The Discrepancy Function: A New Way of Disentangling Mark/Point Interaction in Marked Point Processes

Eric Renshaw

University of Strathclyde, UK

Smogen Workshop, August 2006

email: eric@stams.strath.ac.uk

web: http://www.stams.strath.ac.uk/∼eric

thanks: Fuensanta Saura & Jorge Mateu
University Jaume I, Castellon, Spain

Overview

In many spatial situations:

- point locations of mark variables (e.g. tree heights) plays key role in the space-time generating mechanism of e.g. forest growth;
- marks and points can be highly inter-dependent;
- current analyses investigate conditional mark structure based on a given point structure.

We shall therefore try to disentangle marks and points through:

- a *Discrepancy Function* which isolates the mark structure;
- it involves a harmonic decomposition of the mark frequencies.

The concept is introduced via:

- cosine wave & thinned point process examples;
- Spanish daily ozone data with missing values;
- a spatial growth-interaction process;
- a classic longleaf pine data set from Georgia, USA.

Summary of the Spectral Approach

Lattice marks: $\{X_{st}\}$ $(s = 1, \ldots, m; t = 1, \ldots, n)$.

Fourier coeffs. $A_{pq}(B_{pq}) = \sum_{s=1}^{m} \sum_{t=1,n} (X_{st} - \overline{X}) \cos\,(\sin)[2\pi(ps/m + qt/n)]$

- Lattice periodogram: $I^L_{pq} = (A^2_{pq} + B^2_{pq})/mn$

  over $p = 1, \ldots, m/2; q = -n/2, \ldots, n/2 - 1$.

Marked Point Process: paired-set $\{(x_i, y_i); X_i\}$ of locations (x_i, y_i) & marks X_i.

Fourier coeffs. $a_{pq}(b_{pq}) = \sum_{i=1}^{N} (X_i - \overline{X}) \cos\,(\sin)[2\pi(px_i + qy_i)]$

- Mark periodogram: $I^M_{pq} = (a^2_{pq} + b^2_{pq})/N$

  over some $p = 1, \ldots, f_x; q = -f_y, \ldots, f_y - 1$.

- Point periodogram replace $X_i - \overline{X}$ by 1.

Square Point Wave Example

Consider the one-dimensional mark function

$$m_i = 2 + \cos(6\pi i) \quad \text{over} \quad i = (0, 1, \ldots, 29)/30 \; .$$

As the periodogram of a regular lattice is zero, the mark periodogram is a spike at frequency $p = 3$ and zero elsewhere, i.e. at $p = 1, 2, 4, 5, \ldots, 15$.

So what happens if we inject point structure by sampling m_i at the 15 non-zero points

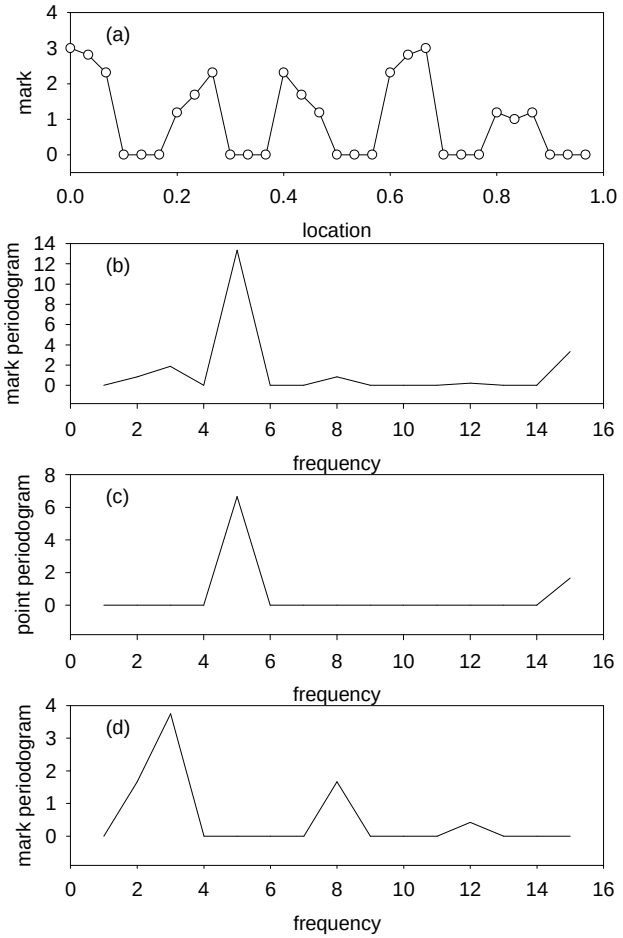
$$\Omega = \{(0, 1, 2, 6, 7, 8, \ldots, 24, 25, 26)/30\} \; ?$$

(a) marks $m_i = 2 + \cos(6\pi i)$ superimposed on a regular square wave of frequency 5 over the points $i = (0, 1, \ldots, 29)/30$; i.e. $m_i = 0$ at missing points.

(b) mark periodogram I_p^L based on all 30 points;

(c) point I_p^P and (d) mark periodograms I_p^M based on the 15 non-empty points.

In (d) note the resonant frequencies at $p = 5 - 3 = 2$, $5 + 3 = 8$ and $3 \times 5 - 3 = 12$.



The Discrepancy Function

- So the fundamental question is whether we can derive a function which better isolates the spatial structure of the marks alone?

The 1-D Fourier mark coeffs. are

$$a_p = \sum_{i=1}^{N} (X_i - \overline{X}) \cos(2\pi p x_i) \quad \text{and} \quad b_p = \sum_{i=1}^{N} (X_i - \overline{X}) \sin(2\pi p x_i) \ ,$$

and the 1-D point Fourier coeffs. are

$$c_p = \sum_{i=1}^{N} \cos(2\pi p x_i) \quad \text{and} \quad d_p = \sum_{i=1}^{N} \sin(2\pi p x_i) \ .$$

Denote I_p^M by I_p and I_p^P by J_p, and consider the simple cosine example

$$m_i = a + \cos(2\pi r i) \ .$$

Then we can write

$$\begin{aligned} NI_p = (1/4)(&NJ_{p+r} + NJ_{r-p} + 2c_{p+r}c_{r-p} - 2d_{p+r}d_{r-p}) \\ &+ (a-\bar{X})^2 NJ_p + (a-\bar{X})c_p(c_{p+r} + c_{r-p}) \\ &+ (a-\bar{X})d_p(d_{p+r} - d_{r-p}) \end{aligned}$$

for $p \in \{1, 2, \ldots, r-1\}$, with similar expressions when $p = r$ and $p \in \{r+1, \ldots, N/2\}$.

- These 3 relationships not only highlight the interplay between the mark and the point spectrum, but they are easily extended to cover the general cosine mark sum

$$m_i = \text{constant} + \sum_{j=1}^{m} k_j \cos(2\pi i p_j + \phi_j) \; . \tag{0.1}$$

Here k_j and ϕ_j denote the amplitude and phase of frequency p_j.

- The underlying principle is that we place each r by a fixed integer s, and then choose that value of s which minimises the distance of the resulting function $S(s, p)$ from I_p.

- Write $S(s,p) \equiv I_p$ at $r = s$.
- Evaluate

$$D(s) \equiv \|\{S(s,p) - I_p\}\|$$

over all possible integers s.

- Then select that value of s which minimises $D(s)$ – we call $D(s)$ the *discrepancy function*.
- Obvious distance measures include

$$D_1(s) = \sum_p |S(s,p) - I_p| \ , \quad D_2(s) = \sum_p [S(s,p) - I_p]^2 \ , \quad D_3(s) = \sqrt{D_2(s)} \ .$$

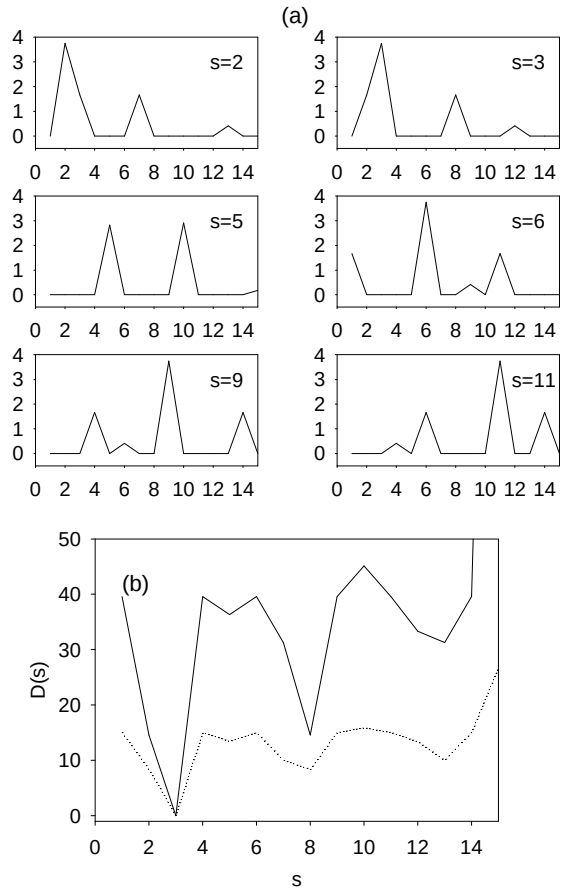
Note that $S(s,p)$ is automatically defined through NI_p: for $p \in \{1, 2, \ldots, s-1\}$

$$\begin{aligned} NS(s,p) &= (1/4)(NJ_{p+s} + NJ_{s-p} + 2c_{p+s}c_{s-p} - 2d_{p+s}d_{s-p}) \\ &+ (a - \bar{X})^2 NJ_p + (a - \bar{X})c_p(c_{p+s} + c_{s-p}) + (a - \bar{X})d_p(d_{p+s} - d_{s-p}) \ ; \end{aligned}$$

with similar expressions for $p \in \{s+1, s+2, \ldots, N/2\}$ and $p = s$.

Example: $m_i = 2 + \cos(6\pi i)$, i.e. frequency 3.

- Clearly, $S(3, p) = I_p$, so a perfect match when $s = 3$.
- Plots for $s \neq 3$ differ substantially from that at $s = 3$; the discrepancy function exploits this difference in shape.
- Both $D_1(s)$ and $D_2(s)$ show a clear minimum at $s = 3$ – the true mark frequency uncorrupted by the point structure.
- As $D_2(s)$ shows the change with p better than $D_1(s)$, we retain it for our study.



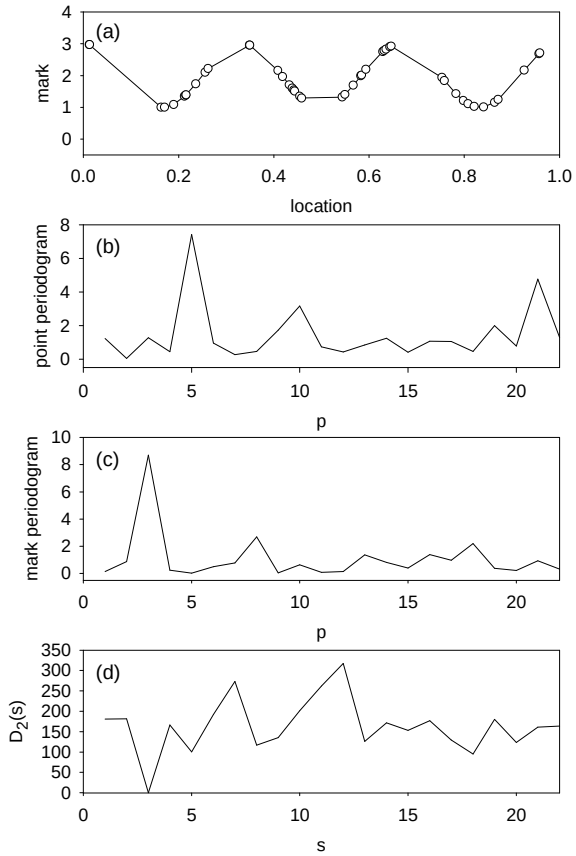
(a) Values of $S(s,p)$ over $p = 1, 2, \ldots, 15$ for $s = 2, 3, 5, 6, 9, 11$.
(b) Values of the discrepancy functions $D_1(s)$ $(\cdots)$ and $D_2(s)$ (—) for $s = 1, \ldots, 15$.

Example: $m_i = 2 + \cos(6\pi i)$ with thinning

- Will such clarity be present when randomness is introduced?
- Replace regular point structure with a thinned point wave with frequency 5.
- Each of 60 random locations were retained only if

$$U_i \leq 1 + \cos(10\pi i) \ .$$

- Point periodogram highlights this frequency 5.
- Mark periodogram highlights frequency $p = 3$, but also the point/mark confounding frequency at $3 + 5 = 8$.
- $D_2(s)$ shows a perfect match at $s = 3$, so rules out $p = 8$



(a) marks $m_i = 2 + \cos(6\pi i)$ on a thinned 44-point cosine wave of frequency 5;
(b) point periodogram J_p; (c) mark periodogram I_p;
(d) values of the discrepancy function $D_2(s)$ for $s = 1, \ldots, 22$.

Two-frequency structures

- In reality would not know either amplitude k or phase ϕ of trial mark structure

$$m_i = a + k\cos(2\pi s x_i + \phi)$$

- So need to minimise over k and ϕ as well.
- Also, no reason why underlying mark frequency $p = 3$ should be integer.
- So full analysis means minimising $D(k, s, \phi)$ over all

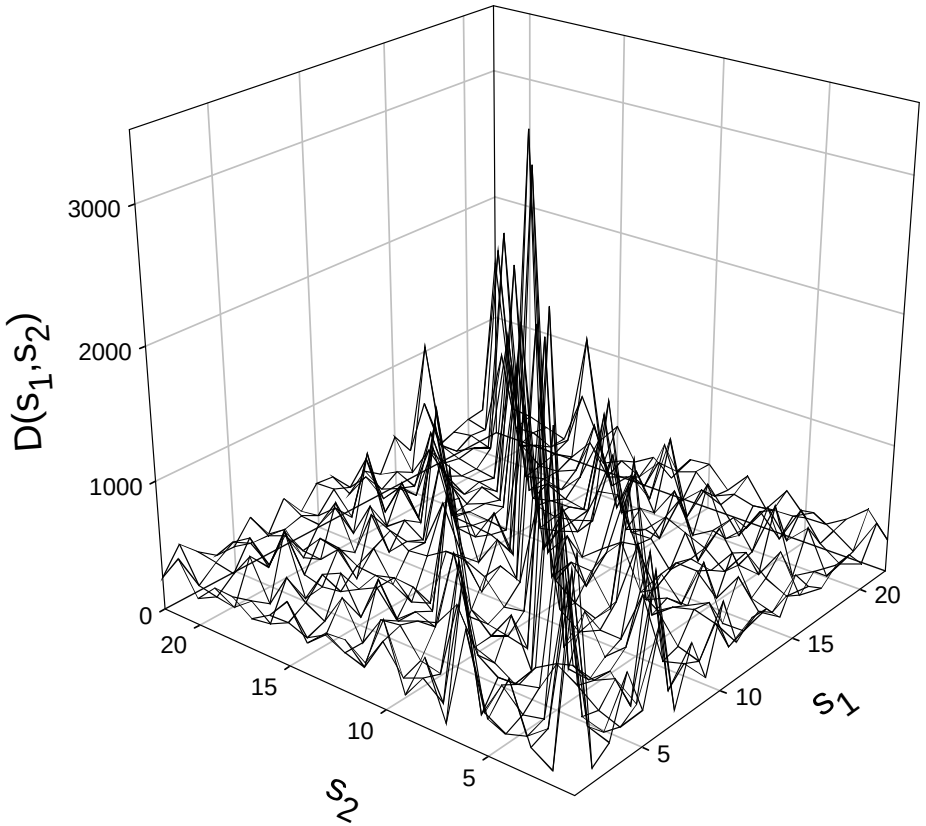
$$0 < |k| \leq k_{max} \quad , \quad 0 \leq \phi < \pi \quad \text{and} \quad 0 < s \leq s_{max} \ .$$

  for some k_{max} and s_{max}.

- m-frequency case is far worse:
  need to minimise over $3m$ parameters;
  discrepancy surface is highly nonlinear.
- Simple double cosine

$$m_i = 2 + \cos(6\pi x_i) + \cos(14\pi x_i)$$

  exhibits high mountain ridges interpersed with local peaks.



Discrepancy function $D(s_1, s_2)$ plotted over $s_1, s_2 = 1, \ldots, 22$ for the candidate mark structure $m_i = 2 + \cos(2\pi s_1 x_i) + \cos(2\pi s_2 x_i)$ and actual mark structure $m_i = 2 + \cos(6\pi x_i) + \cos(14\pi x_i)$ superimposed on a thinned random set of 44 points with frequency 5 in $[0, 1]$.

Optimisation Algorithms

- This highly nonlinear structure means that a general simultaneous procedure in which $D(\cdot)$ is evaluated first over a coarse grid of points, say

$$k_i, k_2 = 0, 0.1, \ldots, 2.0 \quad \text{and} \quad \phi_1, \phi_2 = 0^o, 10^o, \ldots, 170^0 \ ,$$

and then over selectively finer grids based around the previous minimum is unlikely to work:

- for it may well become trapped in a local, high-dimensional, minimum.
- We shall therefore adopt an alternative sequential procedure.
- Use a grid-based search to find amplitude-phase-frequency triple $(\hat{k}_1, \hat{\phi}_1, \hat{s}_1)$ that minimises $D(k_1, \phi_1, s_1)$ based on

$$m_i = \hat{a} + k_1 \cos(2\pi s_1 x_i + \phi_1) \ .$$

- Then determine the second-order triple $(\hat{k}_2, \hat{\phi}_2, \hat{s}_2)$ which minimises $D(k_2, \phi_2, s_2 | \hat{k}_1, \hat{\phi}_1, \hat{s}_1)$ based on

$$m_i = \hat{a} + \hat{k}_1 \cos(2\pi \hat{s}_1 x_i + \hat{\phi}_1) + k_2 \cos(2\pi s_2 x_i + \phi_2) \ .$$

- Repeat, obtaining sequential estimates $\Omega_j = \{\hat{a}, \hat{k}_i, \hat{\phi}_i, \hat{s}_i; i = 1, 2, \ldots, j\}$, until the rate of decrease in $D(\cdot|\Omega_j)$ reaches some preassigned arbitrary level, or 'flattens off'.

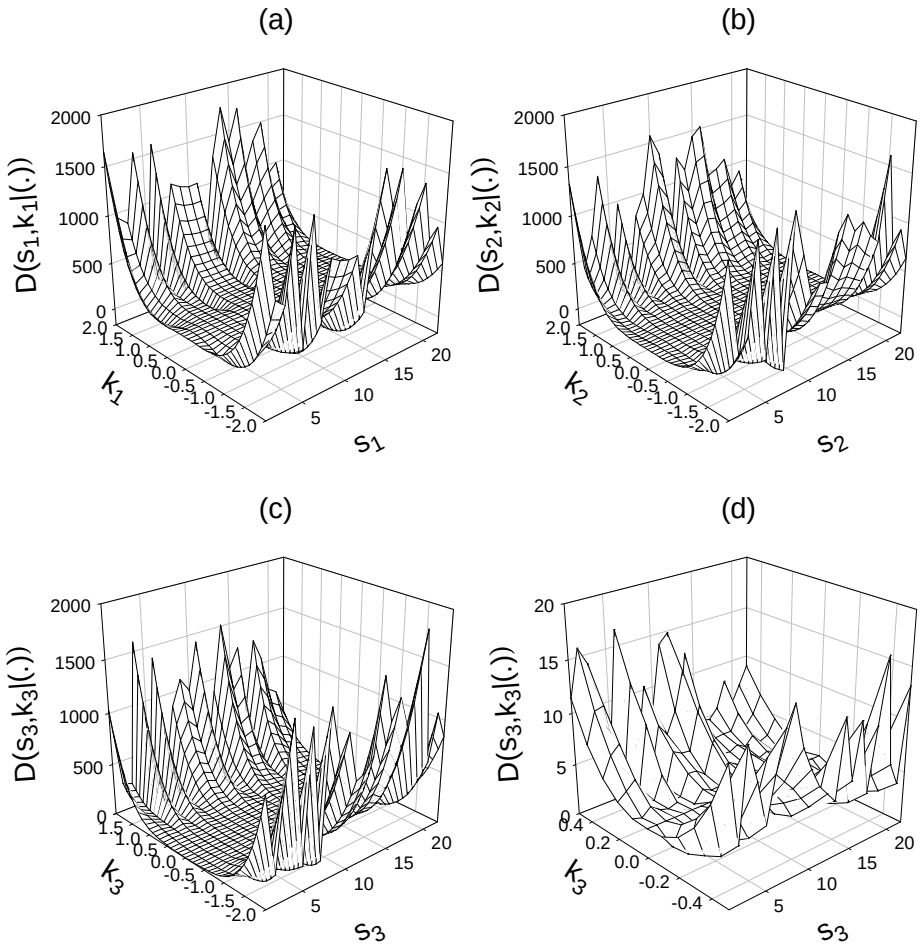
Example: $m_i = 2 + 0.8\cos(6\pi x_i) + \cos(14\pi x_i) + U(-0.2, 0.2)$ on a thinned 44-point cosine wave of frequency 5

- Note frequency 7 dominates.
- This mark/point structure provides a sharp test – we have to differentiate between three frequencies based on a relatively small data set of only 44 points.
- Coarse grid gives:

s_1	$\hat{k}_1$	$\hat{\phi}_1$	$D(s_1 \mid \hat{k}_1, \hat{\phi}_1)$
2	0.9	10^o	66.70
3	0.6	80^o	136.68
7	0.9	10^o	29.15
12	1.1	80^o	83.76

So best choice is frequency 7.

- Then a medium + fine grid search gives $D = 28.4848$ at $\hat{s}_1 = s_1 = 7$ for $\hat{k}_1 = 0.878$ and $\hat{\phi}_1 = 3^o$.
- Figure (a) shows the first-order discrepancy function.



Discrepancy function plotted over $s_1 = 1, \ldots, 22$ and $k_1 = -2.0, -1.9, \ldots, 2.0$.

- Figure (b) shows the second-order function.
- Figures (c) & (d) show the third-order function is effectively flat, i.e. there is no further structure.
- So the 'best' estimated sequential function is

$$\hat{m}^{seq}(t) = 1.901 + 0.878\cos(14\pi t + 3^o) - 0.771\cos(6\pi t + 152^o) \ ,$$

which agrees quite closely with the generating process.

- A follow-on simultaneous search over a finer and finer grid produced

$$\hat{m}^{sim}(t) = 1.901 + 1.000\cos(14\pi t + 182^{\mathrm{O}}) - 0.870\cos(6\pi t - 4^{\mathrm{O}}) \ ,$$

with $D = 0.116$ now considerably lower.

- So $\hat{m}^{sim}(t)$ gives a much closer fit to the true function $m(t)$ than $\hat{m}^{seq}(t)$.
- Clearly, this sequential→simultaneous approach provides an efficient procedure for determining the minimising function.

Non-integer Frequencies

- Now grid-based search procedures are grossly inefficient.
- So use established nonlinear iterative algorithms:

1. Nelder-Mead (N-M) is based on evaluating a function at the vertices of a simplex, which is then iteratively shrunk until some desired bound is obtained – can be slow to converge, but it is robust.

2. Sann is a variant of simulated annealing. Uses the Metropolis function for the acceptance probability: candidate points are generated from a Gaussian Markov kernel with scale proportional to the actual 'temperature'.

3. BFGS is a Quasi-Newton method based on fast descents. The next candidate point is $x_{k+1} = x_k + \alpha_k d^k$, where α_k denotes jump length and d^k is the descent direction given by $-D^k \nabla f(x^k)$ for D^k a symmetric and positive definite matrix. Sophisticated ways reconstruct the curvature in D^k so that D^k approximates the $(\nabla^2 f(\cdot))^{-1}$ in Newton's method; BFGS currently appears to be the best one.

4. CG uses conjugated gradients – well-behaved but inferior to the BFGS method. Accelerates the speed of convergence of the fastest descent method whilst trying to avoid the problems associated with Newton's method (very good when minimising a quadratic function).

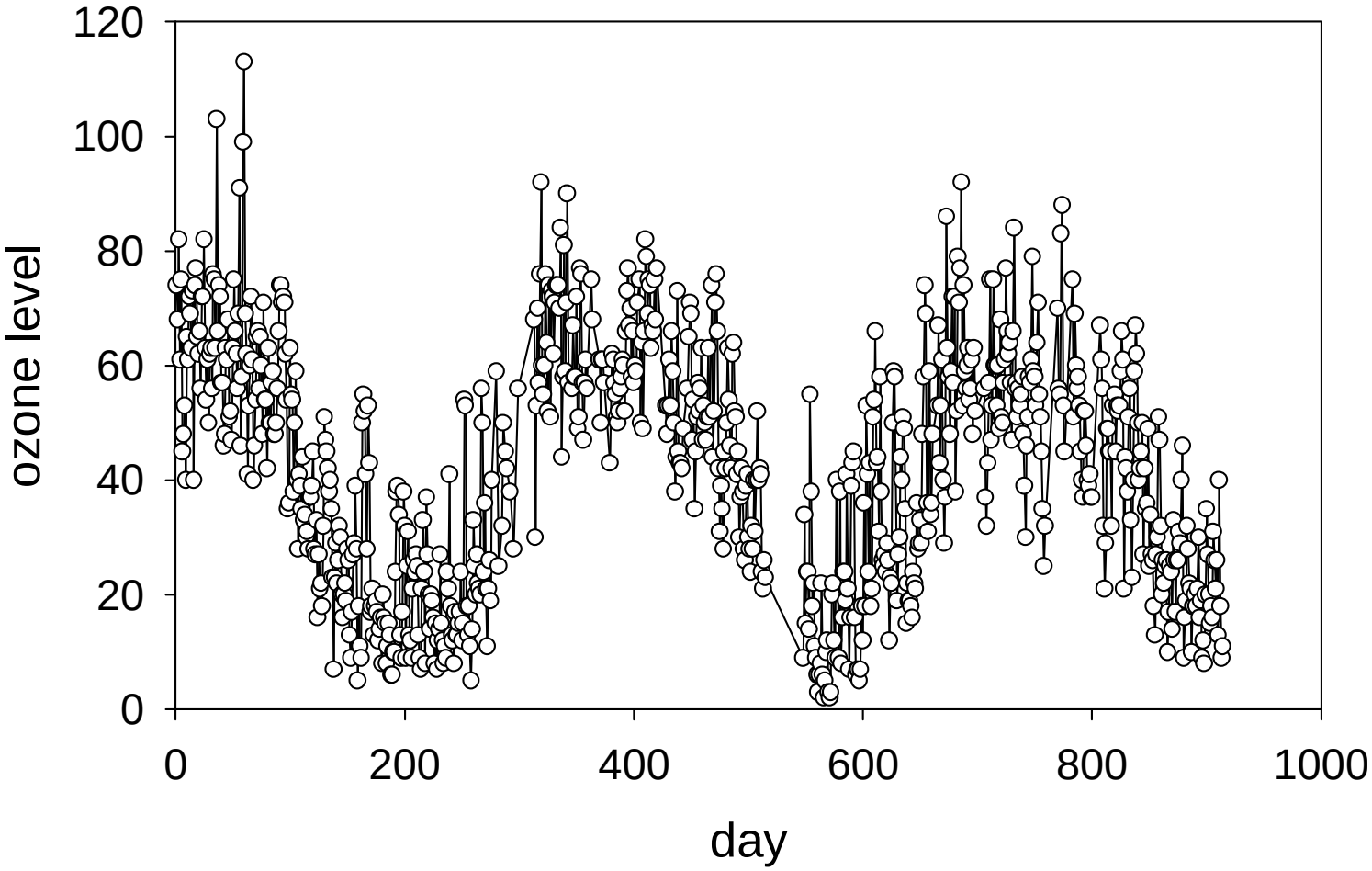
Result of using these four approaches with initial values given by the sequential estimates $\hat{m}^{seq}(t)$ above and convergence tolerance 10^{-6}.

	N-M	BFGS	CG	Sann
s_1	6.989	6.995	6.995	7.074
k_1	0.966	0.990	0.990	0.948
ϕ_1	5.01^o	3.25^o	3.25^o	-16.25^o
s_2	3.060	3.021	3.021	3.125
k_2	-0.835	-0.852	-0.852	-0.787
ϕ_2	166.70^o	171.71^o	171.71^o	140.33^o
s_3	-0.1	0.000	0.000	1.858
k_3	0.543	0.000	0.000	-0.072
ϕ_3	-28.54^o	0.000^o	0.000^o	84.13^o
discrepancy	0.066074	0.076921	0.076921	1.316153
time	2.95	1	5.914	27.939

- So discard Sann.
- N-M incorrectly suggests a possible frequency $s_{13} = 0.1$.
- CG converges 6 times slower than BFGS, so choose BFGS.

Analysis of Spanish Ozone Data

- Daily ozone levels for Castellon de la Plana station from 31/3/01 to 30/11/03.
- Strong seasonal component + missing values (87 & 34 for 2003 & 2004).

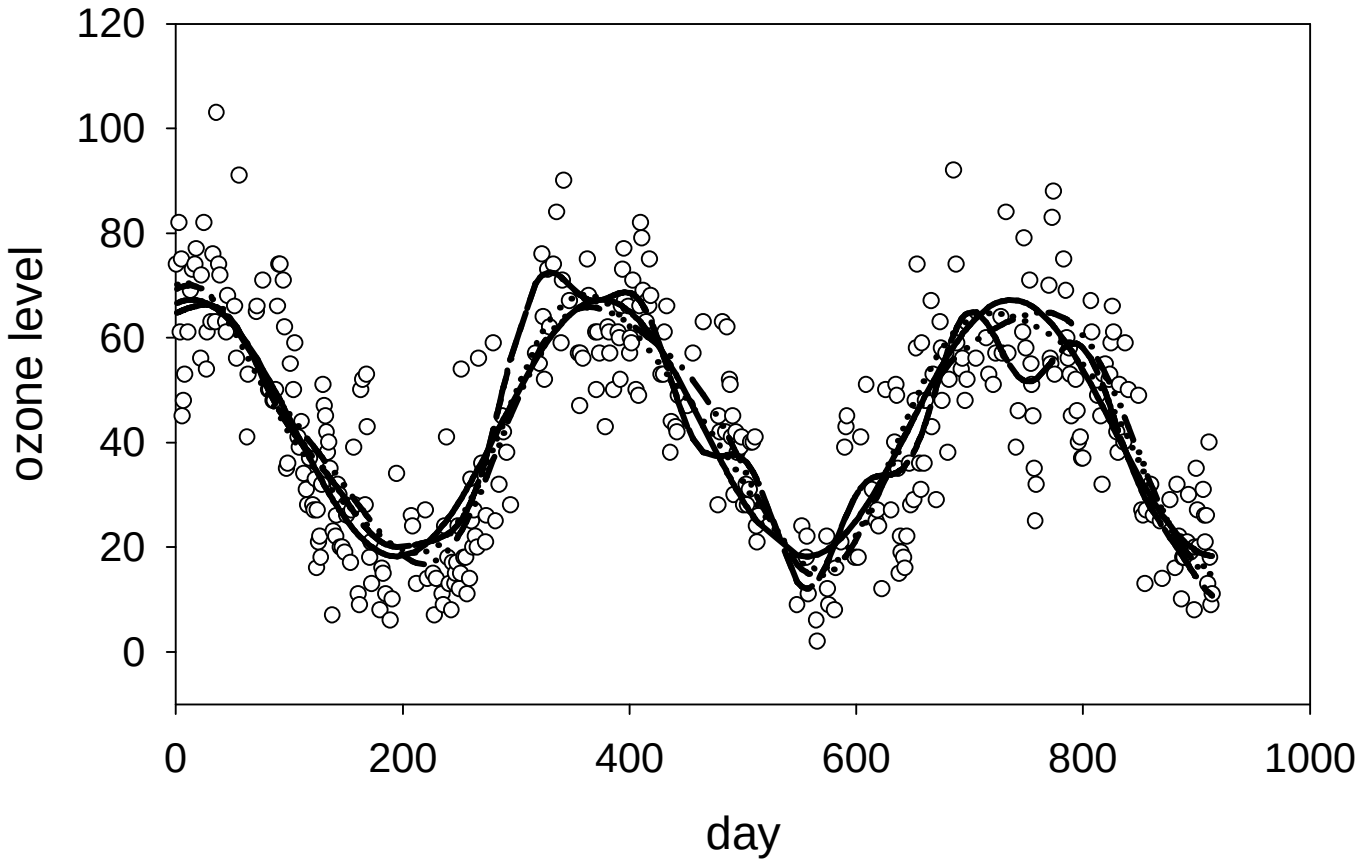


- Note point and mark structures are not related: so provides a 'mid-way' illustration.
- Strengthen missing point structure: $\Pr(\text{point accepted}) = [2 + \cos(14\pi i/N)]/4$
- Single wave spans 914 days, i.e. a frequency of 2.51.

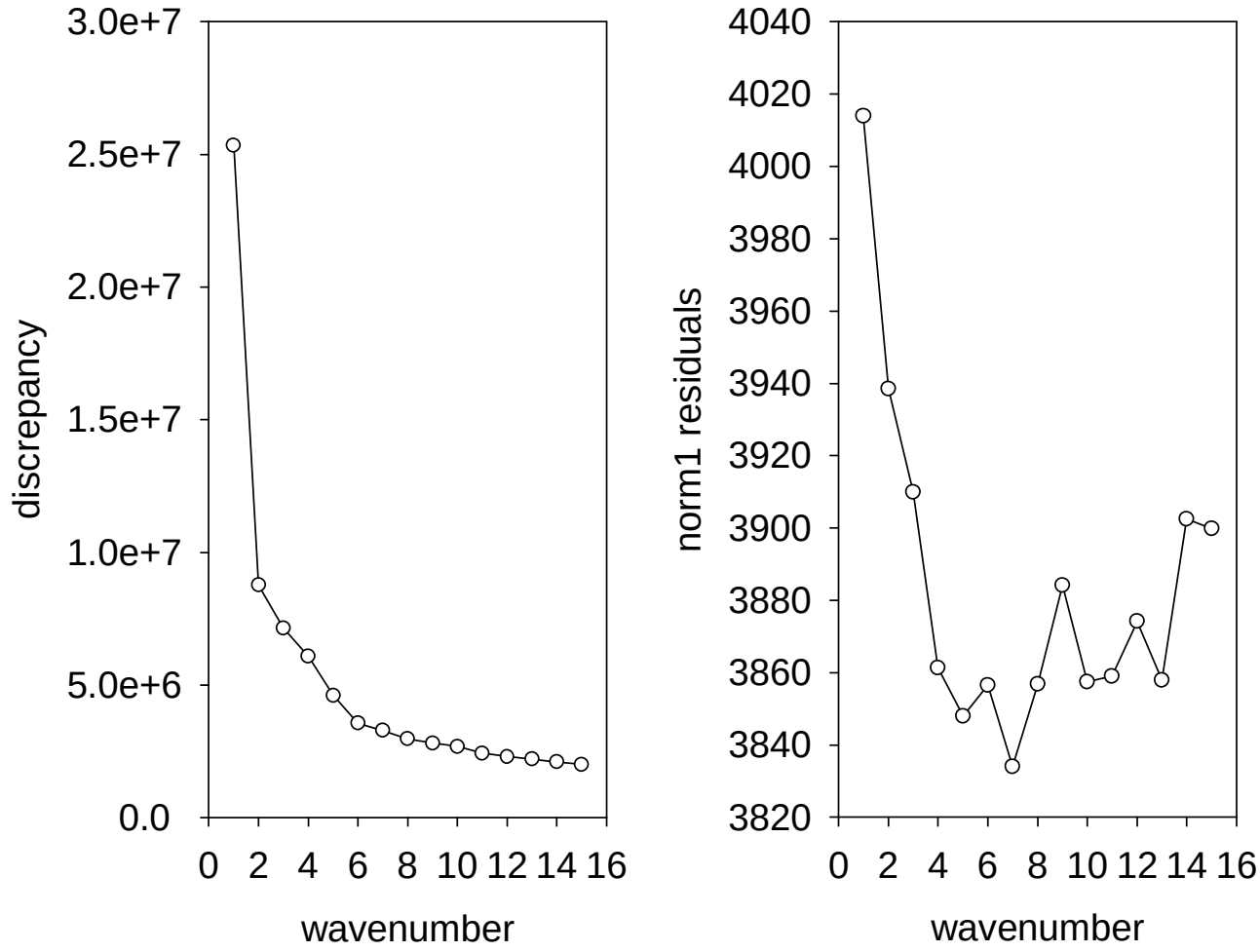
wavenumber	sequential		Sann		BFGS	
1	-24.50	2.53	-24.50	2.53	-25.86	2.50
2	3.51	5.49	3.51	5.49	5.61	5.40
3	-3.00	9.55	-3.00	9.55	-2.57	9.23
4	-1.80	7.00	-1.80	7.00	0.27	6.48
5	-2.10	3.79	-2.10	3.79	-6.15	3.47
6	-1.50	1.06	-1.50	1.06	-5.75	1.04
7	1.19	8.44	1.19	8.44	0.90	8.21

Table : First 7 of 15 estimated amplitudes and frequencies showing sequential estimates together with the Sann and BFGS simultaneous estimates.

- Only primary frequency 2.53 conveys genuine mark information.
- So point structure has been eliminated.



Thinned ozone data showing: sequential fit with 1 wave (——), 2 waves (· · · · ·) and 15 waves (– – –); Sann fit with 2 waves (–··–··–) and 15 waves (as per sequential).



Thinned ozone data showing discrepancy and L_1norm-residual values for the sequential fit with increasing wavenumber.

- note L_1norm $\downarrow$ for first 2 frequencies only.

Application to Space-Time Growth-Interaction Analyses

- Highly interdependent mark/point structures are generated in Renshaw & Särkkä (2001) Särkkä & Renshaw (2006).
- Stochastically: points $\{i\}$ arrive at rate α with $U(0,1)$ marks (say); die at rate μ.
- Deterministically: In $(t, t+dt)$ each mark $m_i(t)$ changes through

$$m_i(t+dt) = m_i(t) + f(m_i(t))dt + \sum_{j \neq i} h(m_i(t), m_j(t); \|x_i - x_j\|)dt \ .$$

- Särkkä & Renshaw (2006) consider the linear growth function

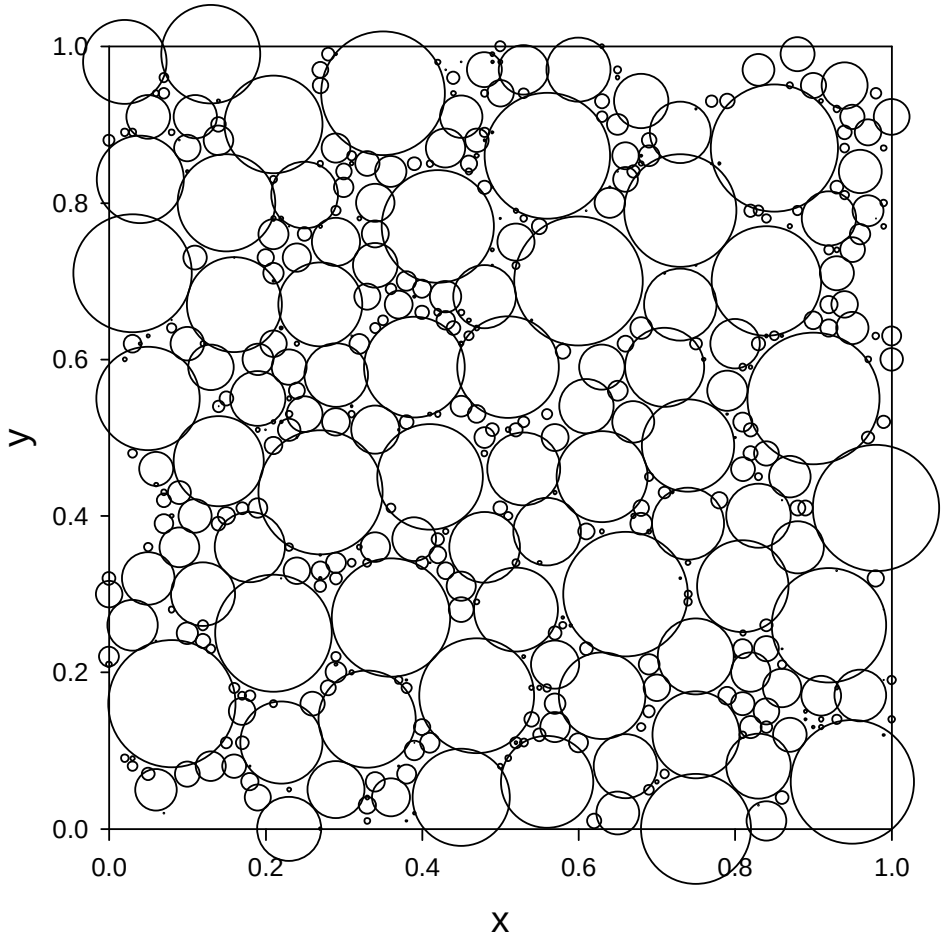
$$f_2(m_i(t)) = \lambda(1 - m_i(t)/K) \ ,$$

in conjunction with the area-interaction function

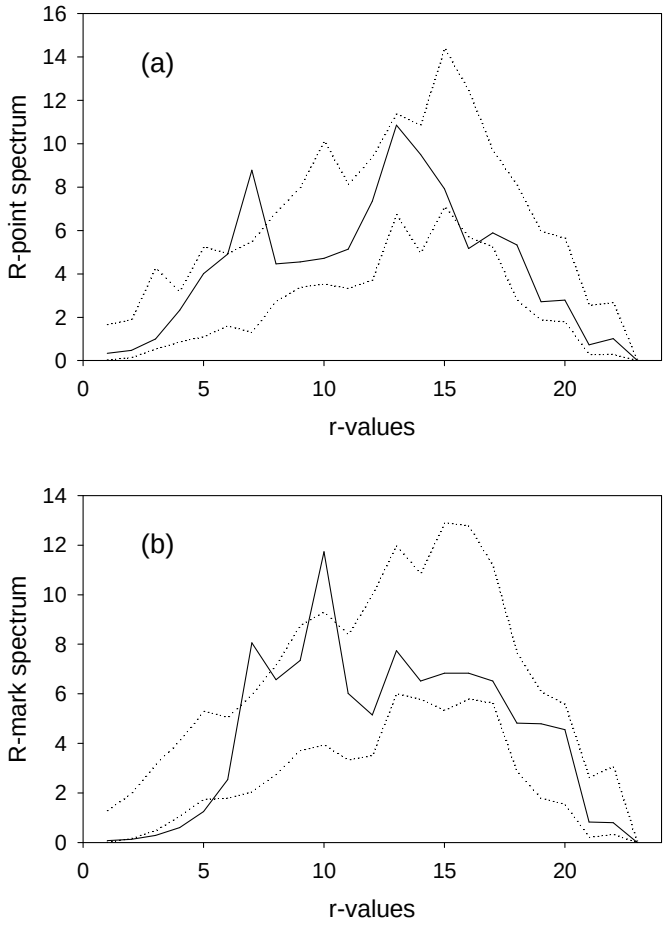
$$h_2(m_i(t), m_j(t)) = -b \ \text{area}\{D(x_i, rm_i(t)) \cap D(x_j, rm_j(t))\}/\pi r^2 m_i^2(t)$$

where $D(x, s)$ denotes the disk with centre x and radius s.

- Note 'soft-core' and asymmetric since a small mark is affected more than a large one.



Realisation of the linear growth and area-interaction process shown at $t = 1000$ for $\alpha = 5$, $\mu = 0$, $\lambda = 1$, $K = 20$, $b = 10$ and $r = 0.005$; bubble radius is $rm_i(t)$



(a) Point and (b) mark R-periodograms with upper and lower envelopes based on 99 simulations of 490 points under HPP, and 99 randomisations of marks on given 490 point locations.

Point spectrum:

- Significant peak at $r = 7$ associated with regular spacing of points having large marks.
- Secondary peak at $r = 13$ probably indicative of the small-mark distance of $\simeq 1/13$ associated with small marks clustering in the gaps between larger marks.

Mark spectrum:

- Similar main peak at $r = 7$.
- plus a secondary peak at $r = 7$ indicative of an overall mark-mark interaction distance of order 0.1.

Bivariate K-function: based on 448 'small' and 42 'large' marks.

- strong lack on inter-point distance < 0.09.
- Significant point-point distance of $\simeq 0.13$, in accord with frequency 7.

The Discrepancy Function enables us to decide whether the $r = 7$ mark effect is an artifact of the point pattern.

Sequential reduction in the discrepancy function based on estimated mark values

$$m_i(1000) = 3.142 + \sum_{j=1}^{u} \hat{k}_j \cos[2\pi(x_{i1}\hat{s}_{j1} + x_{i2}\hat{s}_{j2}) + \hat{\phi}_j] + \epsilon_{im}$$

for $u = 1, \ldots, 15; i = 1, \ldots, 490$ at locations $x_j = (x_{j1}, x_{j2})$ as the number of fitted frequencies $(\hat{s}_{j1}, \hat{s}_{j2})$ is increased.

- Primary wave causes a large reduction in discrepancy.
- First 4 wavenumbers all relate to high-frequency, i.e. highly local, mark pattern with frequencies of $20 - 25$, i.e. spatial structure at distances $\simeq 0.04$ parallel to axes ($\simeq 0.07$ to diagonals).
- Remaining 10- wavenumbers relate to low-frequency mark structure in range $1 - 3$, i.e. to spatial pattern caused by large well-established marks.

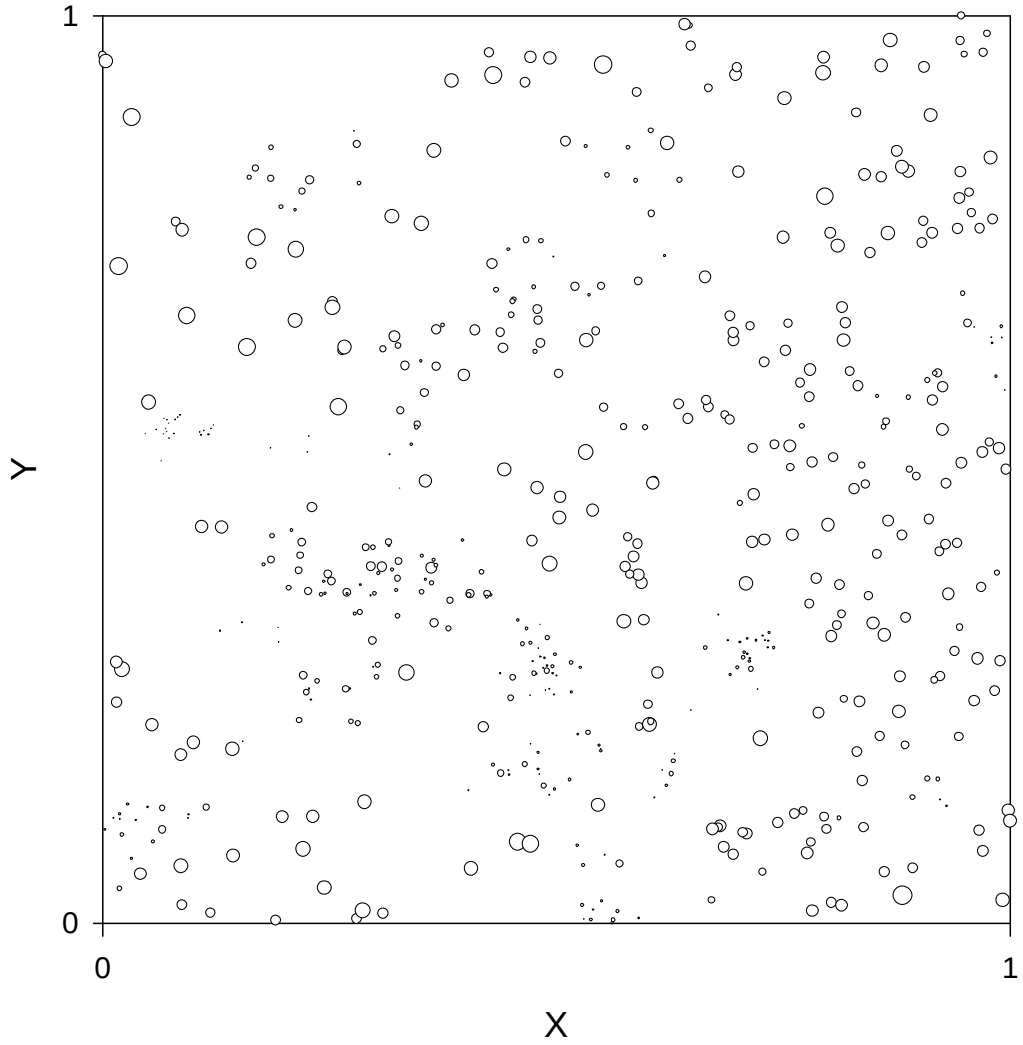
Since the earlier (conditional) $r = 7$ mark effect does *not* feature in the discrepancy analysis, it must be an artefact of the point pattern.

discrepancy	k_j	s_{1j}	s_{2j}	r_j	ϕ_j
355	3.142	0	0	0.0	0
250	4.6	18	9	20.1	88.5
233	2.1	14	20	24.4	13.0
224	1.5	10	22	24.2	21.0
210	1.4	20	3	20.2	−5.5
207	0.5	0	−1	1.0	−15.0
205	0.5	−1	1	1.4	15.0
203	−0.5	2	0	2.0	15.0
201	0.5	2	1	2.2	15.0
199	−0.4	2	3	3.6	15.0
198	−0.4	1	3	3.2	−15.0
197	−0.3	−1	−1	1.4	15.0
196	0.3	−1	0	1.0	12.0
195	0.2	−1	2	2.2	10.5
195	0.2	2	−1	2.2	15.0

Sequential estimates based on the discrepancy function, showing $D(\cdot) \times 10^{-3}$, amplitude k_j, frequencies s_{j1} and s_{j2}, isotropic frequency $r_j = \sqrt{(s_{1j}^2 + s_{2j}^2)}$, and phase ϕ_j $(j = 0, \ldots, 15)$

Analysis of Longleaf Pine Data

Taken from the Wade Tract in Georgia, U.S.A. (1979): coordinates and diameters of trees ≥ 2 cm d.b.h. in 4ha of old-growth forest.



- Point R-periodogram shows frequencies $r \leq 12$ are highly significant – indicative of clustering with intercluster distances ≥ 20 m.
- Agrees with Cressie's nested block quadrat effect.
- Point Θ-periodogram highlights a row effect, and a possible column effect.
- Mark periodogram shows high agreement with these point results.

Suggests it is the tree locations, and not their diameters, that forms the main component of spatial pattern.

Using the Discrepancy Function allows us to explore this further since it provides useful information on the unconditional mark structure.

discrepancy	k_j	s_{1j}	s_{2j}	r_j	ϕ_j
9999	26.84366	0	0	0	0
1000	22.0	18	1	18.03	53
808	6.6	0	−1	1.00	83
667	5.8	1	1	1.41	108.3
606	12.2	18	3	18.25	47
560	5.7	4	2	4.47	153
503	7.6	18	1	18.03	173
457	6.0	2	4	4.47	173
392	4.7	1	2	2.24	−3
369	4.9	12	8	14.42	147
346	3.4	1	4	4.12	173

Sequential estimates, showing $D(\cdot)\times 10^{-6}$, amplitude k_j, frequencies s_{j1} and s_{j2}, isotropic frequency $r_j = \sqrt{(s_{1j}^2 + s_{2j}^2)}$, and phase ϕ_j $(j = 0, \ldots, 10)$.

- Highlights a strong left-right local structure with frequency 18.
- Note $r_1 = 18 >$ spectral 12 and Cressie quadrat 8, so implies considerably shorter interaction distances or, more likely, a different spatial effect.

- Adding secondary waves detects zones with differing levels of mark variability. Main wavelength of 18.03 agrees with the visual degree of dispersion shown by clusters with small mark values.

- So by admitting a much greater degree of 'DISENTANGLEMENT' between marks and points, the discrepancy approach enables us to make a far more definitive, and quantifiably precise, assessment of spatial structure than that provided by standard space or frequency domain techniques.