Bandwidth selection for kernel estimators of spatial intensity functions

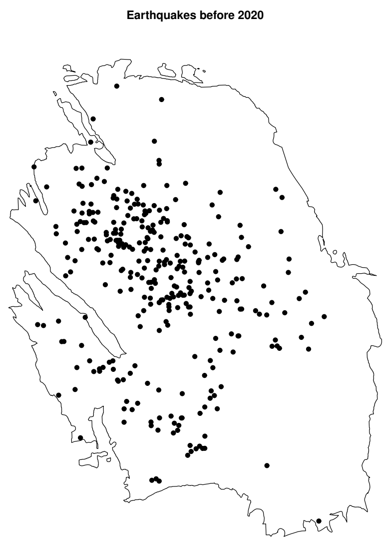
M.N.M. van Lieshout

colette@cwi.nl

CWI & University of Twente
The Netherlands

Moment measure of a point process

A realisation of a point process Φ on $\mathbb{R}^d$ is a **pattern:** an **unordered set** of points such that any **bounded** set $A \subset \mathbb{R}^d$ contains only **finitely many** of them.



Let $N(A)$ be the number of points of Φ in $A \subset \mathbb{R}^d$ and write

$$M(A) = \mathbb{E}N(A),$$

the **expected number of points in** A.

Intensity function

Often

$$M(A) = \int_A \lambda(x)\,dx$$

for some function $\lambda(x) \geq 0$, the **intensity function** of Φ.

Goal: estimate λ based on a realisation $\Phi \cap W$ in a bounded Borel set W (assumed to be open and non-empty).

For $x_0 \in W$, set (Berman and Diggle, 1985, 1989)

$$\widehat{\lambda_{BD}(x_0; h, \Phi, W)} := \frac{N(b(x_0, h) \cap W)}{|b(x_0, h) \cap W|}$$

where $b(x_0, h)$ is the closed ball around x_0 with radius h and $|\cdot|$ denotes area. The **bandwidth** parameter $h > 0$ determines the smoothness.

Kernels

The box kernel may be replaced by any kernel (symmetric pdf), e.g. the **Gaussian kernel**

$$\kappa^{\infty}(x) = \frac{1}{(2\pi)^{d/2}} \exp\left(-x^T x/2\right), \quad x \in \mathbb{R}^d,$$

or the **Beta kernel**

$$\kappa^{\gamma}(x) = \frac{1}{c(d,\gamma)} (1 - x^T x)^{\gamma}\, 1\{x \in b(0,1)\}, \quad x \in \mathbb{R}^d,$$

for $\gamma \geq 0$, where

$$c(d,\gamma) = \int_{b(0,1)} (1 - x^T x)^{\gamma} dx = \frac{\pi^{d/2}\Gamma(\gamma+1)}{\Gamma(d/2+\gamma+1)}, \quad d \in \mathbb{N}, \gamma \geq 0.$$

Beta kernels are **compactly supported**, the box kernel has $\gamma = 0$. For $\gamma > k$, κ^{γ} is C^k.

Selecting the bandwidth

- Let Φ a **stationary, isotropic Cox process** driven by Λ. For the box kernel in $\mathbb{R}^2$ and $w_h \equiv 1$, minimise

$$\mathbb{E}\left[\{\lambda(0;\widehat{h,\Phi,}W) - \Lambda(0)\}^2\right] =$$

$$\rho^{(2)}(0)+\frac{\lambda^2}{\pi^2 h^4}\int_0^{2h}\left\{2h^2\arccos\left(\frac{t}{2h}\right)-\frac{t}{2}(4h^2-t^2)^{1/2}\right\}dK(t)+\lambda\frac{1-2\lambda K(h)}{\pi h^2}$$

where $\lambda K(h) = \mathbb{E}\left[N(b(0,h)|0 \in \Phi\right]$.

(Diggle, 1985)

- Let Φ be an **inhomogeneous Poisson process**. Maximise the leave-one-out cross-validation log likelihood

$$\sum_{x\in\Phi\cap W}\log\lambda(x;\widehat{h,\Phi\setminus\{x\}},W) - \int_W \lambda(u;\widehat{h,\Phi,}W)\,du.$$

Non-parametric bandwidth selection

By the Campbell–Mecke formula

$$\mathbb{E}\left\{\sum_{x\in\Phi\cap W}\frac{1}{\lambda(x)}\right\}=\int_W\frac{1}{\lambda(x)}\lambda(x)\,dx=|W|$$

so minimise the discrepancy between $|W|$ and

$$T_\kappa(h;\Phi,W)=\begin{cases}\displaystyle\sum_{x\in\Phi\cap W}\frac{1}{\widehat{\lambda}(x;h,\Phi,W)}, & \Phi\cap W\neq\emptyset,\\ |W|, & \text{otherwise.}\end{cases}$$

(Cronie and Van Lieshout, 2018)

No model assumptions required!

Computationally straightforward, no numerical approximation of integrals.

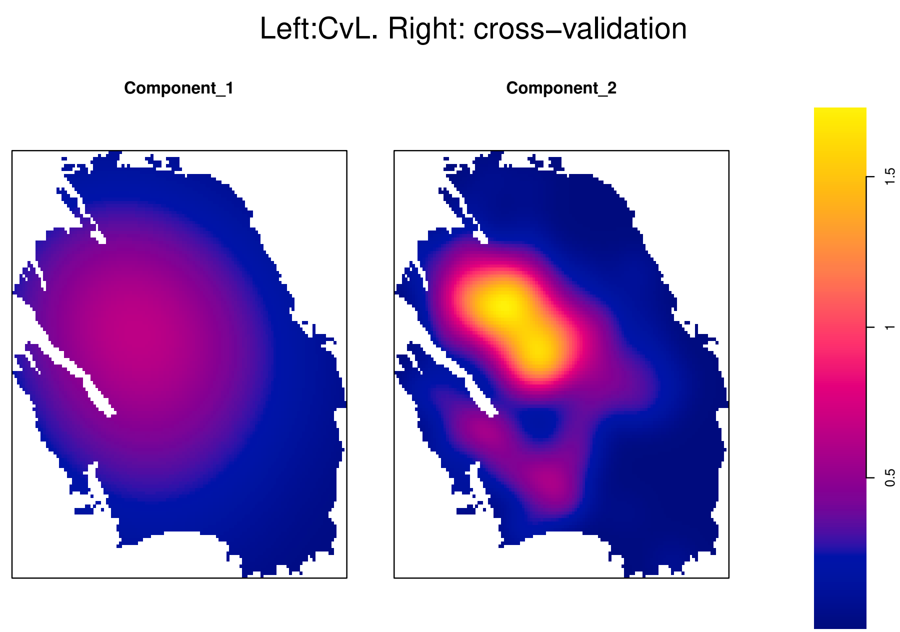
Conclusions

Based on a simulation study, we reach the following conclusions.

- For **clustered** patterns with a moderate number of points, the new method performs the best.
- For **Poisson** processes with a moderate number of points, likelihood based cross-validation performs the best.
- For **regular** patterns with a moderate number of points, the new and the likelihood-based methods give good results.
- For large patterns, the Diggle method seems best.

For details:
O. Cronie and M.N.M. van Lieshout. A non-model based approach to bandwidth selection for kernel estimators of spatial intensity functions. *Biometrika* 105:455–462, 2018.

Back to Groningen

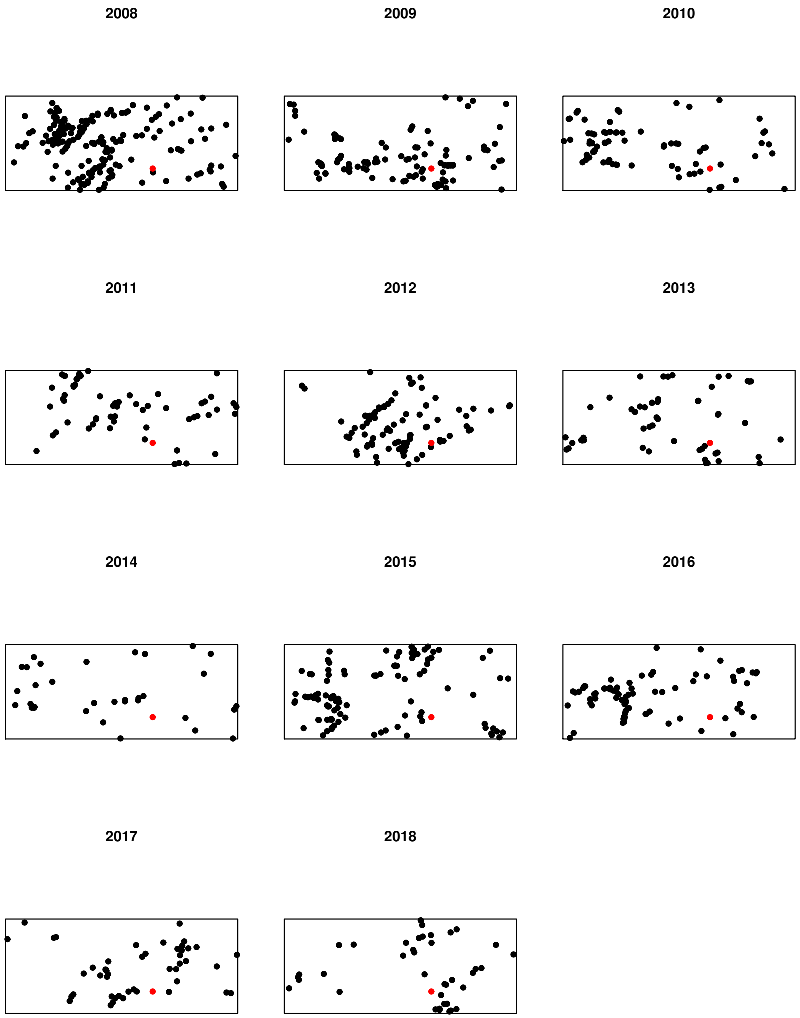


Asymptotic theory: which way to infinity?

Ripley (1988) discusses two asymptotic regimes.

- **Increasing domain:** $W_n \to \mathbb{R}^d$. Not applicable
  - when the point process is defined on a fixed domain;
  - unless strong ergodicity conditions are imposed such as stationarity.
- **Infill asymptotics:** replicated patterns in the same window.

Example: tornadoes in Kansas



Tornadoes in Kansas during the Spring seasons of 2008–2018.

Complications

If each replicate contains **a single point**,

- the union is a Poisson process;
- classic probability density estimation results apply;
- asymptotics are in terms of the number of points.

(Lo, 2017).

In general, however,

- λ is not normalised;
- the number of points is random;
- and their locations are not necessarily independent.

Infill asymptotics regime

Let $\Phi_1, \Phi_2, \ldots$ be i.i.d. simple point processes observed in a bounded open subset $\emptyset \neq W \subset \mathbb{R}^d$ with **intensity function** λ and **pair correlation function** g.

Write

$$Y_n = \bigcup_{i=1}^{n} \Phi_i$$

and set

$$\widehat{\lambda_n(x_0)} := \frac{\lambda(x_0; \widehat{h_n, Y_n}, W)}{n} = \frac{1}{n} \sum_{i=1}^{n} \lambda(x_0; \widehat{h_n, \Phi_i}, W).$$

Side remark: the bandwidths may differ per component. If so, replace h_n by a diagonal matrix H_n.

Mean squared error

Assume that

- product densities exist up to second order;
- $\lambda > 0$ on W, so g is well-defined.

Then, for a Beta kernel κ^γ, $\gamma \geq 0$,

$$\begin{aligned} \mathrm{mse}\,\widehat{\lambda_n(x_0)} &= \left(\frac{1}{h^d} \int_{b(x_0,h)\cap W} \kappa^\gamma \left(\frac{x_0 - u}{h} \right) \lambda(u)\, du - \lambda(x_0) \right)^2 \\ &+ \frac{1}{nh^{2d}} \int_{b(x_0,h)\cap W} \kappa^\gamma \left(\frac{x_0 - u}{h} \right)^2 \lambda(u)\, du \end{aligned}$$

$$+ \frac{1}{nh^{2d}} \int_{(b(x_0,h)\cap W)^2} \kappa^\gamma \left(\frac{x_0 - u}{h} \right) \kappa^\gamma \left(\frac{x_0 - v}{h} \right) (g(u,v) - 1)\, \lambda(u)\, \lambda(v)\, du\, dv.$$

Asymptotic expansion

Impose the **technical conditions**

- $h_n > 0$, $h_n \to 0$ and $nh_n^d \to \infty$ as $n \to \infty$;
- $g : W \times W \to \mathbb{R}$ is bounded;
- $\lambda : W \to (0, \infty)$ is C^2 with $\lambda_{ij} = D_{ij}\lambda$, $i, j = 1, \ldots, d$, Hőlder continuous with exponent $\alpha \in (0, 1]$ on W, that is, $\exists C > 0$ such that $\forall i, j = 1, \ldots, d$:

$$|\lambda_{ij}(x) - \lambda_{ij}(y)| \leq C||x - y||^{\alpha}, \quad x, y \in W.$$

Then, using a Beta kernel κ^{γ}, $\gamma \geq 0$,

1. $\text{bias}\,\widehat{\lambda_n(x_0)} = \frac{h_n^2 \sum_{i=1}^{d} \lambda_{ii}(x_0)}{2(d+2\,\gamma+2)} + O(h_n^{2+\alpha})$
2. $\text{Var}\,\widehat{\lambda_n(x_0)} = \frac{\lambda(x_0)\,c(d,2\gamma)}{n\,h_n^d\,c(d,\gamma)^2} + O\left(\frac{1}{nh_n^{d-1}}\right)$

as $n \to \infty$.

Asymptotically optimal bandwidth

The **mean squared error** of $\widehat{\lambda_n(x_0)}$ can be expanded as

$$h_n^4 \frac{\left(\sum_{i=1}^d \lambda_{ii}(x_0)\right)^2}{4\,(d+2\gamma+2)^2} + \frac{1}{n\,h_n^d} \frac{\lambda(x_0)\,c(d,2\gamma)}{c(d,\gamma)^2} + O\left(h_n^{4+\alpha}\right) + O\left(\frac{1}{n\,h_n^{d-1}}\right).$$

Hence, provided $\sum_i \lambda_{ii}(x_0) \neq 0$, the asymptotic mse is minimal for

$$h_n^*(x_0) = \frac{1}{n^{1/(d+4)}} \left(\frac{d\,\lambda(x_0)\,c(d,2\gamma)\,(d+2\gamma+2)^2}{c(d,\gamma)^2 \left(\sum_{i=1}^d \lambda_{ii}(x_0)\right)^2} \right)^{1/(d+4)}.$$

For details:
M.N.M. van Lieshout. Infill asymptotics and bandwidth selection for kernel estimators of spatial intensity functions. *Methodology and Computing in Applied Probability* 22:995–1008, 2020.

Abramson principle

Idea: in sparse regions more smoothing is necessary then in regions that are rich in points. (Abramson, 1982)

Assume that $\lambda(x_0) > 0$.

Definition

$$\widehat{\lambda(x_0; h, \Phi}, W) = \sum_{y \in \Phi} \frac{c(y)^d}{h^d} \kappa\left(\frac{x_0 - y}{h} c(y)\right)$$

based on a **weight function** $c : W \to (0, \infty)$ on W. In our context,

$$c(x) = \sqrt{\lambda(x)/\lambda(x_0)}.$$

Bandwidth $h/c(y)$ is larger in sparser regions.

Asymptotic expansion

Impose the **technical conditions**

- $h_n > 0$, $h_n \to 0$ and $nh_n^d \to \infty$ as $n \to \infty$;
- $g : W \times W \to \mathbb{R}$ is bounded;
- $\lambda : W \to [\underline{\lambda}, \bar{\lambda}]$ is **bounded and bounded away from zero**, $\underline{\lambda} > 0$, and C^2.

Then, using a Beta kernel κ^γ **with** $\gamma > 2$,

1. $\text{bias}\,\widehat{\lambda_n(x_0)} = o(h_n^2)$.
2. $\text{Var}\,\widehat{\lambda_n(x_0)} = \frac{\lambda(x_0)\,c(d,2\gamma)}{n\,h_n^d\,c(d,\gamma)^2} + O\left(\frac{1}{n\,h_n^{d-1}}\right)$

as $n \to \infty$.

Note: variance unchanged, bias of smaller order $o(h_n^2)$ compared to

$$h_n^2 \sum_{i=1}^{d} \frac{\lambda_{ii}(x_0)}{2\,(d + 2\gamma + 2)}$$

for a fixed bandwidth.

Infill asymptotics for Abramson estimator

To obtain a leading bias term, impose the **stronger conditions:**

- $h_n > 0$, $h_n \to 0$ and $nh_n^d \to \infty$ as $n \to \infty$;
- $g : W \times W \to \mathbb{R}$ is bounded;
- $\lambda : W \to [\underline{\lambda}, \bar{\lambda}]$ is bounded, bounded away from zero **and** C^4 **such that** $c_{ijkl} = D_{ijkl}c$, $i, j, k, l = 1, \ldots, d$, is Hőlder continuous with exponent $\alpha \in (0, 1]$ on W, that is, there exists some $C > 0$ such that for all $i, j, k, l = 1, \ldots, d$:

$$|c_{ijkl}(x) - c_{ijkl}(y)| \leq C||x - y||^{\alpha}, \quad x, y, \in W.$$

Then, for a Beta kernel κ^{γ} **with** $\gamma > 5$

$$\text{bias}\, \widehat{\lambda_n(x_0)} = \lambda(x_0)\, h_n^4 \int_{\mathbb{R}^d} A(u; x_0)\, du + O(h_n^{4+\alpha}),$$

for an integrable function $A(\cdot; x_0)$ (defined in terms of partial derivatives of λ up to fourth order).

Asymptotically optimal bandwidth

The mse of $\widehat{\lambda_n(x_0)}$ can be expanded as

$$h_n^8 \lambda(x_0)^2 \left(\int_{\mathbb{R}^d} A(u; x_0)\, du \right)^2 + \frac{1}{n\, h_n^d} \frac{\lambda(x_0)\, c(d, 2\gamma)}{c(d,\gamma)^2} + O(h_n^{8+\alpha}) + O\left(\frac{1}{n\, h_n^{d-1}} \right).$$

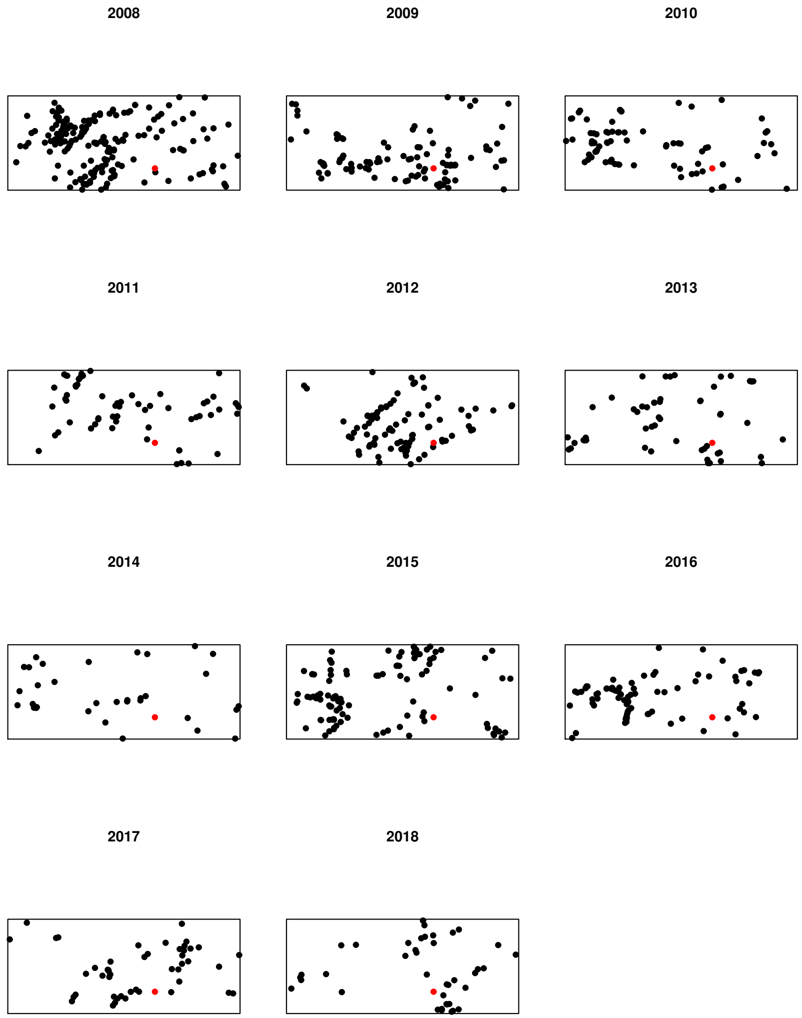
Hence, provided $\int_{\mathbb{R}^d} A(u; x_0)\, du \neq 0$, the asymptotic mse is minimal for

$$h_n^*(x_0) = \frac{1}{n^{1/(d+8)}} \left(\frac{d\, c(d, 2\gamma)}{8\, \lambda(x_0)\, c(d,\gamma)^2 \left(\int_{\mathbb{R}^d} A(u; x_0)\, du \right)^2} \right)^{1/(d+8)}.$$

Remarks:

- the squared bias is reduced from $O(h_n^4)$ to $O(h_n^8)$;
- $h_n^*(x_0) \to 0$ at rate $n^{-1/(d+8)}$ compared to $n^{-1/(d+4)}$;
- λ is assumed C^4 rather than C^2.

Tornadoes



Tornado intensity at Wichita

Take $x_0 = (-97.33, 37.68)$ and $\gamma = 6$.

To evaluate $h_n^*(x_0)$, we use a classic **pilot** kernel estimator with fixed bandwidth chosen non-parametrically.

$h_n^*(x_0) \approx 0.8$ and $\widehat{\lambda_n(x_0)} \approx 2.6$. compared to 2.9 for fixed bandwidth 0.8.

Conclusions

- For local bandwidth estimation the asymptotically optimal bandwidth in an infill regime is given by

$$h_n^*(x_0) = \frac{1}{n^{1/(d+4)}} \left(\frac{d\,\lambda(x_0)\,c(d,2\gamma)\,(d+2\gamma+2)^2}{c(d,\gamma)^2 \left(\sum_{i=1}^{d} \lambda_{ii}(x_0)\right)^2} \right)^{1/(d+4)}.$$

- For adaptive local bandwidth estimation,

$$h_n^*(x_0) = \frac{1}{n^{1/(d+8)}} \left(\frac{d\,c(d,2\gamma)}{8\,\lambda(x_0)\,c(d,\gamma)^2 \left(\int_{\mathbb{R}^d} A(u;x_0)du\right)^2} \right)^{1/(d+8)}.$$

- If a single pattern only is observed, non-parametric bandwidth selectors may be used.

For details:
M.N.M. van Lieshout. Infill asymptotics and bandwidth selection for kernel estimators of spatial intensity functions. ArXiv 1904.05095, April 2019.

Non-parametric adaptive bandwidth selection

1. Choose a non-adaptive pilot bandwidth h_g by minimising $|T_\kappa(h;\Phi,W) - |W||$ over h.
2. Optimise

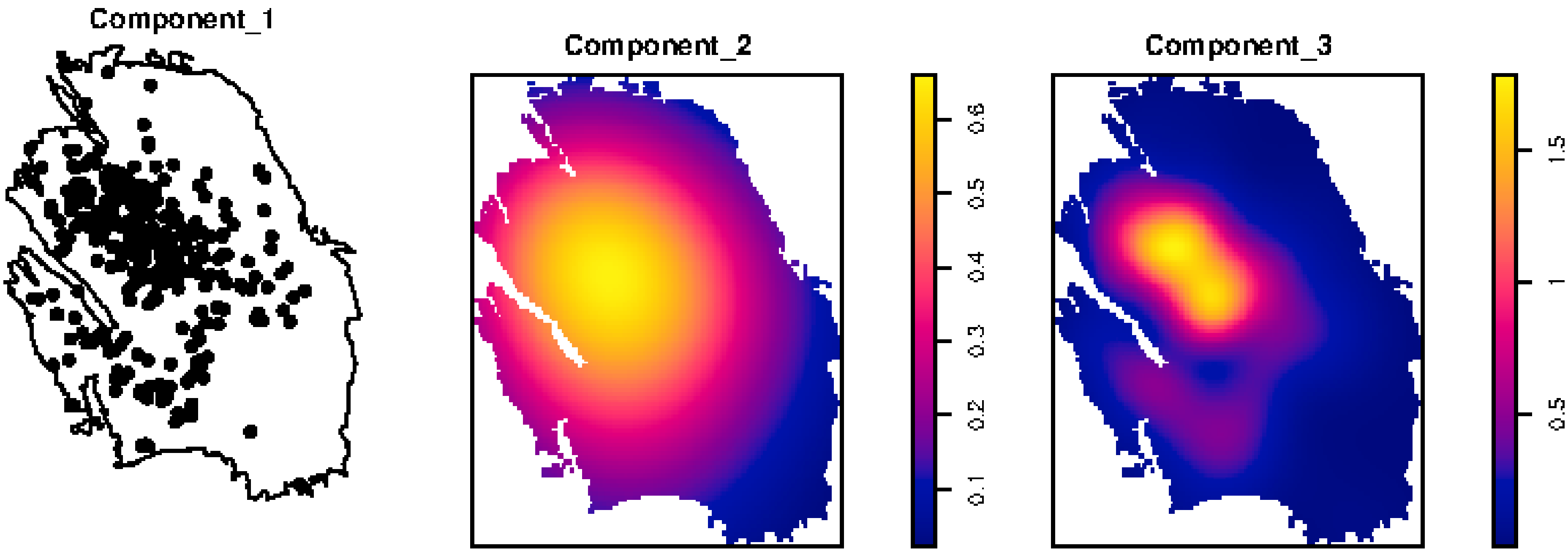
$$\left| \sum_{x \in \Phi \cap W} \frac{1}{\hat{\lambda}(x; h, \Phi, W)} - |W| \right|$$

over $h > 0$ where $\hat{\lambda}$ is the adaptive kernel estimator with

$$c(y)^2 = \frac{1}{h_g^d} \sum_{z \in \Phi \cap W} \kappa\left(\frac{y-z}{h_g}\right).$$

Back to Groningen once more

Fixed and adaptive bandwidth.



Thank you for your attention!