 1979: Tests of 'randomness' for spatial point patterns. *Journal of the Royal Statistical Society. Series* B **41**, 368-374.

Thönnes E. and van Lieshout, M. C., 1999: A comparative study on the power of van Lieshout and Baddeley's J-function. *Biometrical Journal* **41**, 721-734.

Vincent, P. J., Howarth, J., Griffiths, J. and Collins, B., 1977: Urban settlement patterns and the properties of the simplicial graph. *The Professional Geographer* **29**, 21-25.

Postal address: *S. N. Chiu*

Department of Mathematics

Hong Kong Baptist University

Kowloon Tong

Hong Kong

E-mail: *snchiu@math.hkbu.edu.hk*

Table 1: Estimated powers (in percent) against simple sequential inhibition alternatives with radius r (K-S = Kolmogorov-Smirnov statistic; C-vM = Cramér-von Mises statistic)

	$r = 0.05$		$r = 0.2$		$r = 0.3$	
Characteristic	K-S	C-vM	K-S	C-vM	K-S	C-vM
L_V	7	5	11	8	10	7
R	8	7	18	14	98	84
A_D	4	5	74	65	100	100
P_D	2	6	9	12	100	99
L_D	5	4	3	9	100	99
α	8	7	6	7	76	76
α_{min}	6	5	82	84	100	100
α_{mid}	4	5	11	10	45	40
α_{max}	5	3	45	42	96	97
K-function	19		100		100	

Table 2: Estimated powers (in percent) against Matérn cluster alternatives with cluster radius r and mean cluster size ρ (K-S = Kolmogorov-Smirnov statistic; C-vM = Cramér-von Mises statistic)

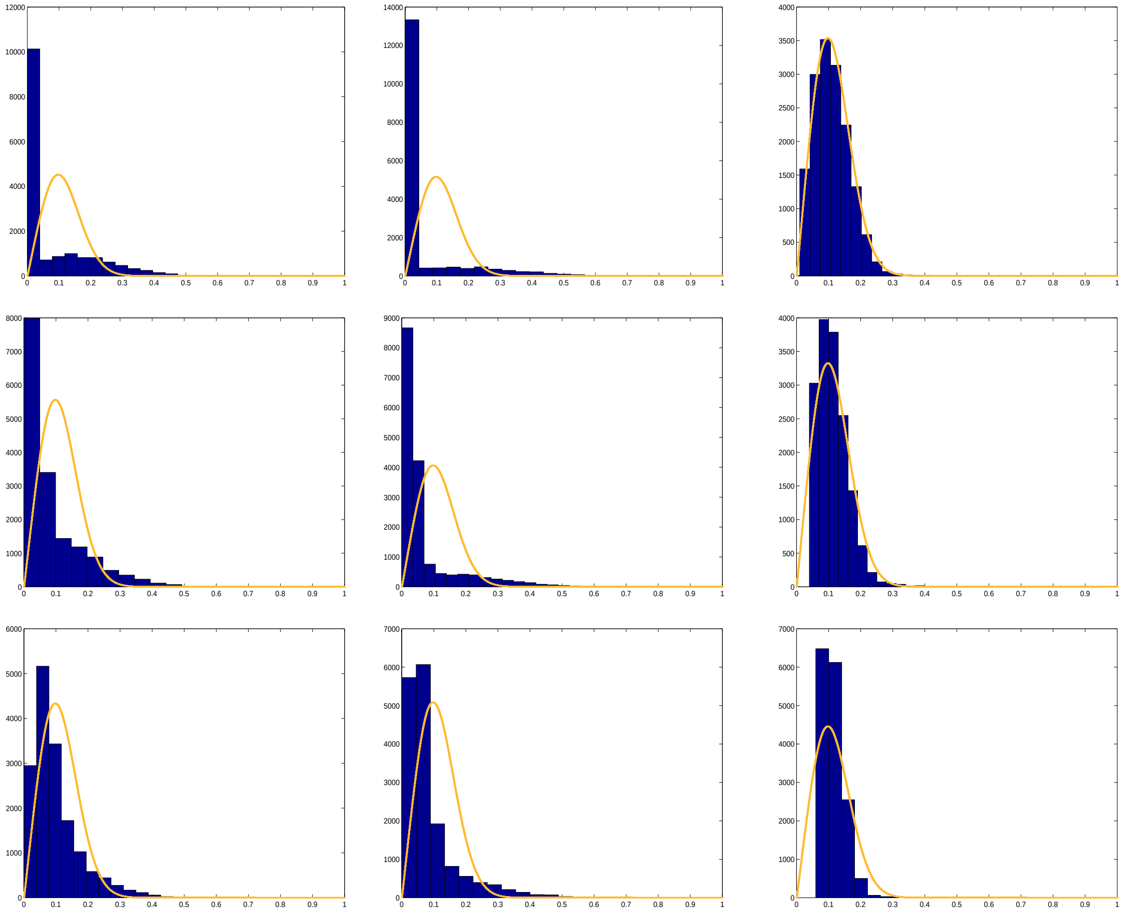
	$r = 0.1,\ \rho = 5$		$r = 0.2,\ \rho = 5$		$r = 0.2,\ \rho = 10$	
Characteristic	K-S	C-vM	K-S	C-vM	K-S	C-vM
L_V	98	92	74	71	98	93
R	100	100	100	94	100	100
A_D	100	100	100	100	100	100
P_D	100	100	100	100	100	100
L_D	100	100	99	99	100	100
α	70	78	22	24	29	31
α_{min}	100	100	73	79	88	90
α_{mid}	67	66	36	30	59	56
α_{max}	100	100	61	68	87	89
K-function	100		100		100	

Table 3: Estimated powers (in percent) against the Baddeley-Silver cell process (K-S = Kolmogorov-Smirnov statistic; C-vM = Cramér-von Mises statistic)

	L_V	R	A_D	P_D	L_D	α	$\alpha_{\min}$	α_{mid}	$\alpha_{\max}$
K-S	16	65	63	45	39	8	68	9	67
C-vM	15	36	58	48	40	8	73	8	62

Table 4: Estimated p-values (K-S = Kolmogorov-Smirnov statistic; C-vM = Cramér-von Mises statistic) for the CSR of the locations of 65 Japanese black pine saplings, 62 redwood seedlings and 42 biological cell centres (Diggle, 1983, pp. 1-2)

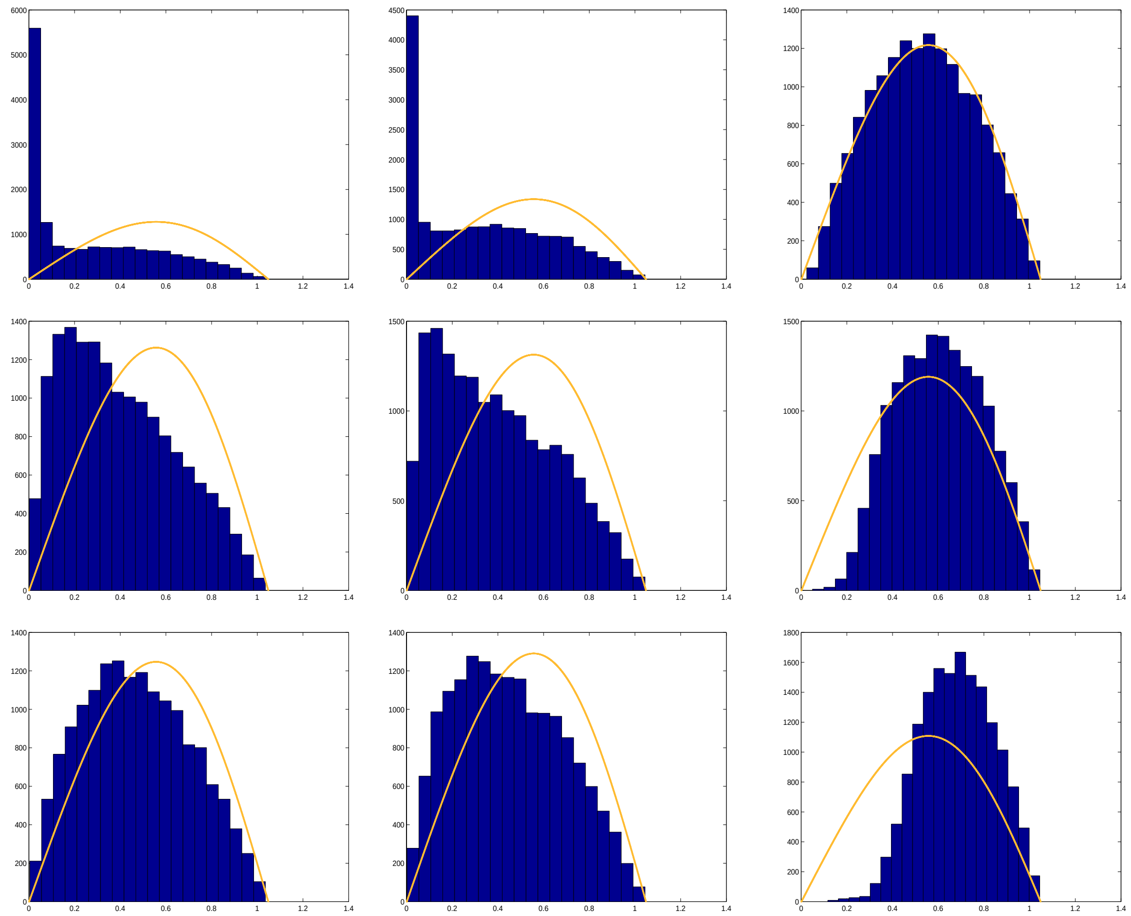
	L_V	R	A_D	P_D	L_D	α	$\alpha_{\min}$	α_{mid}	$\alpha_{\max}$
Japanese black pine saplings									
K-S	0.078	0.235	0.967	0.958	0.755	0.176	0.622	0.798	0.803
C-vM	0.094	0.154	0.966	0.990	0.818	0.229	0.706	0.612	0.798
redwood seedlings									
K-S	0.207	0.023	0.000	0.000	0.000	0.073	0.000	0.000	0.000
C-vM	0.136	0.024	0.000	0.000	0.000	0.095	0.000	0.009	0.000
biological cell centres									
K-S	0.062	0.001	0.000	0.000	0.001	0.022	0.000	0.125	0.000
C-vM	0.048	0.006	0.000	0.000	0.000	0.008	0.000	0.045	0.000



$\rho = 5$ $\rho = 10$

Matérn cluster Simple sequential inhibition

Figure 1: Empirical frequency distribution of L_D in Matérn cluster processes with cluster radii (from top to bottom) 0.01, 0.1 and 0.2 and mean cluster size ρ and simple sequential inhibition processes with radii (from top to bottom) 0.05, 0.2 and 0.3. The solid curves represent the corresponding frequency distribution for a stationary Poisson process.



$\rho = 5$ $\rho = 10$

Matérn cluster Simple sequential inhibition

Figure 2: Empirical frequency distribution of $\alpha_{\min}$ in Matérn cluster processes with cluster radii (from top to bottom) 0.01, 0.1 and 0.2 and mean cluster size ρ and simple sequential inhibition processes with radii (from top to bottom) 0.05, 0.2 and 0.3. The solid curves represent the corresponding frequency distribution for a stationary Poisson process.

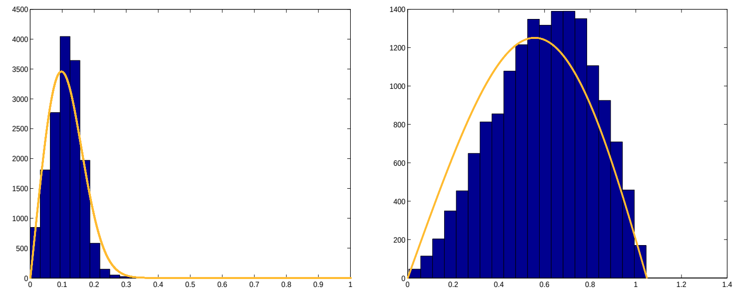
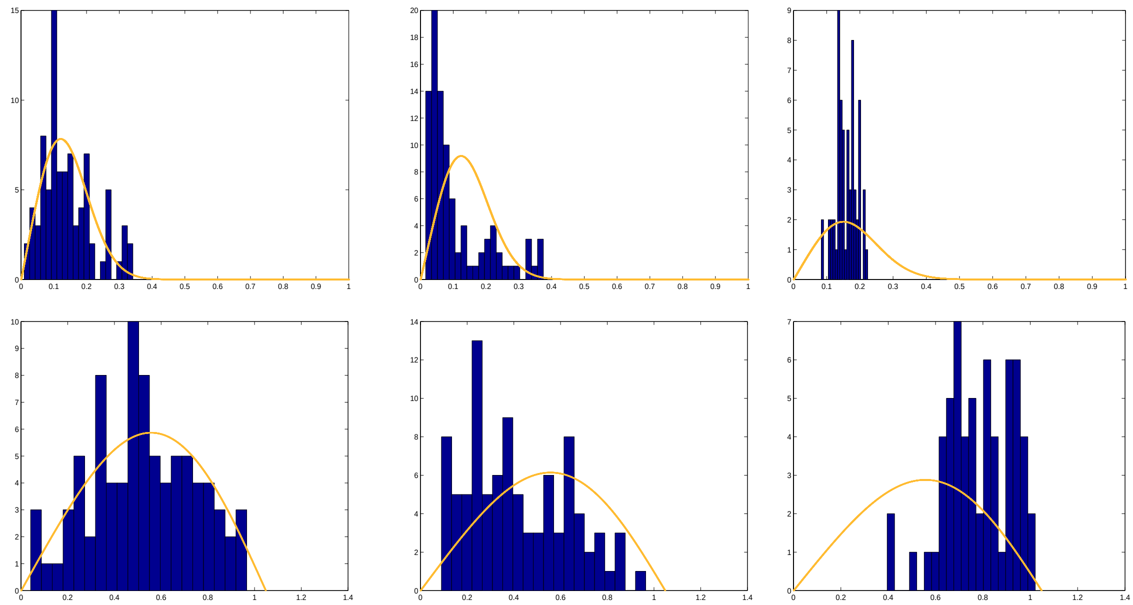


Figure 3: Empirical frequency distribution of L_D and $\alpha_{\min}$ of 100 realisations of the cell process.



Japanese black pine saplings Redwood seedlings Biological cells

Figure 4: Empirical frequency distribution of L_D (upper row) and $\alpha_{\min}$ (lower row) of the locations of 65 Japanese black pine saplings, 62 redwood seedlings and 42 biological cell centres (Diggle, 1983, pp. 1-2). The solid curves represent the corresponding frequency distribution for a stationary Poisson process.