Estimating the pair correlation function from images of epidermal nerve fibers

Lance A. Waller, Traci Leong

Emory University

William Kennedy, Gwen Wendelschafer-Crabb

University of Minnesota

Aila Särkkä

Chalmers University

lwaller@sph.emory.edu

Location matters!

Outline

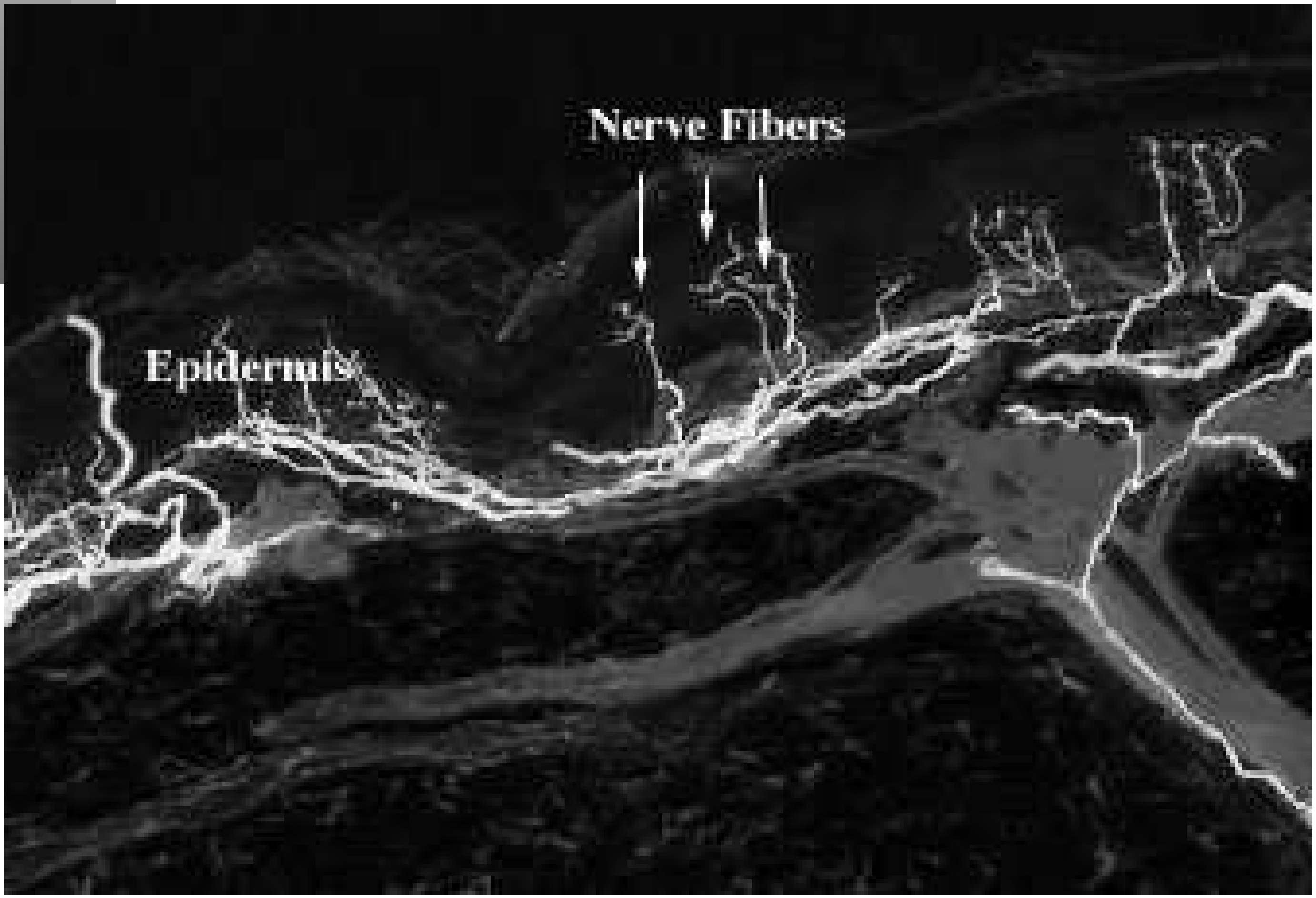
- What do we have? (epidermal nerve fiber images)
- What do we want? (clusters versus clustering)
- How do we do it? (K functions and pair correlation functions)
- Estimating pair correlation functions.
- Preliminary results.
- Conclusions/questions.

What do we have?

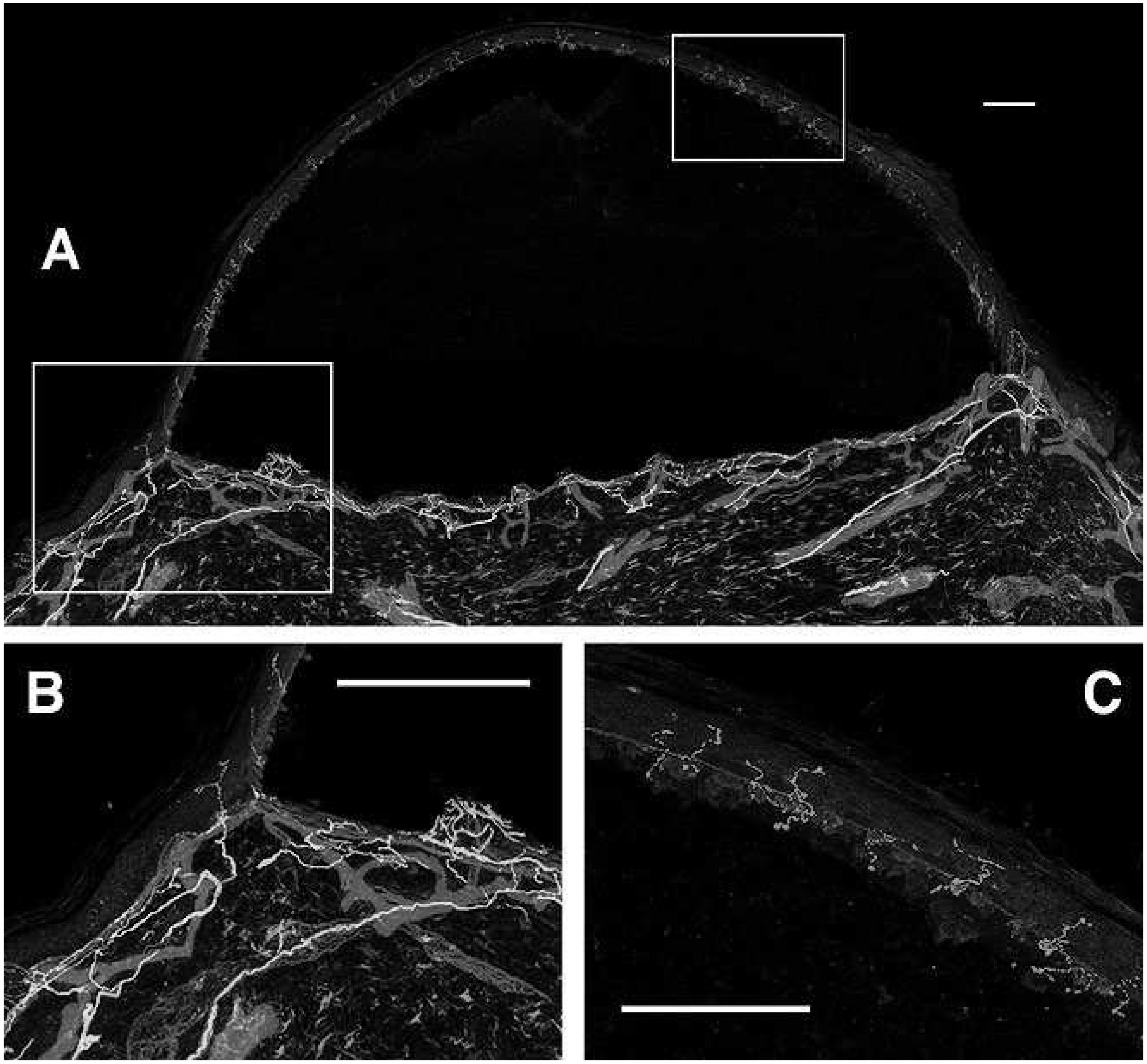
Epidermal nerve fibers (ENFs)

- Living nerve fibers extending from the dermis into the epidermis.
- Transmit heat, cold, pain.
- First imaged by Kennedy, Wendelschafer-Crabb, and Johnson (1996, *Neurology*).
- In *neuropathy*, ENFs “die off”, resulting in reduced nerve density.
- But seem to die off in a pattern, leaving a “clustered” pattern.

Image of ENFs



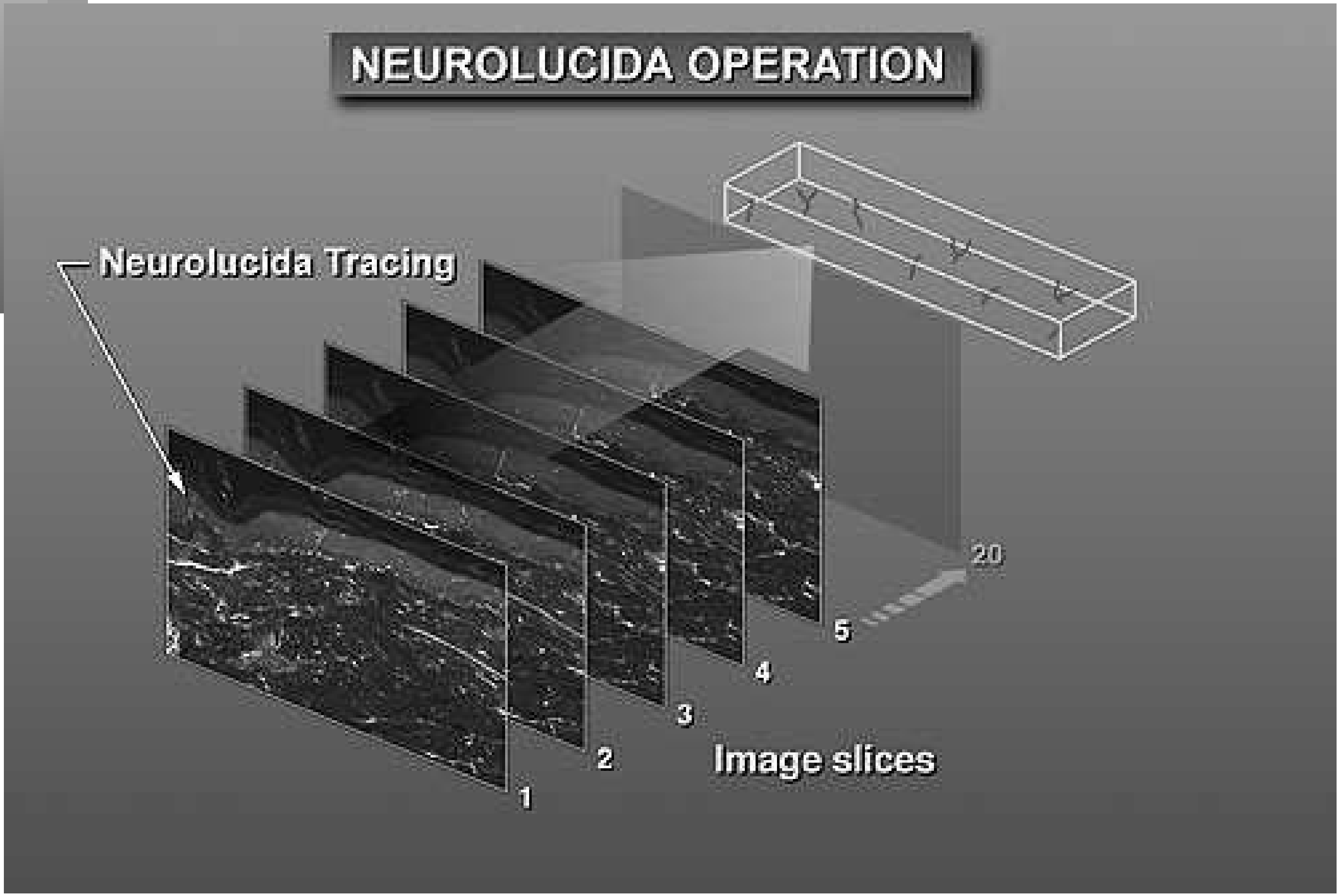
Skin Blister Biopsy



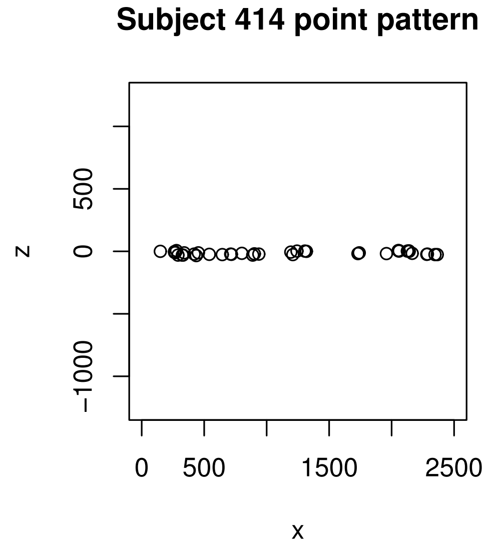
Types of biopsy

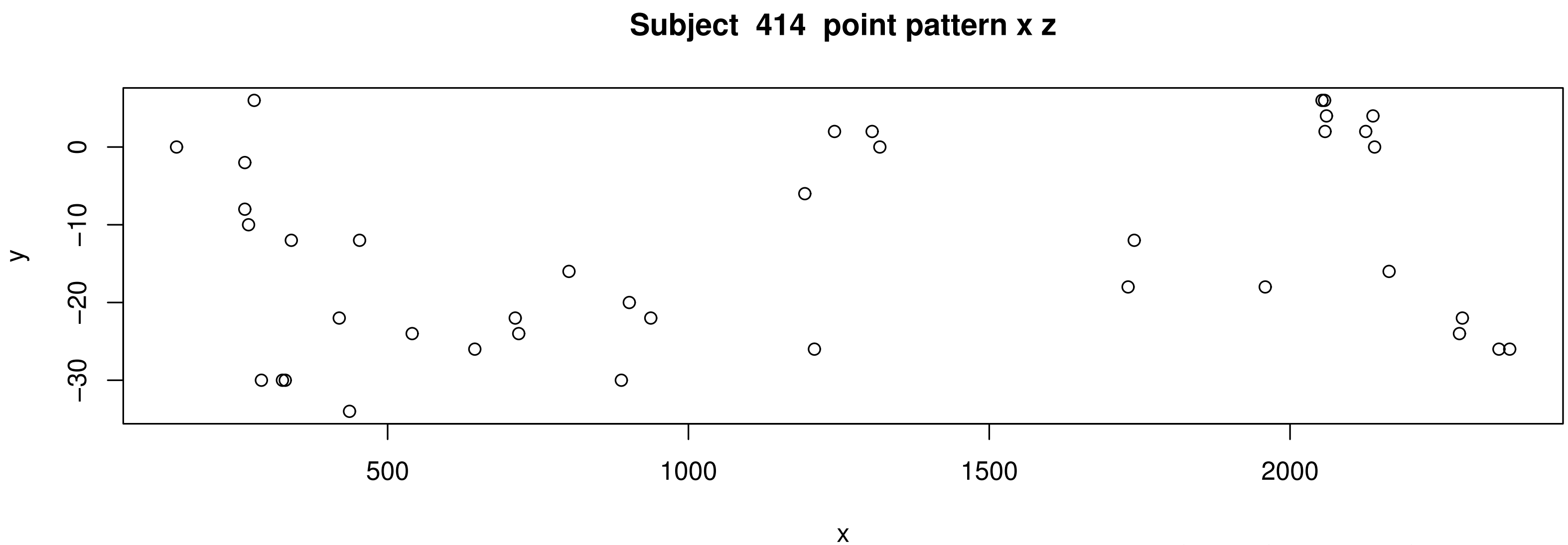
- *Skin blister biopsy*: Suction-induced 3mm sample of epidermis only.
- Flattened and imaged in confocal microscope from above (horizontal “layers”).
- *Skin punch biopsy*: Epidermis and dermis.
- Confocal microscopy from side (vertical “layers”).
- Trace each fiber using Neurolucida software.
- Map of “trunk” of each “tree”.
- We project to 2-dimensions (from 3).

Confocal microscopy



Data from Subject 414





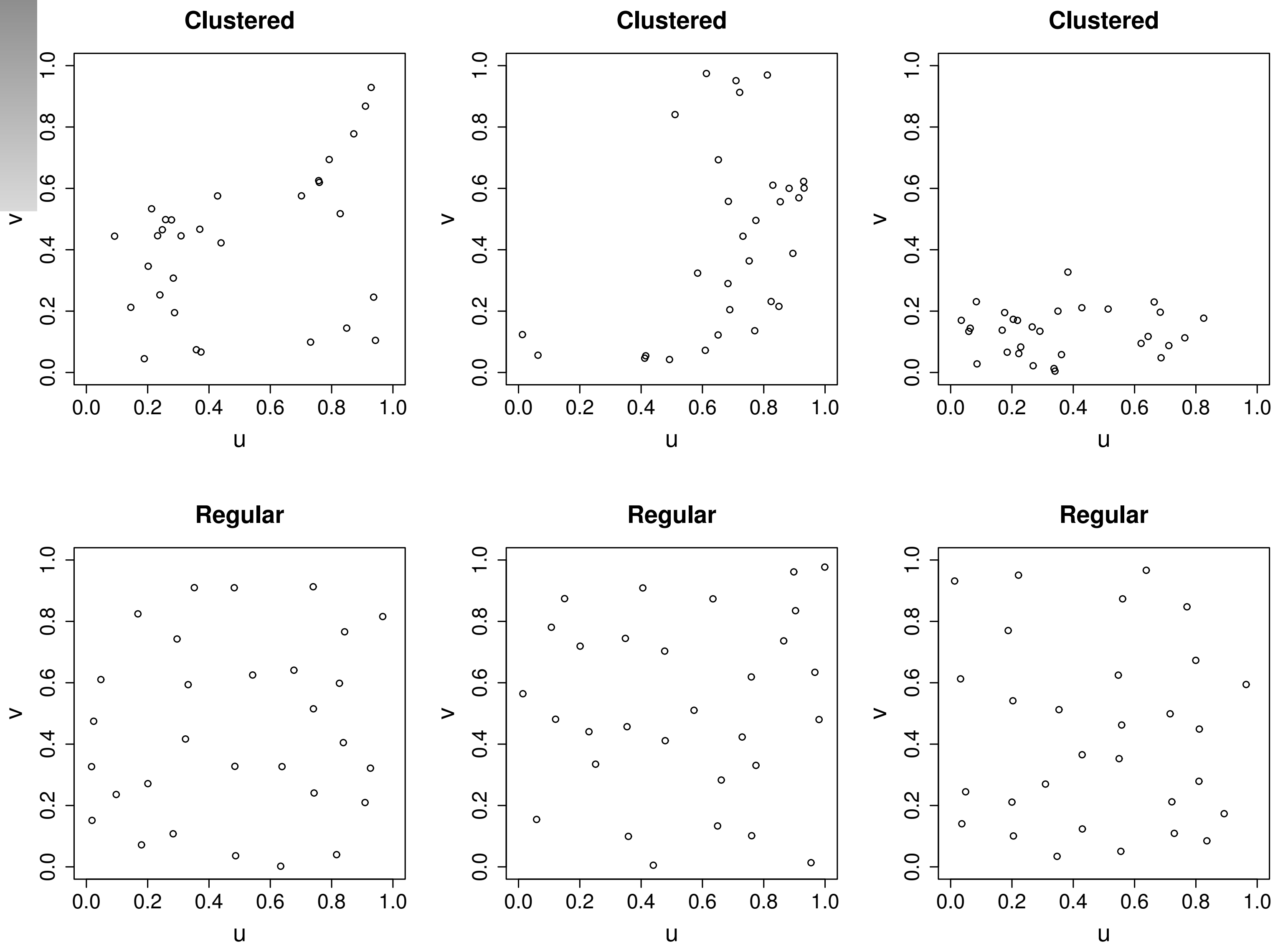
What do we want?

Contrast ideas of:

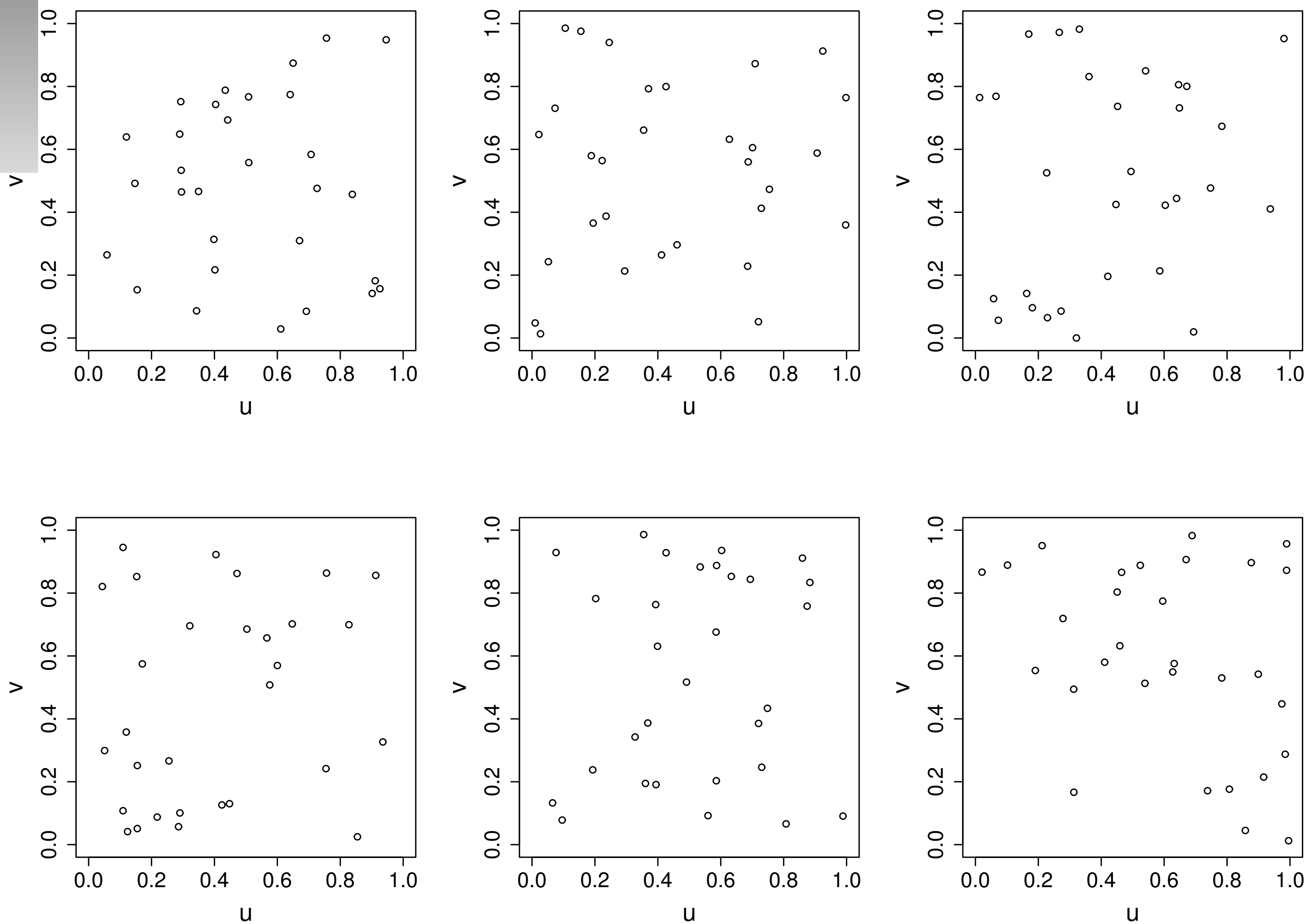
- *Cluster*: Single anomoly.
- *Clustering*: A tendency for observed events to occur near other events.
- *Regularity*: A tendency for observed events to avoid other events.

We want to identify whether one observed pattern is more clustered than another.

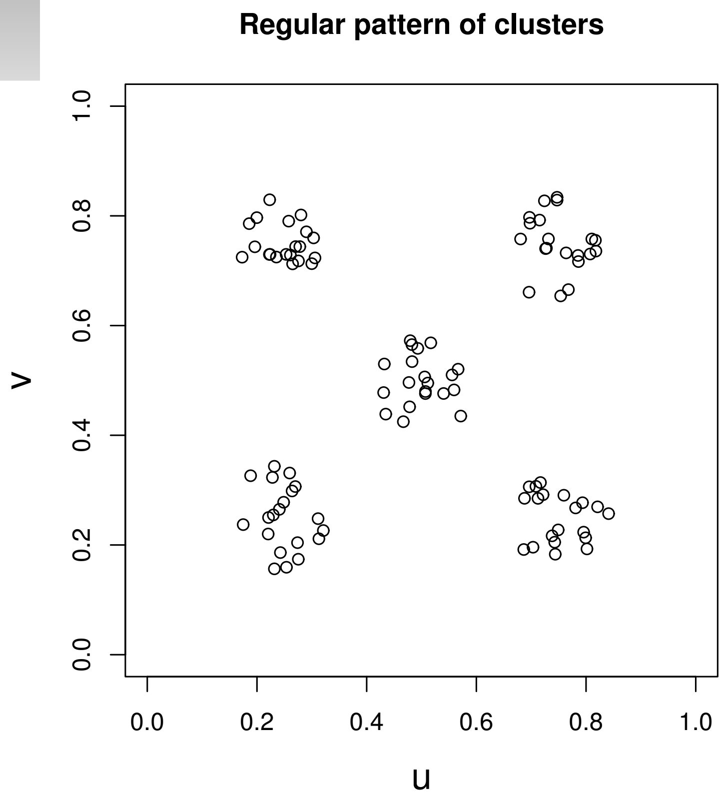
Too Clustered (top), Too Regular (bottom)

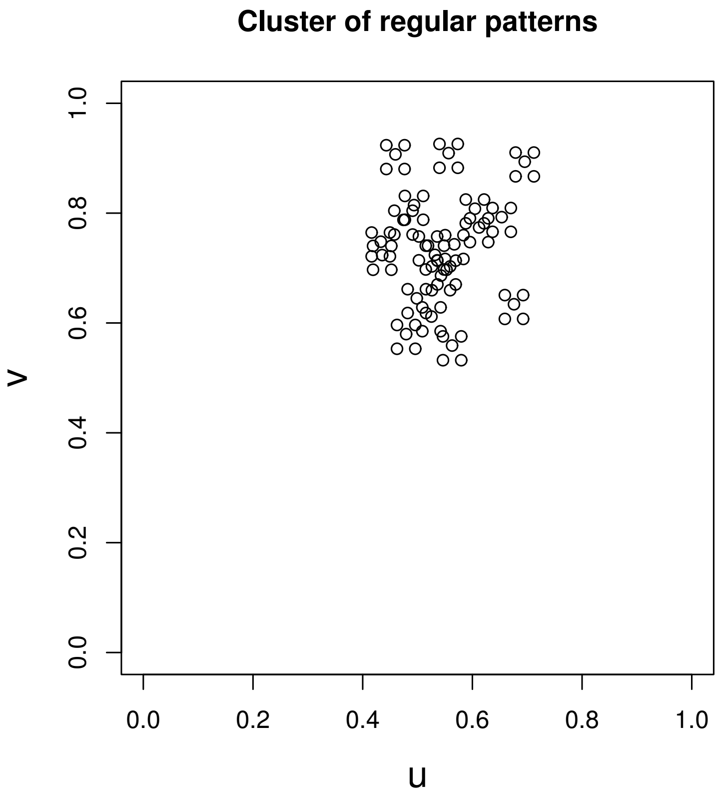


Complete Spatial Randomness



Spatial Scale Matters!





2nd Order Property: K ***function***

Ripley (1976, 1977) introduced the *reduced second moment measure* or K *function*

$$K(h) = \frac{E[\text{\# events within } h \text{ of a } \textit{randomly} \text{ chosen event}]}{\lambda},$$

for any positive *spatial lag* h.

NOTE: Use of λ implies assumption of stationary process!

Properties of $K(h)$

- Ripley (1977) shows specifying $K(h)$ for all $h > 0$, equivalent to specifying Var$[N(A)]$ for any subregion A.
- Under CSR, $K(h) = \pi h^2$ (area of circle of with radius h).
- Clustered? $K(h) > \pi h^2$.
- Regular? $K(h) < \pi h^2$.

Estimation

$$\widehat{K}_{ec}(h) = \widehat{\lambda}^{-1} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} (w_{ij})^{-1} \delta(d(i,j) < h)$$

where w_{ij} = proportion of the circumference of circle centered at event i, radius $d(i,j)$ within the study area.

Plots with $K(h)$

- Plotting $(h, K(h))$ for CSR is a parabola.
- $K(h) = \pi h^2$ implies

$$\left(\frac{K(h)}{\pi}\right)^{1/2} = h.$$

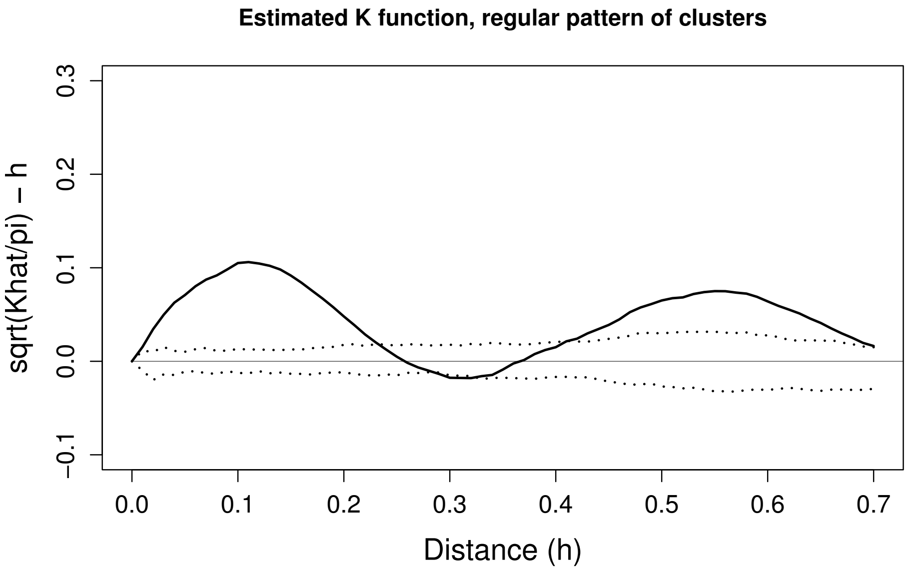
- Besag (1977) suggests plotting

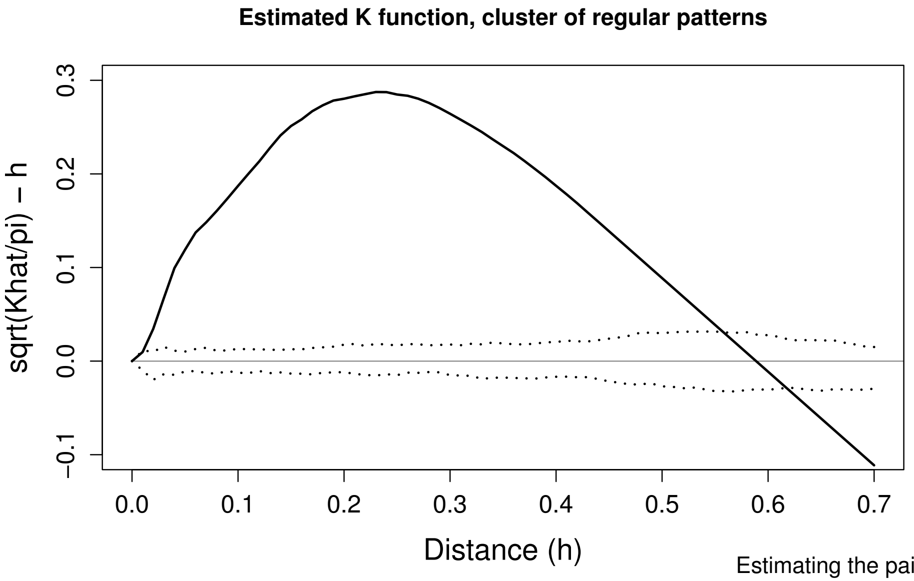
$$h \text{ versus } \widehat{L}(h)$$

where

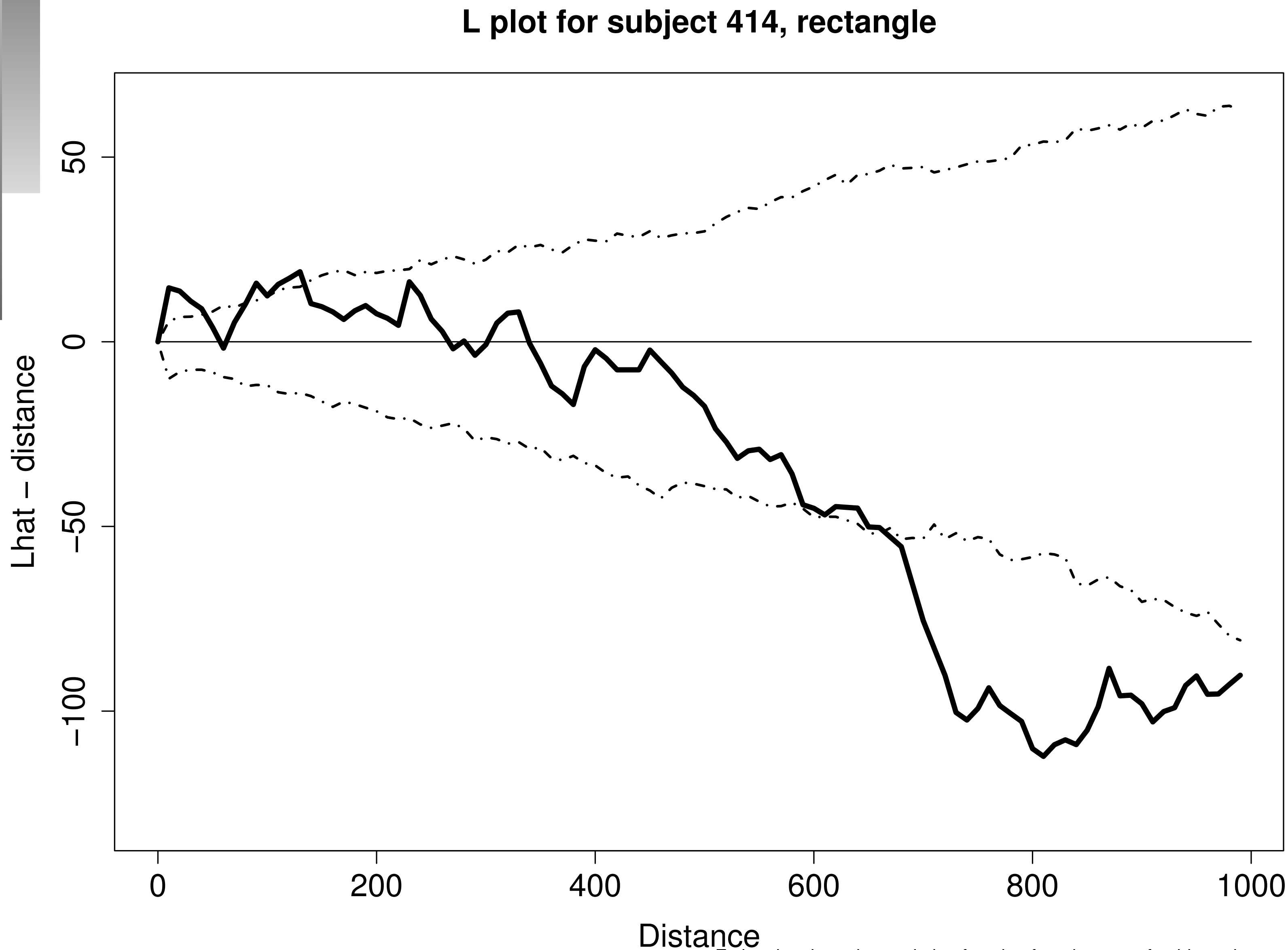
$$\widehat{L}(h) = \left(\frac{\widehat{K}_{ec}(h)}{\pi}\right)^{1/2} - h$$

Clusters of regular points...





K function for Subject 414



Pair correlation function

Note that $K(h)$ measures the *cumulative* amount of clustering/regularity up to distance h.

What about the *instantaneous* amount of clustering at distance h? Better idea of *scale* of clustering.

Consider the *pair correlation function*:

$$g(h) = \frac{1}{2\pi h}\frac{dK(h)}{dh}$$

Estimation of $g(h)$

Fiksel (1988, *Statistics*) proposed an edge-corrected estimator

$$\tilde{g}(h) = \frac{1}{2\pi h} \sum_{i} \sum_{j \neq i} \frac{k_h(||x_i - x_j|| - h)}{|W_{x_i} \cap W_{x_j}|}, h > 0,$$

where $W_x = W + x = \{y : y = z + x, z \in W\}$,
and $k_h(\cdot)$ is the Epanechnikov kernel,

$$k(s) = (1 - s^2/5)\frac{3}{4\sqrt{5}}, |s| \leq \sqrt{5}.$$

More estimation of $g(h)$

`spatstat` library for `R` (A. Baddeley and R. Turner)

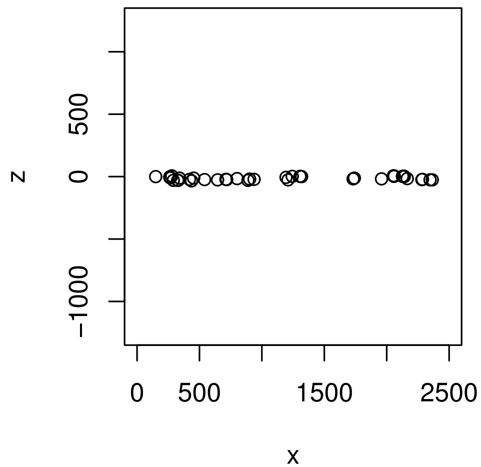
- Step 1: Estimate $K(h)$ via Ripley's correction.
- Fit smoothing spline to $\widehat{K}_{ec}(h)$ via `Kest`.
- Smoothing spline provides derivative (hence $g(h)$).
- From documentation for `Kest` function: "For a rectangular window it is prudent to restrict the r values to a maximum of 1/4 of the smaller side length of the rectangle. Bias may become appreciable for point patterns consisting of fewer than 15 points."

Initial thoughts

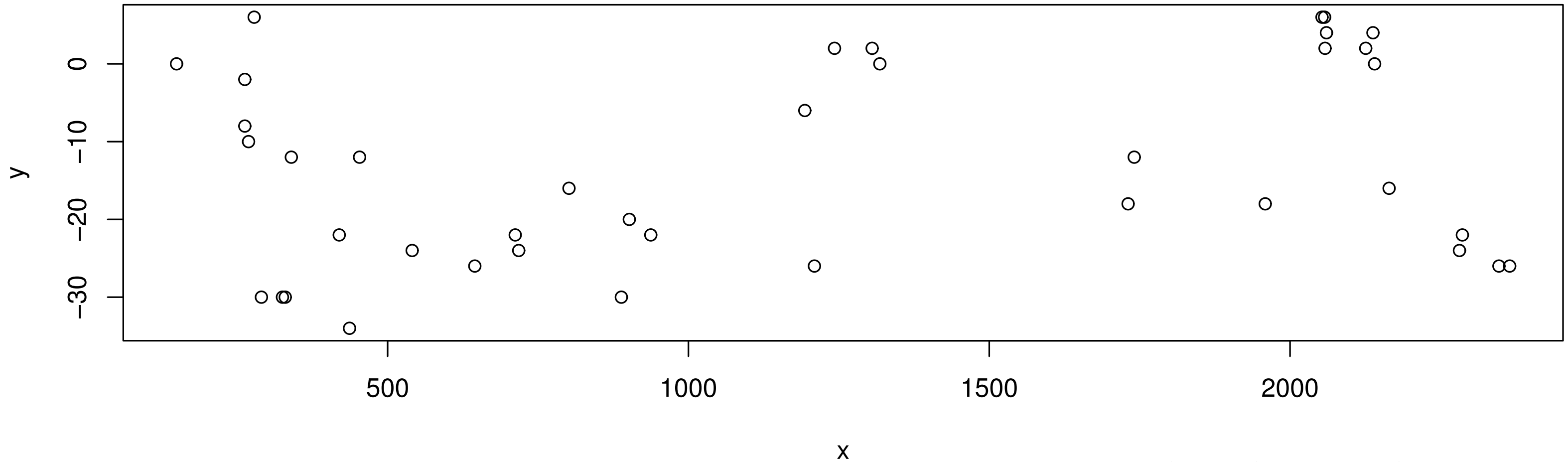
- Which do we expect to work better?
- Subject 414: x range: ≈ 2200, y range: 40.
- `spatstat` requires accuracy of `Kest` and of smoothing spline (control smoothness through spline parameters).
- Fiksel requires accuracy of kernel estimate (control smoothness through bandwidth).

Subject 414: Data

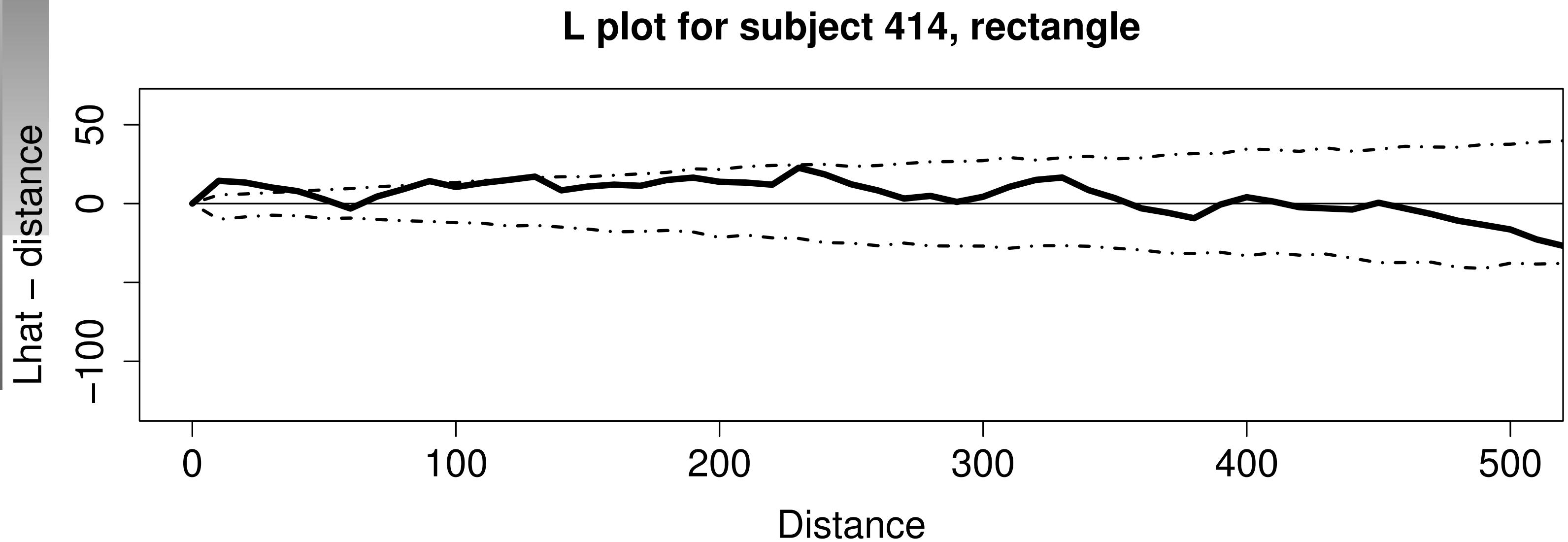
Subject 414 point pattern



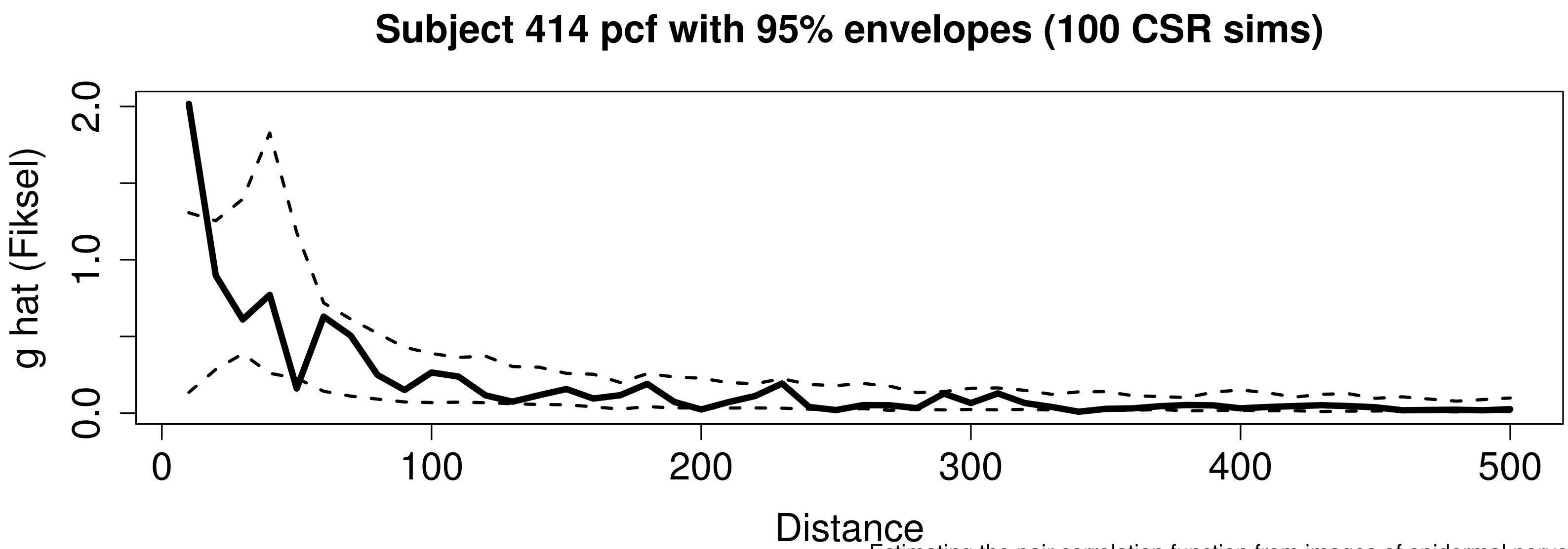
Subject 414 point pattern x z



Subject 414

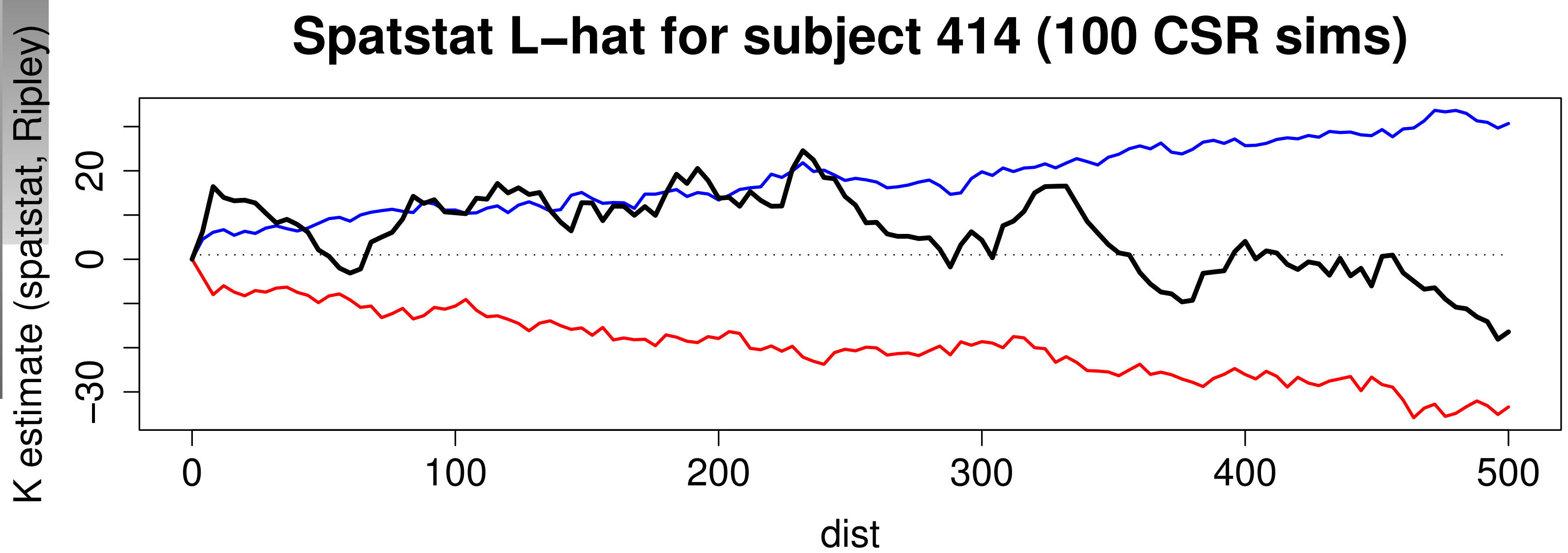


L plot for subject 414, rectangle
Lhat – distance
50
0
−100
0
100
200
300
400
500
Distance

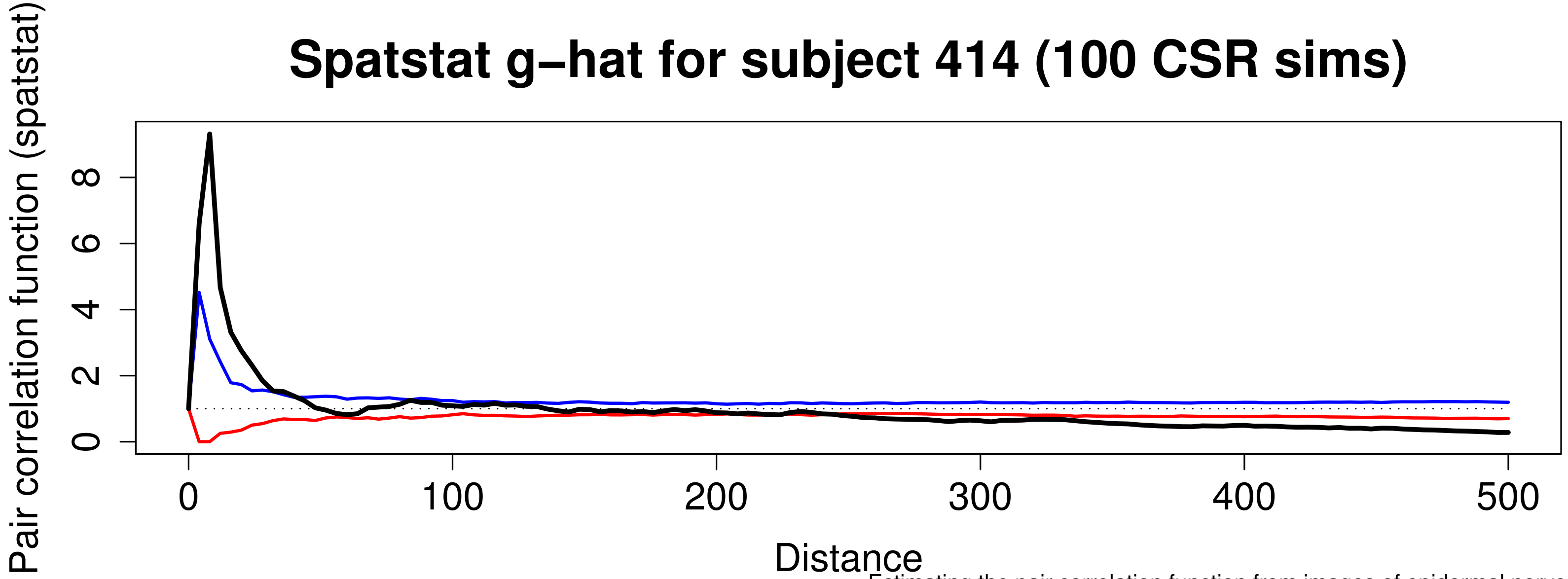


Subject 414 pcf with 95% envelopes (100 CSR sims)
g hat (Fiksel)
0.0
1.0
2.0
0
100
200
300
400
500
Distance

Subject 414

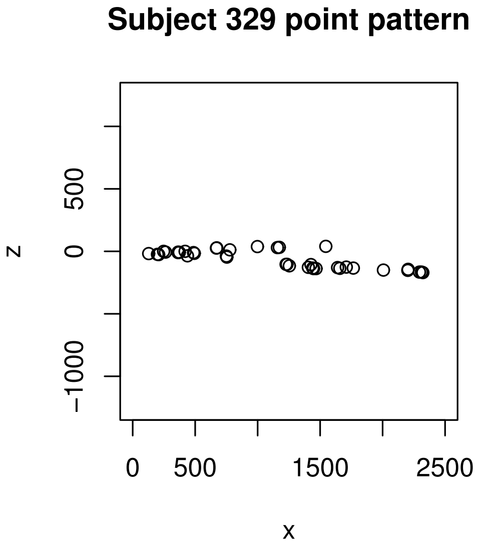


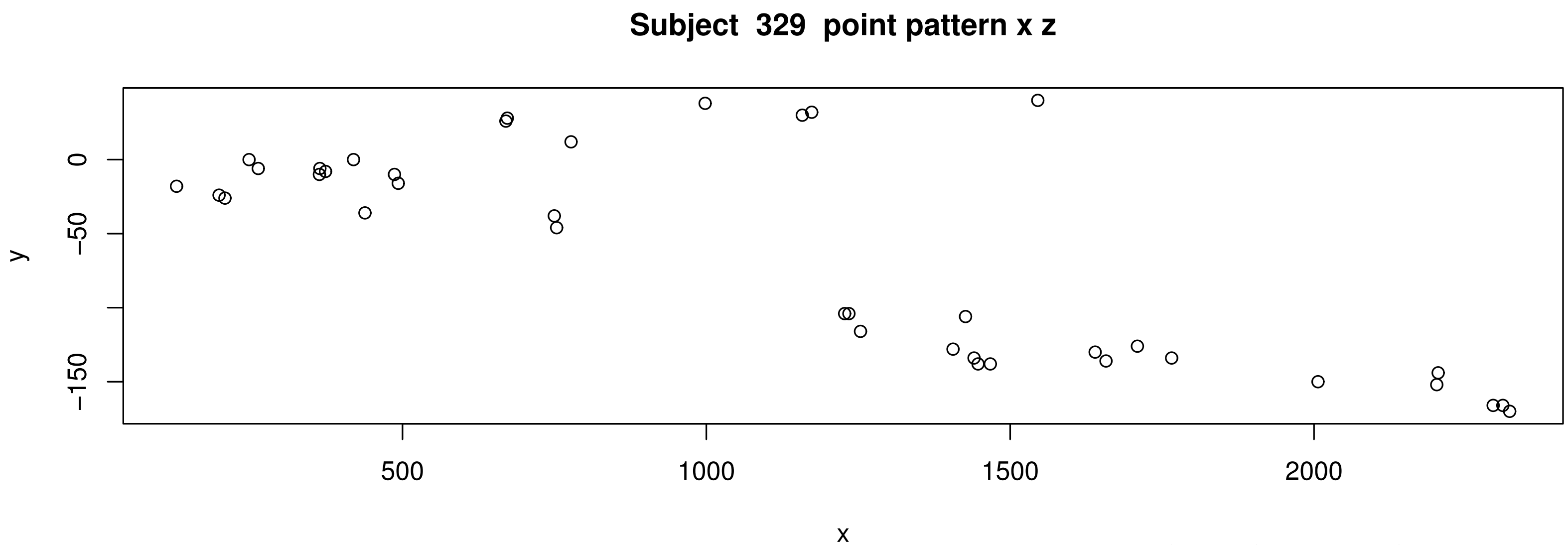
Spatstat L−hat for subject 414 (100 CSR sims)
K estimate (spatstat, Ripley)
20
0
−30
0
100
200
300
400
500
dist



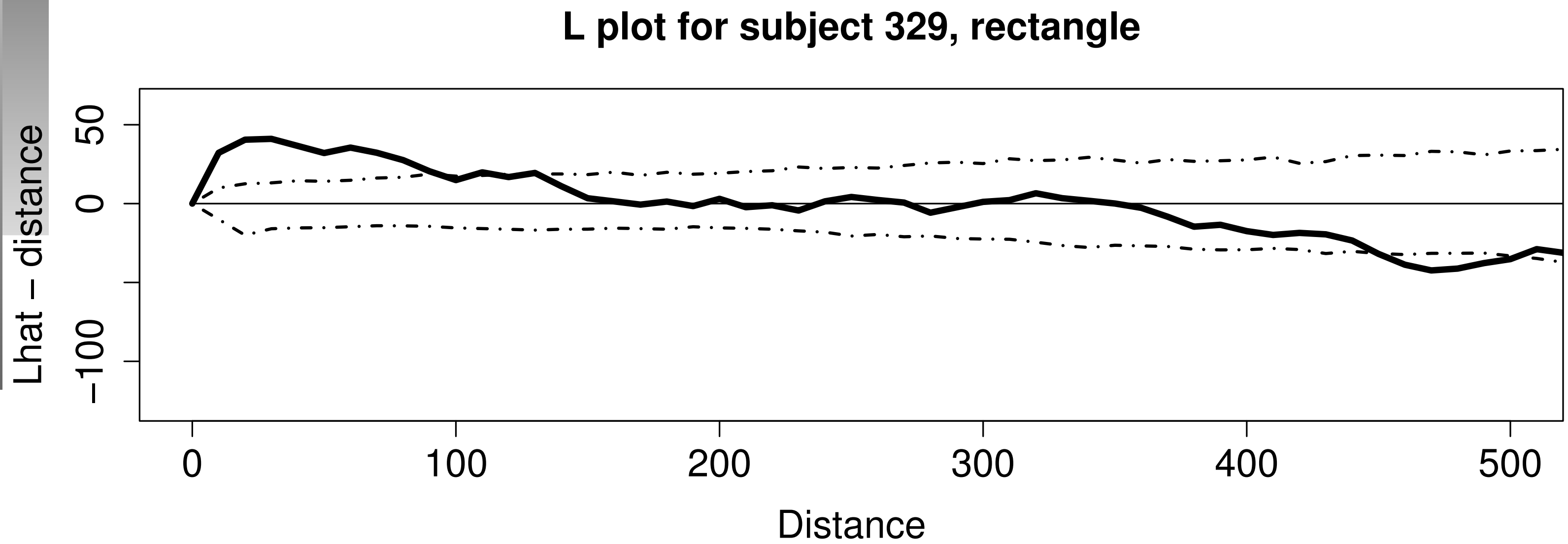
Spatstat g−hat for subject 414 (100 CSR sims)
Pair correlation function (spatstat)
8
6
4
2
0
0
100
200
300
400
500
Distance

Subject 329: Data

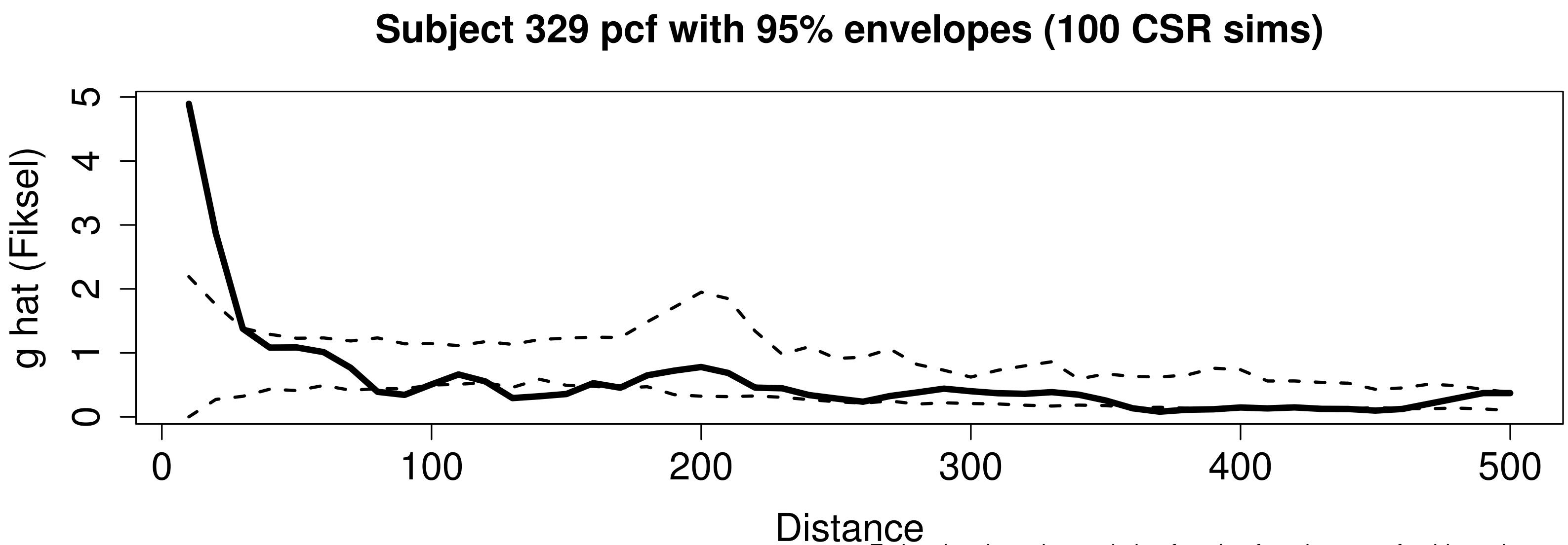




Subject 329

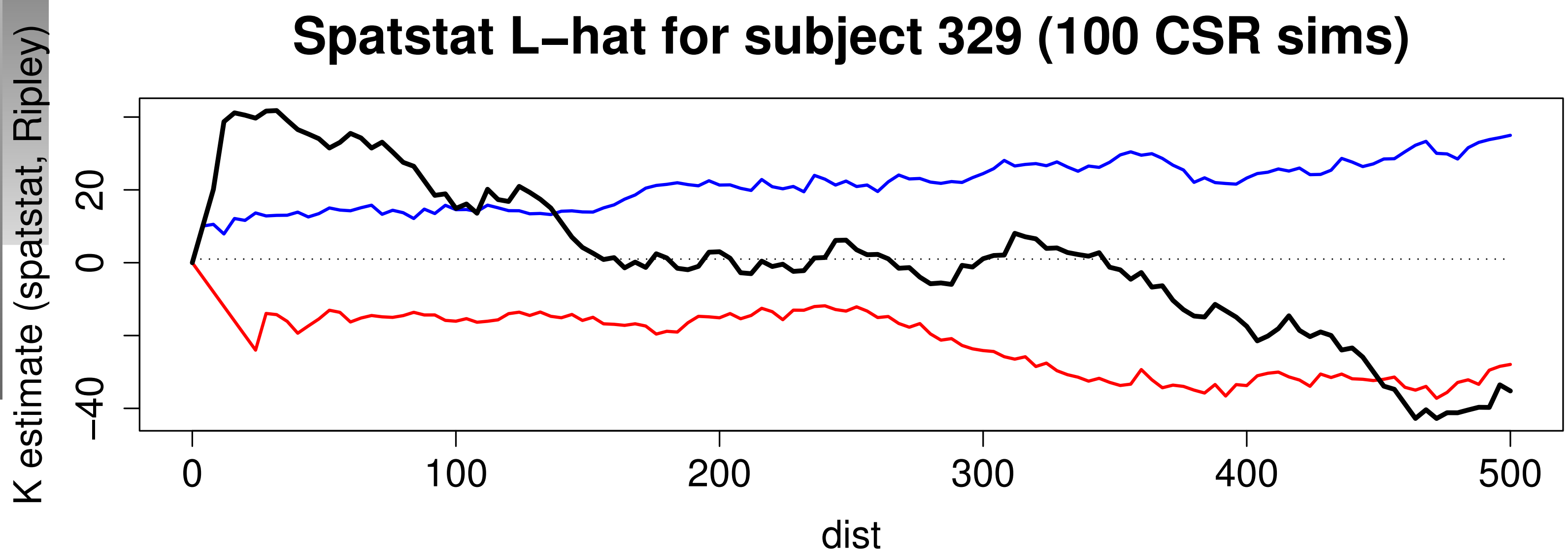


L plot for subject 329, rectangle
Lhat – distance
50
0
−100
0
100
200
300
400
500
Distance

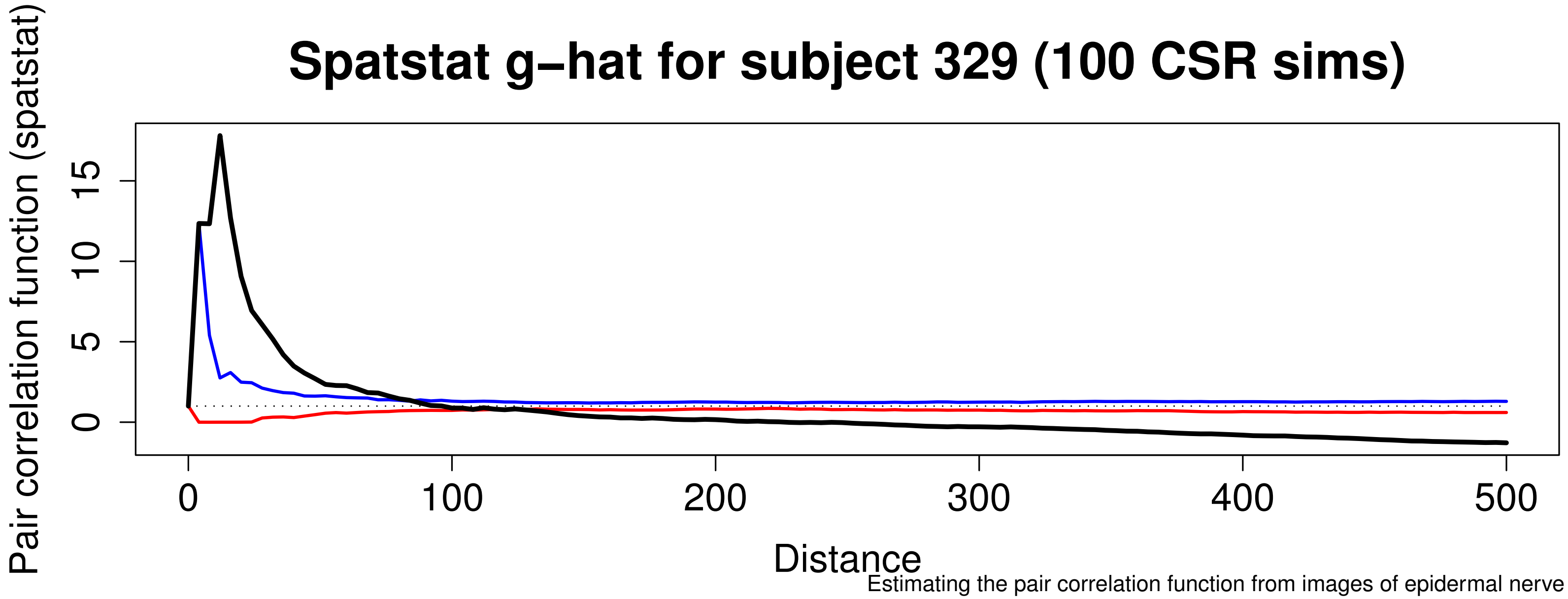


Subject 329 pcf with 95% envelopes (100 CSR sims)
g hat (Fiksel)
0
1
2
3
4
5
0
100
200
300
400
500
Distance

Subject 329

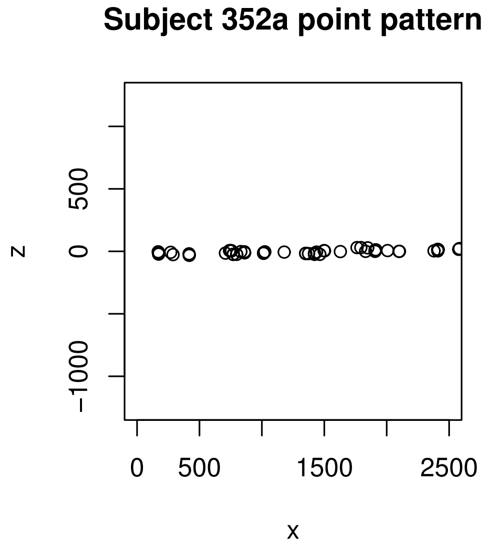


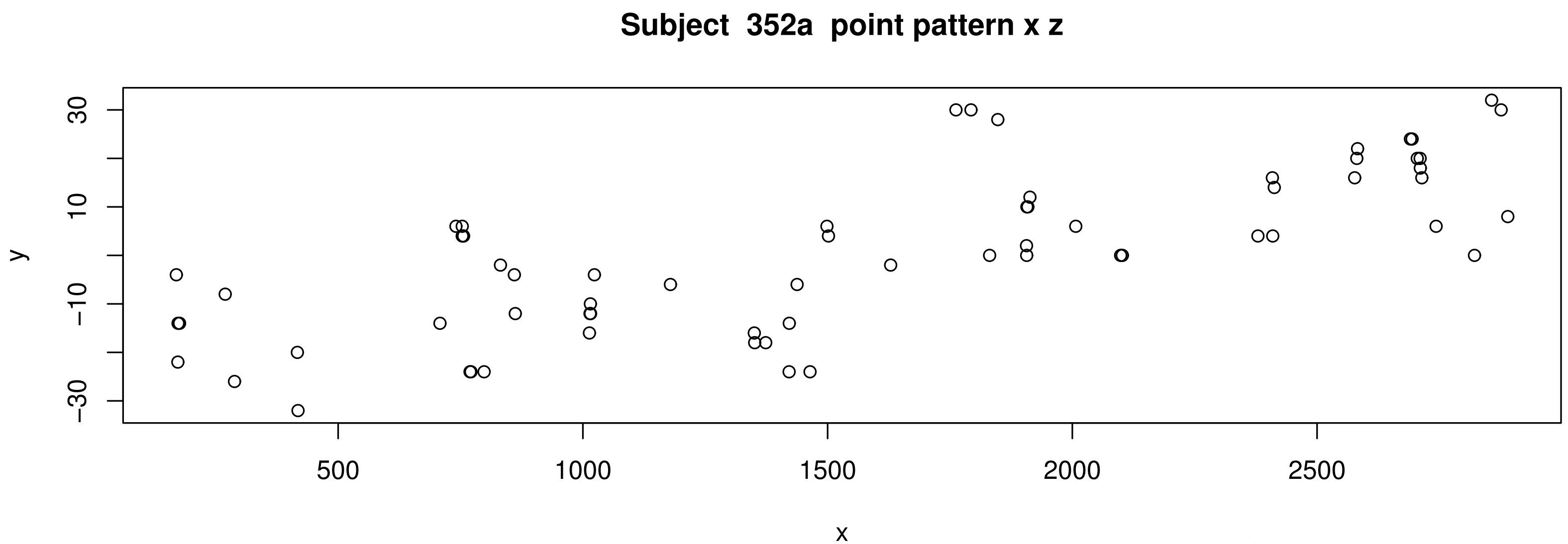
Spatstat L−hat for subject 329 (100 CSR sims)
K estimate (spatstat, Ripley)
−40
0
20
0
100
200
300
400
500
dist



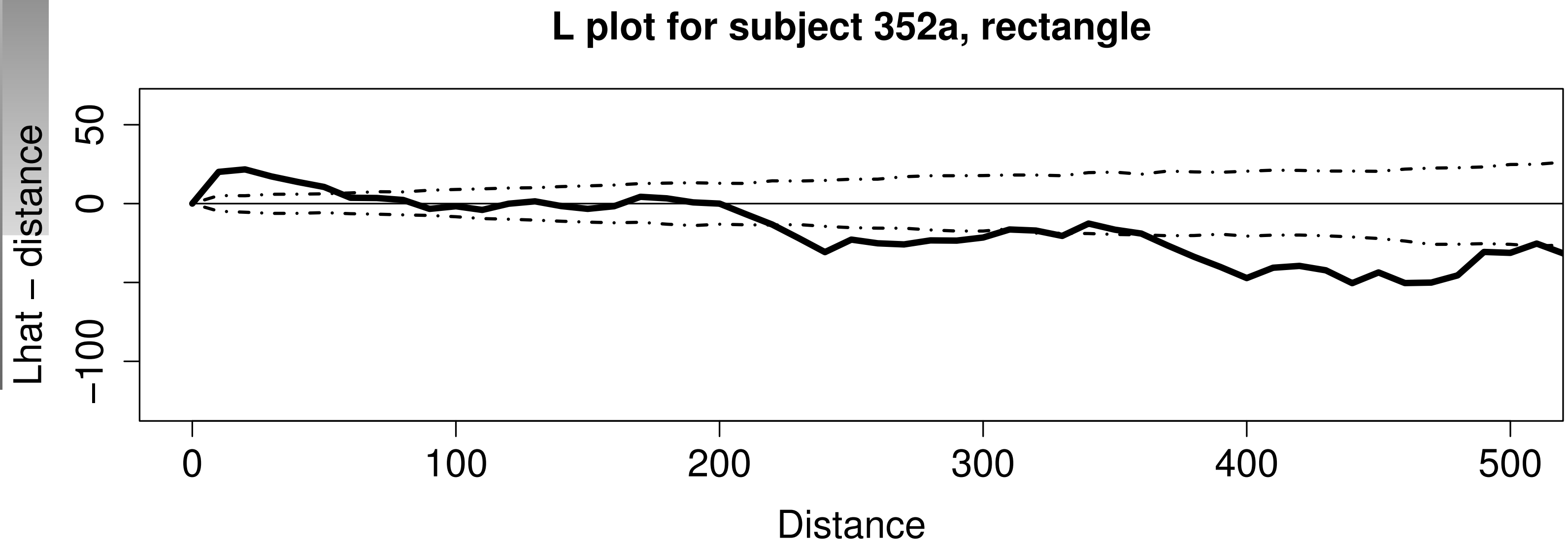
Spatstat g−hat for subject 329 (100 CSR sims)
Pair correlation function (spatstat)
0
5
10
15
0
100
200
300
400
500
Distance

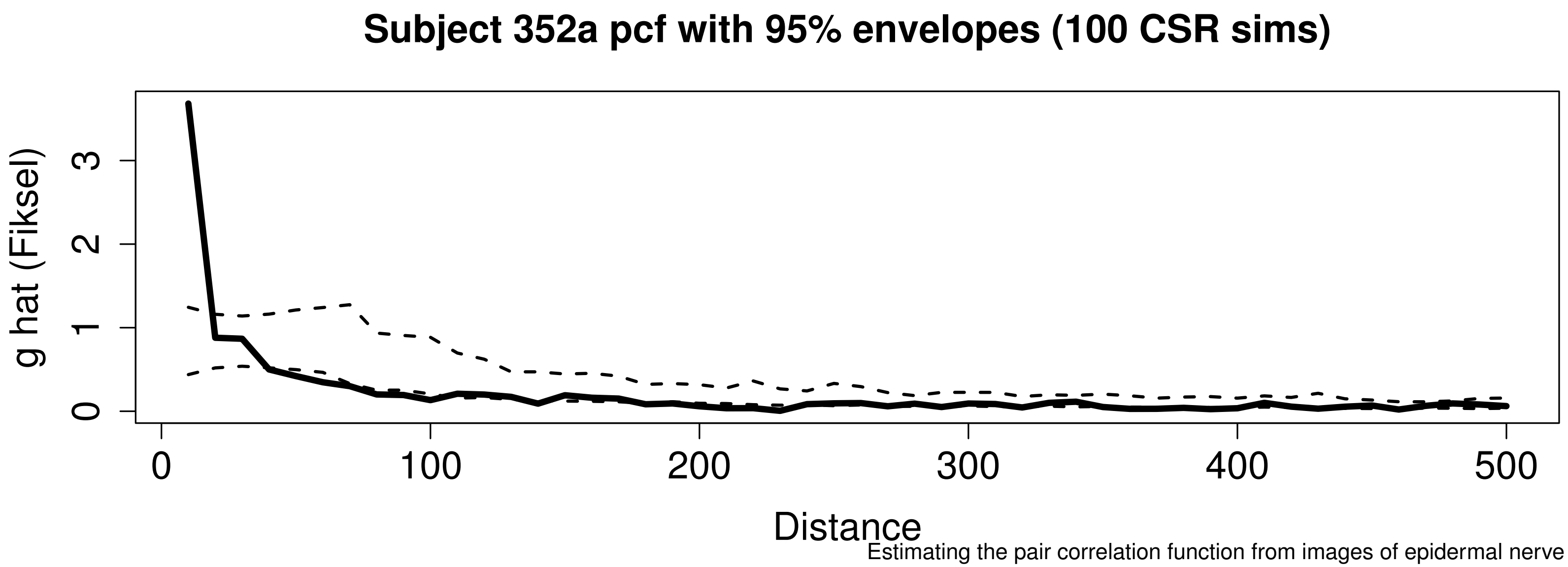
Subject 352a: Data



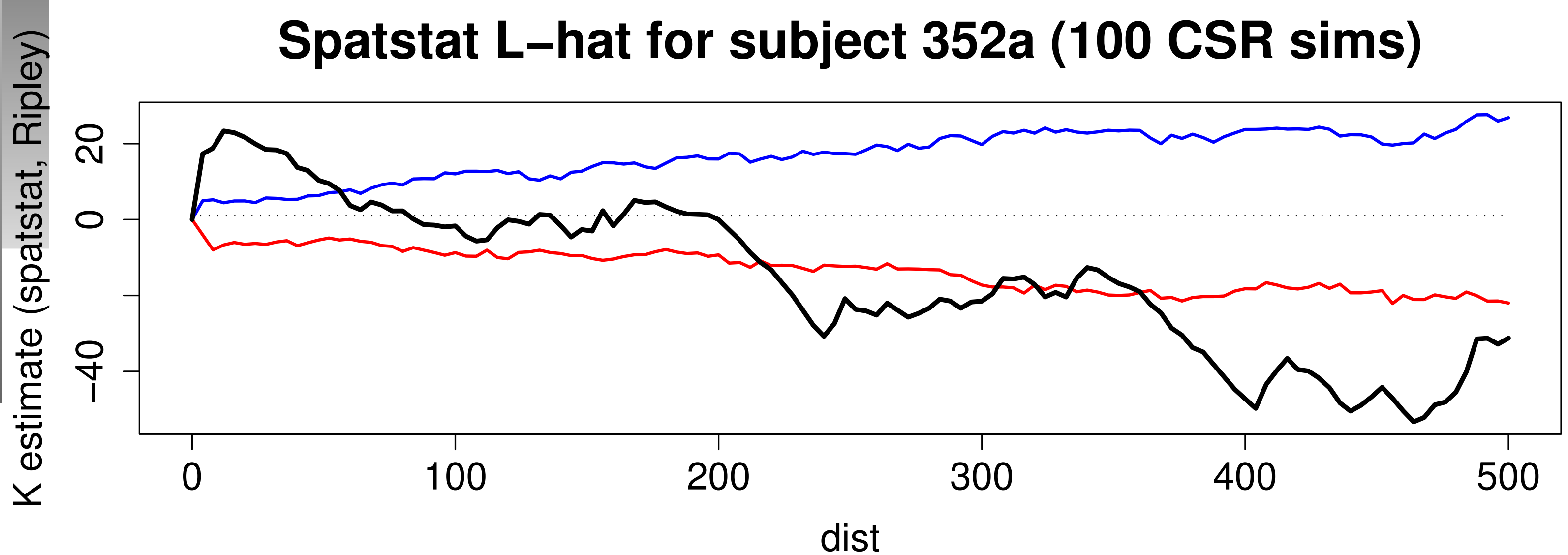


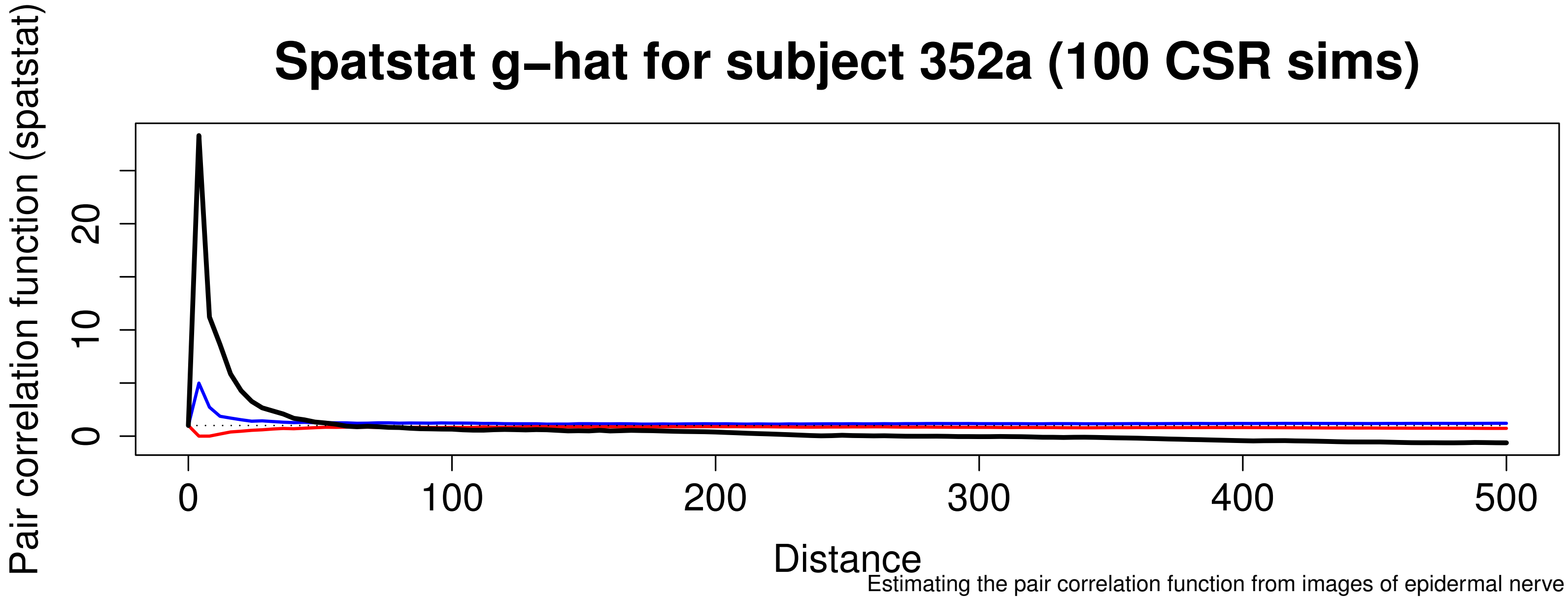
Subject 352a



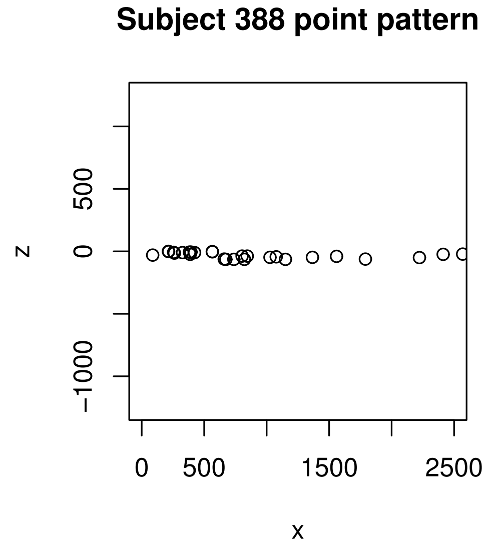


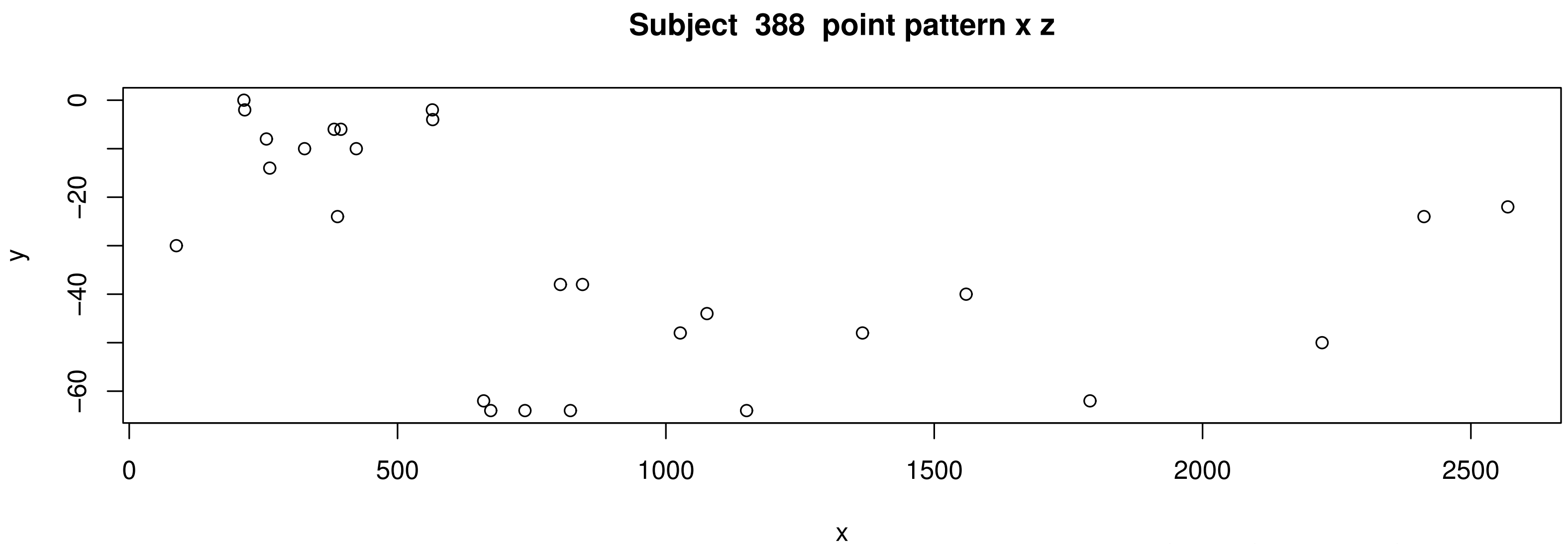
Subject 352a



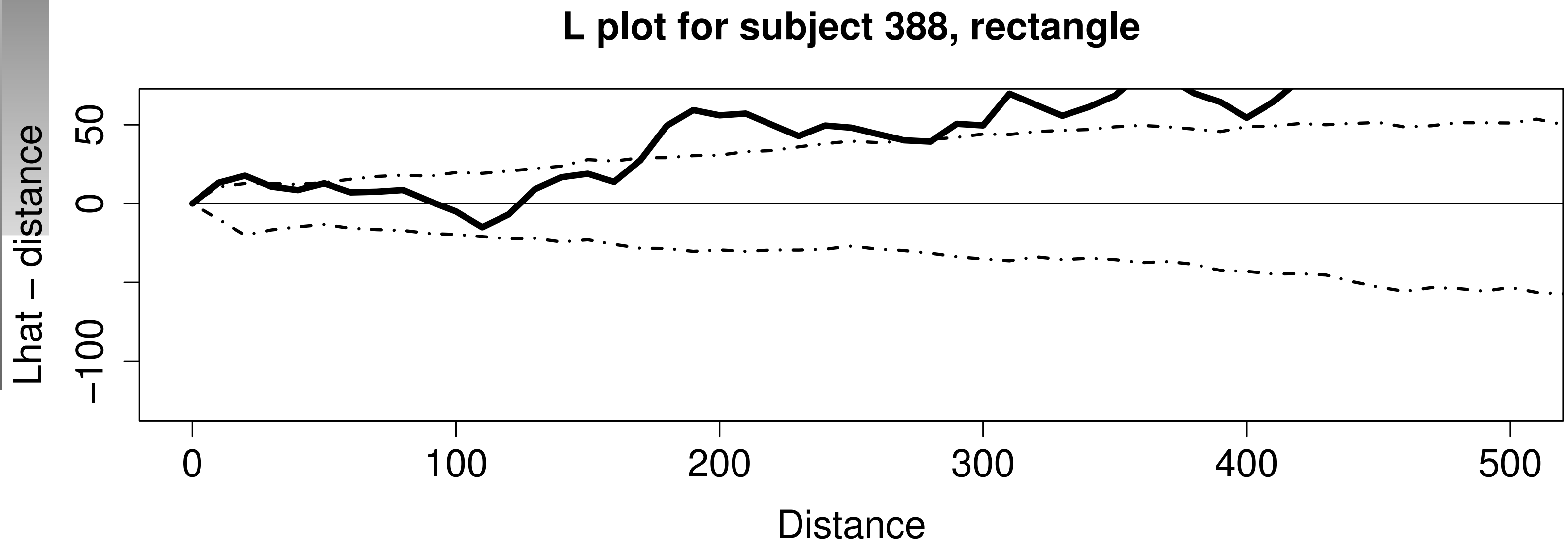


Subject 388: Data

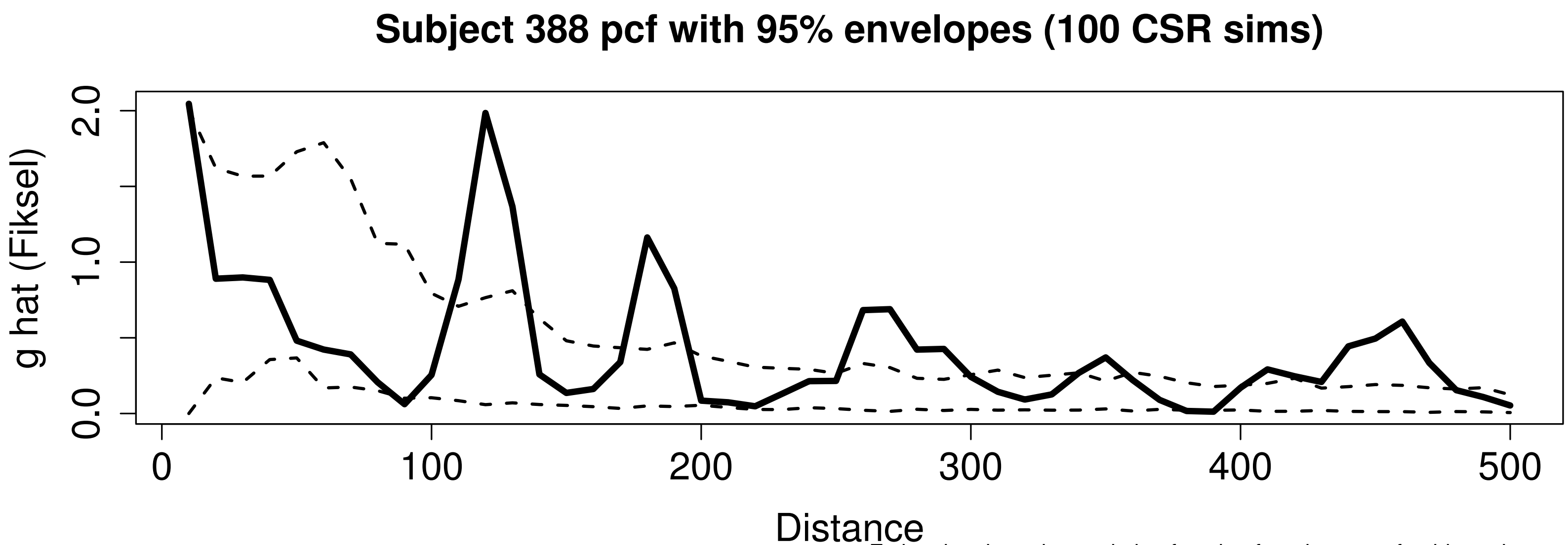




Subject 388

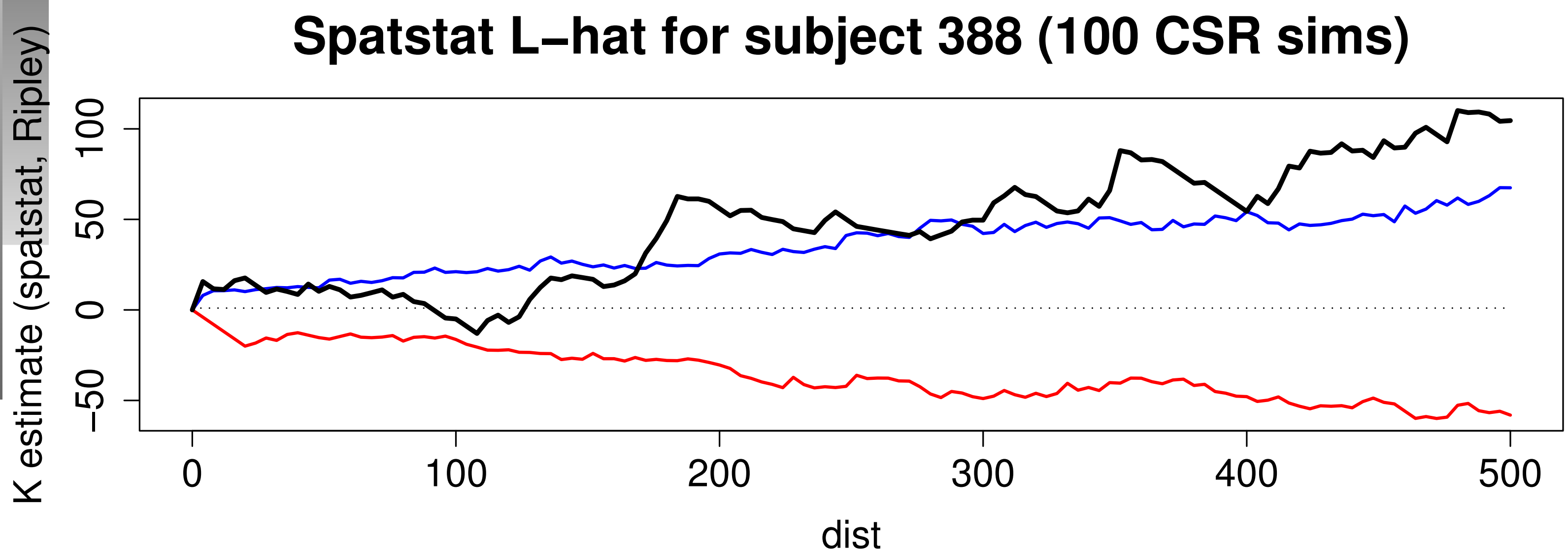


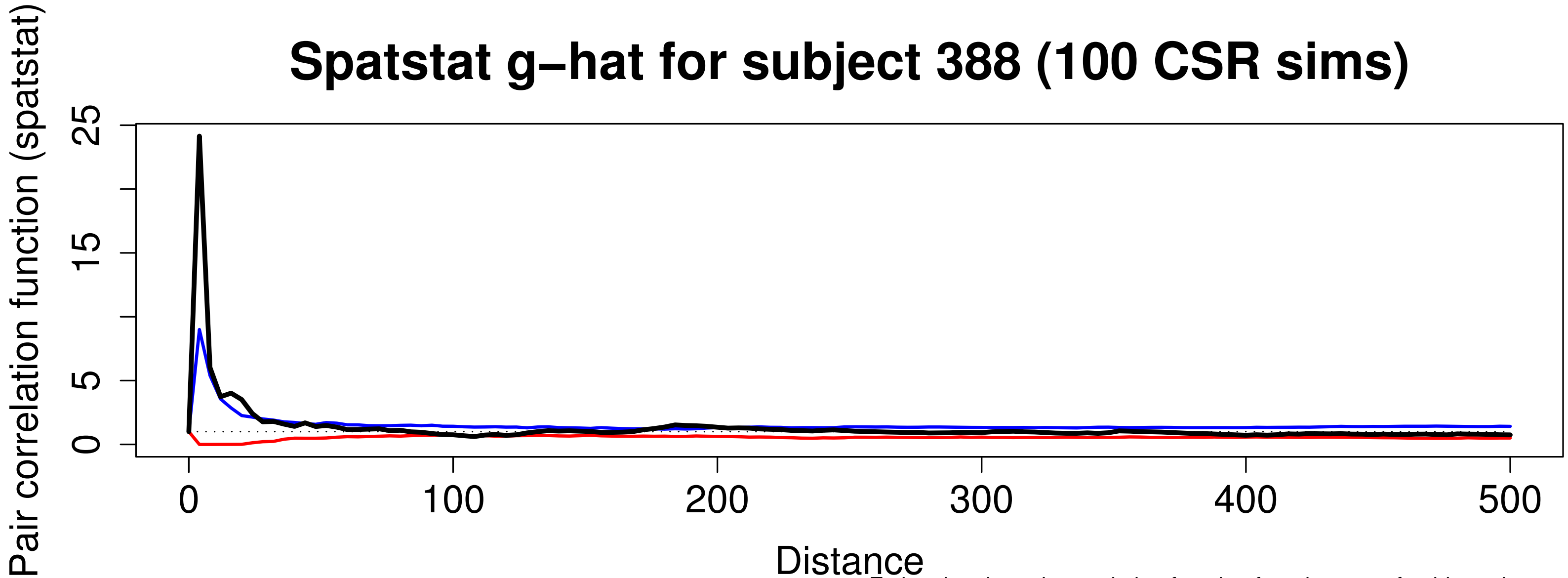
L plot for subject 388, rectangle
Lhat – distance
−100
0
50
0
100
200
300
400
500
Distance



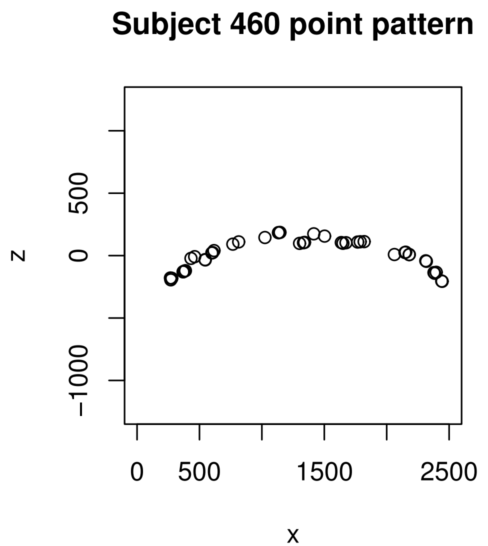
Subject 388 pcf with 95% envelopes (100 CSR sims)
g hat (Fiksel)
0.0
1.0
2.0
0
100
200
300
400
500
Distance

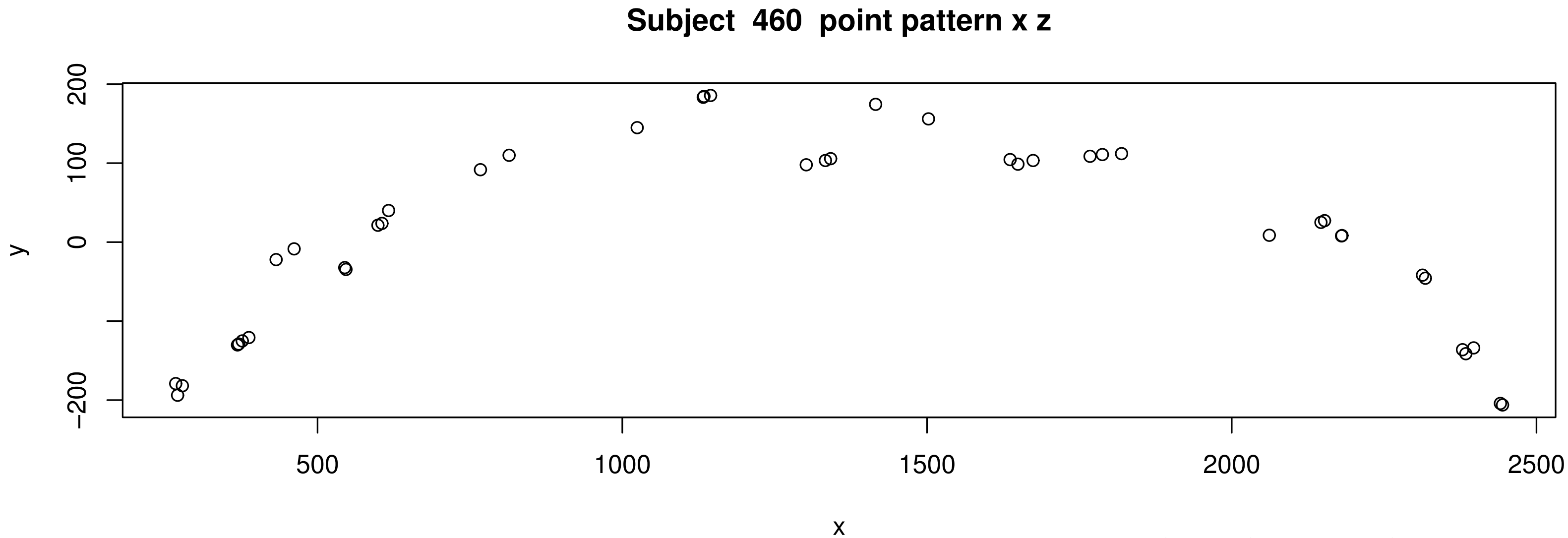
Subject 388


Spatstat L−hat for subject 388 (100 CSR sims)
K estimate (spatstat, Ripley)
−50
0
50
100
0
100
200
300
400
500
dist

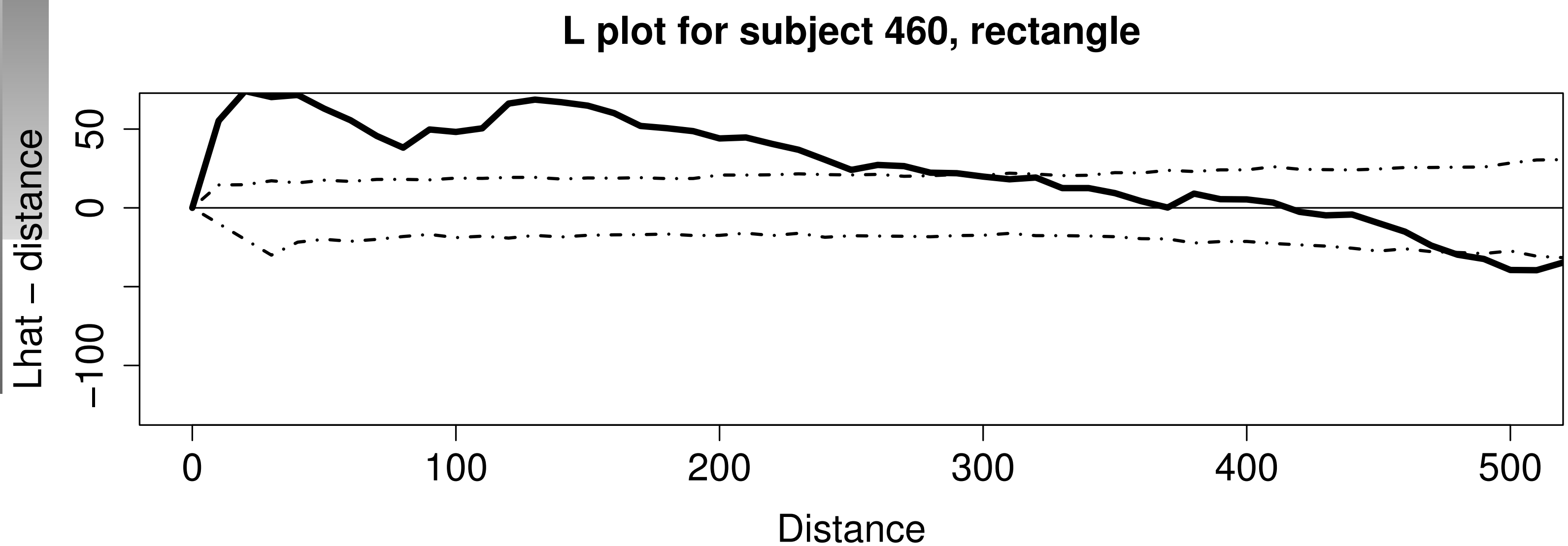

Spatstat g−hat for subject 388 (100 CSR sims)
Pair correlation function (spatstat)
0
5
15
25
0
100
200
300
400
500
Distance

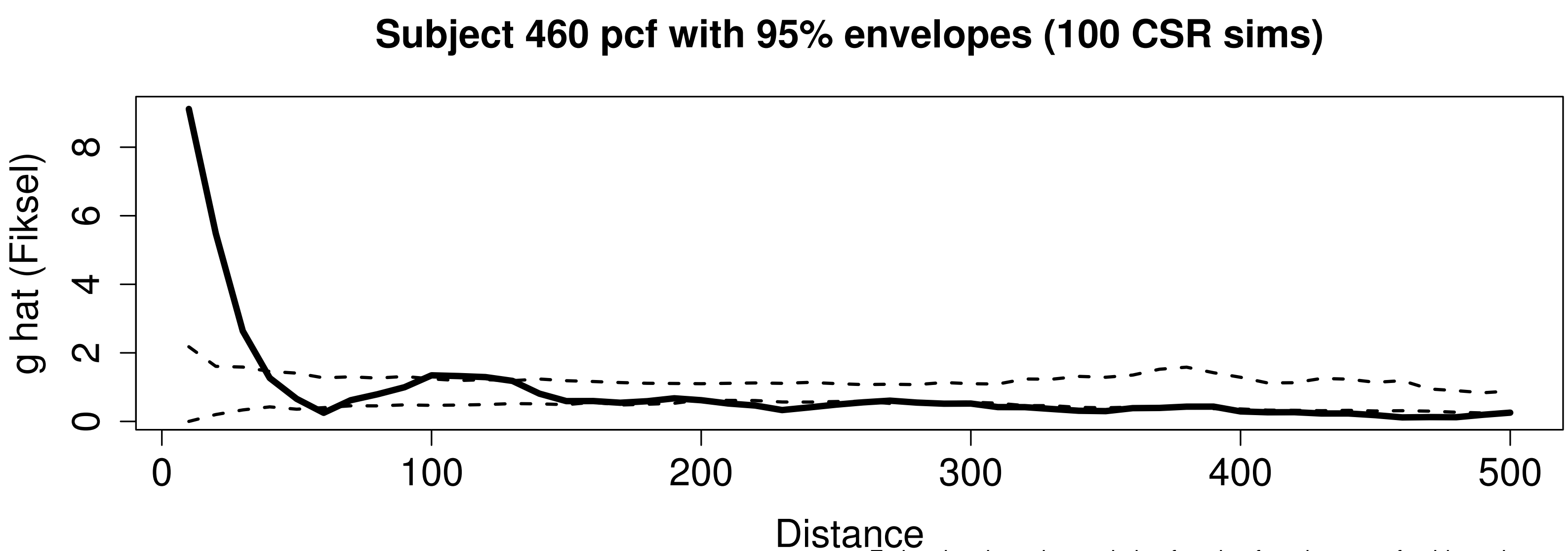
Subject 460: Data



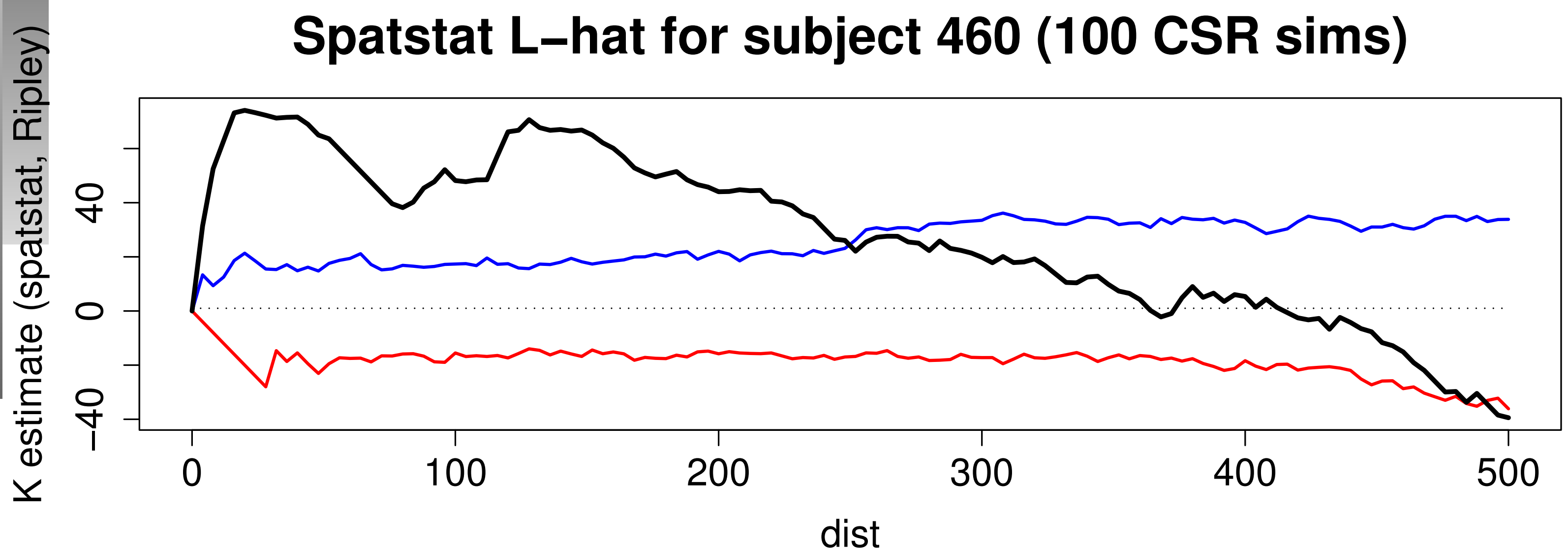


Subject 460

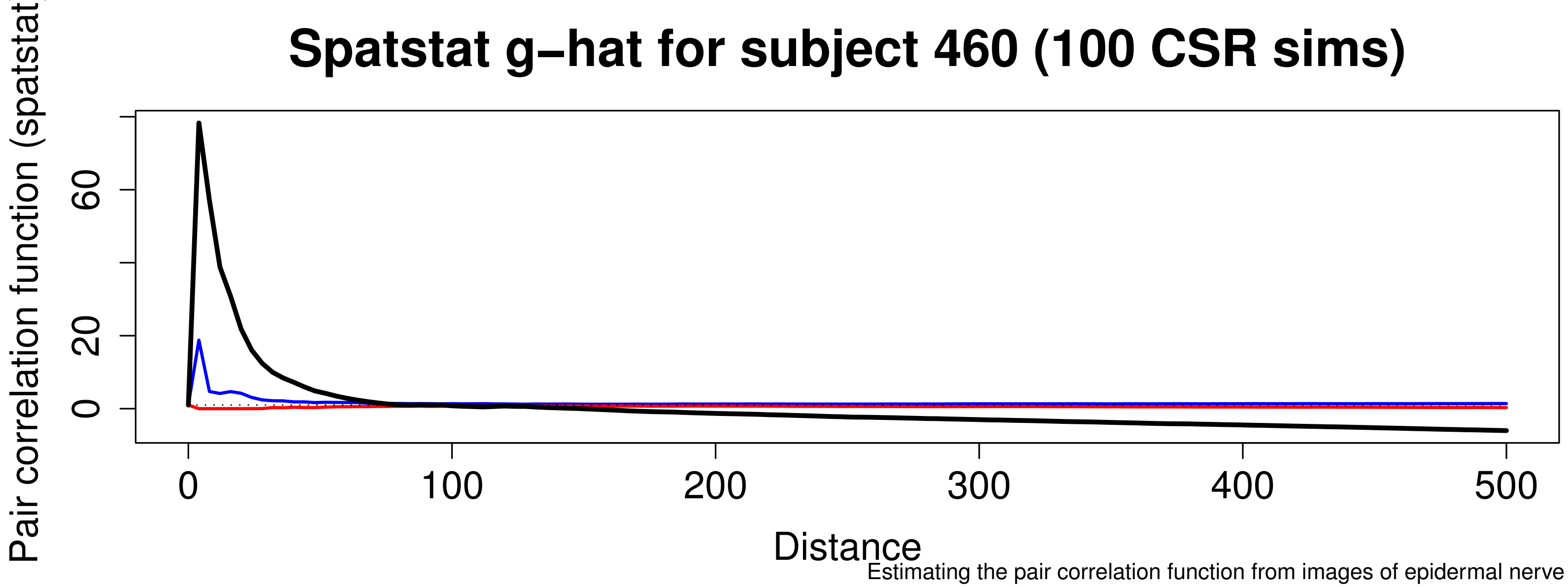

L plot for subject 460, rectangle
Lhat – distance
−100
0
50
0
100
200
300
400
500
Distance


Subject 460 pcf with 95% envelopes (100 CSR sims)
g hat (Fiksel)
0
2
4
6
8
0
100
200
300
400
500
Distance

Subject 460



Spatstat L−hat for subject 460 (100 CSR sims)
K estimate (spatstat, Ripley)
−40
0
40
0
100
200
300
400
500
dist



Spatstat g−hat for subject 460 (100 CSR sims)
Pair correlation function (spatstat)
0
20
60
0
100
200
300
400
500
Distance

Comments

- Clear short distance clustering in all cases.
- My Fiksel code is suspicious.
- Ripley's edge-correction seems to yield stable $\widehat{K}_{ec}(h)$.
- Two problems: sample size and edge effects.
- Subject patterns differ from CSR (but we suspect this in *healthy* patients).
- What kind of pattern is observed in healthy patients?

Conclusions

- Add more images per site per patient (intra-patient variability).
- Add more patients (diseased, non-diseased, inter-patient variability).
- Quantify “scale of clustering”.

- Supported by U.S. NIH, NINDS grant 1-R21 NS46258-01