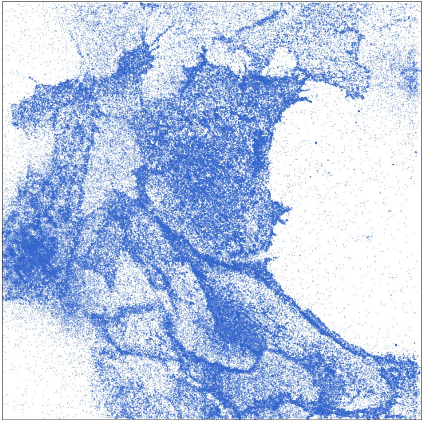
Cluster marked cluster point processes

Ute Hahn

CENTRE FOR **STOCHASTIC GEOMETRY** AND ADVANCED **BIOIMAGING**

Aarhus University,
Denmark

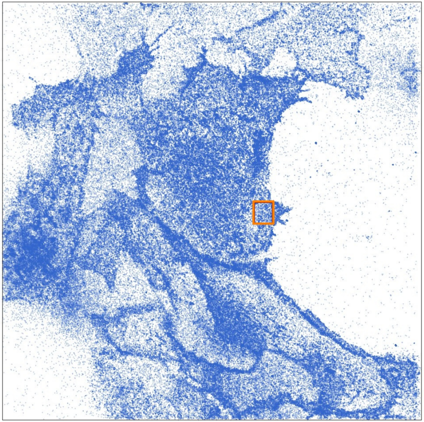
19th Workshop on
Stochastic Geometry, Stereology and Image Analysis
Luminy 15.05.2017



Positions of aquaporin molecules on kidney cell membrane

Lene N. Nejsum,
Eva A. Christensen

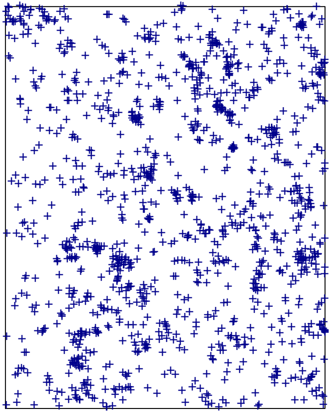
Department of
Molecular Biology and Genetics,
AU, Denmark



Positions of aquaporin molecules on kidney cell membrane

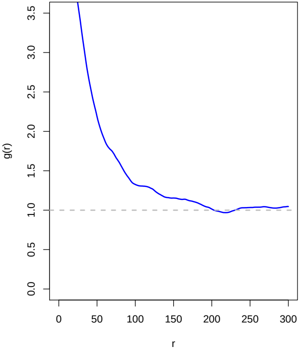
Lene N. Nejsum,
Eva A. Christensen

Department of
Molecular Biology and Genetics,
AU, Denmark

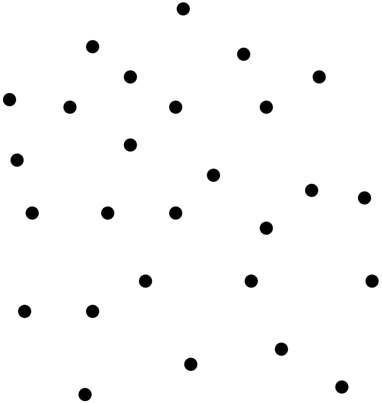


ca. 3300 × 4100 nm

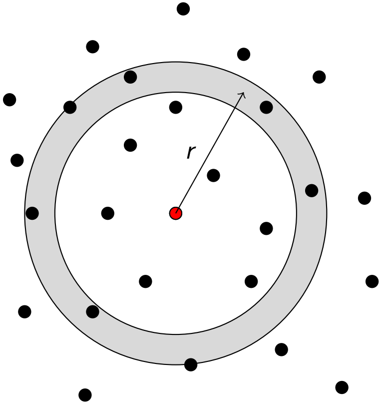
Aim: characterize clustering by pair correlation function



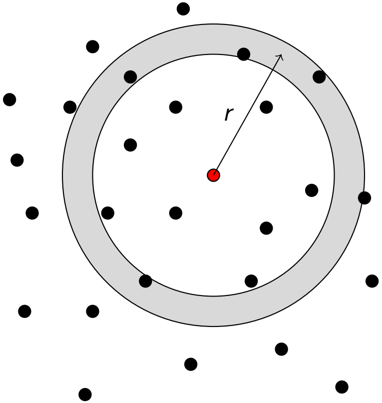
$$g(r) = \frac{\text{Palm intensity of } \boldsymbol{X} \text{ at distance } r \text{ to typical point}}{\text{Intensity of } \boldsymbol{X}}$$



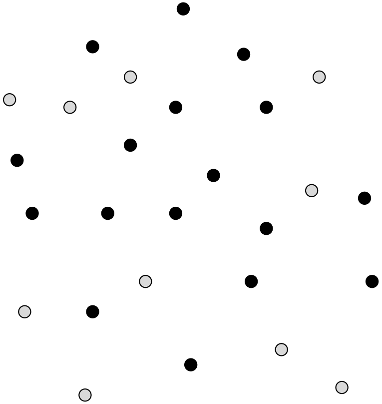
$$g(r) = \frac{\text{Palm intensity of } \boldsymbol{X} \text{ at distance } r \text{ to typical point}}{\text{Intensity of } \boldsymbol{X}}$$



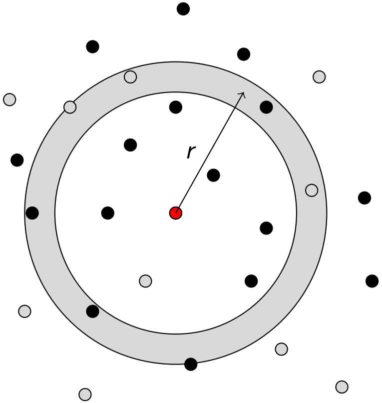
$$g(r) = \frac{\text{Palm intensity of } \boldsymbol{X} \text{ at distance } r \text{ to typical point}}{\text{Intensity of } \boldsymbol{X}}$$



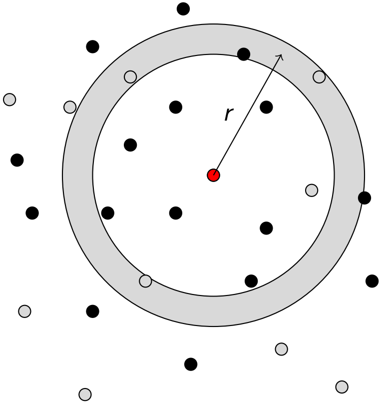
The pair correlation function is invariant to independent thinning of the process.

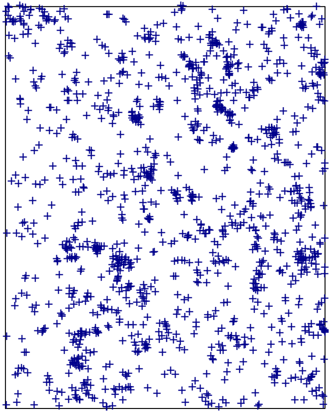


The pair correlation function is invariant to independent thinning of the process.



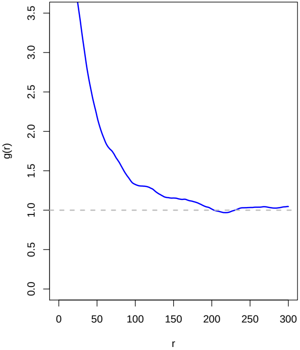
The pair correlation function is invariant to independent thinning of the process.





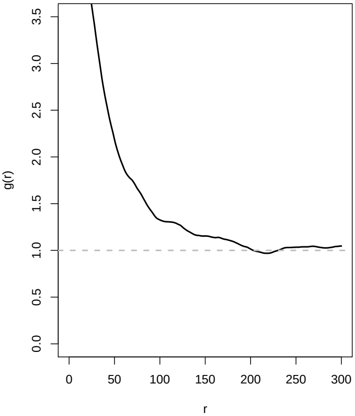
ca. 3300 × 4100 nm

Aim: characterize clustering by pair correlation function



original pair correlation of $\boldsymbol{X}$:

`pcf(X)`

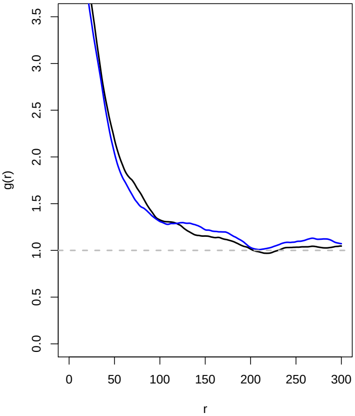


original pair correlation of $\boldsymbol{X}$:

```
pcf(X)
```

indep.thinning, retain 500 points:

```
pcf(X[sample(npoints(X), 500)])
```



original pair correlation of ***X***:

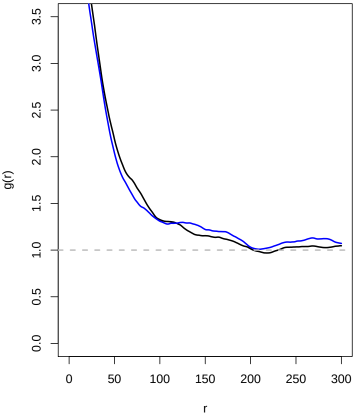
```
pcf(X)
```

indep.thinning, retain 500 points:

```
pcf(X[sample(npoints(X), 500)])
```

first 500 points:

```
pcf(X[1:500])
```



original pair correlation of ***X***:

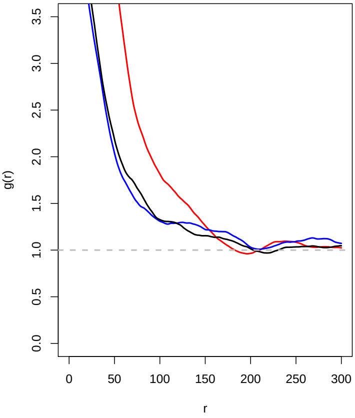
```
pcf(X)
```

indep.thinning, retain 500 points:

```
pcf(X[sample(npoints(X), 500)])
```

first 500 points:

```
pcf(X[1:500])
```



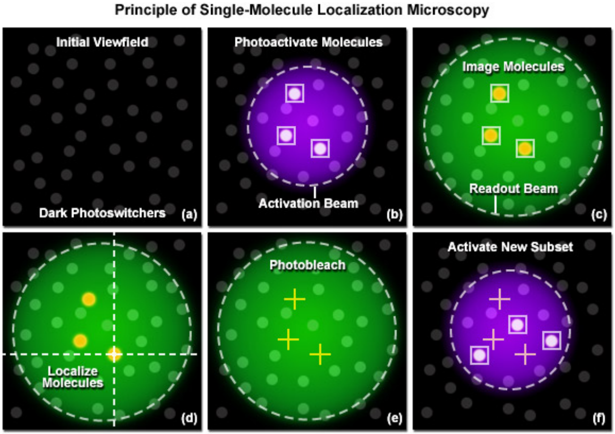
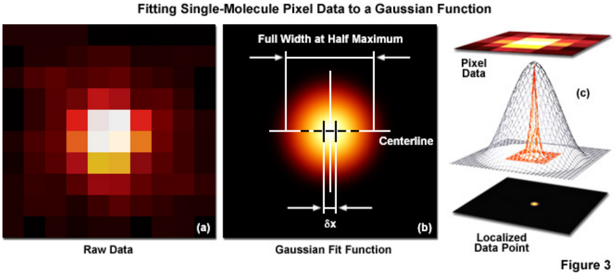


Figure 1

Published 2006, Method of the Year 2008, Nobel Prize in Chemistry 2014 (Betzig, Hell, Moerner)

Source: http://zeiss-campus.magnet.fsu.edu

Molecule positions estimated from digital images by maximum likelihood



Source: http://zeiss-campus.magnet.fsu.edu

Scientific hypothesis on PALM images:

- Points represent positions of fluorescent molecules,
- molecules fluoresce independently of each other.

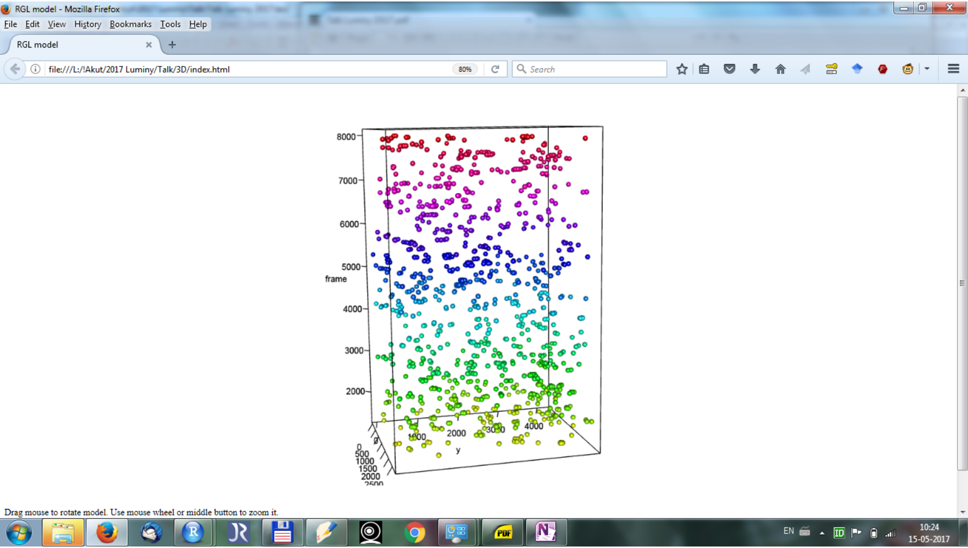
Scientific hypothesis on PALM images:

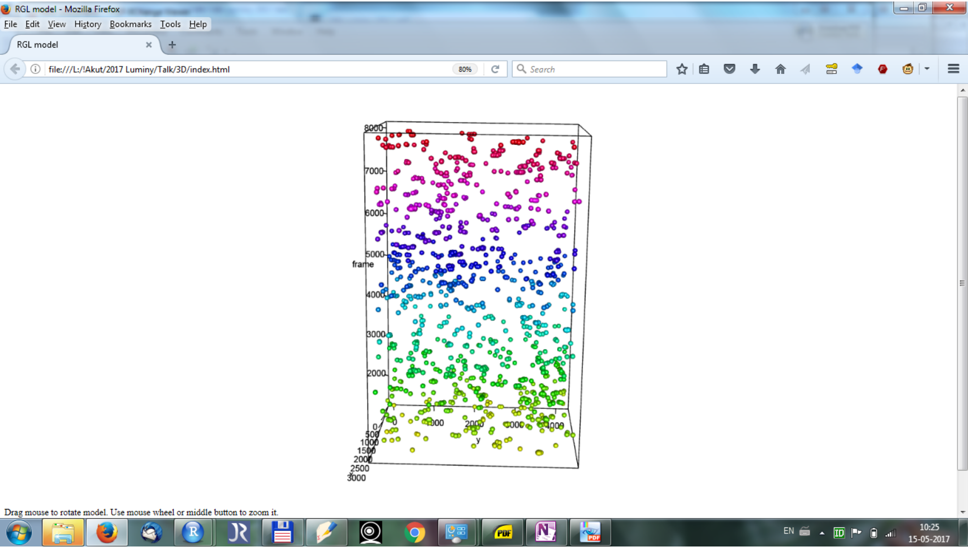
- Points represent positions of fluorescent molecules,
- molecules fluoresce independently of each other.

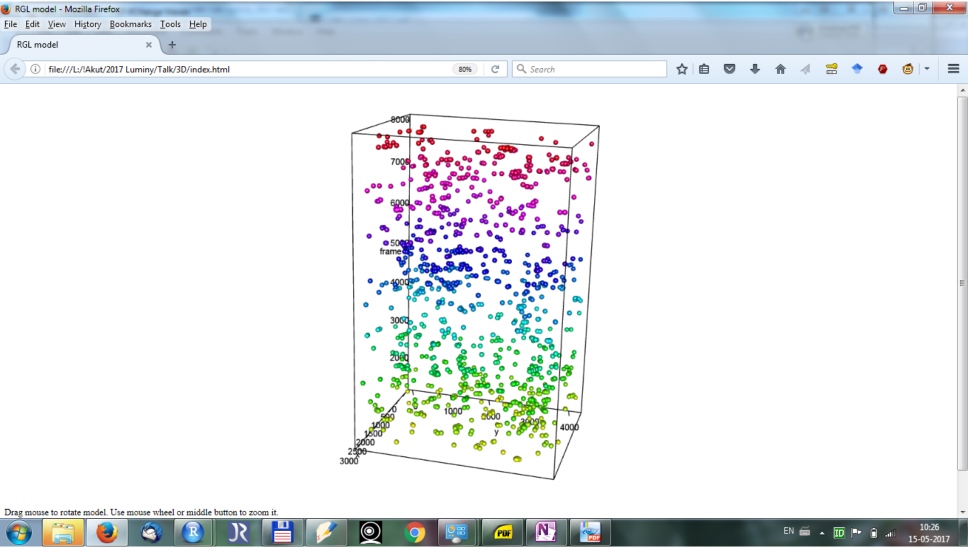
Data:

- points registered in sequence of video frames,
- frame number available (range:$[1400, 8000]$).

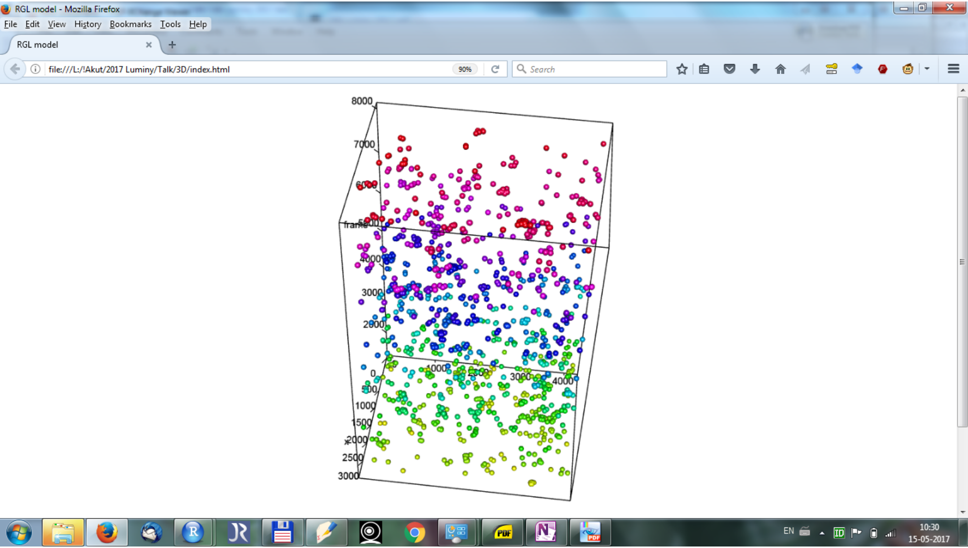
☞ analyse as space-time or marked point pattern;
inspect data in 3D, as points in $\mathbb{R}^2 \times$ frames.

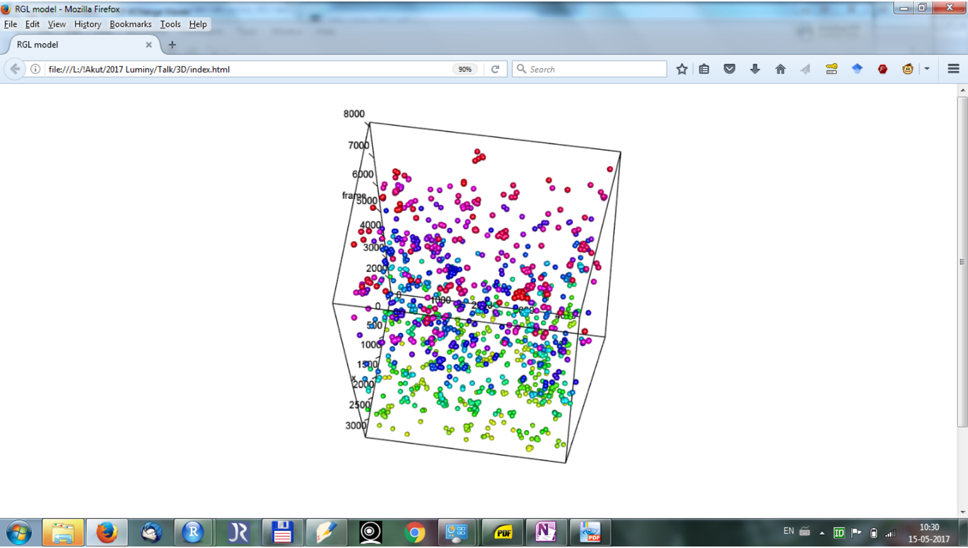

RGL model - Mozilla Firefox
File Edit View History Bookmarks Tools Help
RGL model
file:///L:/!Akut/2017 Luminy/Talk/3D/index.html
80%
Search
8000
7000
6000
5000
frame
4000
3000
2000
1000
2000
3000
4000
y
0
500
1000
1500
2000
Drag mouse to rotate model. Use mouse wheel or middle button to zoom it.
EN
10:24
15-05-2017

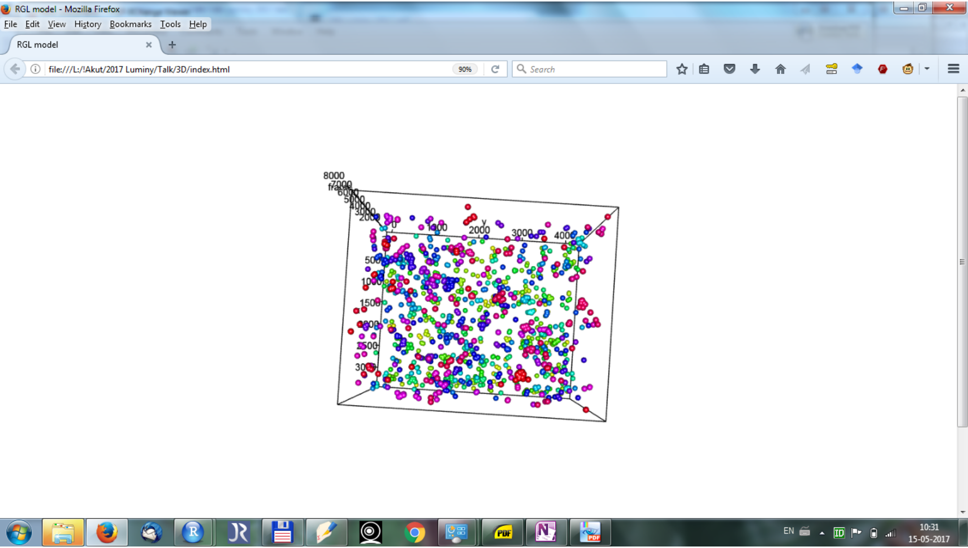

RGL model - Mozilla Firefox
File Edit View History Bookmarks Tools Help
RGL model
file:///L:/!Akut/2017 Luminy/Talk/3D/index.html
80%
Search
8000
7000
6000
5000
frame
4000
3000
2000
0
500
1000
1500
2000
2500
3000
2000
4000
y
Drag mouse to rotate model. Use mouse wheel or middle button to zoom it.
EN
10:25
15-05-2017



RGL model - Mozilla Firefox
File Edit View History Bookmarks Tools Help
RGL model
file:///L:/!Akut/2017 Luminy/Talk/3D/index.html
80%
Search
8000
7000
6000
5000
frame
4000
3000
2000
0
500
1000
1500
2000
2500
3000
1000
2000
3000
4000
y
Drag mouse to rotate model. Use mouse wheel or middle button to zoom it.
EN
10:26
15-05-2017


RGL model - Mozilla Firefox
File Edit View History Bookmarks Tools Help
RGL model
file:///L:/!Akut/2017 Luminy/Talk/3D/index.html
90%
Search
8000
7000
6000
5000
frame
4000
3000
2000
1000
0
500
1000
1500
2000
2500
3000
x
3000
4000
EN
10:30
15-05-2017

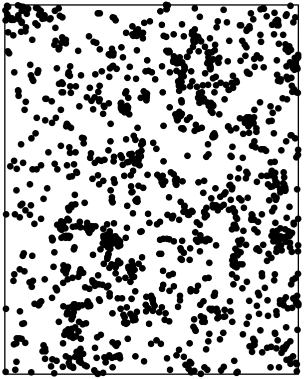

RGL model - Mozilla Firefox
File Edit View History Bookmarks Tools Help
RGL model
file:///L:/!Akut/2017 Luminy/Talk/3D/index.html
90%
Search
8000
7000
6000
frame
5000
4000
3000
2000
0
1000
0
500
1000
1500
x
2000
2500
3000
EN
10:30
15-05-2017


RGL model - Mozilla Firefox
File Edit View History Bookmarks Tools Help
RGL model
file:///L:/!Akut/2017 Luminy/Talk/3D/index.html
90%
Search
8000
7000
6000
5000
4000
3000
2000
0
1000
2000
3000
4000
y
500
1000
1500
2500
EN
10:31
15-05-2017

Ground process: cluster process $\boldsymbol{X}$

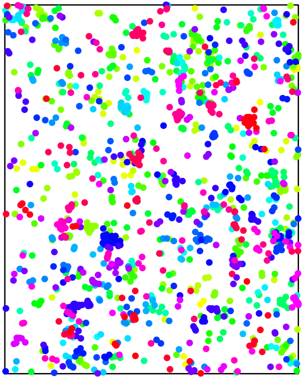
- Parent process $\boldsymbol{Y}$
- cluster template $\boldsymbol{C}_0$
- independent clusters $\boldsymbol{C}(y) \sim \boldsymbol{C}_0 + y$, for all $y \in \boldsymbol{Y}$
- superposition of clusters

$$\boldsymbol{X} = \bigcup_{y \in \boldsymbol{Y}} \boldsymbol{C}(y).$$



Parent dependent marking

- Notation: $m(u)$ mark of point u, mark space $\mathbb{M}$.
- Location family of mark distributions for the clusters: $\{\mathcal{M}_\theta, \theta \in \Theta\}$.
- Parent process is i.i.d. marked, $m(y) \sim \mathcal{T}$, with values in Θ.
- Points $u \in \boldsymbol{C}(y)$ are i.i.d. marked with $m(u) \sim \mathcal{M}_{m(y)}$.



Resulting cluster marked cluster process (CMCpp) is mark stationary: Palm mark distribution

$$\mathcal{M}(\mathrm{d}\, m) = \mathbf{E}\mathcal{M}_T(\mathrm{d}\, m), \quad T \sim \mathcal{T},$$

is independent of location.

Appropriate model for the data:

- Mark distributions $\mathcal{M}_\theta$ have small variance,
- distribution $\mathcal{T}$ of location parameter spreads widely.

☞ Points close to each other in space have preferentially similar frame marks.

Want: a statistics that shows that spatial clusters cover only a close range of marks (frames).

Stoyan's mark correlation function for stationary isotropic point process $\boldsymbol{X}$: f measureable function on $\mathbb{M} \times \mathbb{M}$

$$k_f(r) := \frac{\mathbf{E}[f(m(u), m(v)) \mid u, v \in \boldsymbol{X}]}{\mathbf{E}[f(M_1, M_2)]},$$
$$\|u - v\| = r,\ M_1, M_2 \sim \mathcal{M}, \text{indep.}$$

- If $\boldsymbol{X}$ is independently marked, $k_f(r) \equiv 1$.
- For (quasi) continuous marks, the usual choice is $f(m_1, m_2) = m_1 m_2$.

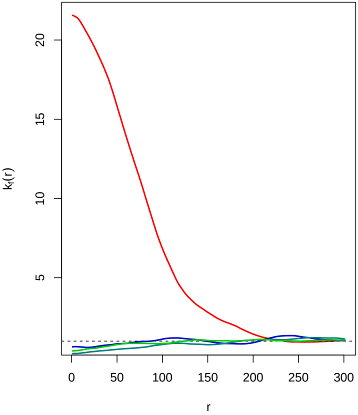
- Under the model, $|m(u) - m(v)|$ is small for points from same cluster, but likely to be large for points from different clusters.
- Choose $f(m_1, m_2) = \mathbf{1}(|m_1 - m_2| \in [t, t + \Delta])$, thus

$$k_f(r) = \frac{\Pr(|m(u) - m(v)| \in [t, t + \Delta] \mid u, v \in \boldsymbol{X}, \|u - v\| = r)}{\Pr(|m(u) - m(v)| \in [t, t + \Delta] \mid u, v \in \boldsymbol{X})}$$

– an odds ratio.

Results, $\Delta = 20$:

- $t = 0$:
  emphasize points close in time
- $t = 20$, $t = 40$, $t = 60$



interpretation: close points (distance $r < 200$), are more likely to have similar frame numbers than different frame numbers.

Theorem.

Let:

$\boldsymbol{X}$ be a cluster marked cluster process with pair correlation function $g_{\boldsymbol{X}}$,
$\boldsymbol{Y}$ its parent process,
$\tilde{\boldsymbol{C}}$ the distribution of a random point in template clusters $\boldsymbol{C}$,
$\mathcal{M}_\theta, \theta \in \Theta$ the mark distributions of the clusters,
$\mathcal{M}$ the overall Palm mark distribution.

Let f be a function on $\mathbb{M}^2$ that fulfils, for independent M_1, M_2:

$$\mathbf{E} f(M_1, M_2) \begin{cases} = 0, & M_1, M_2 \sim \mathcal{M}_\theta, \text{ for all } \theta \in \Theta, \\ > 0, & M_1, M_2 \sim \mathcal{M} \end{cases}$$

then for the mark correlation function k_f of $\boldsymbol{X}$,

$$k_f(r) g_{\boldsymbol{X}}(r) = g_{\tilde{\boldsymbol{Y}}}(r),$$

where $\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{C}} * \boldsymbol{Y}$ is the randomly $\tilde{\boldsymbol{C}}$-displaced parent process.

f-weighted second order factorial moment measure

$$\alpha_f^{(2)}(B_1, B_2) = \mathbf{E} \sum_{u,v \in \boldsymbol{X}}^{\neq} f(m(u), m(v)) \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v)$$

independent marking:

$$\alpha_f^{(2)}(B_1, B_2) = \underbrace{\mathbf{E} f(M_1, M_2)}_{=:\mu} \underbrace{\mathbf{E} \sum_{u,v \in \boldsymbol{X}}^{\neq} \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v)}_{=\alpha^{(2)}(B_1, B_2)}, \quad M_1, M_2 \sim \mathcal{M}, \text{indep.}$$

☞ mark correlation function

$k_f(\,,\,)$ is the density of $\alpha_f^{(2)}$ wrt. $\mu\alpha^{(2)}$.

independent locations (Poisson process)

$$\alpha^{(2)}(B_1, B_2) = \Lambda(B_1)\Lambda(B_2) =: \Lambda^2(B1, B2) \qquad \Lambda : \text{intensity measure}$$

☞ pair correlation function

$g(\,,\,)$ is the density of $\alpha^{(2)}$ wrt. Λ^2

independent locations (Poisson process)

$$\alpha^{(2)}(B_1, B_2) = \Lambda(B_1)\Lambda(B_2) =: \Lambda^2(B1, B2) \qquad \Lambda : \text{intensity measure}$$

☞ pair correlation function

$$g(\,,\,) \text{ is the density of } \alpha^{(2)} \text{ wrt. } \Lambda^2$$

thus, with k_f being the density of $\alpha_f^{(2)}$ wrt. $\mu\alpha^{(2)}$,

$$k_f \cdot g \text{ is the density of } \alpha_f^{(2)} \text{ wrt. } \mu\,\Lambda^2.$$

Cluster marked process

$$
\begin{aligned}
\alpha_f^{(2)}(B_1, B_2) &= \mathbf{E} \sum_{u,v \in \boldsymbol{X}}^{\neq} f(m(u), m(v)) \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v) \\
&= \mathbf{E} \sum_{y \in \boldsymbol{Y}} \sum_{u,v \in \boldsymbol{C}(y)}^{\neq} f(m(u), m(v)) \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v) \\
&\quad + \mathbf{E} \sum_{y,y' \in \boldsymbol{Y}}^{\neq} \sum_{u \in \boldsymbol{C}(y)} \sum_{v \in \boldsymbol{C}(y')} f(m(u), m(v)) \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v)
\end{aligned}
$$

Cluster marked process

$$\begin{aligned}\alpha_f^{(2)}(B_1, B_2) &= \mathbf{E} \sum_{u,v \in \boldsymbol{X}}^{\neq} f(m(u), m(v)) \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v) \\ &= \mathbf{E} \sum_{y \in \boldsymbol{Y}} \sum_{u,v \in \boldsymbol{C}(y)}^{\neq} \mathbf{E} f(m(u), m(v)) \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v) \\ &\quad + \mathbf{E} \sum_{y,y' \in \boldsymbol{Y}}^{\neq} \sum_{u \in \boldsymbol{C}(y)} \sum_{v \in \boldsymbol{C}(y')} \mathbf{E} f(m(u), m(v)) \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v)\end{aligned}$$

Cluster marked process

$$\alpha_f^{(2)}(B_1, B_2) = \mathbf{E} \sum_{u,v \in \boldsymbol{X}}^{\neq} f(m(u), m(v)) \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v)$$

$$= \mathbf{E} \sum_{y \in \boldsymbol{Y}} \sum_{u,v \in \boldsymbol{C}(y)}^{\neq} \mu_{C(y)} \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v)$$

$$+ \mathbf{E} \sum_{y,y' \in \boldsymbol{Y}}^{\neq} \sum_{u \in \boldsymbol{C}(y)} \sum_{v \in \boldsymbol{C}(y')} \mu \, \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v)$$

Cluster marked process

$$\alpha_f^{(2)}(B_1, B_2) = \mathbf{E} \sum_{u,v \in \boldsymbol{X}}^{\neq} f(m(u), m(v)) \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v)$$

$$= \mathbf{E} \sum_{y \in \boldsymbol{Y}} \sum_{u,v \in \boldsymbol{C}(y)}^{\neq} \mu_{C(y)} \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v)$$

$$+ \mathbf{E} \sum_{y,y' \in \boldsymbol{Y}}^{\neq} \sum_{u \in \boldsymbol{C}(y)} \sum_{v \in \boldsymbol{C}(y')} \mu\, \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v)$$

with f such that $\mu_{\boldsymbol{C}(y)} = 0$,

$$= \mu\ \mathbf{E} \sum_{y,y' \in \boldsymbol{Y}}^{\neq} \sum_{u \in \tilde{\boldsymbol{C}}} \sum_{v \in \tilde{\boldsymbol{C}}'} \mathbf{1}_{B_1}(y+u) \mathbf{1}_{B_2}(y'+v)$$

Cluster marked process

$$\alpha_f^{(2)}(B_1, B_2) = \mathbf{E} \sum_{u,v \in \boldsymbol{X}}^{\neq} f(m(u), m(v)) \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v)$$

$$= \mathbf{E} \sum_{y \in \boldsymbol{Y}} \sum_{u,v \in \boldsymbol{C}(y)}^{\neq} \mu_{C(y)} \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v)$$

$$+ \mathbf{E} \sum_{y,y' \in \boldsymbol{Y}}^{\neq} \sum_{u \in \boldsymbol{C}(y)} \sum_{v \in \boldsymbol{C}(y')} \mu \, \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v)$$

with f such that $\mu_{\boldsymbol{C}(y)} = 0$,

$$= \mu \; \nu^2 \mathbf{E} \sum_{y,y' \in \tilde{\boldsymbol{Y}}}^{\neq} \mathbf{1}_{B_1}(y) \mathbf{1}_{B_2}(y')$$

ν: mean number of points in $\boldsymbol{C}$

Cluster marked process

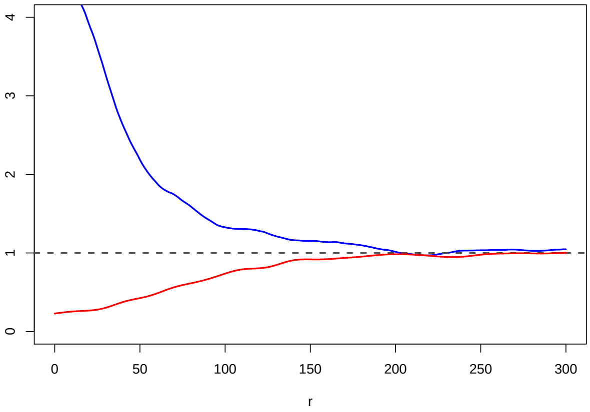
$$
\begin{aligned}
\alpha_f^{(2)}(B_1, B_2) &= \mathbf{E} \sum_{u,v \in \boldsymbol{X}}^{\neq} f(m(u), m(v)) \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v) \\
&= \mathbf{E} \sum_{y \in \boldsymbol{Y}} \sum_{u,v \in \boldsymbol{C}(y)}^{\neq} \mu_{C(y)} \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v) \\
&\quad + \mathbf{E} \sum_{y,y' \in \boldsymbol{Y}}^{\neq} \sum_{u \in \boldsymbol{C}(y)} \sum_{v \in \boldsymbol{C}(y')} \mu \, \mathbf{1}_{B_1}(u) \mathbf{1}_{B_2}(v)
\end{aligned}
$$

with f such that $\mu_{\boldsymbol{C}(y)} = 0$,

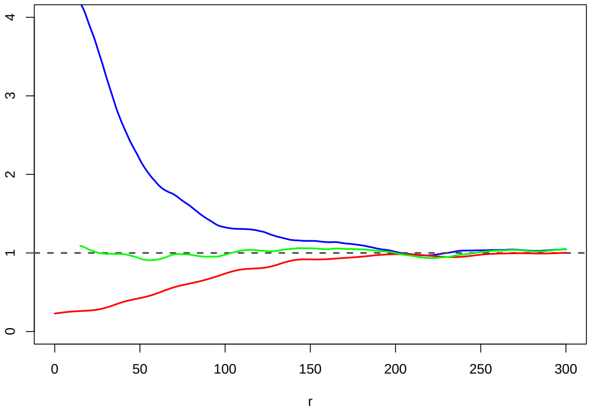
$$
\begin{aligned}
&= \mu \, \nu^2 \mathbf{E} \sum_{y,y' \in \tilde{\boldsymbol{Y}}}^{\neq} \mathbf{1}_{B_1}(y) \mathbf{1}_{B_2}(y') \\
&= \mu \nu^2 \, \alpha_{\tilde{\boldsymbol{Y}}}^{(2)}(B_1, B_2).
\end{aligned}
$$

pair correlation function $g(r)$

mark correlation function $k_f(r)$ with $f(m_1, m_2) = |m_1 - m_2|$

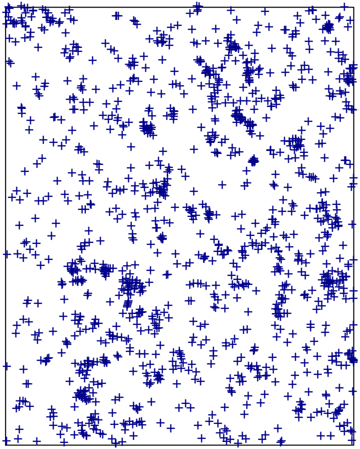


pair correlation function $g(r)$
mark correlation function $k_f(r)$ with $f(m_1, m_2) = |m_1 - m_2|$
$k_f(r)g(r)$



Model for ground process:

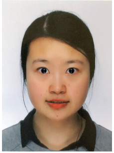
Matérn cluster process superimposed with Poisson "noise".



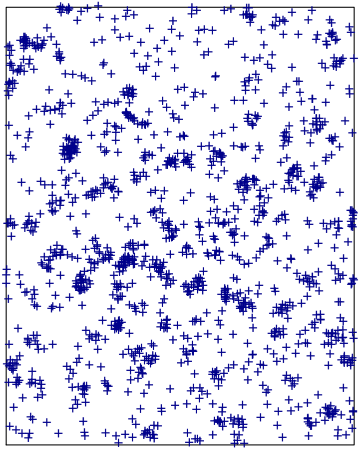
data

Model for ground process:

Matérn cluster process superimposed with Poisson “noise”.



Jiaxin Chen



fittet model