Statistical modelling and inference for spatio-temporal disease processes

Martin Hazelton[1]

Statistics and Bioinformatics Group
Institute of Fundamental Sciences, Massey University

24 October 2012

IDReC
Infectious Disease Research Centre

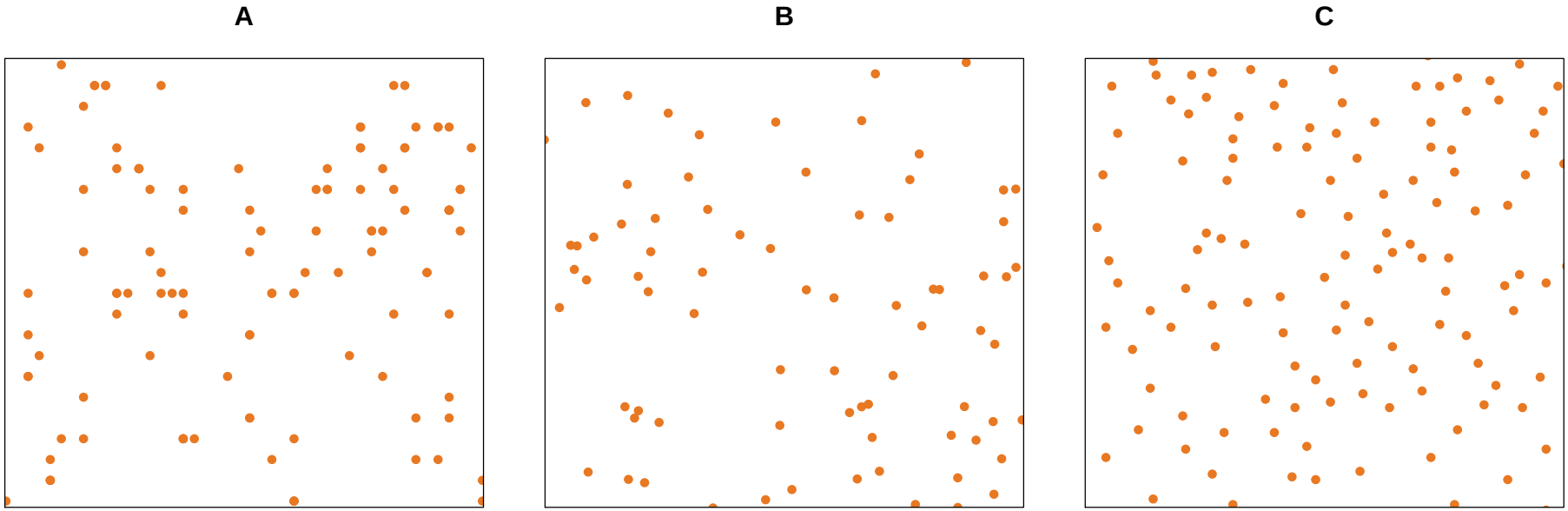
[1] Presenter: m.hazelton@massey.ac.nz

Disease, Prediction and Chance

- Spread of infectious disease is a complex process typically involving a plethora of factors.
- Some are explicable/predictable.
- Others are essentially random.
- Plenty of work for statisticians deciding which is which.

IDReC
Infectious Disease Research Centre

Spatial Patterns ... of Disease?

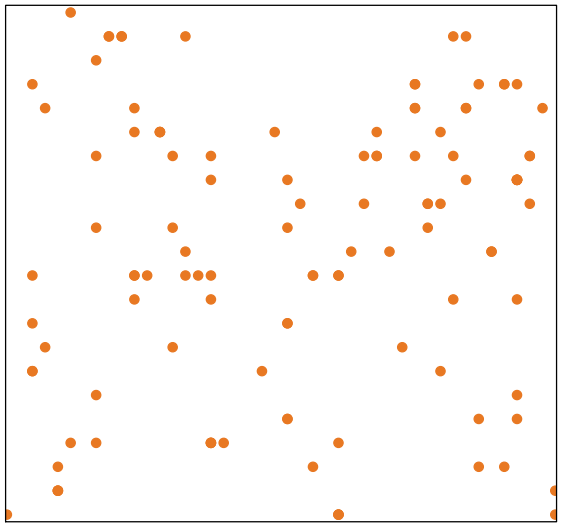


- Complete spatial randomness
- Locations of gastroenteritis cases
- Mystery spatial point pattern

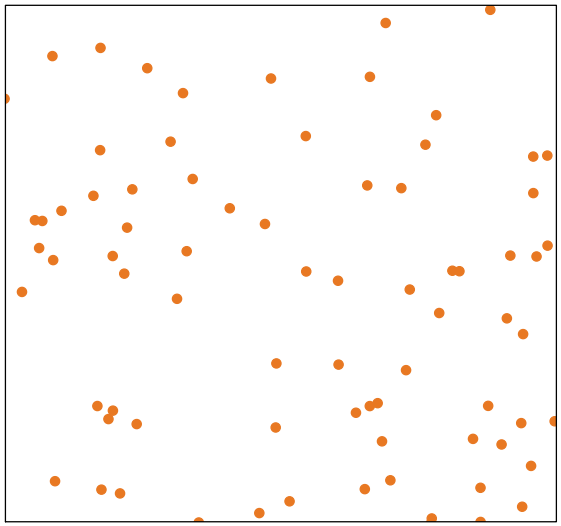
IDReC
Infectious Disease Research Centre

Spatial Patterns ... of Disease?

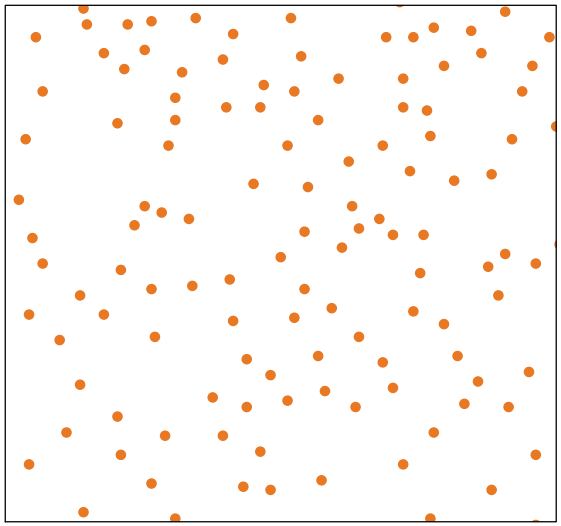
Gastroenteritis cases



Complete spatial randomness



Sea anemone locations



- Complete spatial randomness (B)
- Locations of gastroenteritis cases (A)
- Mystery spatial point pattern (C)

IDReC
Infectious Disease Research Centre

Measuring Spatial Risk

- To understand spatial patterns of disease we must adjust for underlying population density.
- To this end, Bithell (1990) defined the **relative risk function**:

$$r(\boldsymbol{z}) = \frac{f(\boldsymbol{z})}{g(\boldsymbol{z})} \qquad \boldsymbol{z} \in \mathcal{R}$$

  where f is density of cases and g density of controls.
- Usual to work on log-scale:

$$\rho(\boldsymbol{z}) = \log\{r(\boldsymbol{z})\} = \log\{f(\boldsymbol{z})\} - \log\{g(\boldsymbol{z})\}$$

Bithell J.F. (1990). An application of density estimation to geographical epidemiology. *Statistics in Medicine* **9**:691–701.

IDReC Infectious Disease Research Centre

Kernel Estimation

- Straightforward approach to estimation of r is to replacing unknown densities by kernel estimates thereof:

$$\hat{r}(\boldsymbol{z}) = \frac{\hat{f}(\boldsymbol{z})}{\hat{g}(\boldsymbol{z})} \qquad \boldsymbol{z} \in \mathcal{R}$$

- **Kernel density estimate** constructed from bivariate data $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$:

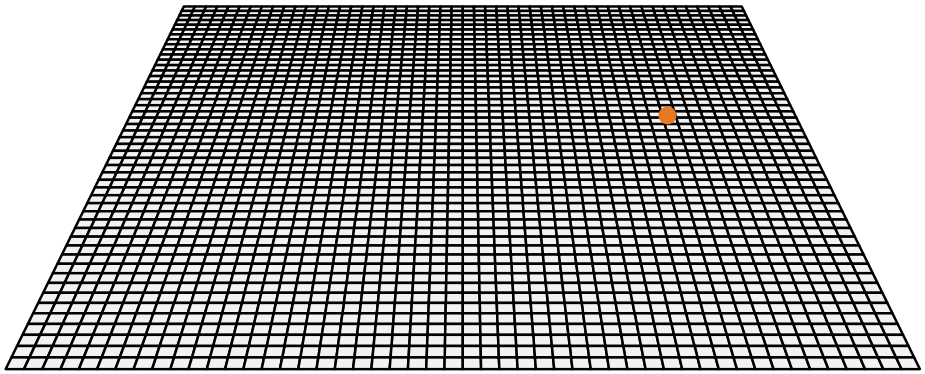
$$\hat{f}(\boldsymbol{z}) = n^{-1} \sum_{i=1}^{n} K_h(\boldsymbol{z} - \boldsymbol{x}_i)$$

  - $K_h(\boldsymbol{z}) = h^{-2} K(\boldsymbol{z}/h)$, with kernel K being spherically symmetric;
  - h is the bandwidth, controlling degree of smoothing.

IDReC Infectious Disease Research Centre

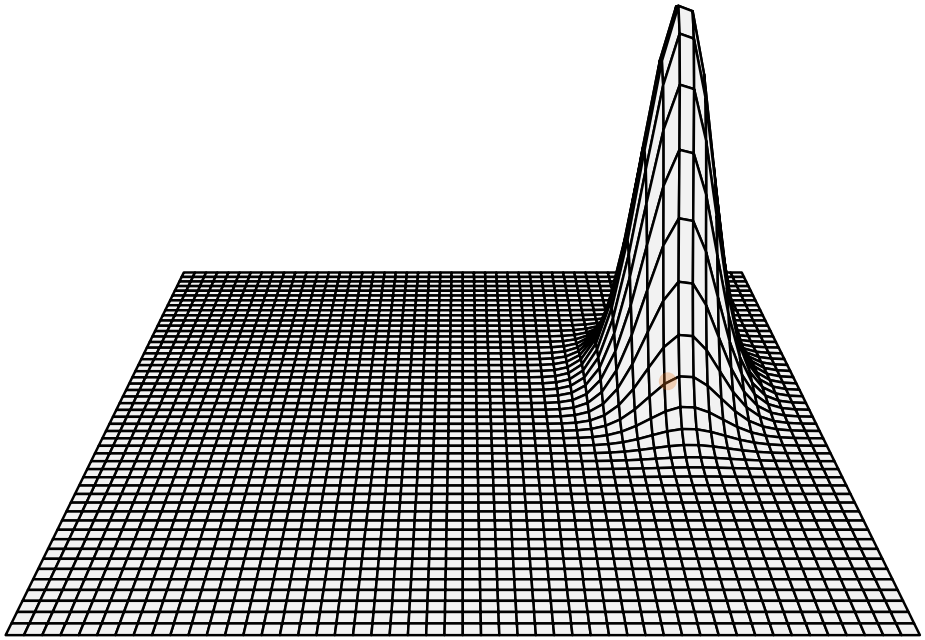
Constructing Kernel Density Estimates

An example using the gastroenteritis data



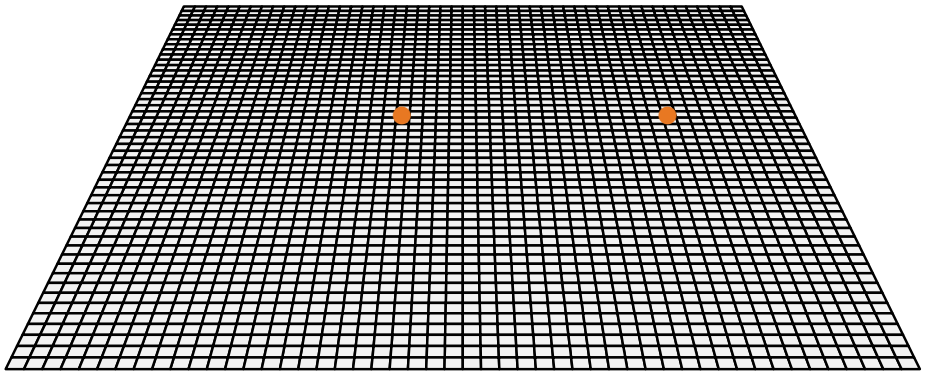
Constructing Kernel Density Estimates

An example using the gastroenteritis data



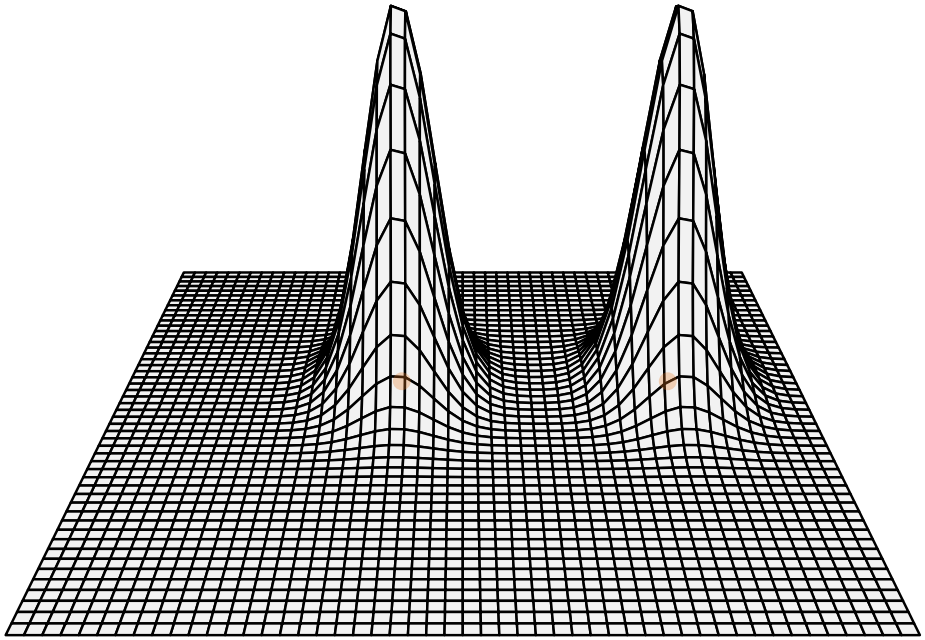
Constructing Kernel Density Estimates

An example using the gastroenteritis data



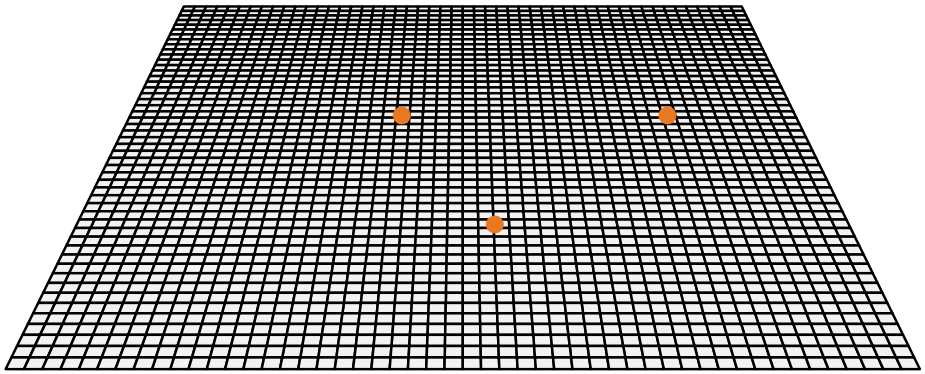
Constructing Kernel Density Estimates

An example using the gastroenteritis data



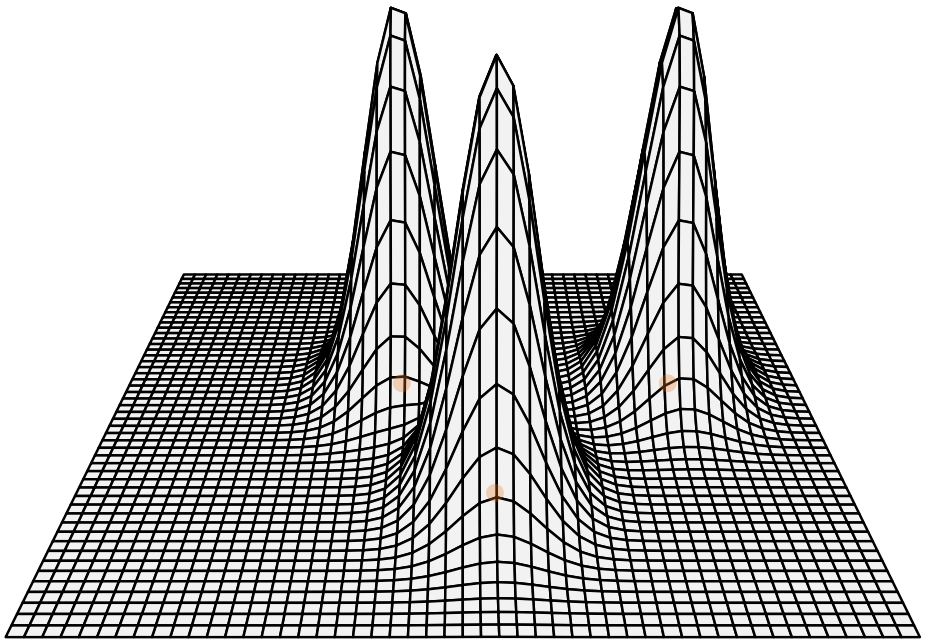
Constructing Kernel Density Estimates

An example using the gastroenteritis data



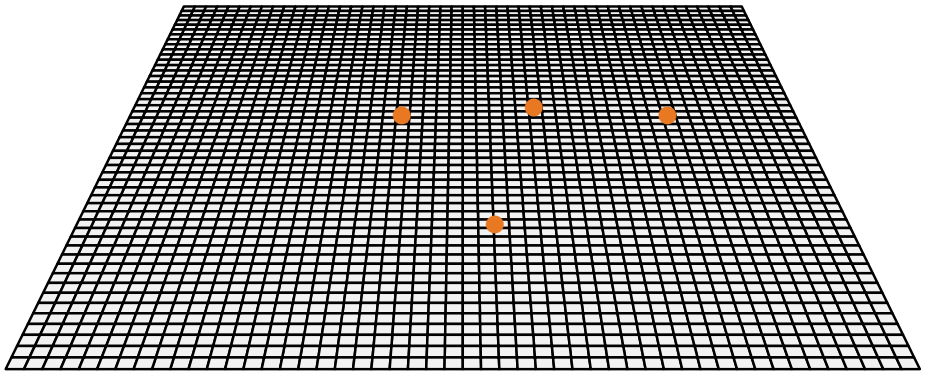
Constructing Kernel Density Estimates

An example using the gastroenteritis data



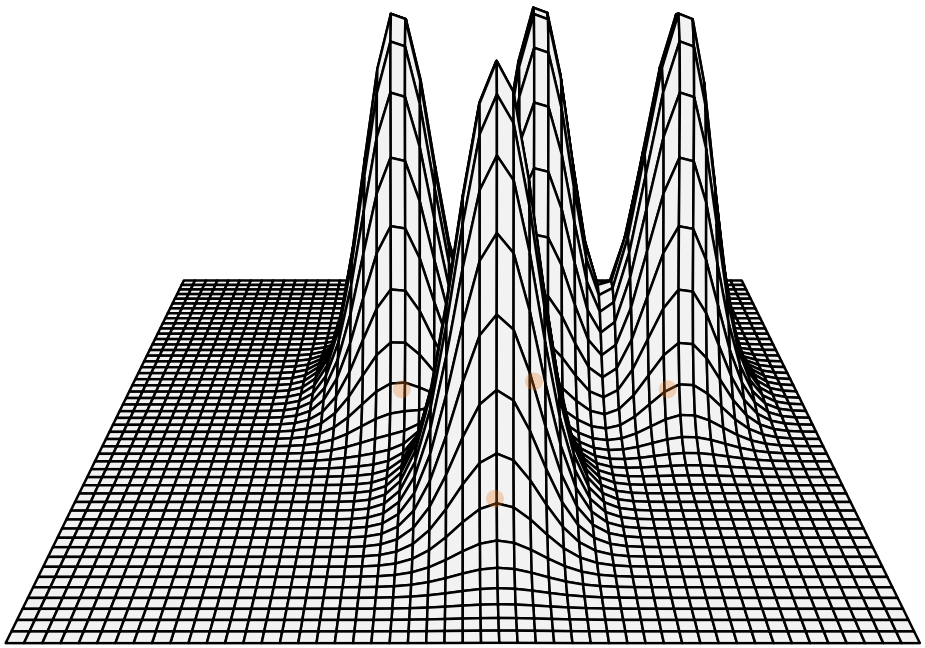
Constructing Kernel Density Estimates

An example using the gastroenteritis data



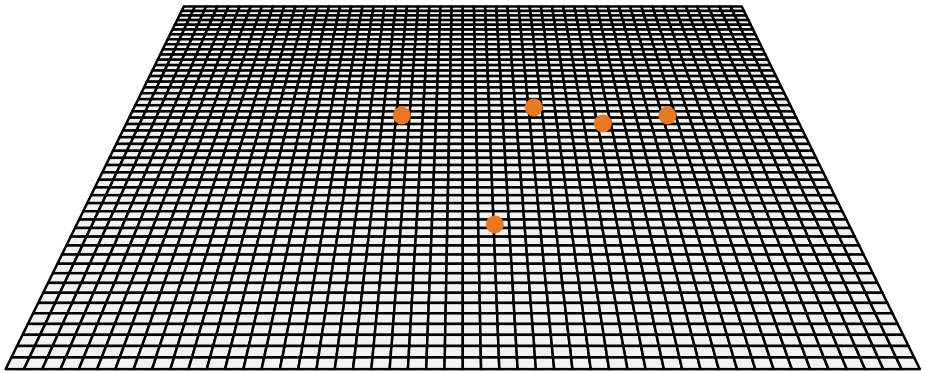
Constructing Kernel Density Estimates

An example using the gastroenteritis data



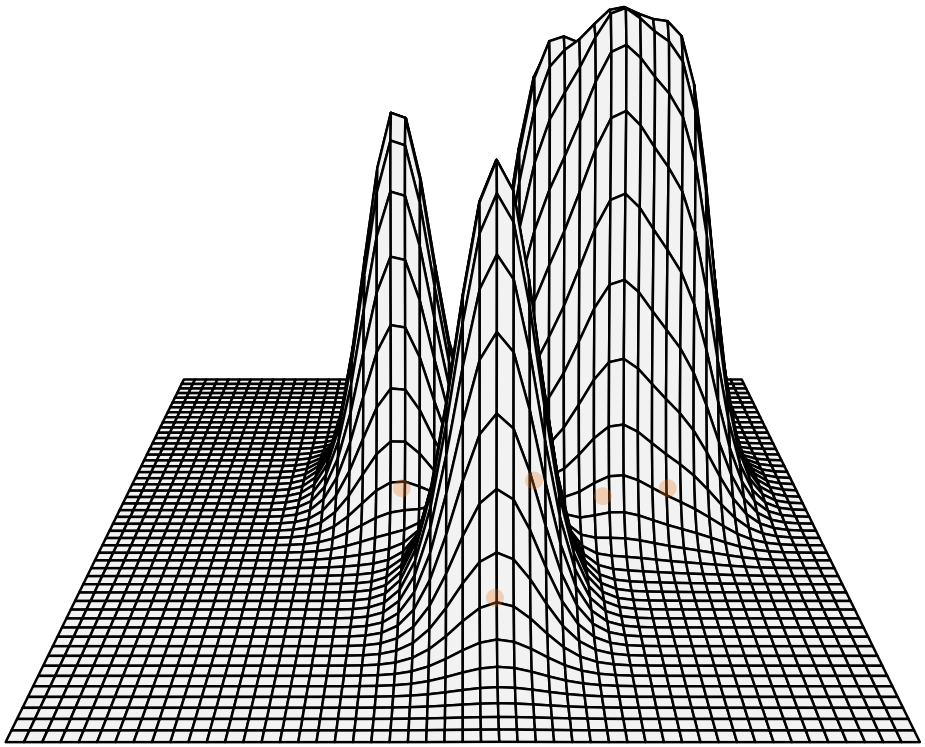
Constructing Kernel Density Estimates

An example using the gastroenteritis data



Constructing Kernel Density Estimates

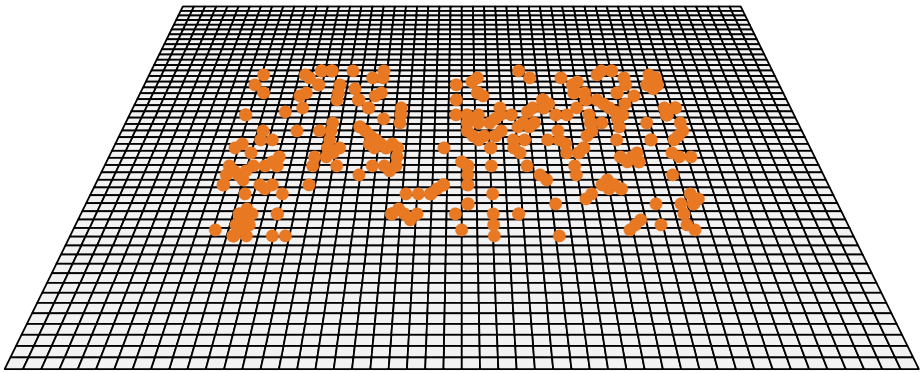
An example using the gastroenteritis data



IDReC
Infectious Disease Research Centre

Constructing Kernel Density Estimates

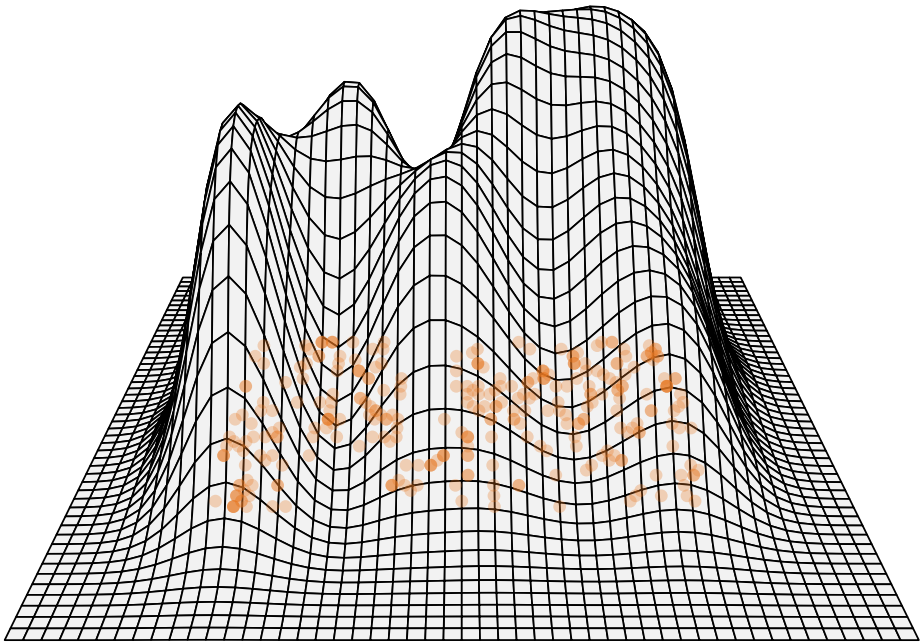
An example using the gastroenteritis data



IDReC
Infectious Disease Research Centre

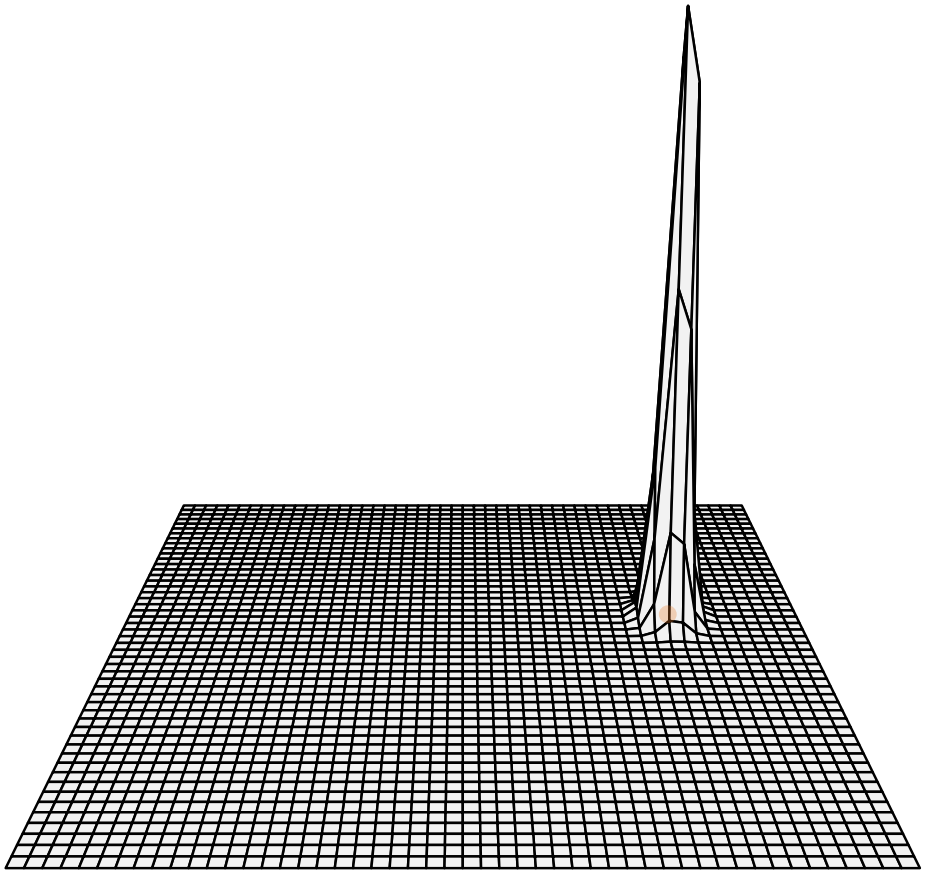
Constructing Kernel Density Estimates

An example using the gastroenteritis data



Smoothing of Kernel Density Estimates

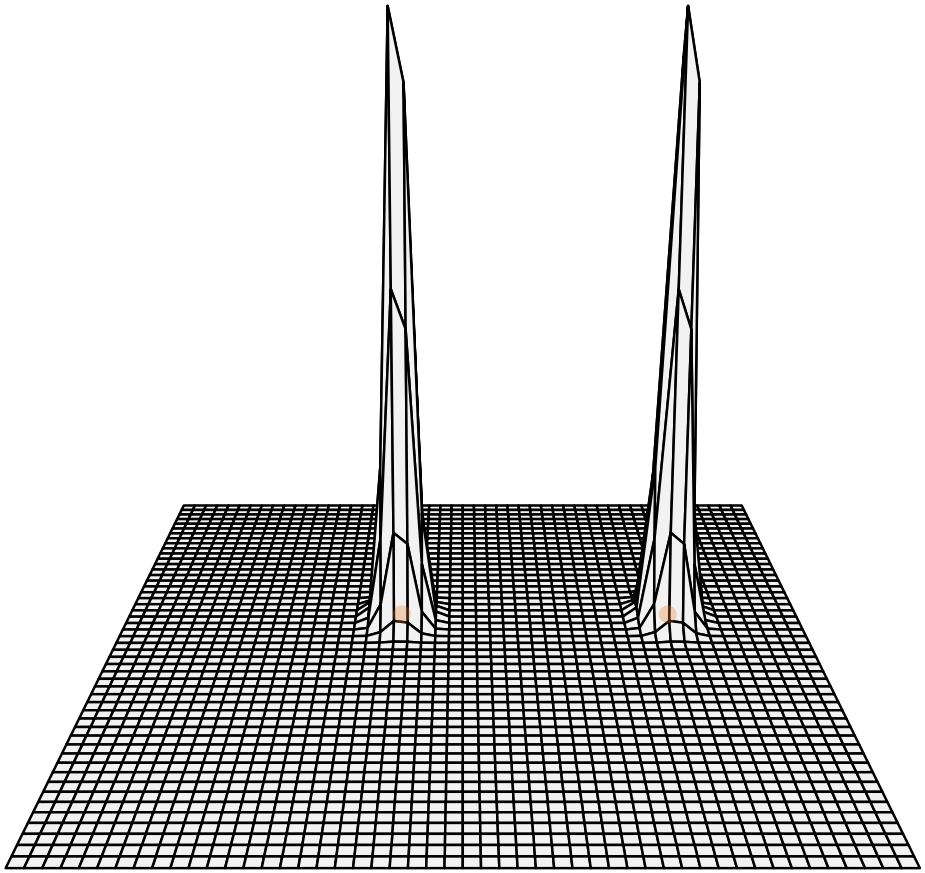
Small bandwidth (undersmoothing)



IDReC
Infectious Disease Research Centre

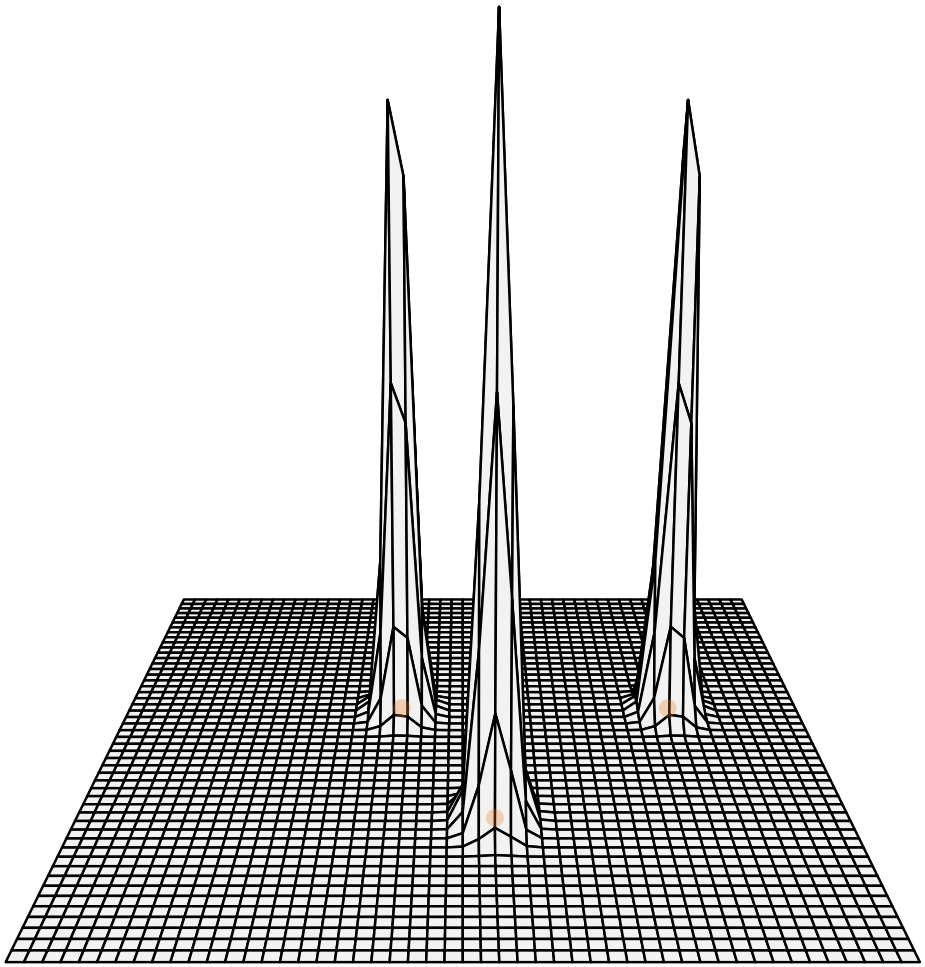
Smoothing of Kernel Density Estimates

Small bandwidth (undersmoothing)



Smoothing of Kernel Density Estimates

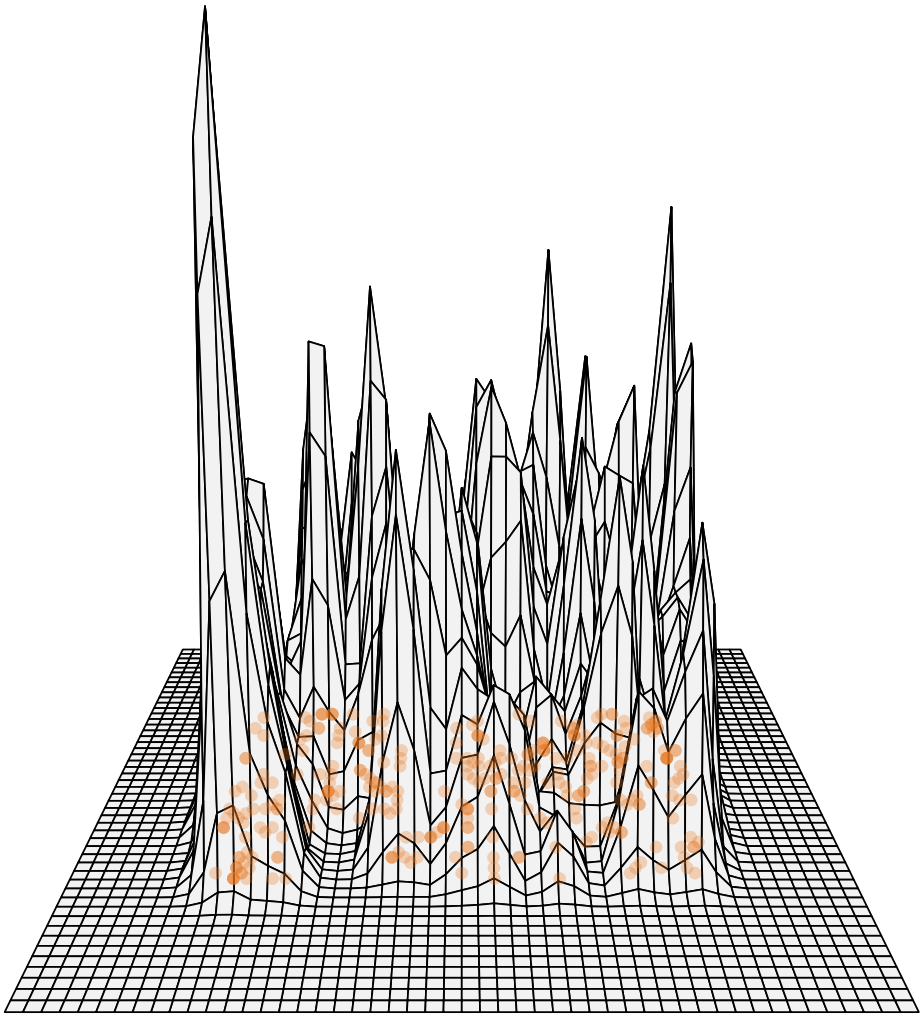
Small bandwidth (undersmoothing)



IDReC
Infectious Disease Research Centre

Smoothing of Kernel Density Estimates

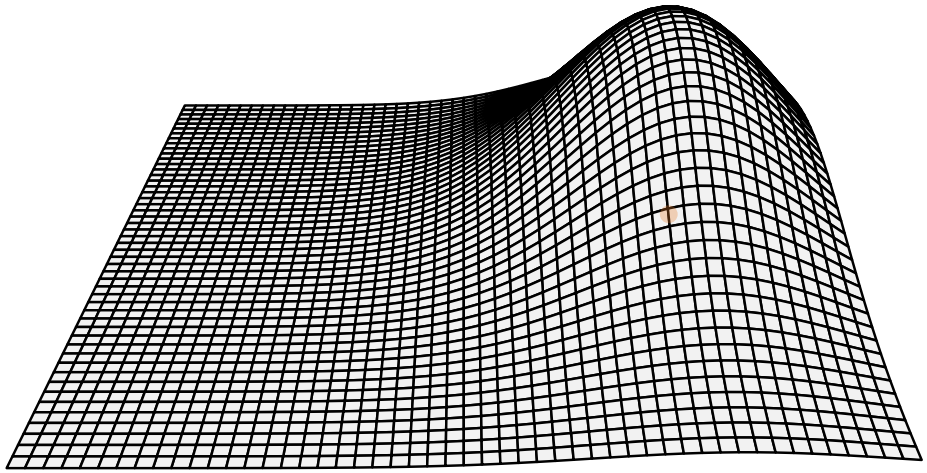
Small bandwidth (undersmoothing)



IDReC
Infectious Disease Research Centre

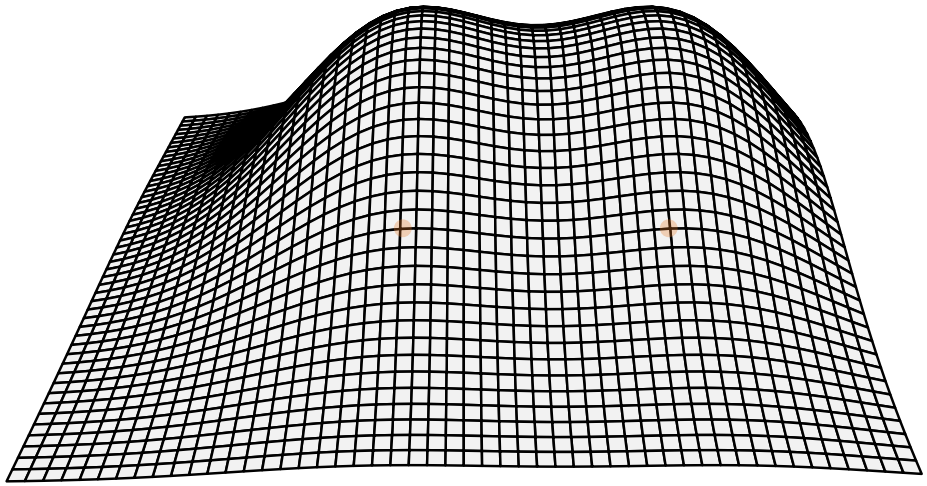
Smoothing of Kernel Density Estimates

Large bandwidth (oversmoothing)



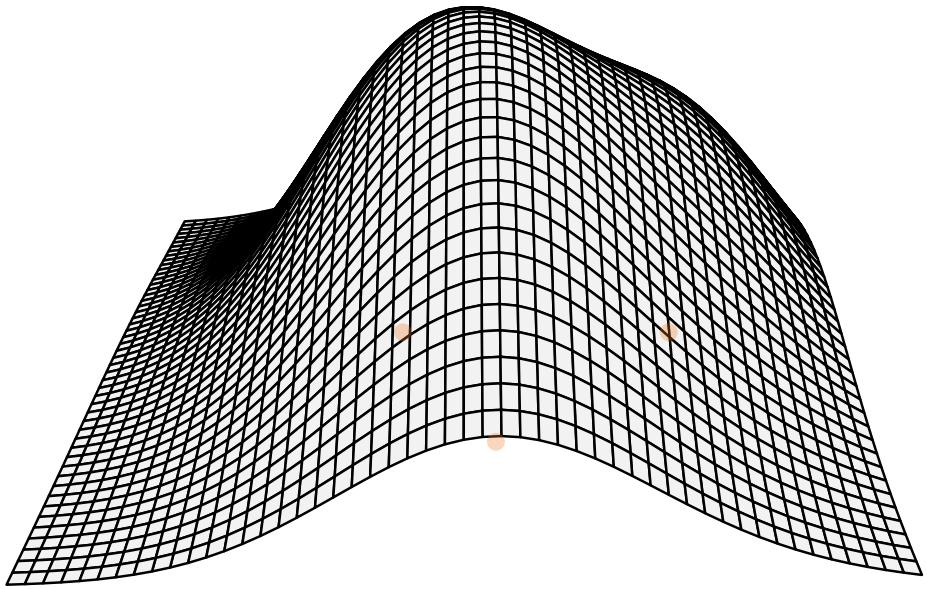
Smoothing of Kernel Density Estimates

Large bandwidth (oversmoothing)



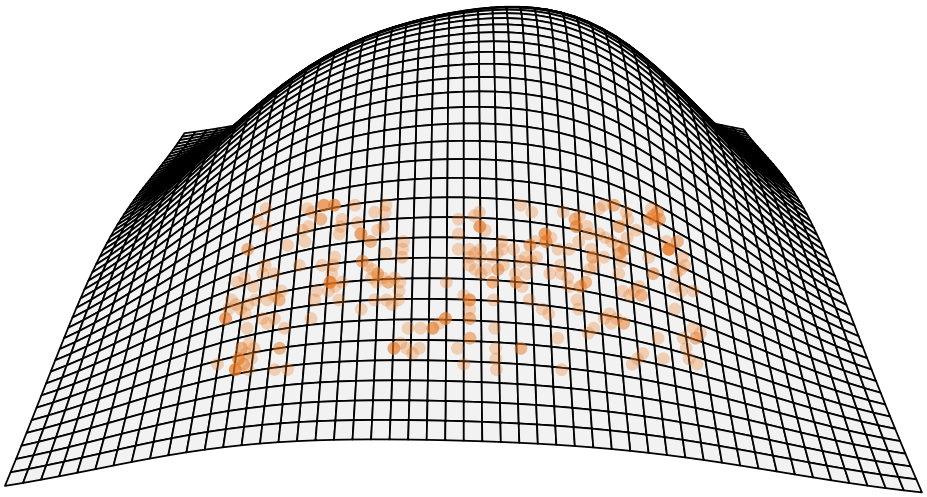
Smoothing of Kernel Density Estimates

Large bandwidth (oversmoothing)



Smoothing of Kernel Density Estimates

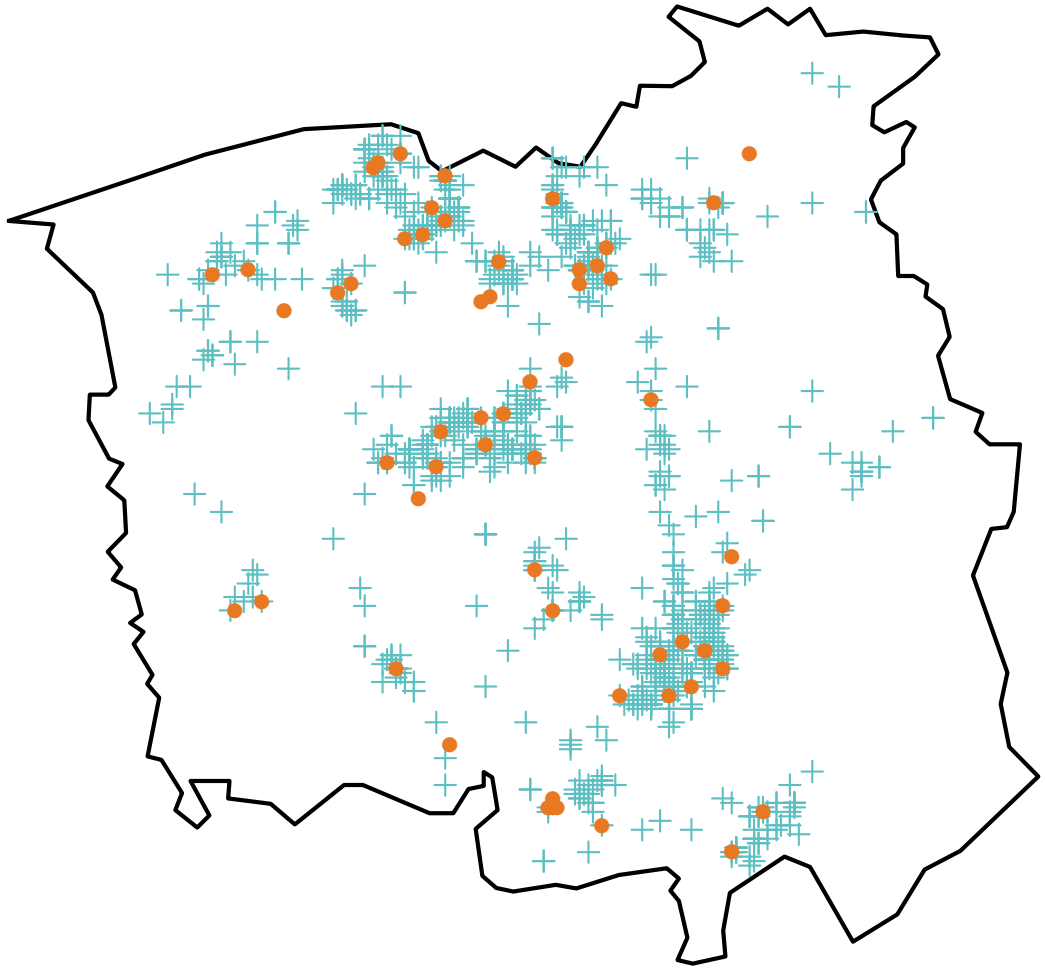
Large bandwidth (oversmoothing)



IDReC
Infectious Disease Research Centre

Example: Cancer of the Larynx in South Lancashire

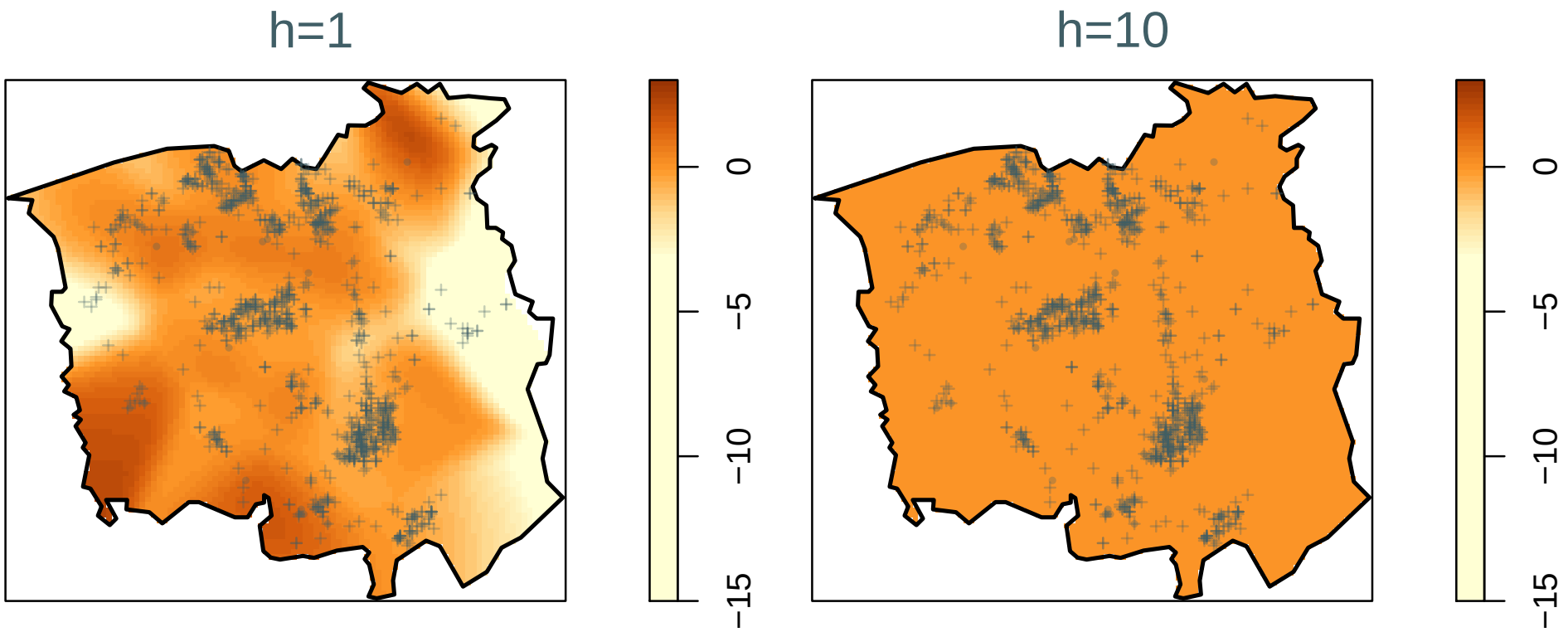
- Data are geographical coordinates for 58 cases of cancer of the larynx, and 978 controls (cases of lung cancer).
- Data collected between 1974 and 1983 by Chorley and South Ribble Health Authority in Lancashire, England.



IDReC
Infectious Disease Research Centre

The Effect of Bandwidth Selection

For the log-relative risk function



IDReC
Infectious Disease Research Centre

Research on Bandwidth Selection

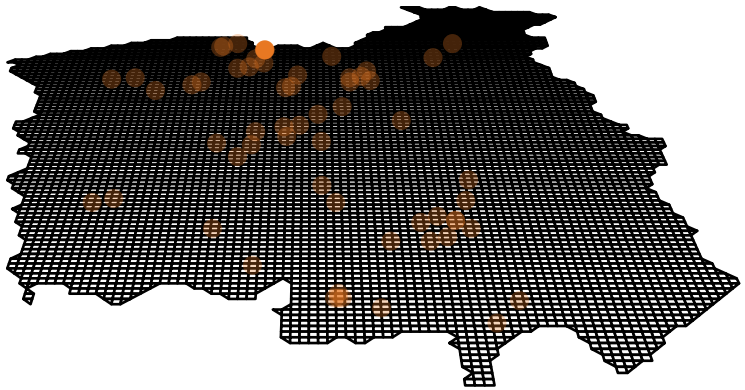
- Bandwidth selection is a challenging problem for both theoretical and practical reasons.
  - Typical inhomogeneity of populations.
  - Want stable estimates of risk even where data are sparse.
  - Asymptotic theory very delicate.
- Recent progress on spatially adaptive smoothing regimens Davies & Hazelton (2010).

Davies, T.M. and Hazelton, M.L. (2010). Adaptive kernel estimation of spatial relative risk. *Statistics in Medicine* **29**, 2423–2437.

IDReC Infectious Disease Research Centre

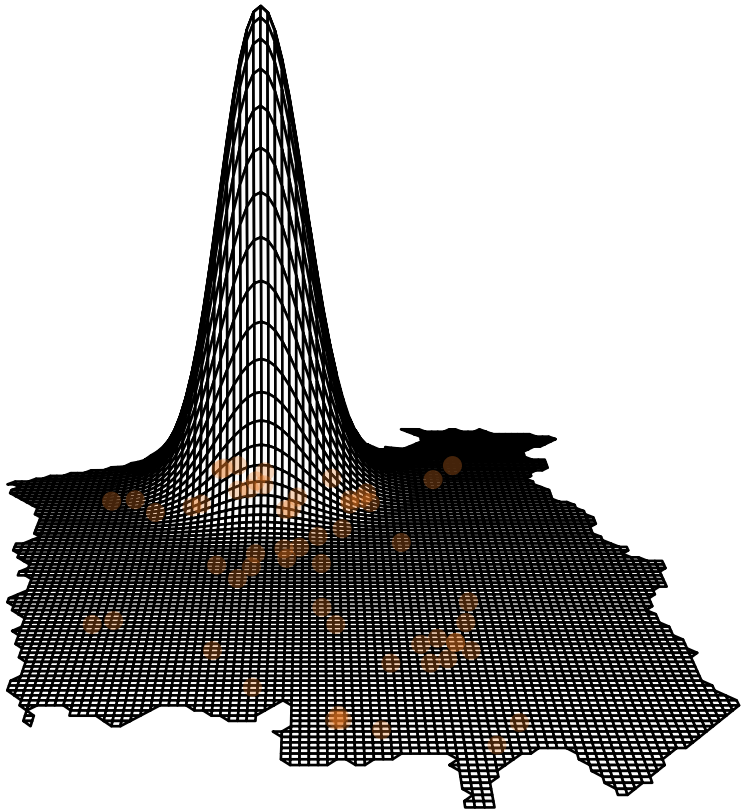
Estimation with Spatially Adaptive Bandwidths

Estimation of case density



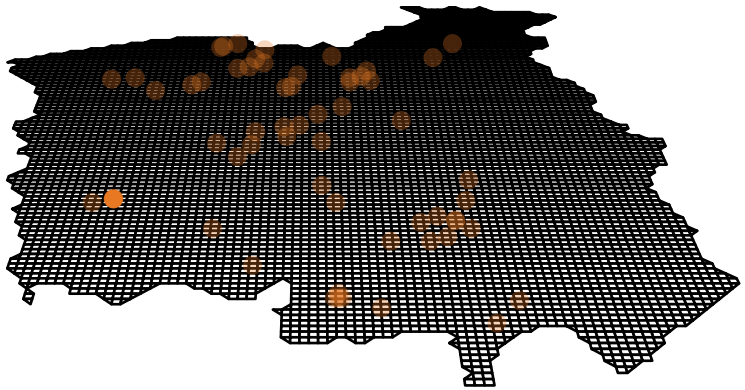
Estimation with Spatially Adaptive Bandwidths

Estimation of case density



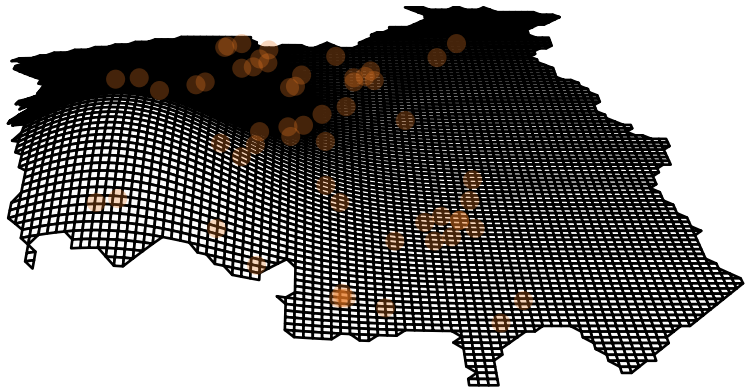
Estimation with Spatially Adaptive Bandwidths

Estimation of case density



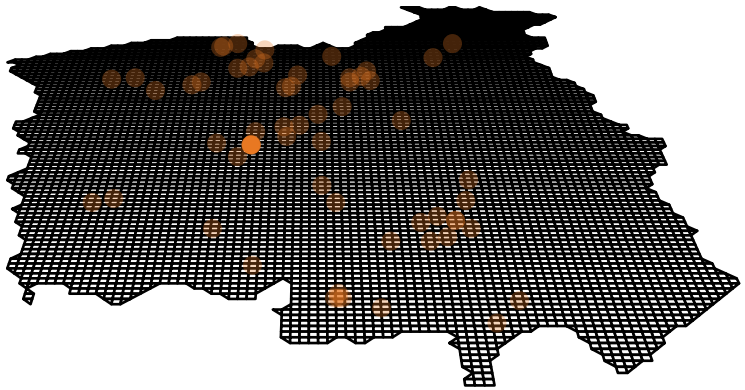
Estimation with Spatially Adaptive Bandwidths

Estimation of case density



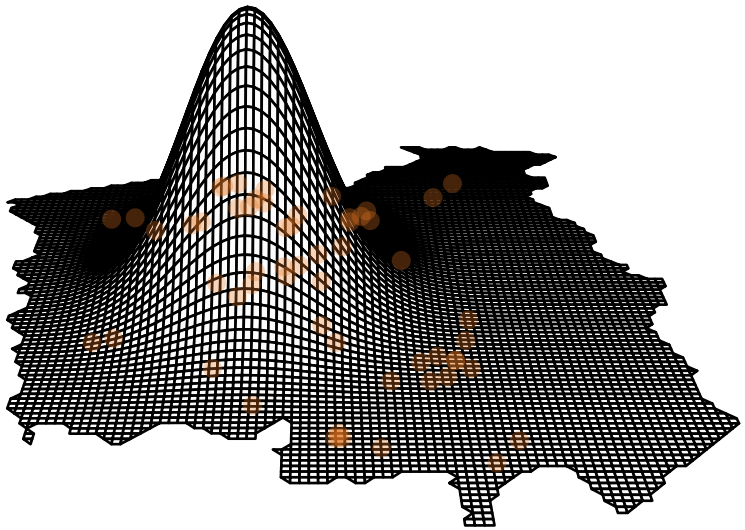
Estimation with Spatially Adaptive Bandwidths

Estimation of case density



Estimation with Spatially Adaptive Bandwidths

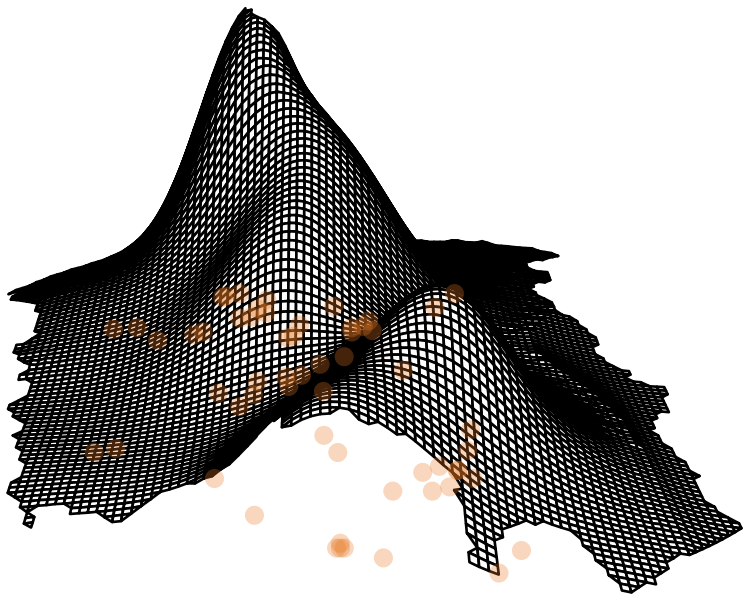
Estimation of case density



IDReC
Infectious Disease Research Centre

Estimation with Spatially Adaptive Bandwidths

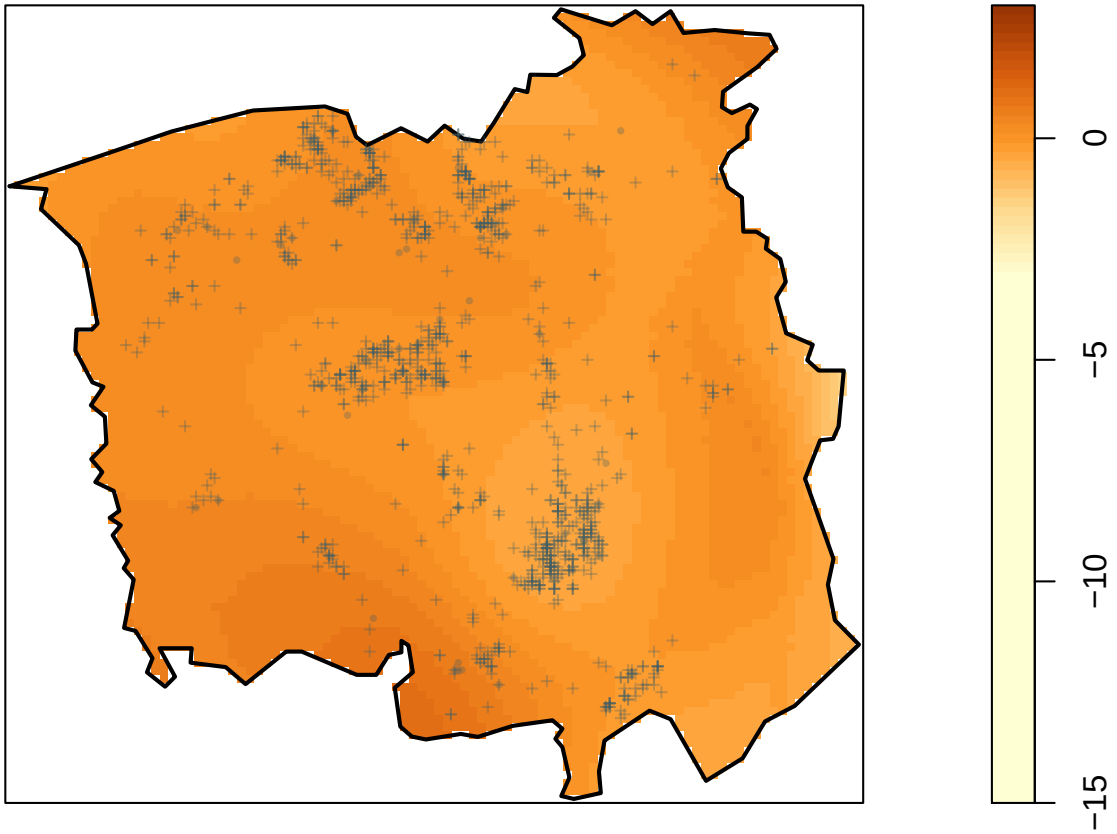
Estimation of case density



IDReC
Infectious Disease Research Centre

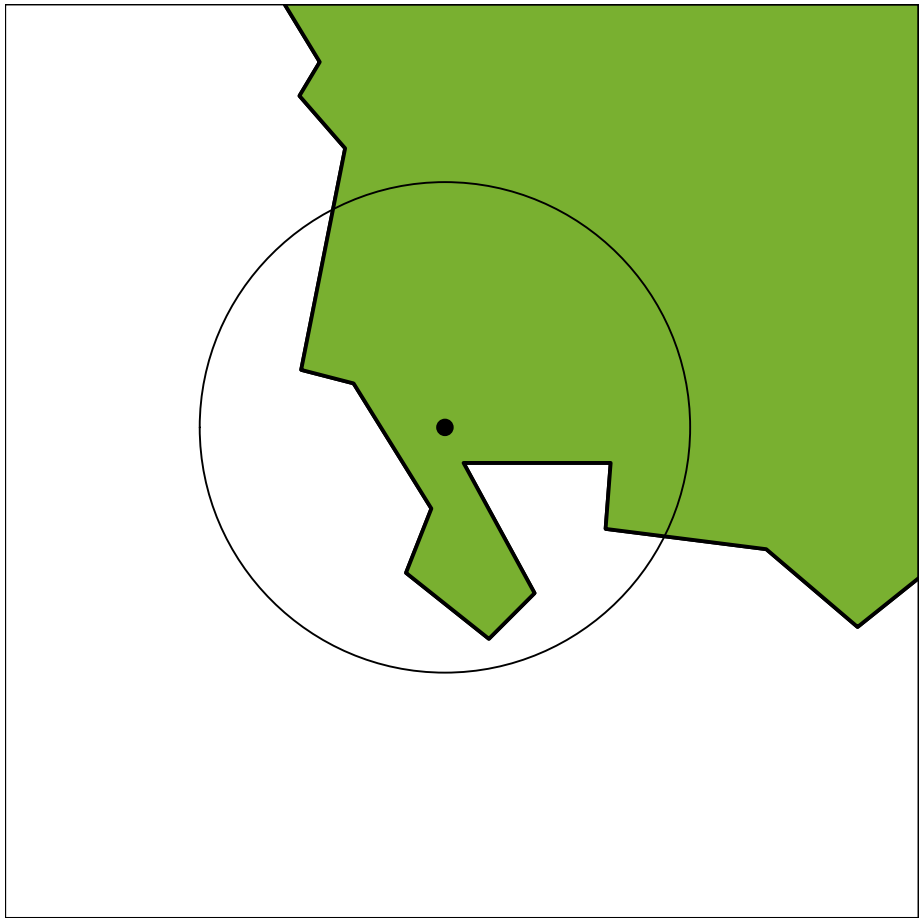
Estimation with Spatially Adaptive Bandwidths

Estimation of log-relative risk surface



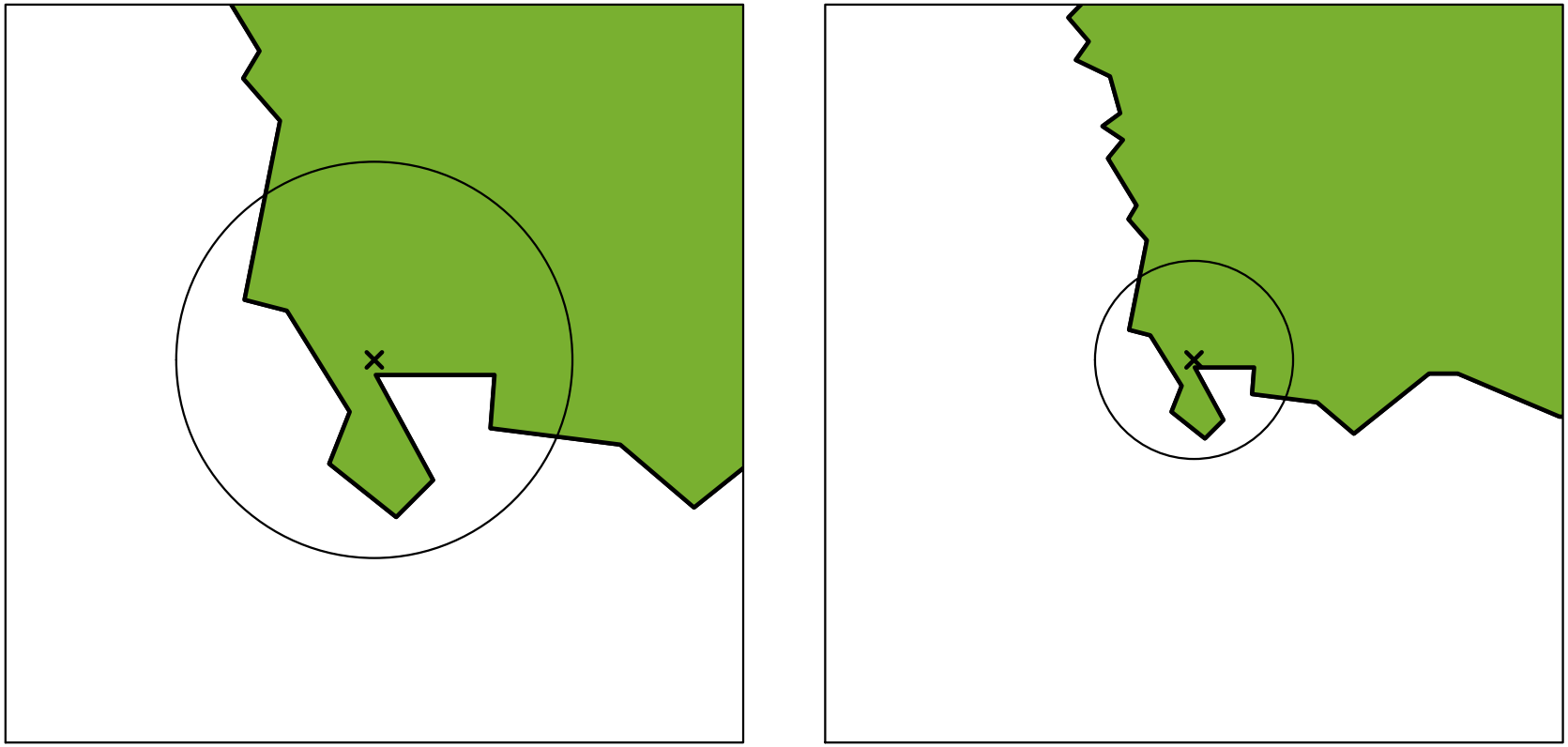
Boundary Bias

- Disease data collected over finite region.
- For points near edge, some of kernel weight spills over boundary and is lost.
- Has significant impact on kernel estimates.
- Theoretical analysis is intricate.



IDReC
Infectious Disease Research Centre

Foundations of Asymptotic Analysis of Boundary Bias



Asymptotics with Adaptive Boundary Kernel

Or ... Why Postdocs are Great

$$
\begin{aligned}
\frac{\text{bias}(\hat{f}(z))}{h_0^2} = \frac{[D^{10}f(z)]^2}{4f(z)} \Bigg[& \frac{b_0}{q}\left(a_{40}^{(20)} + 2a_{31}^{(11)} + a_{22}^{(02)} + 7a_{30}^{(10)} + 7a_{21}^{(01)} + 8a_{20}\right) \\
& + \frac{b_1}{q}\left(a_{50}^{(20)} + 2a_{41}^{(11)} + a_{32}^{(02)} + 9a_{40}^{(10)} + 9a_{31}^{(01)} + 15a_{30}\right) \\
& + \frac{b_2}{q}\left(a_{41}^{(20)} + 2a_{32}^{(12)} + a_{23}^{(02)} + 9a_{31}^{(10)} + 9a_{22}^{(01)} + 15a_{21}\right)\Bigg] \\
+ \frac{D^{10}f(z)D^{01}f(z)}{4f(z)} \Bigg[& \frac{b_0}{q}\left(2a_{31}^{(20)} + 4a_{22}^{(11)} + 2a_{13}^{(02)} + 14a_{21}^{(10)} + 14a_{12}^{(01)} + 16a_{11}\right) \\
& + \frac{b_1}{q}\left(2a_{41}^{(20)} + 4a_{22}^{(11)} + 2a_{23}^{(02)} + 18a_{31}^{(10)} + 18a_{22}^{(01)} + 30a_{21}\right) \\
& + \frac{b_2}{q}\left(2a_{32}^{(20)} + 4a_{23}^{(11)} + 2a_{14}^{(02)} + 18a_{22}^{(10)} + 18a_{13}^{(01)} + 30a_{12}\right)\Bigg] \\
+ \frac{[D^{01}f(z)]^2}{4f(z)} \Bigg[& \frac{b_0}{q}\left(a_{22}^{(20)} + 2a_{13}^{(11)} + a_{04}^{(02)} + 7a_{12}^{(10)} + 7a_{03}^{(01)} + 8a_{02}\right) \\
& + \frac{b_1}{q}\left(a_{32}^{(20)} + 2a_{23}^{(11)} + a_{14}^{(02)} + 9a_{22}^{(10)} + 9a_{13}^{(01)} + 15a_{12}\right) \\
& + \frac{b_2}{q}\left(a_{23}^{(20)} + 2a_{14}^{(11)} + a_{05}^{(02)} + 9a_{13}^{(10)} + 9a_{04}^{(01)} + 15a_{03}\right)\Bigg] \\
+ \frac{D^{20}f(z)}{4} \Bigg[& \frac{b_0}{q}\left(2a_{30}^{(10)} + 2a_{21}^{(01)} + 8a_{20}\right) + \frac{b_1}{q}\left(2a_{40}^{(10)} + 2a_{31}^{(01)} + 10a_{30}\right) \\
& + \frac{b_2}{q}\left(2a_{31}^{(10)} + 2a_{22}^{(01)} + 10a_{21}\right)\Bigg]
\end{aligned}
$$

IDReC Infectious Disease Research Centre

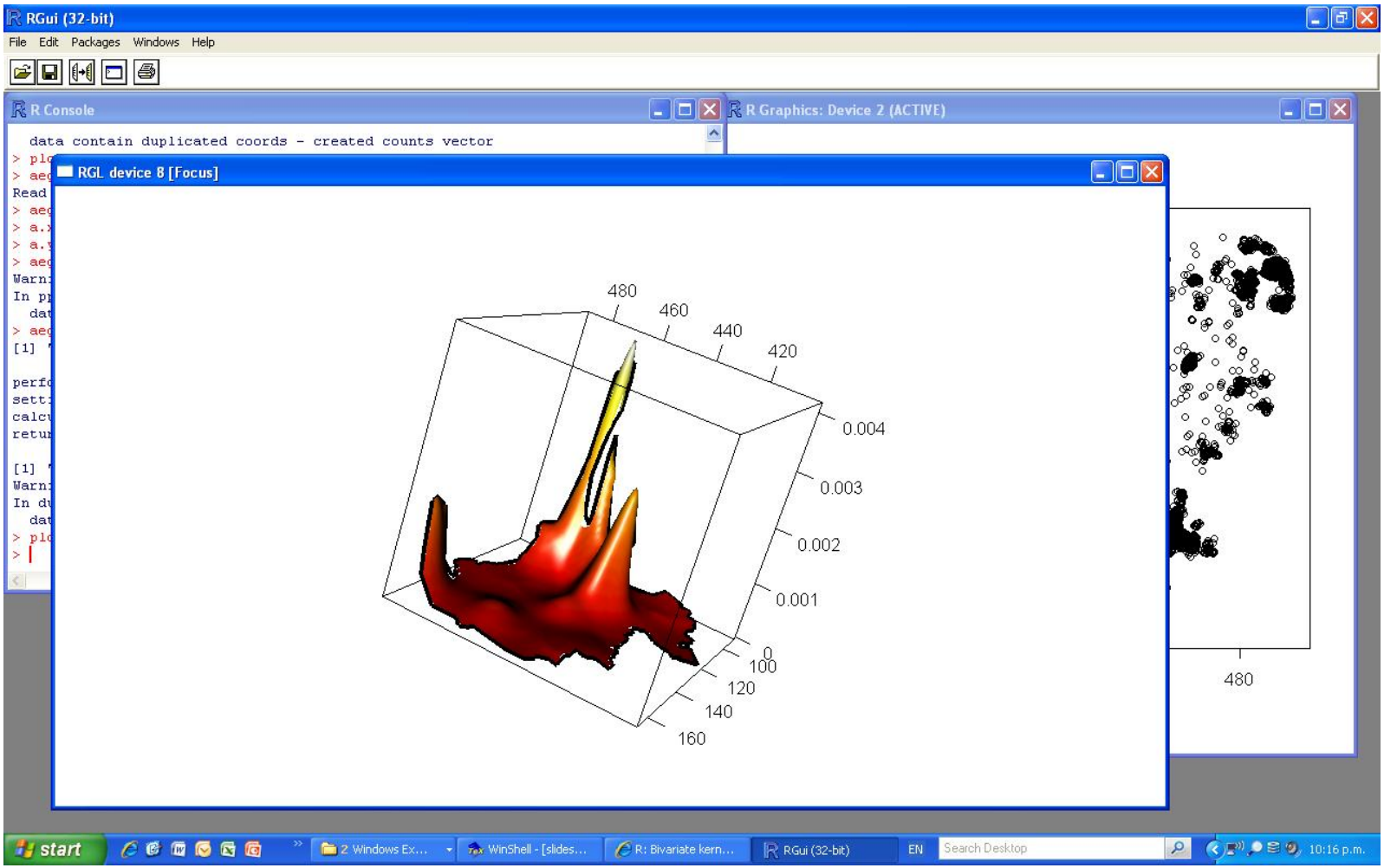
Computer Implementation

- Important practical problems in spatial epidemiology can quickly lead to fascinating theoretical work.
- Need to make this stuff available to users.
- Released R package `sparr`. See Davies, Hazelton & Marshall (2011).

Davies, T.M., Hazelton, M.L. and Marshall, J.C. (2011). sparr: Analyzing spatial relative risk using fixed and adaptive kernel density estimation in R. *Journal of Statistical Software* **39**, 1–14.

IDReC Infectious Disease Research Centre

Screenshot of R Package `Sparr`



IDReC
Infectious Disease Research Centre

Statistics and Bioinformatics Group Research

- Methodological research with applications to infectious disease
  - Spatial statistics
  - Markov chain Monte Carlo methods
  - Simulation based inference
  - Smoothing methods
  - Semiparametric modelling
  - Statistical software development
- Variety of applications

IDReC
Infectious Disease Research Centre

Some Members of the Research Group

Postgraduate students

Sarojinie Fernando
PhD (recently submitted)

Kate Richards
PhD (year 2)

Lyndal Henden
Honours

Brigid Betz-Stablein
PhD (year 3)

Khair Jones
PhD (year 1)

Staff with interest in infectious disease applications

Prof Martin Hazelton
A/Prof Geoff Jones
Dr Chris Jewell
Dr Jonathan Marshall

IDReC
Infectious Disease Research Centre

Some Members of the Research Group

Postgraduate students

Sarojinie Fernando

PhD (recently submitted)

Kate Richards

PhD (year 2)

Lyndal Henden

Honours

Brigid Betz-Stablein

PhD (year 3)

Khair Jones

PhD (year 1)

Staff with interest in infectious disease applications

Prof Martin Hazelton

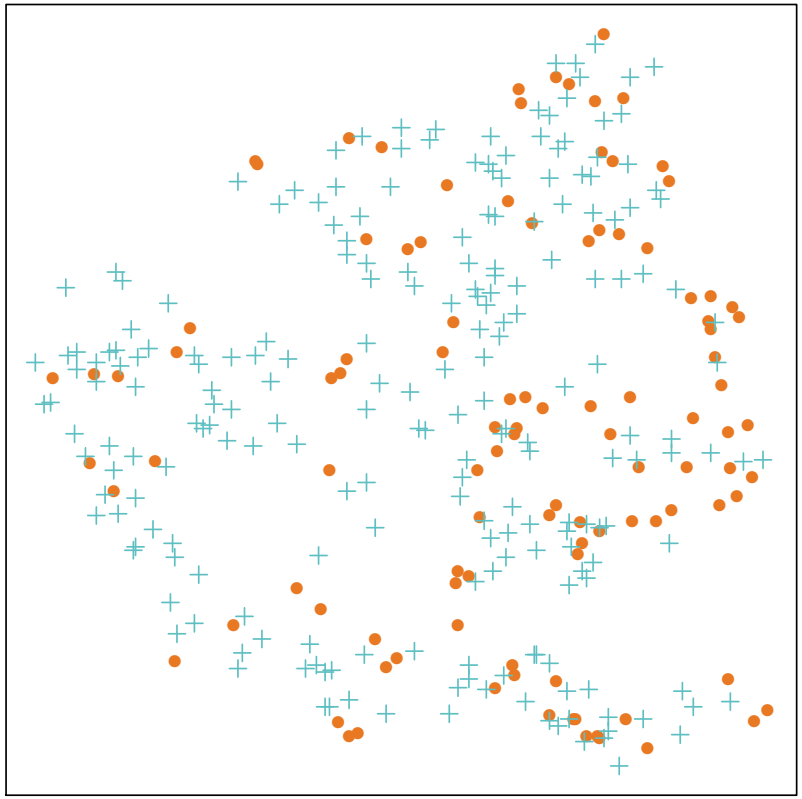
A/Prof Geoff Jones

Dr Chris Jewell

Dr Jonathan Marshall

IDReC Infectious Disease Research Centre

Improved Spatio-Temporal Risk Estimation



- Local linear versus density ratio estimation of relative risk.
- Application to myrtle wilt in Tasmanian myrtle beech (*Nothofagus cunninghamii*).

IDReC
Infectious Disease Research Centre

Some Members of the Research Group

Postgraduate students

Sarojinie Fernando

PhD (recently submitted)

Kate Richards

PhD (year 2)

Lyndal Henden

Honours

Brigid Betz-Stablein

PhD (year 3)

Khair Jones

PhD (year 1)

Staff with interest in infectious disease applications

Prof Martin Hazelton

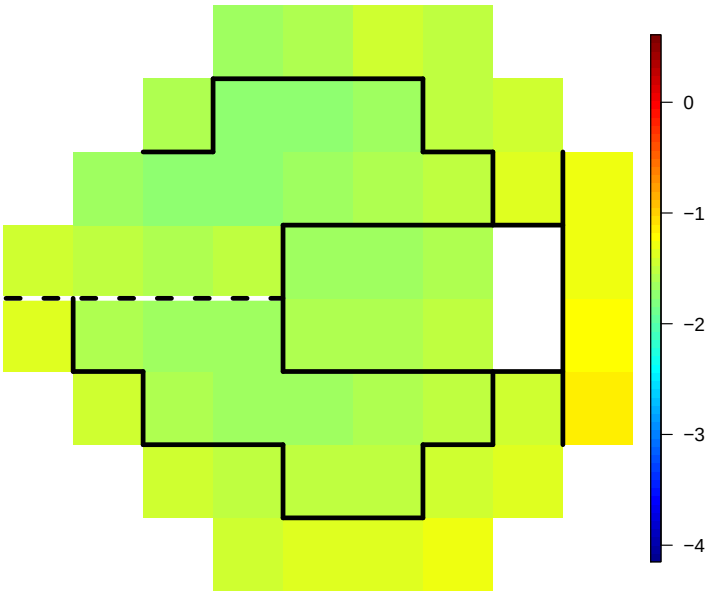
A/Prof Geoff Jones

Dr Chris Jewell

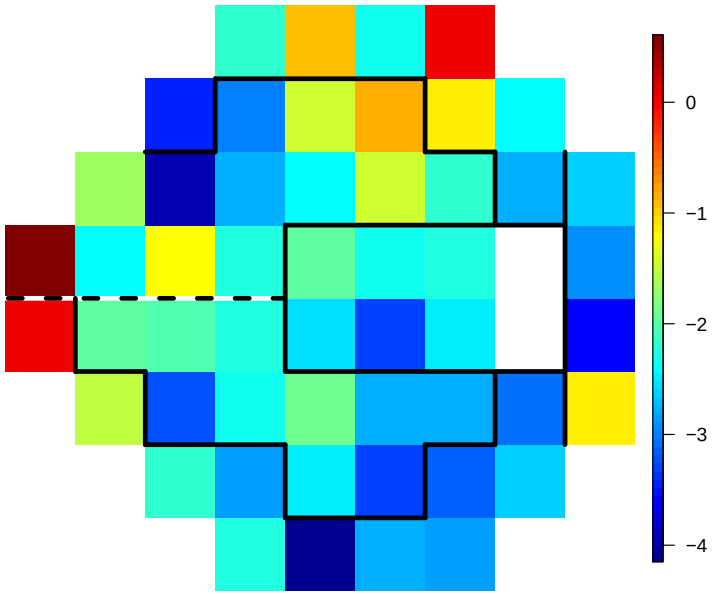
Dr Jonathan Marshall

IDReC Infectious Disease Research Centre

Modelling Progression of Eye Disease



(a) SPROG method



(b) PLR method

- Using ideas from geographical disease mapping to examine spatial patterns of damage over retina.
- Interpretation of visual field data must account for physiology and optics.

IDReC
Infectious Disease Research Centre

Some Members of the Research Group

Postgraduate students

Sarojinie Fernando
PhD (recently submitted)

Kate Richards
PhD (year 2)

Lyndal Henden
Honours

Brigid Betz-Stablein
PhD (year 3)

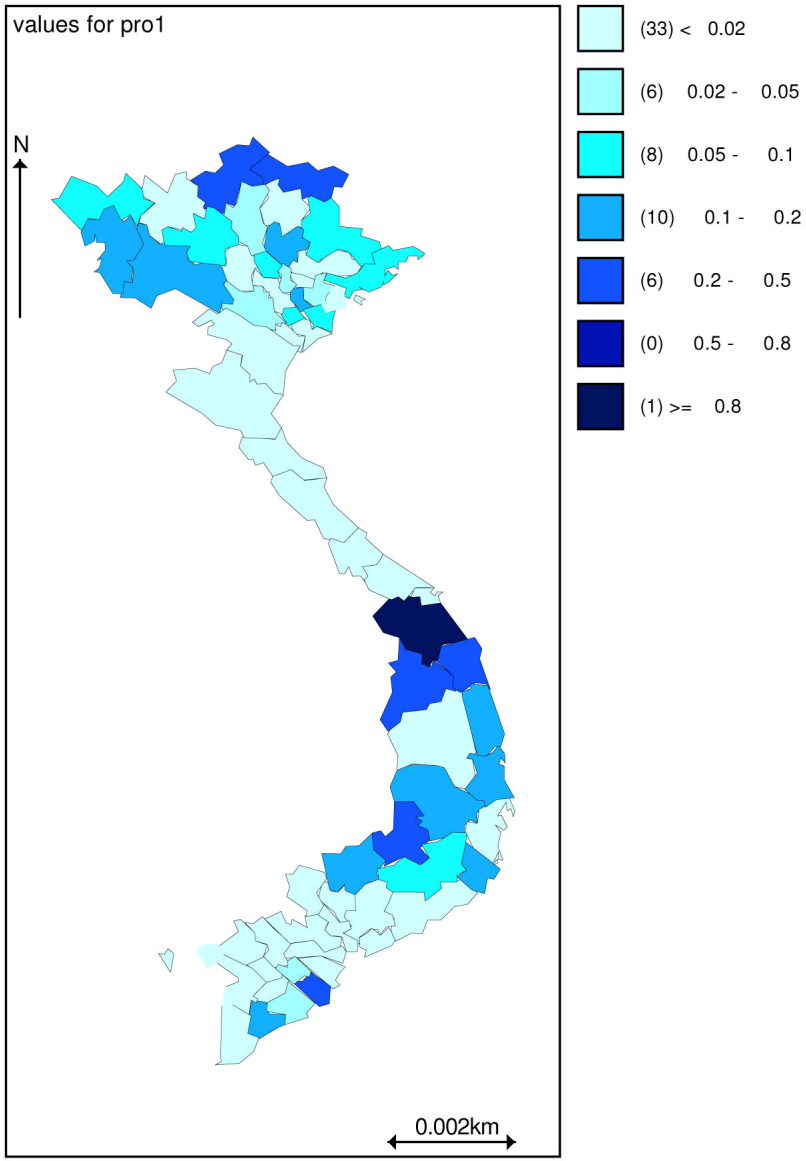
Khair Jones
PhD (year 1)

Staff with interest in infectious disease applications

Prof Martin Hazelton
A/Prof Geoff Jones
Dr Chris Jewell
Dr Jonathan Marshall

IDReC
Infectious Disease Research Centre

Modelling Foot and Mouth Disease in Vietnam



- Data quality issues for FMD in Vietnam.
- Appropriate level of data aggregation for modelling?
- Developing methods for identifying anomalies
  - High risk provinces.
  - Under-reporting.

IDReC
Infectious Disease Research Centre

Some Members of the Research Group

Postgraduate students

Sarojinie Fernando
PhD (recently submitted)

Kate Richards
PhD (year 2)

Lyndal Henden
Honours

Brigid Betz-Stablein
PhD (year 3)

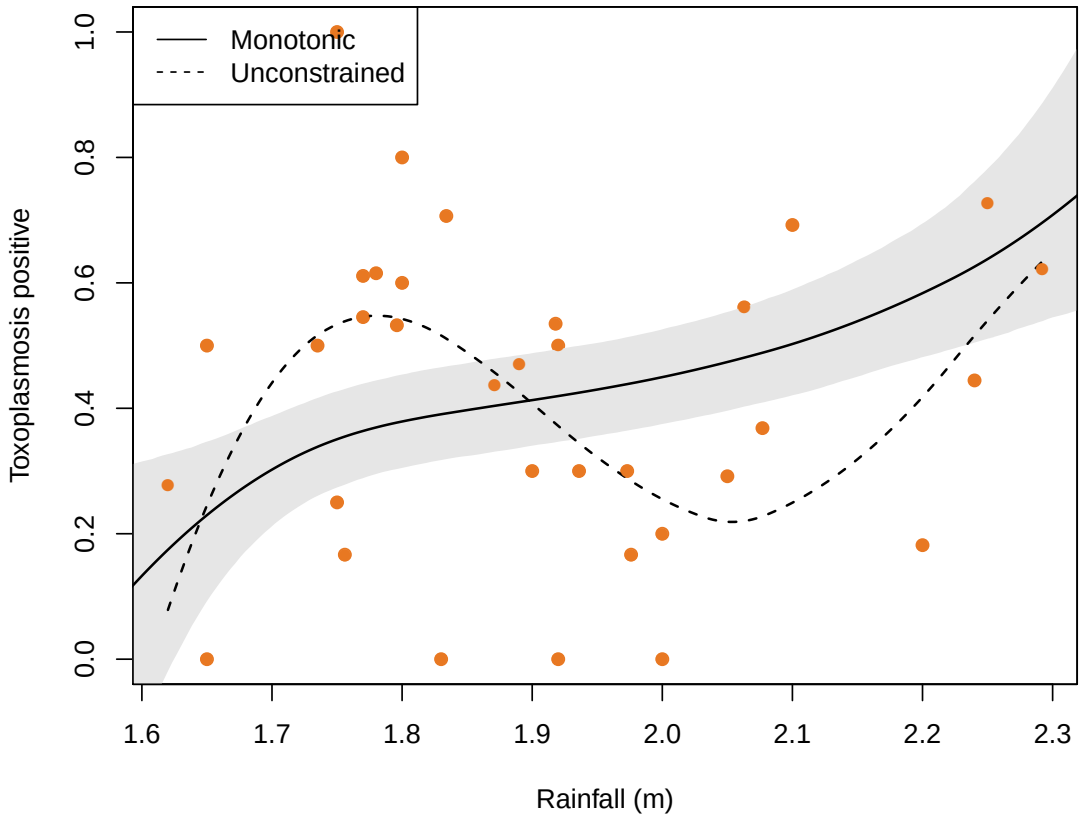
Khair Jones
PhD (year 1)

Staff with interest in infectious disease applications

Prof Martin Hazelton
A/Prof Geoff Jones
Dr Chris Jewell
Dr Jonathan Marshall

IDReC
Infectious Disease Research Centre

Bayesian Constrained Smoothing Problems



- In some problems want flexible smooth fit under shape constraints.
- Application to toxoplasmosis data from 34 cities in El Salvador.

IDReC
Infectious Disease Research Centre

Some Members of the Research Group

Postgraduate students

Sarojinie Fernando
PhD (recently submitted)

Kate Richards
PhD (year 2)

Lyndal Henden
Honours

Brigid Betz-Stablein
PhD (year 3)

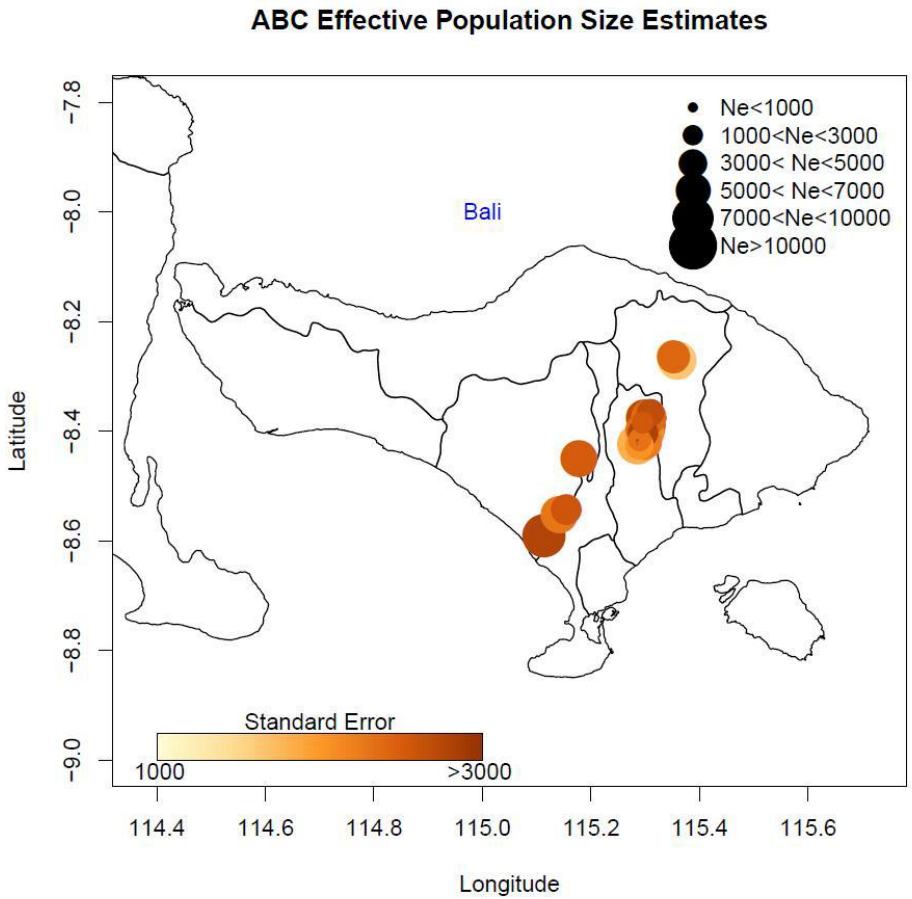
Khair Jones
PhD (year 1)

Staff with interest in infectious disease applications

Prof Martin Hazelton
A/Prof Geoff Jones
Dr Chris Jewell
Dr Jonathan Marshall

IDReC
Infectious Disease Research Centre

Methods of Simulation Based Inference



- Modern complex stochastic models often have intractable distribution theory.
- Standard methods of inference infeasible.
- We are working on improving simulation based approaches (e.g. ABC).
- Application is estimation of effective population sizes in Bali using coalescent model.

IDReC
Infectious Disease Research Centre

TO BE CONTINUED...