SPRING SCHOOL OF TOLEDO:
Advances and challenges in space-time modelling of natural events

MODEL BASED GEOSTATISTICS [1]

Paulo Justiniano Ribeiro Jr

LEG: Laboratório de Estatística e Geoinformação
Universidade Federal do Paraná and ESALQ/USP (Brasil)

in collaboration with

Peter J Diggle

Lancaster University (UK) and
Johns Hopkins University School of Public Health (USA)

Universidad de Castilla la Mancha
Toledo, Spain, March, 17–21, 2010

---

[1] *modified from previous course notes from PJD & PJRJr and (mostly) based on Diggle & Ribeiro Jr (2007) "Model Based Geostatistics", Springer.*

An announcement:

IBC-FLORIPA-2010

International Biometrics Conference

- Organised by: IBS, Rbras, RArg
- 05-10 december de 2010, Florianópolis, SC, Brasil
- satelite events (opened to proposals)
- free day/excursions on wednesday, 07/12
- http://www.ibc-floripa-2010.org
  http://www.tibs.org
- submissions now opened

IBC-FLORIPA-2010: Dec, 5-12, 2010

SPRING SCHOOL OF TOLEDO:
Advances and challenges in space-time modelling of natural events

MODEL BASED GEOSTATISTICS [1]

Paulo Justiniano Ribeiro Jr

LEG: Laboratório de Estatística e Geoinformação
Universidade Federal do Paraná and ESALQ/USP (Brasil)

in collaboration with

Peter J Diggle

Lancaster University (UK) and
Johns Hopkins University School of Public Health (USA)

Universidad de Castilla la Mancha
Toledo, Spain, March, 17–21, 2010

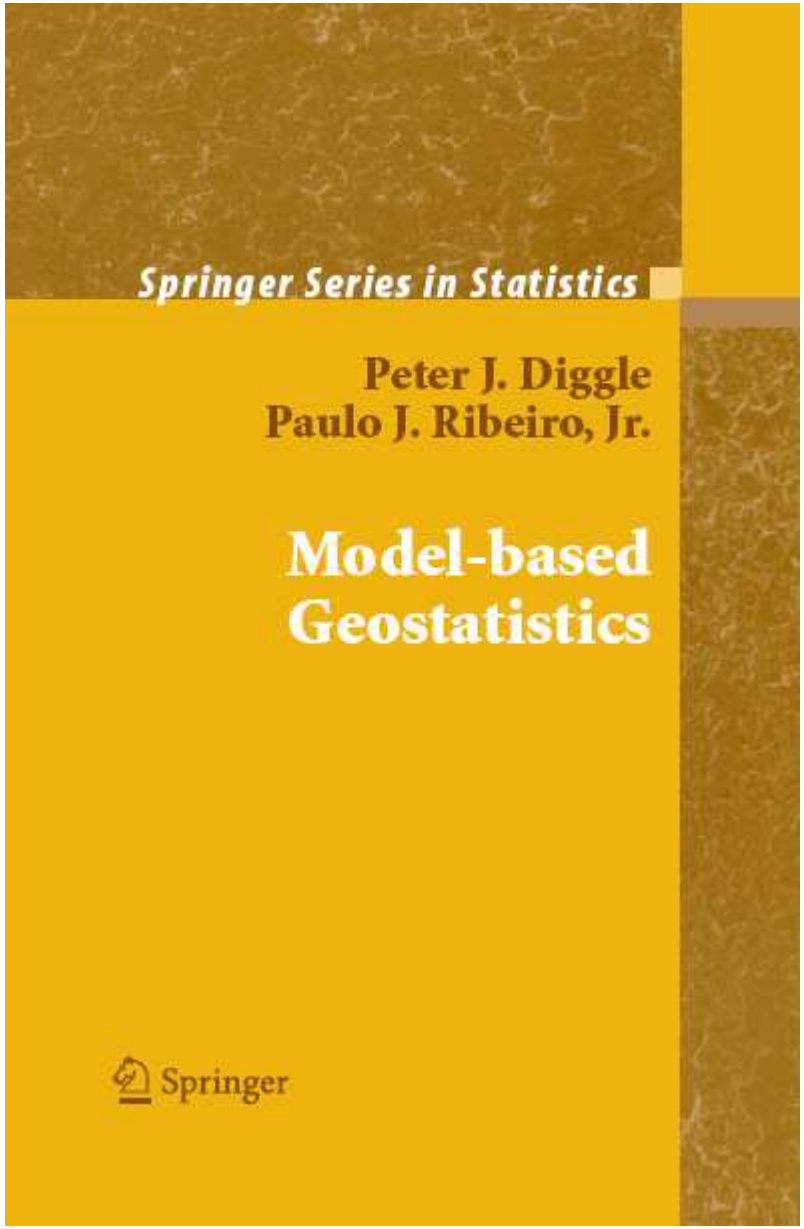
---

[1] *modified from previous course notes from PJD & PJRJr and (mostly) based on Diggle & Ribeiro Jr (2007) "Model Based Geostatistics", Springer.*

Outline

An overview of
Diggle, P.J. & Ribeiro Jr, P.J. *Model Based Geostatistics*, Springer, 2007

- Basics of geoestatistical models
- Inference and prediction
- Some topics and extensions
- Case studies



PART 1

Introduction, examples and modelling

Context for spatial data

- (geo)-referenced data
- GIS - Geographical Information Systems
- Spatial Analysis
- Spatial Statistics
  - discrete spatial variation
  - continuous spatial variation
    - point process
    - geostatistical or point referenced data
  - mixed and . . . more complex structures
- methods now expanded to a wider context

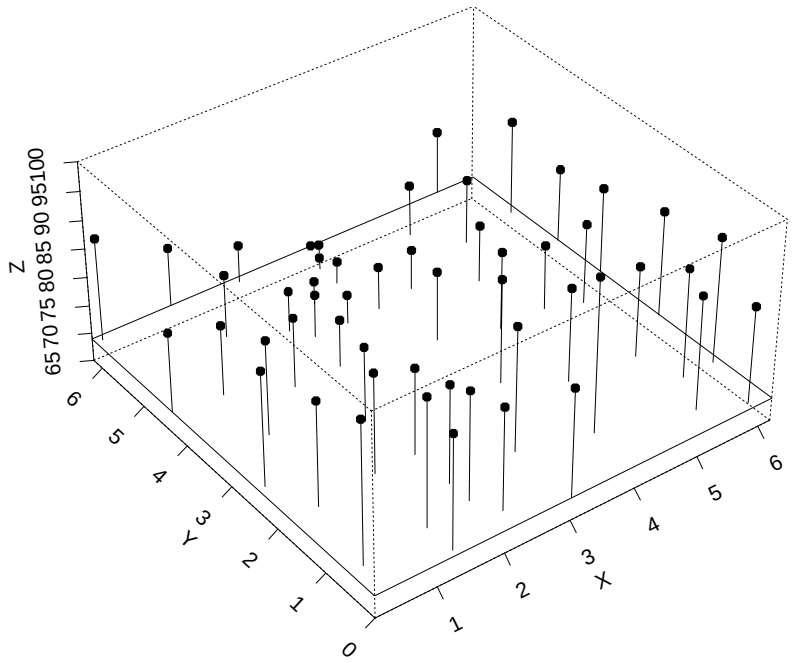
Geostatistics

- traditionally, a self-contained methodology for spatial prediction, developed at École des Mines, Fontainebleau, France
- nowadays, that part of spatial statistics which is concerned with data obtained by spatially discrete sampling of a spatially continuous process

Model-based Geostatistics

- the application of general principles of statistical modelling and inference to geostatistical problems
- Example: kriging as minimum mean square error prediction under Gaussian modelling assumptions
- framework for tackling problems by exploring and extending the basic model

Example 1.1: Measured surface elevations



`require(geoR) ; data(elevation) ; ?elevation`

Potential distinction between $S(x)$ and $Y(x)$

Summary(I): Terminology and Notation

- $(Y_i, x_i) : i = 1, \ldots, n$ basic format for geostatistical data
- $\{x_i : i = 1, \ldots, n\}$ is the sampling design
- in principle x_i is fixed or stochastically independent of Y_i
- $\{Y(x) : x \in A\}$ is the measurement process
- $\{S(x) : x \in A\}$ is the signal process, assumed underlying stochastic process
- $T = \mathcal{F}(S)$ is the target for prediction
- $[S, Y] = [S][Y|S]$: specification of the geostatistical model

Summary(II):A canonical geostatistical data analysis

Basic steps:

- exploratory data analysis
- model choice
- inference on the model parameters
- spatial prediction

Assumptions:

- stationarity (translation) global mean, variance and spatial correlation
- isotropy (rotation)
- Gaussianity

Summary(III):Core Geostatistical Problems

Design

- how many locations?
- how many measurements?
- spatial layout of the locations?
- what to measure at each location?

Modelling

- probability model for the signal, $[S]$
- conditional probability model for the measurements, $[Y|S]$

Estimation

- assign values to unknown model parameters
- make inferences about (functions of) model parameters

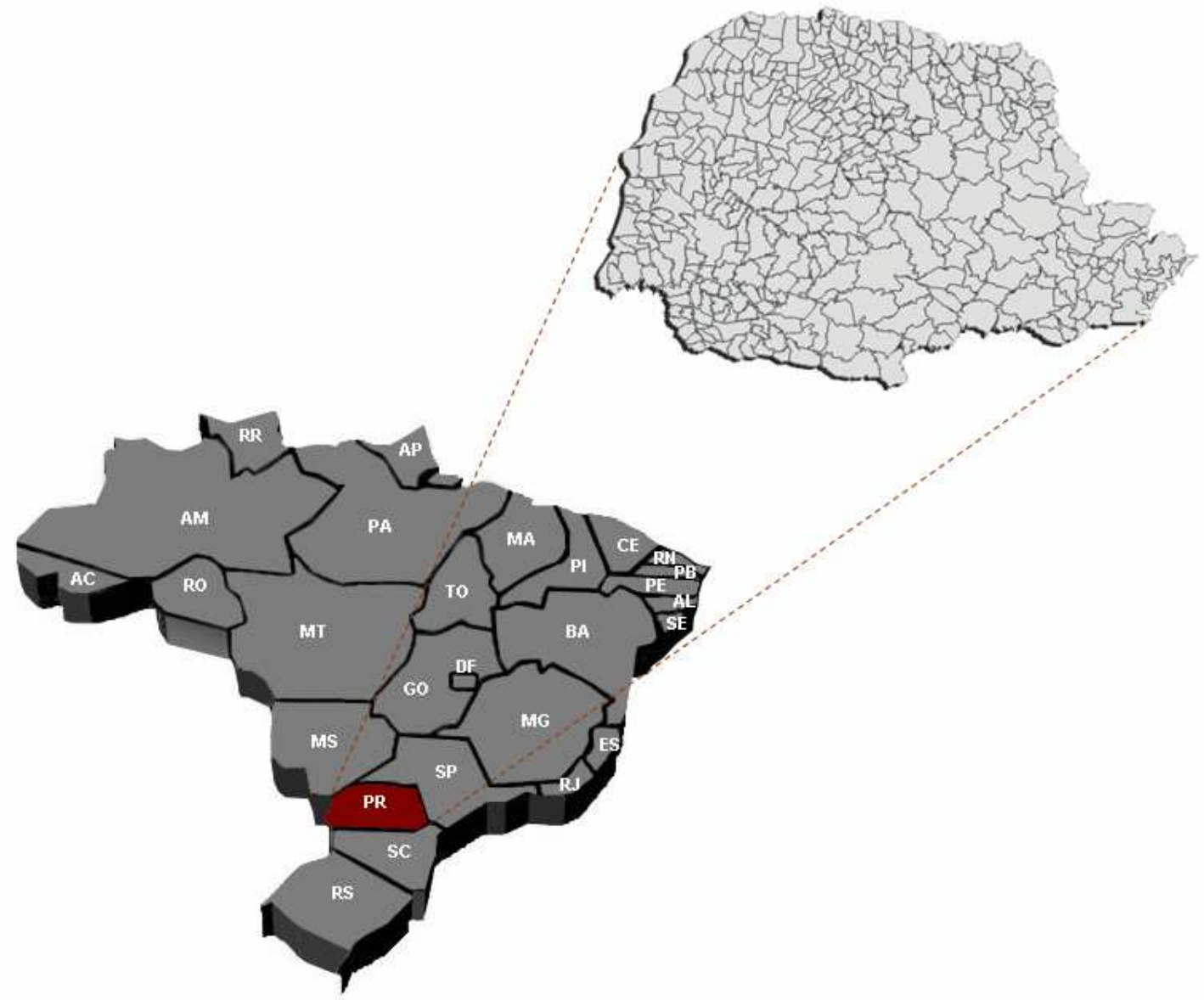
Prediction

- evaluate $[T|Y]$, the conditional distribution of the target given the data

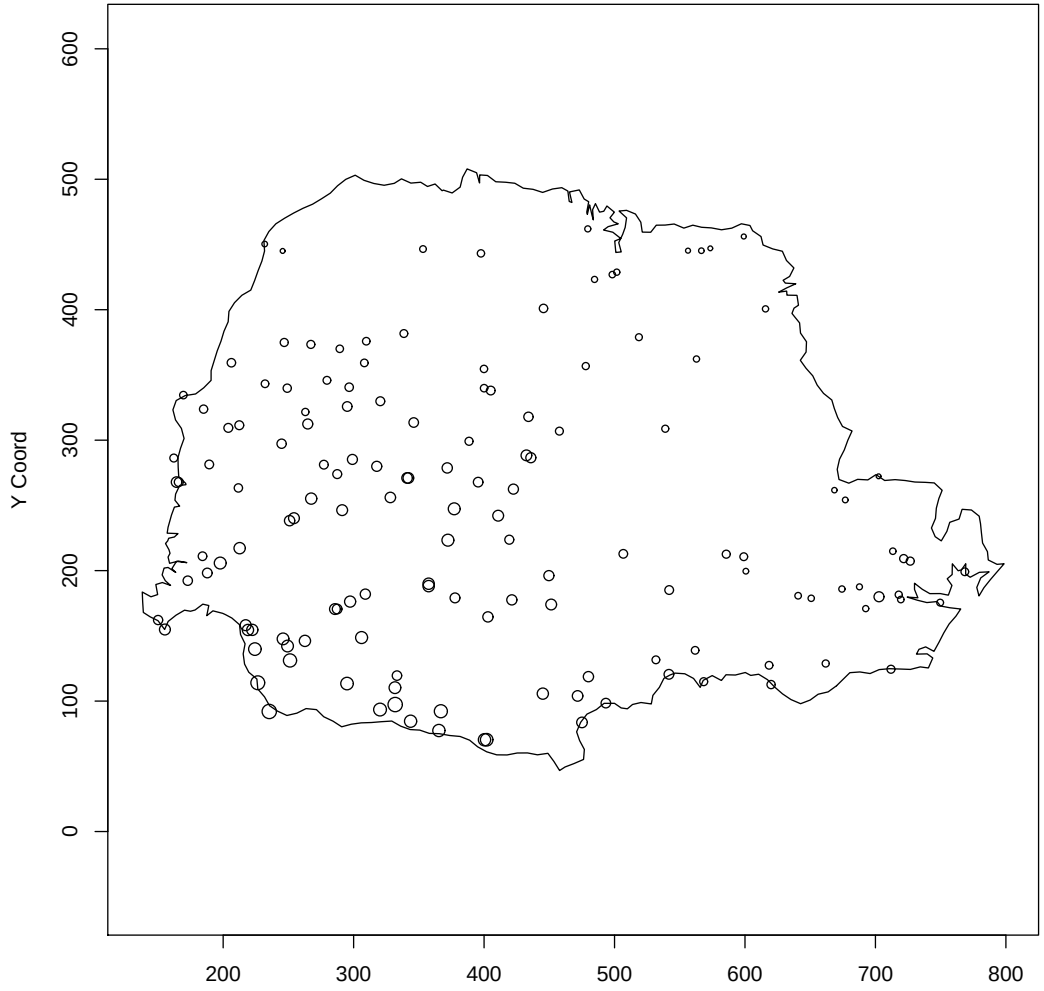
Motivating examples

In the following examples we should identify:

- **the structure of the available data**
- **the nature of the response variable(s)**
- **potential covariates**
- **the underlying (latent) process(es)**
- **the scientific objectives**
- **combine elements/features for a possible statistical model**

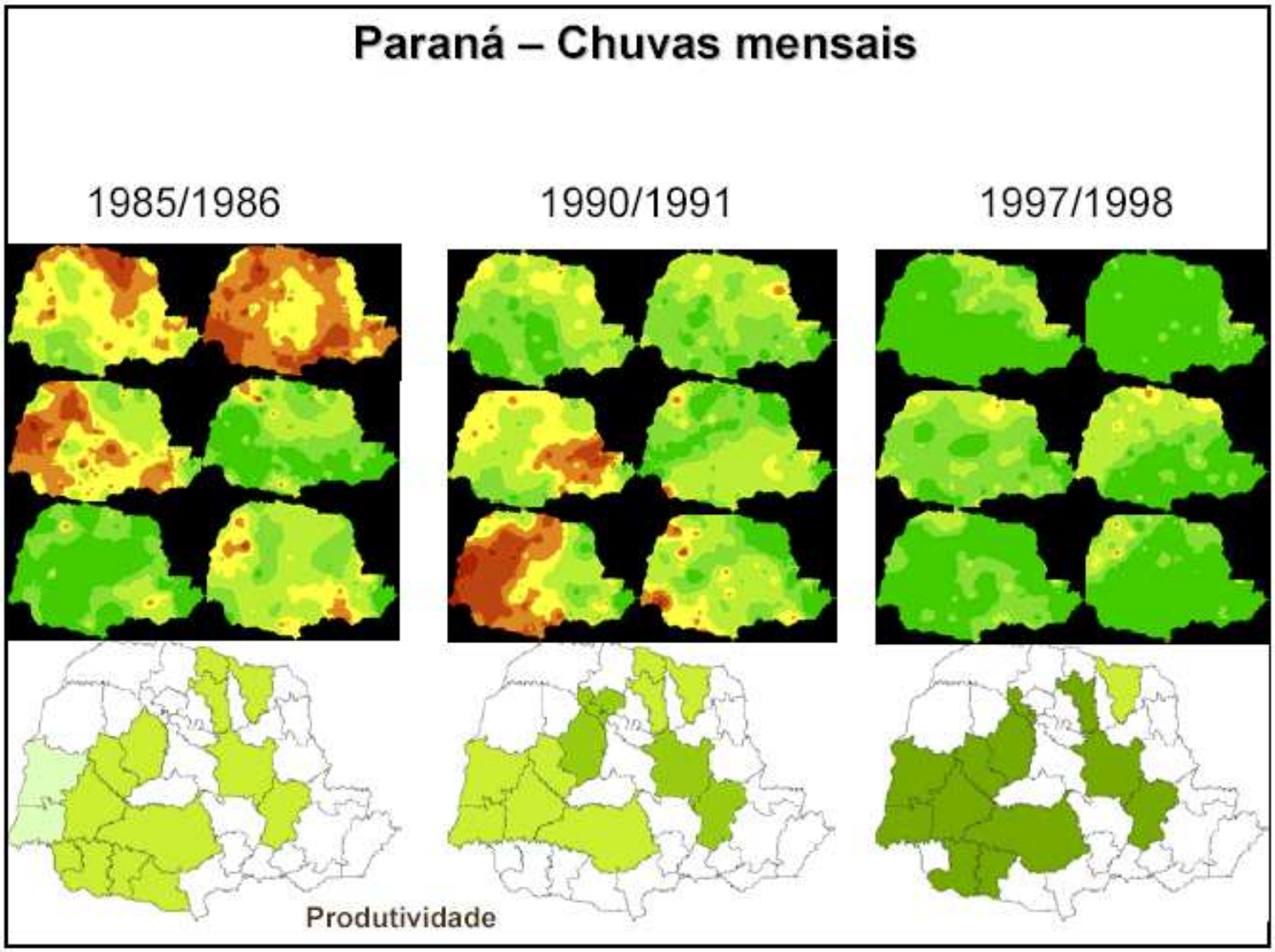

RR
AP
AM
PA
MA
CE
RN
PB
PI
PE
AL
AC
RO
TO
SE
MT
BA
DF
GO
MG
MS
ES
SP
RJ
PR
SC
RS

Example 1.2.1: Paraná rainfall data

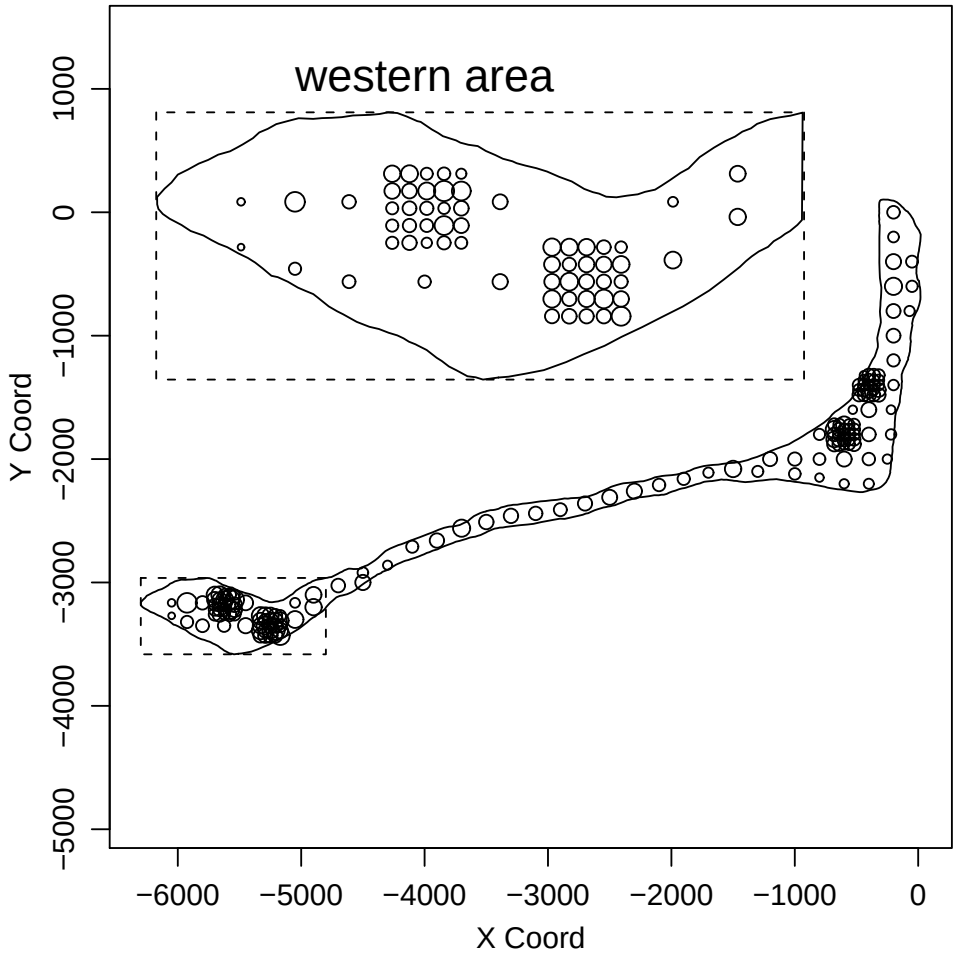



```
require(geoR) ; data(parana) ; ?parana ; points(parana)
```

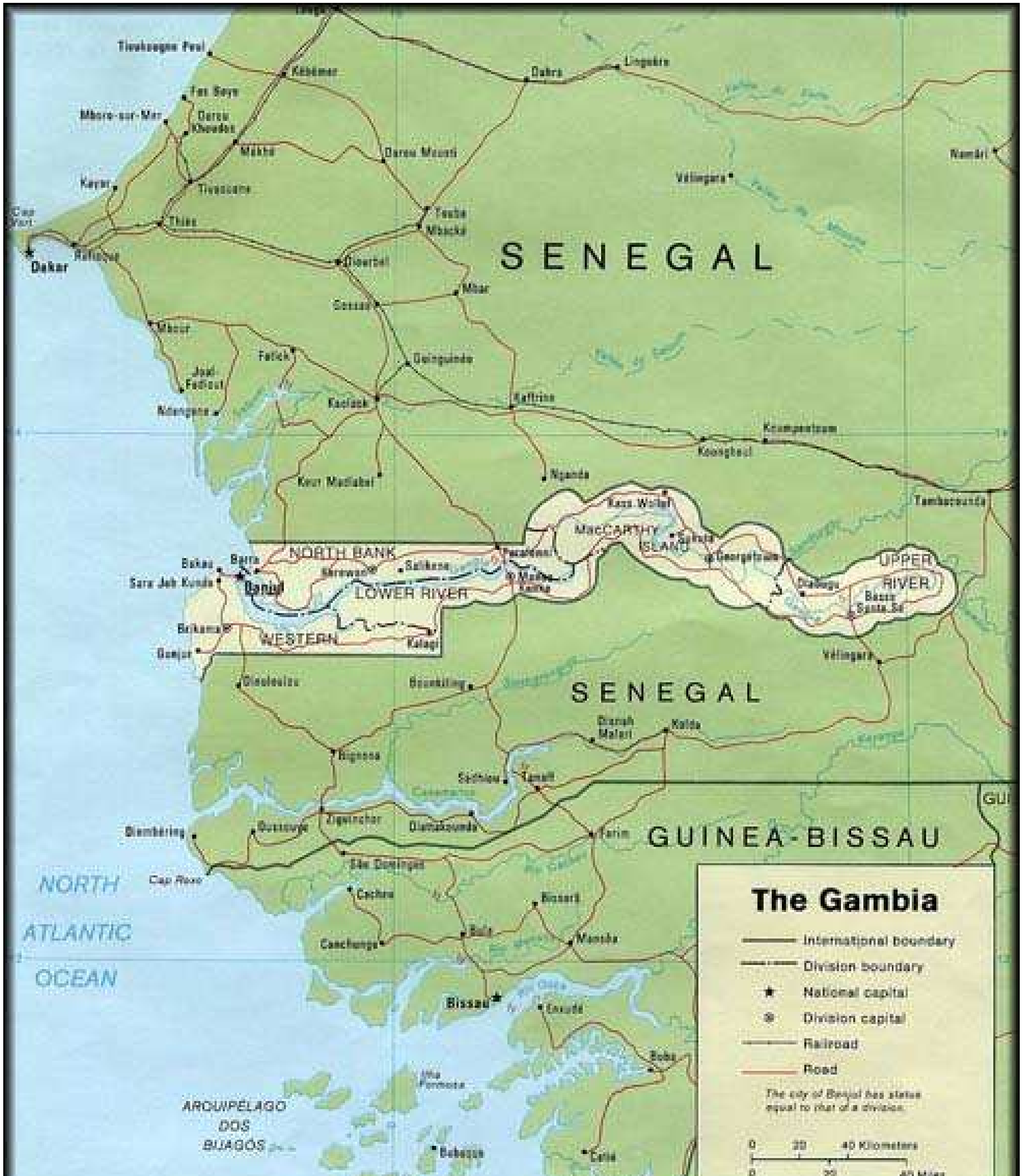
Example 1.2.3: Favorable/risk zones



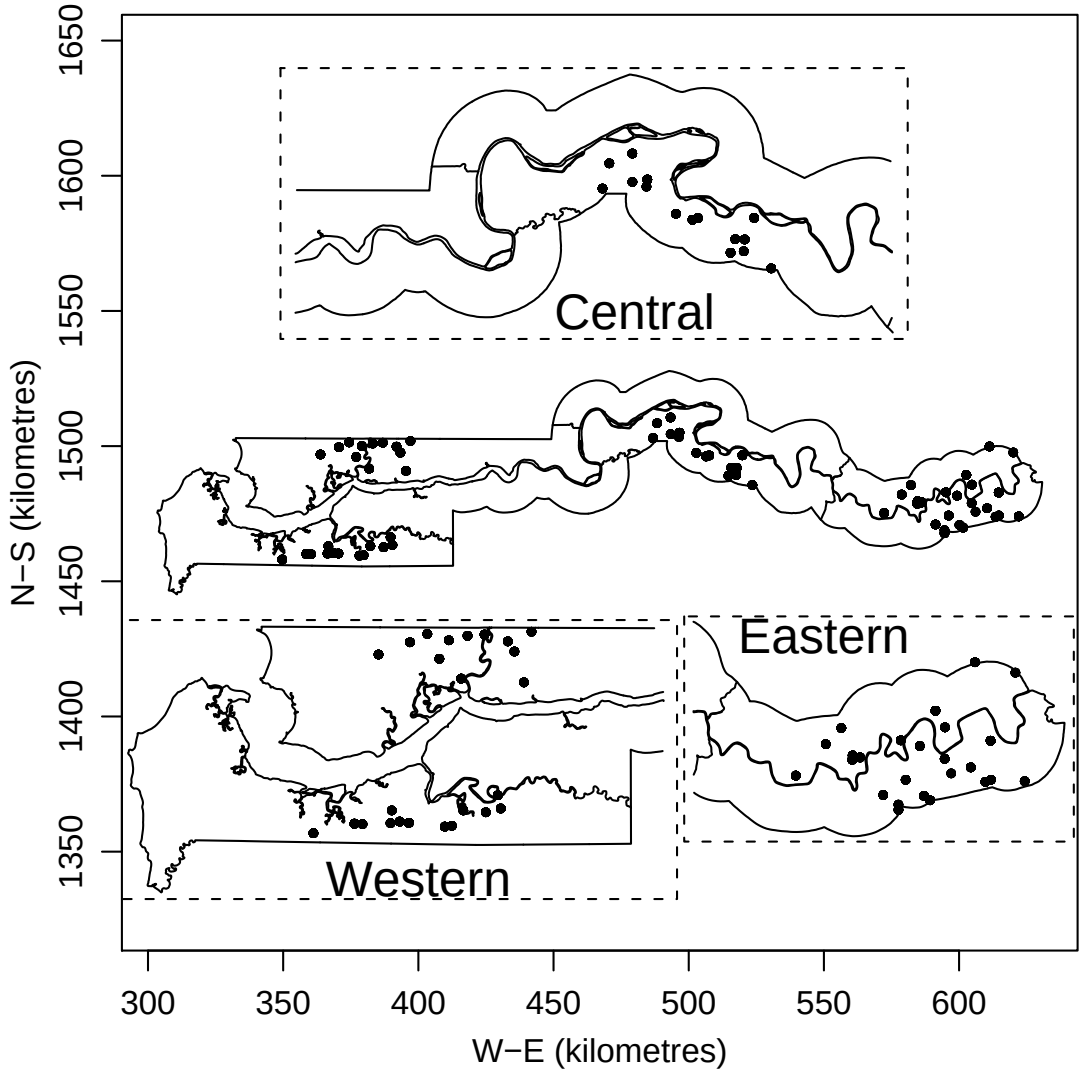
Example 1.3: Residual contamination from nuclear weapons testing



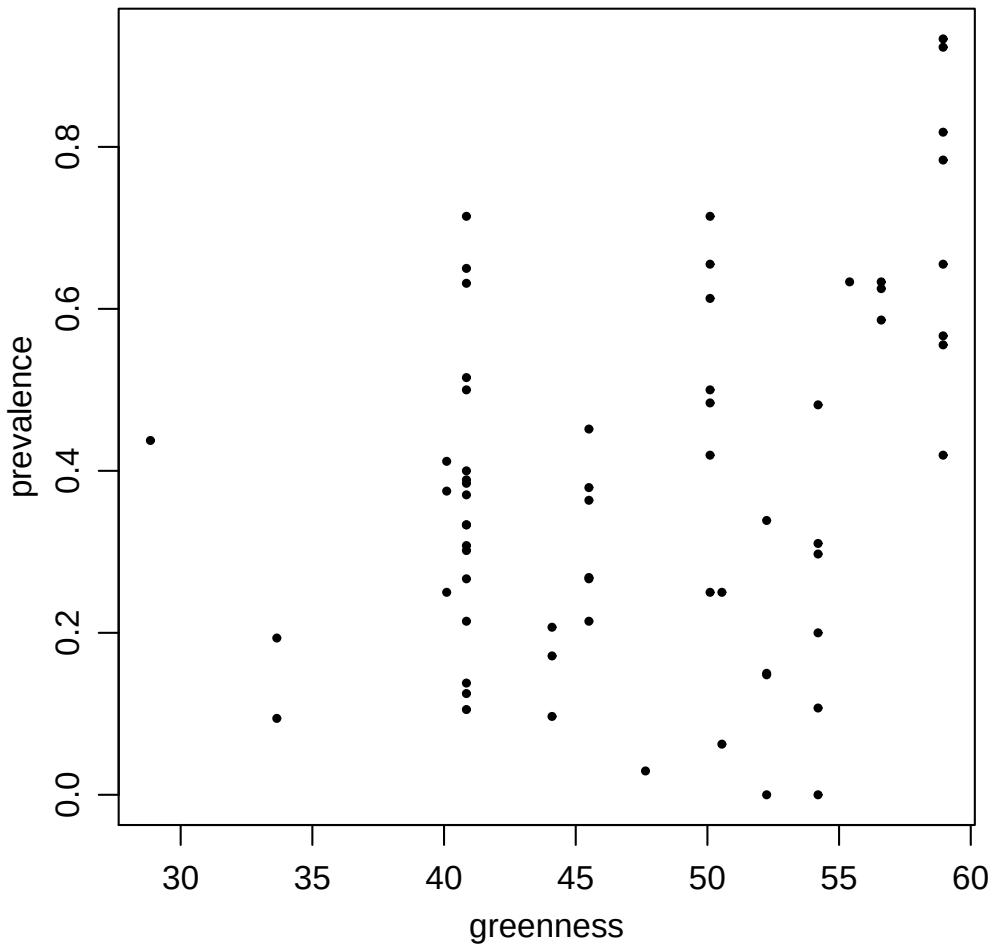

Sunset at Georgetown


SENEGAL
SENEGAL
GUINEA-BISSAU
NORTH ATLANTIC OCEAN
The Gambia
International boundary
Division boundary
National capital
Division capital
Railroad
Road
The city of Banjul has status equal to that of a division.
NORTH BANK
LOWER RIVER
WESTERN
MacCARTHY ISLAND
UPPER RIVER
Banjul
Dakar
Bissau
Kaolack
Thiès
Tambacounda
Ziguinchor
Kolda
Dahra
Linguère
ARQUIPÉLAGO DOS BIJAGÓS

Example 1.4: Childhood malaria in Gambia

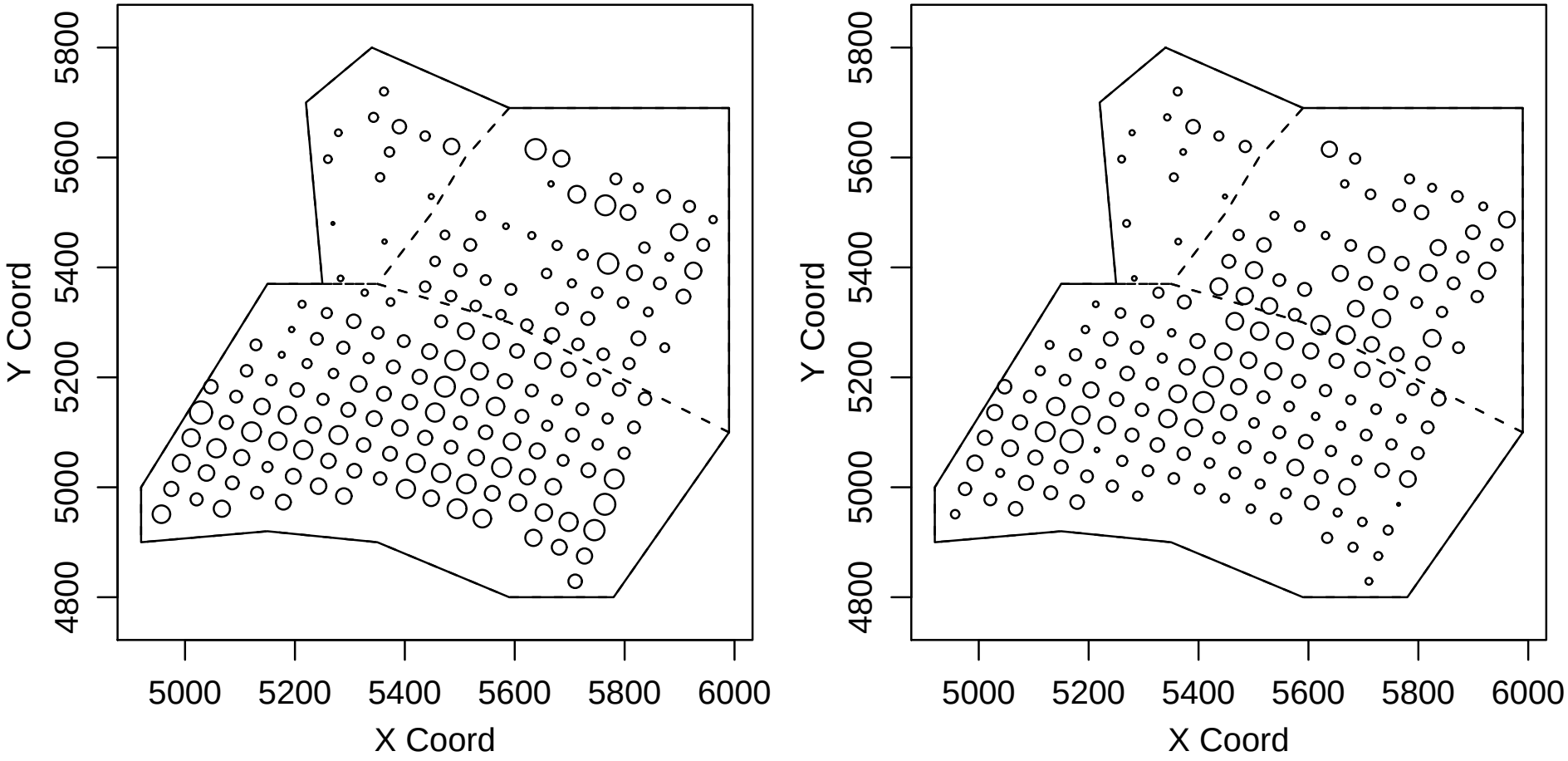


Example 1.4: continued



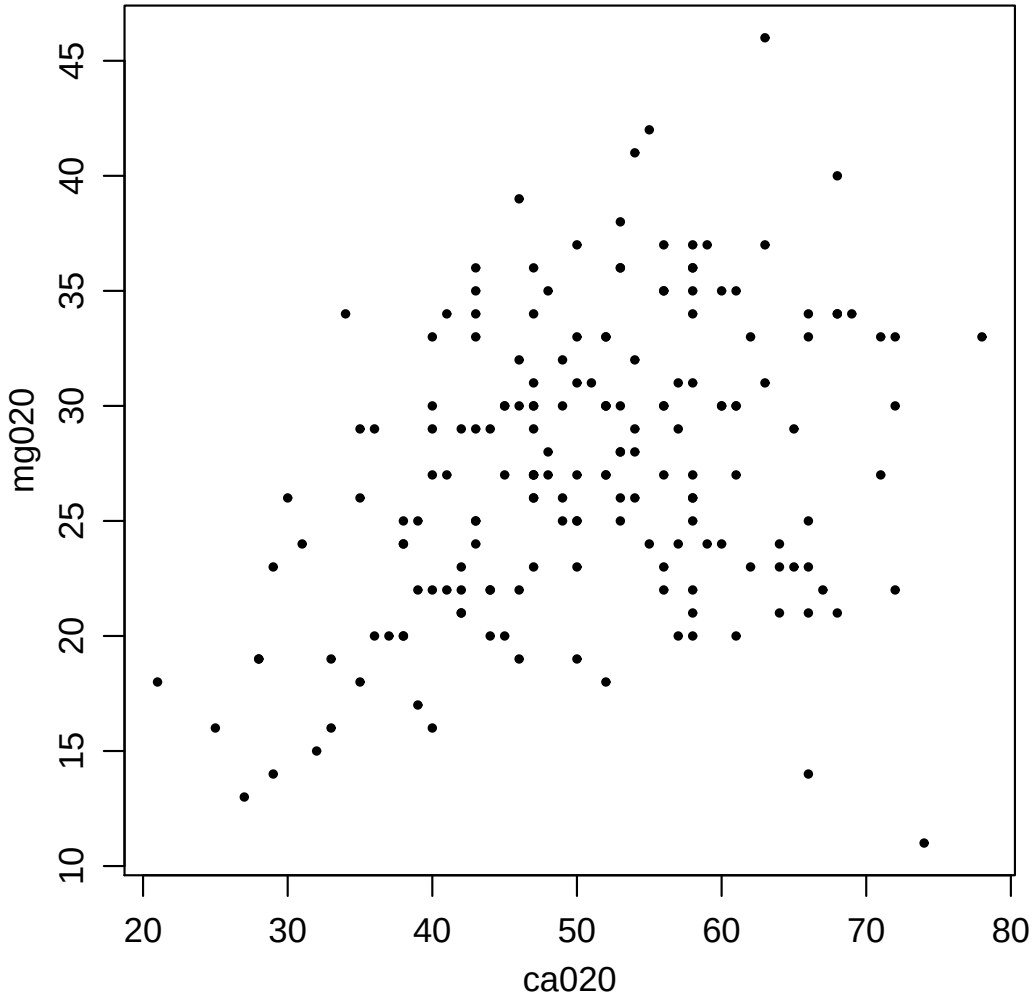
Correlation between prevalence and green-ness of vegetation?

Example 1.5: Soil data



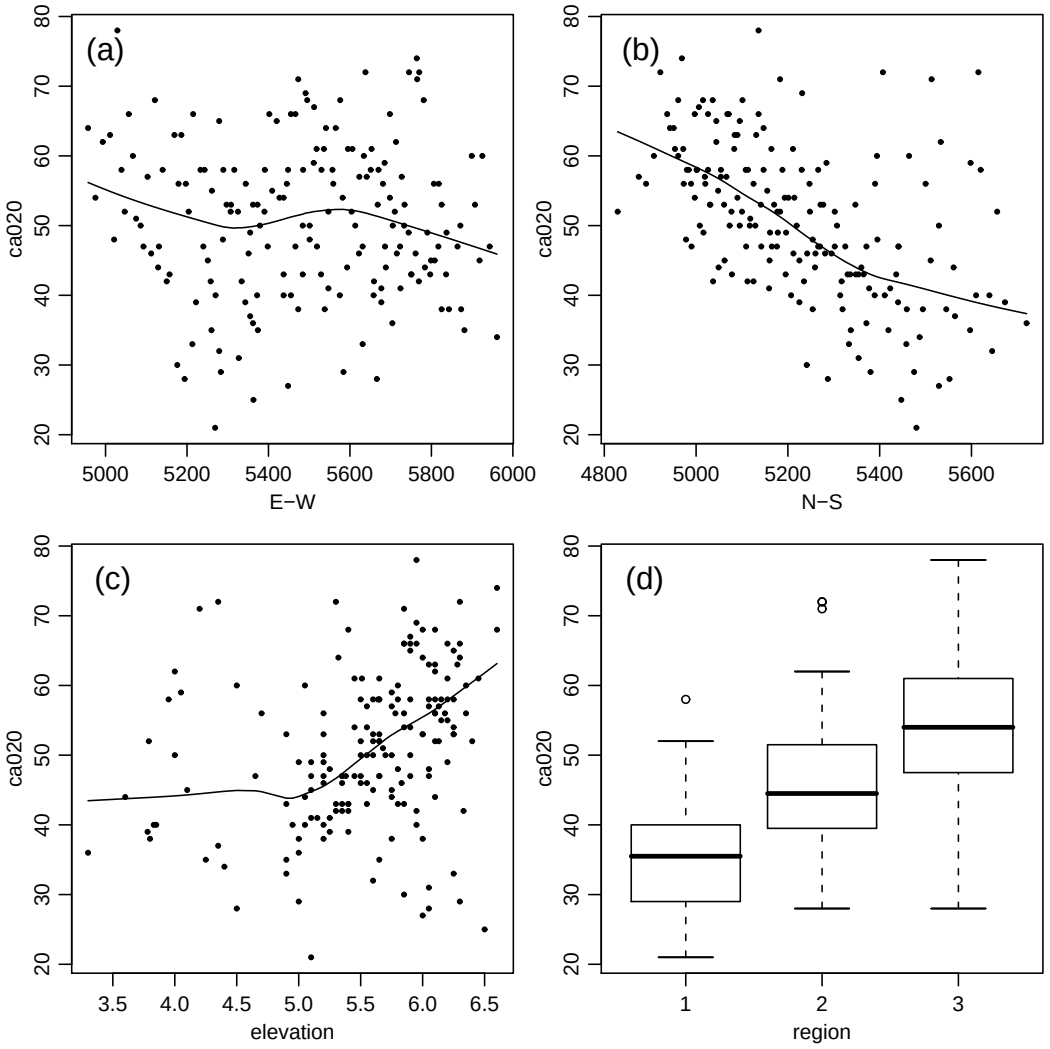
Ca (left-panel) and Mg (right-panel) concentrations

Example 1.5: Continued



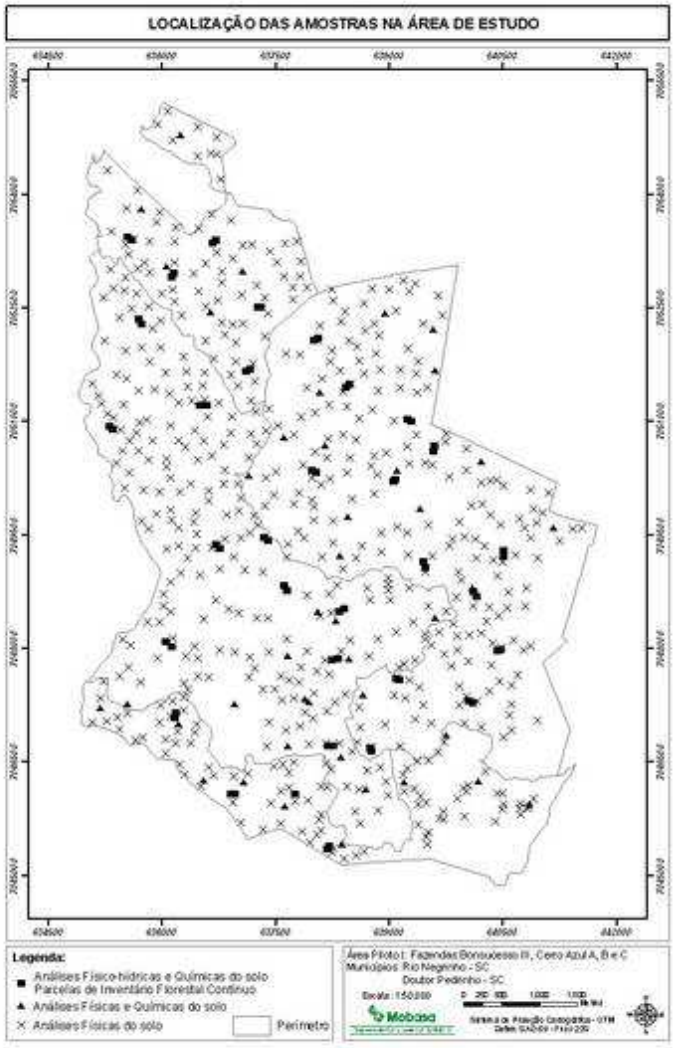
Correlation between local Ca and Mg concentrations.

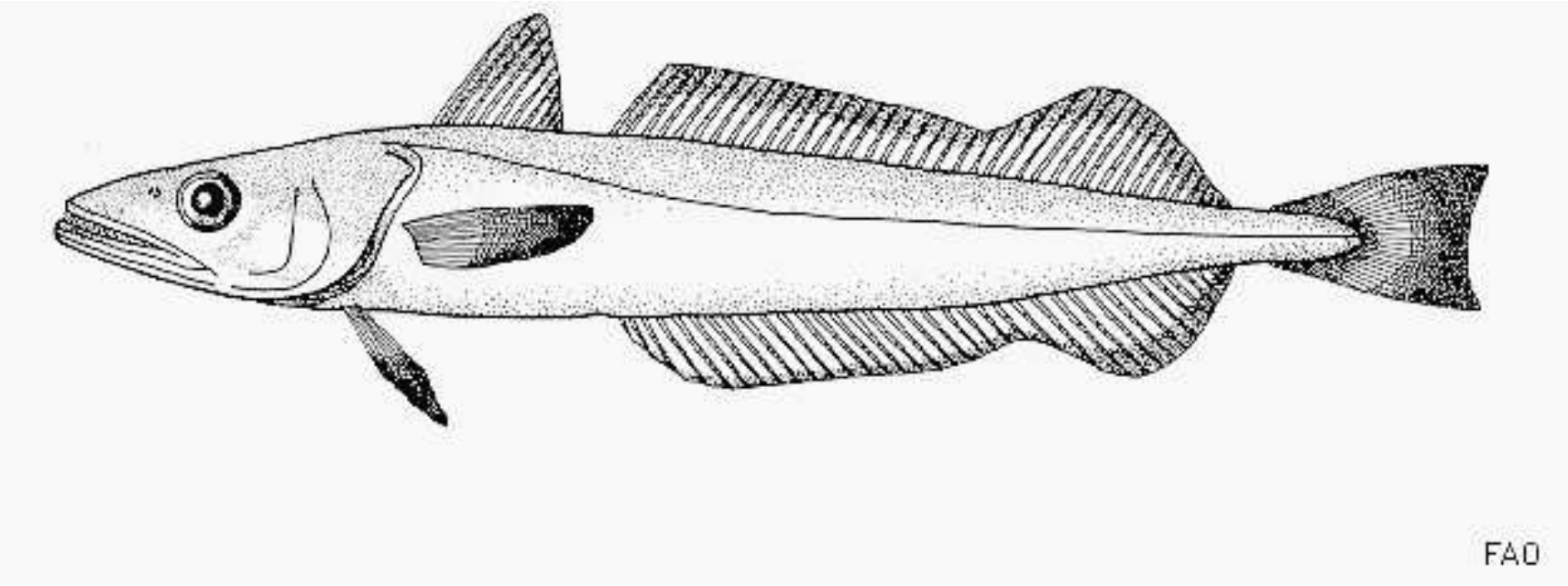
Example 1.5: Continued



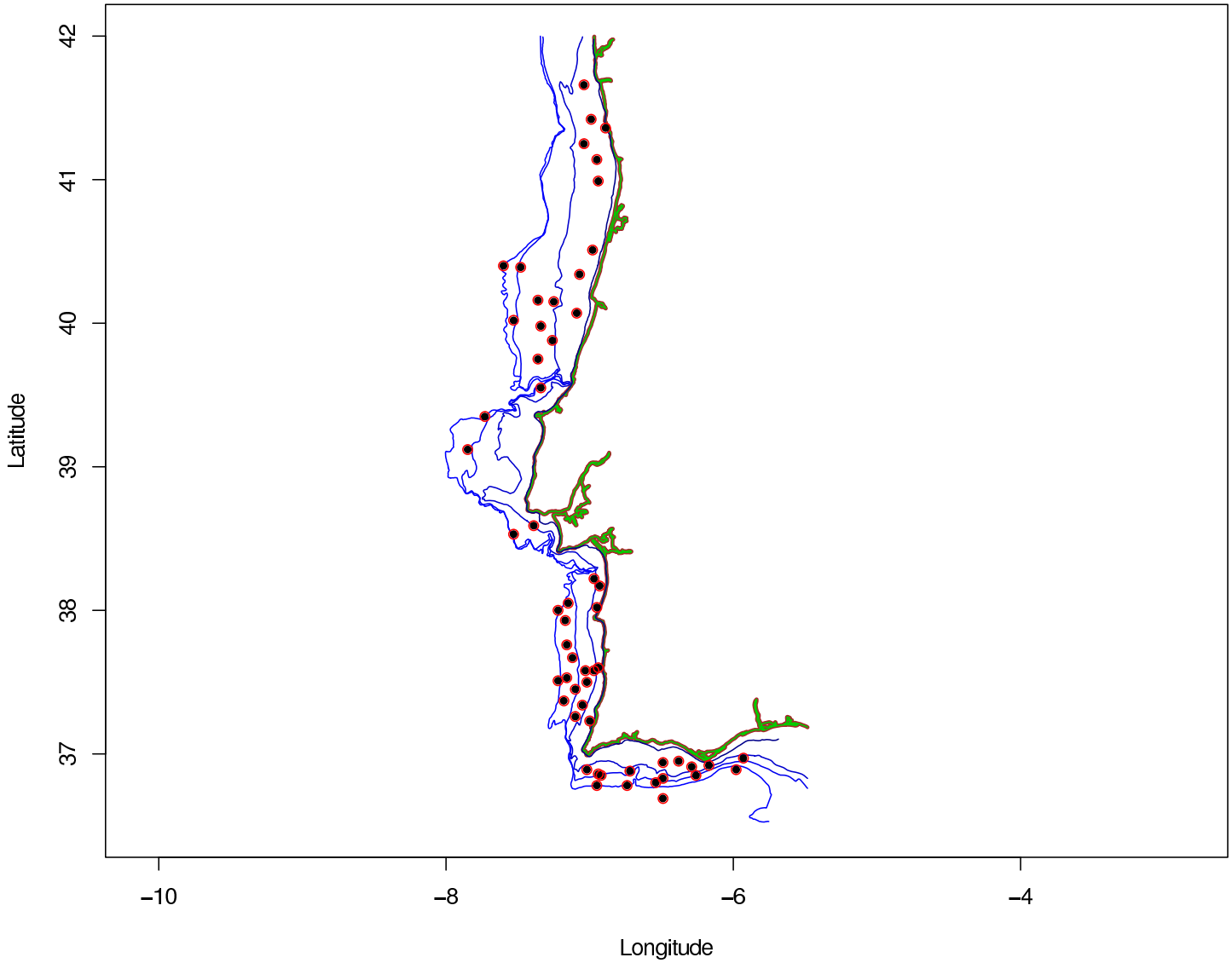
Covariate relationships for Ca concentrations.

Example 1.5: Forestry inventory



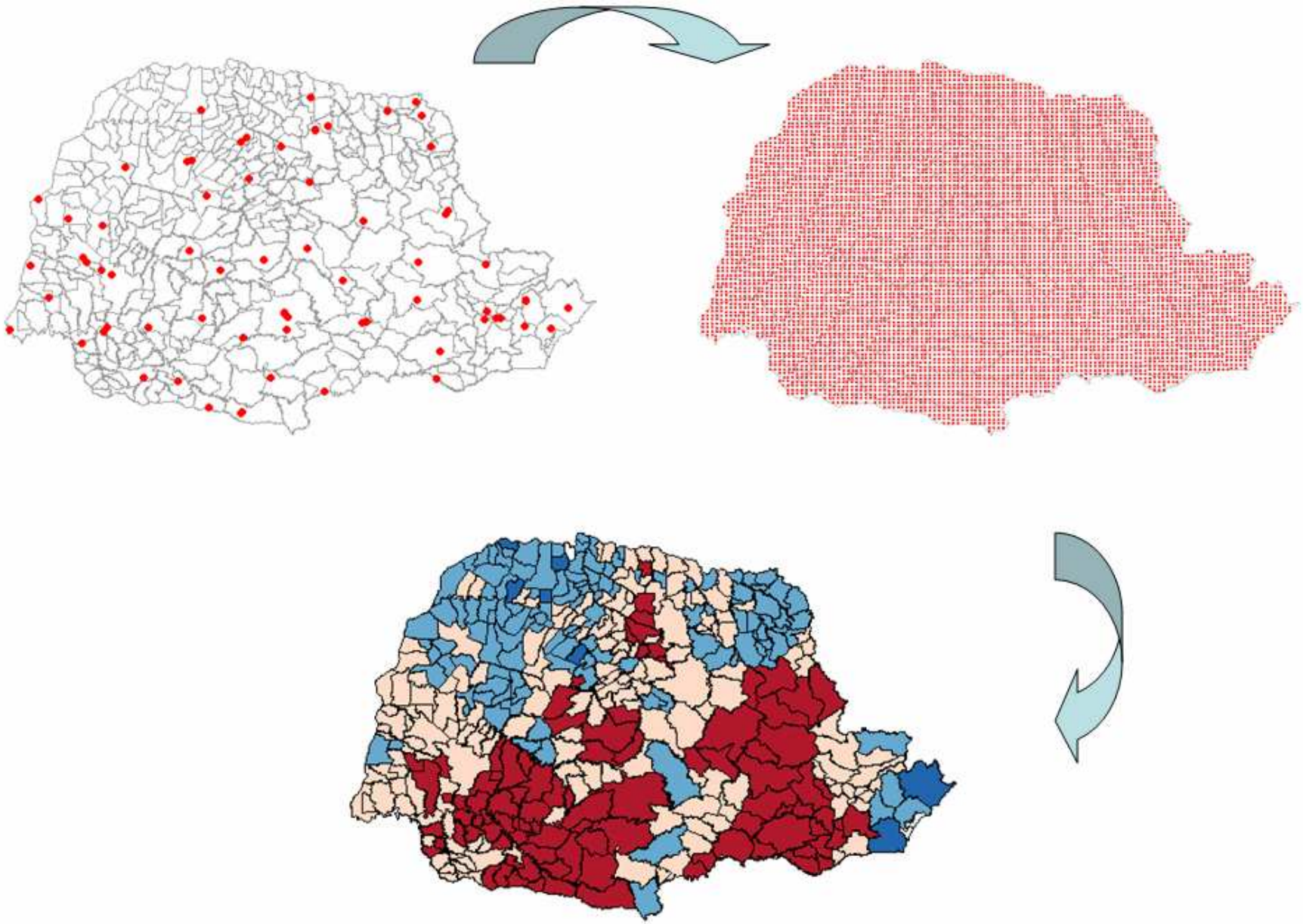

FAO

Example 1.6: Fish stocks (Hake)



Support

- x_i is in principle a point, but sometimes measurements are taken on (maybe small) portions
- revisiting the examples (e.g. elevation and rongelap) we can see contrasting situations
- $S(x) = \int w(r) S^*(x - r) dr$
- smoothness of $w(s)$ constrains allowable forms for the correlation function
- support *vs* data from discrete spatial variation
- (in)compatible supports for different sources of data



Multivariate responses and covariates

- $Y(x_i)$ can be a vector of observable variable
- measurements not necessarily taken at coincident locations
- data structure (x_i, y_i, d_i) can include covariates (potential explanatory variables)
- jargon: *external trend* and *trend surface* (coordinates or functions of them as covariates)
- distinction between multivariate responses and covariates is not aways sharp and pragmatically, it may depend on the objectives and/or availability of data
- revisiting examples

Design

What and where to address questions of scientific interest

- ***elevation data*: map the true surface**
- ***Rongelap data*: short range variation and "*max*"**
- ***Paraná and Gambia data*: locations given**

How many: sample size

- **statistical criteria**
- **pratical contraints: time, costs, operational, ...**

Where: design locations

- **completely *random vs* completely *regular***
- **different motivations (e.g. estimation/prediction), need to compromise**
- ***oportunistic* designs: concerns about preferential sampling and impact on inferences**

A basic reference model

Gaussian/linear geostatistics

The model:

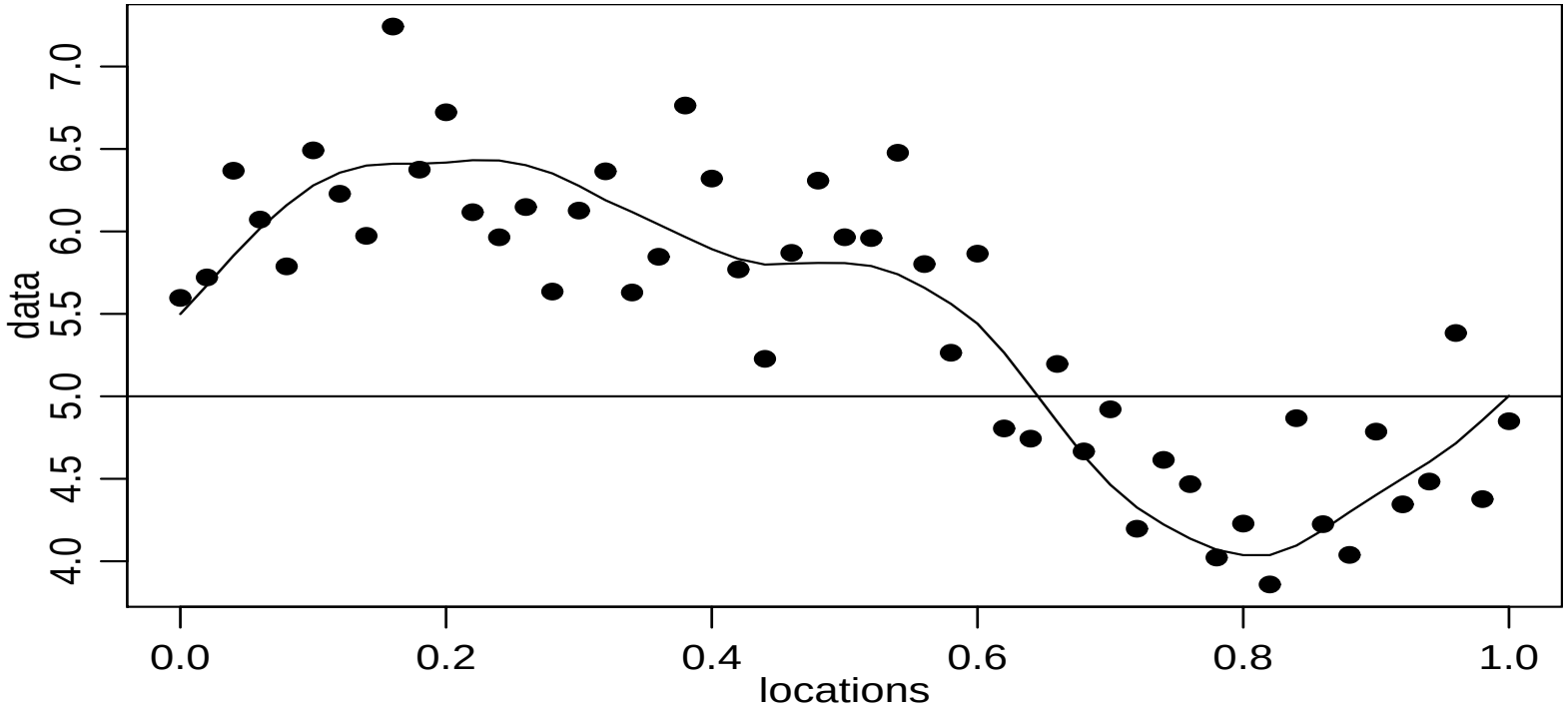
- $[Y, S] = [S][Y|S]$
- **Stationary Gaussian process** $S(x) : x \in \mathbb{R}^2$
  - $\mathrm{E}[S(x)] = \mu$
  - $\mathrm{Cov}\{S(x), S(x')\} = \sigma^2 \rho(\|x - x'\|)$
- $Y_i|S(\cdot) \overset{\text{ind}}{\sim} \mathrm{N}(S(x_i), \tau^2)$ **(conditional independence)**

Is equivalent to:

$$Y(x) = S(x) + \epsilon$$

Gaussian (linear) model

1-D Schematic representation:



- Gaussian stochastic process widely used:
  physical representation, behaviour and tractability
- underlying structure many geoestatistical methods
- benchmark for hierarquical models
- GRF or GAM ?

Some extensions

- non-constant mean model (covariates or trend surface)
  - $\mathrm{E}\left[S(x)\right] = 0$
  - $Y_i | S(\cdot) \overset{\text{ind}}{\sim} \mathrm{N}(\mu(x_i) + S(x_i), \tau^2)$
  - $Y(x) = \sum_{k=1}^{p} \beta_k d_k(x_i) + S(x) + \epsilon$
- transformation of the response variable (Box-Cox)

$$Y^* = \begin{cases} (Y^\lambda - 1)/\lambda & : \quad \lambda \neq 0 \\ \log Y & : \quad \lambda = 0 \end{cases}$$

- more general covariance functions
- non-stationary covariance structure

Comments

- word of caution:
  decision on one will probably affect the other
- spatially varying mean *vs* correlation in the response variable around the mean
- whenever possible, keep it simple!
- likelihood based measures can guide model choice

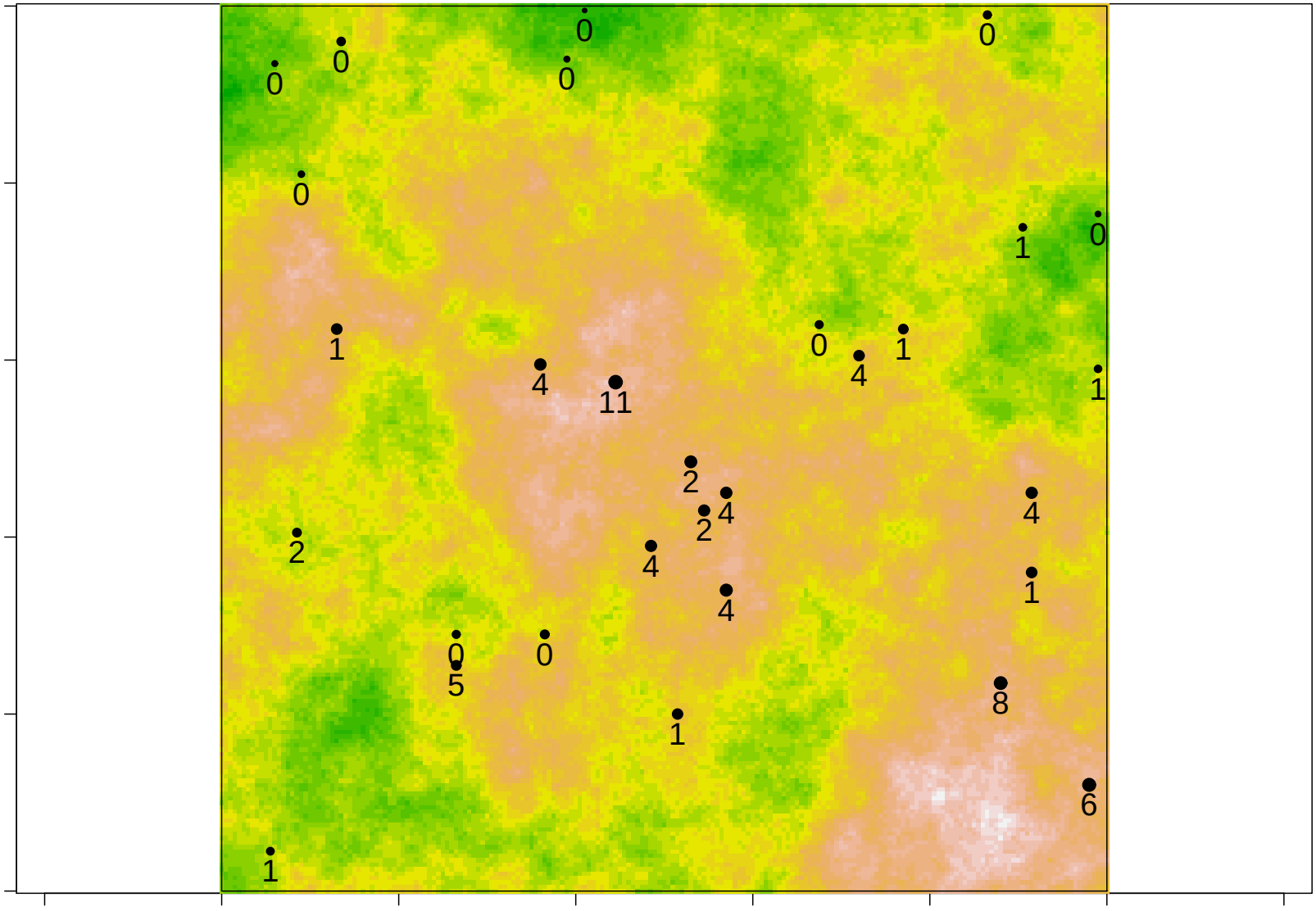
Generalised linear models

- GLM's and marginal and mixed models
- GLGM: Generalized linear geostatistical models
- Model elements:
  1. a Gaussian process $S(x)$, the *signal*
  2. data generating mechanism given the signal
  3. $Y_i|S(\cdot) \overset{\text{ind}}{\sim} \text{EF}(\mu_i, \tau^2)$ (conditional independence)
  4. relation to explanatory variables

$$h(\mu_i) = \sum_{k=1}^{p} \beta_k d_k(x_i) + S(x_i)$$

- Nugget: clear distinction between micro-scale variation and measurement error

GLGM



Characterising $S(x)$: correlation function

- a function $\mathrm{Cov}(\cdot) : \mathbb{R} \to \mathbb{R}$ is a valid covariance function iff $\mathrm{Cov}(\cdot)$ is positive definite
- core of the spatially continuous models
  - $\mathrm{Cov}\{S(x), S(x')\} = \sigma^2 \rho(\|x - x'\|) = \sigma^2 \rho(\|u\|)$
- $\rho(u)$ is positive definite if
  - $\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \mathrm{Cov}(x_i - x_j) \geq 0$ for all $a_i \in \mathbb{R}$, $x_i \in \mathbb{R}^d$
  - then, any $\sum_{i=1}^{m} a_i S(x_i)$ has a non-negative variance
- typically assuming a parametric form for $\rho(\cdot)$
- Example: exponential model $\rho(u) = \exp\{-u/\phi\}$
- stationarity assumption

Properties

1. $\mathrm{Cov}\,[\boldsymbol{Z}(\mathbf{x}), \boldsymbol{Z}(\mathbf{x}+\mathbf{0})] = \mathrm{Var}\,[\boldsymbol{Z}(\mathbf{x})] = \mathrm{Cov}\,(\mathbf{0}) \geq \mathbf{0}$

2. $\mathrm{Cov}\,(\mathbf{u}) = \mathrm{Cov}\,(\mathbf{-u})$

3. $\mathrm{Cov}\,(\mathbf{0}) \geq |\ \mathrm{Cov}\,(\mathbf{u})\ |$

4. $\mathrm{Cov}\,(\mathbf{u}) = \mathrm{Cov}\,[\boldsymbol{Z}(\mathbf{s}), \boldsymbol{Z}(\mathbf{x}+\mathbf{u})] = \mathrm{Cov}\,[\boldsymbol{Z}(\mathbf{0}), \boldsymbol{Z}(\mathbf{u})]$

5. **If** $\mathrm{Cov}_j(\mathbf{u})$, $j = 1, 2, \ldots, k$, **are valid cov. fc. then** $\sum_{j=1}^{k} b_j \mathrm{Cov}_j(\mathbf{u})$ **is valid for** $b_j \geq 0 \forall j$

6. **If** $\mathrm{Cov}_j(\mathbf{u})$, $j = 1, 2, \ldots, k$, **are valid cov. fc. then** $\prod_{j=1}^{k} \mathrm{Cov}_j(\mathbf{u})$ **is valid**

7. **If** $\mathrm{Cov}\,(\mathbf{u})$ **is valid in** $\mathbf{R}^d$**, then is also valid in** $\mathbf{R}^p$, $p < d$

Continuity and Smoothness ...

- **A process $S(x)$ is** *mean-square continuous* **if, for all x,**

$$\mathrm{E}\left[\{S(x+u) - S(x)\}^2\right] \to 0 \text{ as } u \to 0$$

- **A formal description of the smoothness of a spatial surface $S(x)$ is its degree of differentiability.**

- **$S(x)$ is** *mean square differentiable* **if there exists a process $S'(x)$ such that, for all x,**

$$\mathrm{E}\left[\left\{\frac{S(x+u) - S(x)}{u} - S'(x)\right\}^2\right] \to 0 \text{ as } u \to 0$$

. . . and the correlation function

- the mean-square differentiability of $S(x)$ is directly linked to the differentiability of its covariance function

- Let $S(x)$ be a stationary Gaussian process with correlation function $\rho(u) : u \in \mathbb{R}$. Then:

  - $S(x)$ is mean-square continuous iff $\rho(u)$ is continuous at $u = 0$;

  - $S(x)$ is k times mean-square differentiable iff $\rho(u)$ is (at least) $2k$ times differentiable at $u = 0$.

Spectral representation

Bochener Theorem (iff):

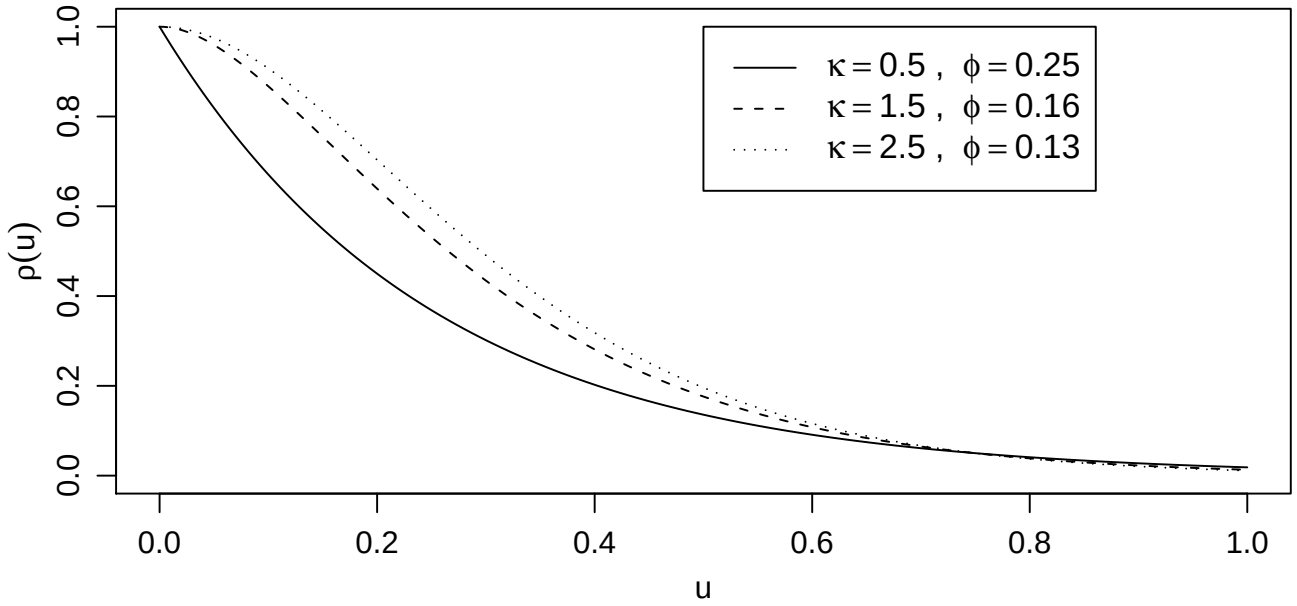
$$\mathbf{Cov}(u) = \int_{-\infty}^{+\infty} \exp\{iu\} s(w) dw$$

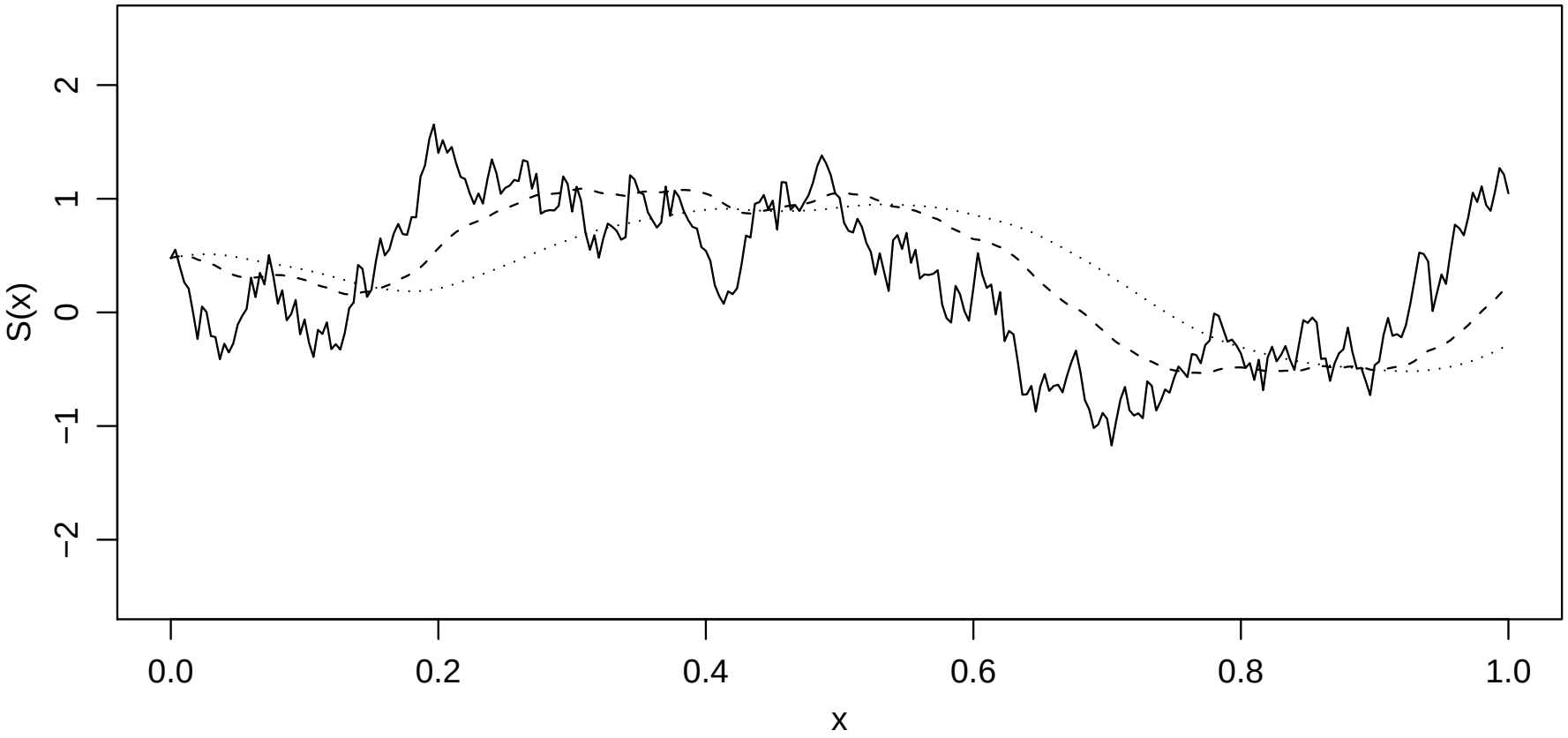
- $s(w)$ **is the** *spectral density function*
- $\mathrm{Cov}\,(u)$ **and** $s(w)$ **form a Fourier pair (the latter can be expressed as a function of the former)**
- **provided an alternative way to estimate covariance structure from the data using** *periodogram* $-\ \hat{s}(w)$
- *in principle* **provides way for testing for valid covariance functions and/or to derive new ones**

The (Whittle-)Matérn family

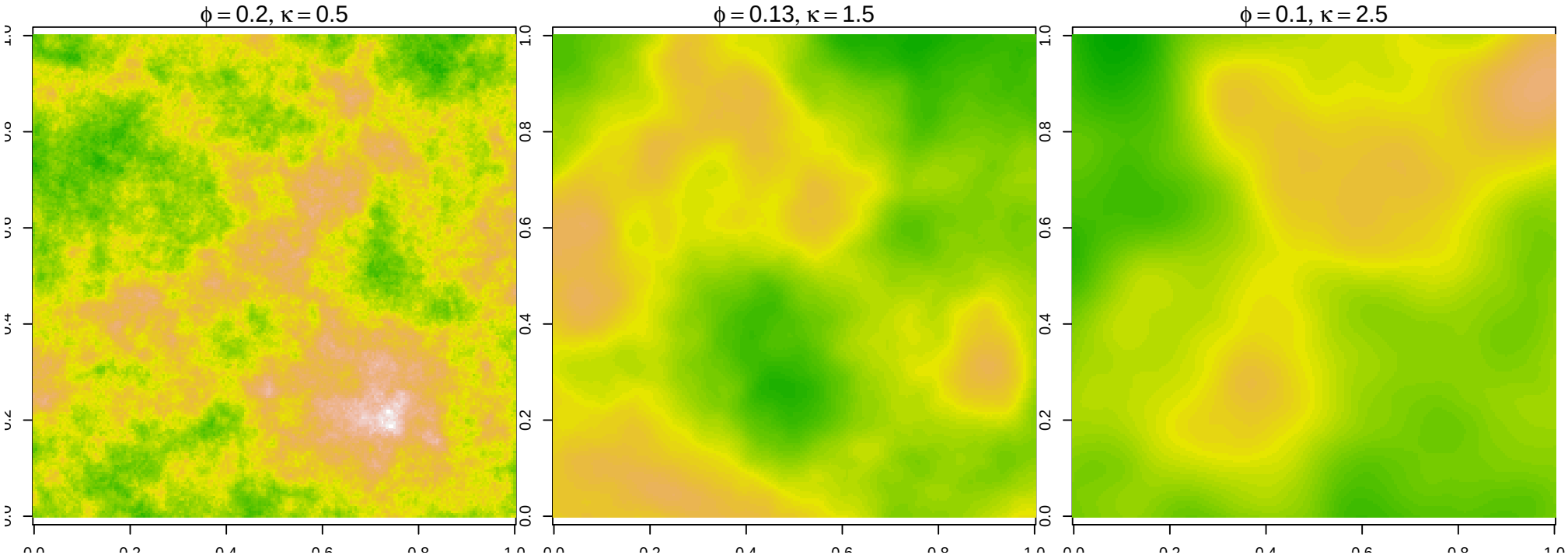
$$\rho(u) = \{2^{\kappa-1}\Gamma(\kappa)\}^{-1}(u/\phi)^{\kappa}K_{\kappa}(u/\phi)$$

- $K_{\kappa}(\cdot)$ denotes modified Bessel function of order κ
- parameters: $\kappa > 0$ (smoothness of $S(x)$) and $\phi > 0$ (extent of the spatial correlation)
  - for $\kappa = 0.5$, $\rho(u) = \exp\{-u/\phi\}$: exponential model
  - for $\kappa = 1$, $\rho(u) = (u/\phi)K_1(u/\phi)$: Whitle, 1954
  - for $\kappa \to \infty$ $\rho(u) = \exp\{-(u/\phi)^2\}$: Gaussian model
- $\lceil \kappa - 1$ times differentiable.
- κ and ϕ are not orthogonal
  - ϕ not comparable for different κ
  - reparametrisation: $\alpha = 2\phi\sqrt{\kappa}$
- A review: Guttorp and Gneiting (2005)


κ = 0.5 , ϕ = 0.25
κ = 1.5 , ϕ = 0.16
κ = 2.5 , ϕ = 0.13
ρ(u)
u
0.0
0.2
0.4
0.6
0.8
1.0


S(x)
x
−2
−1
0
1
2
0.0
0.2
0.4
0.6
0.8
1.0

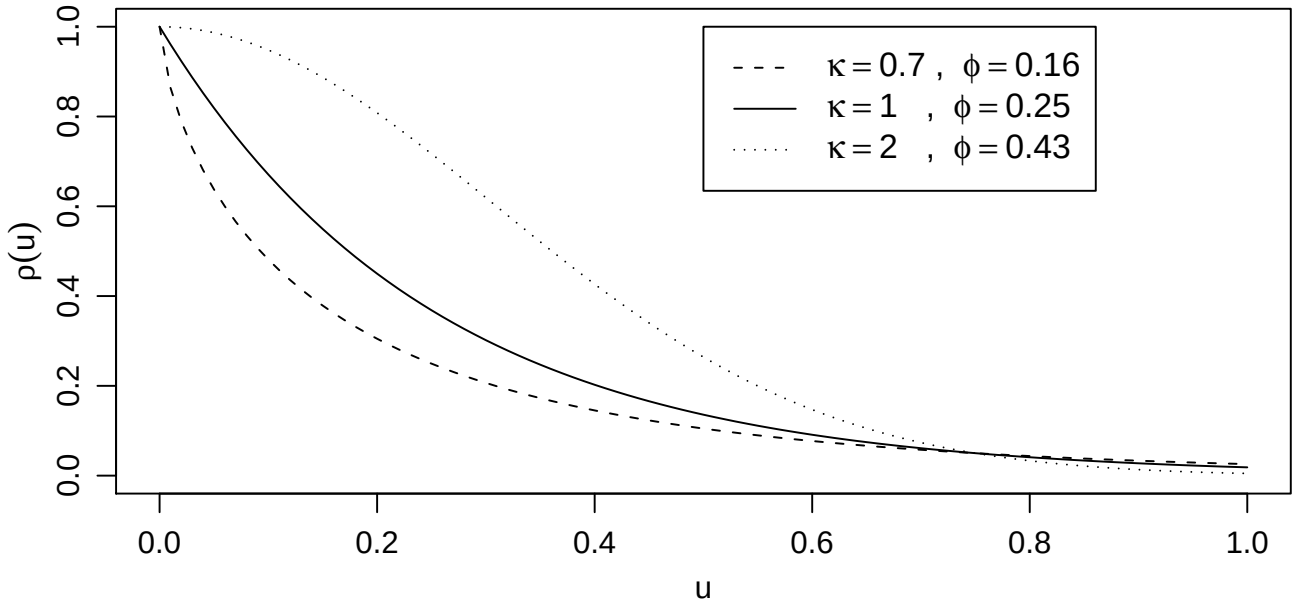
The (Whittle-)Matérn family (2D)

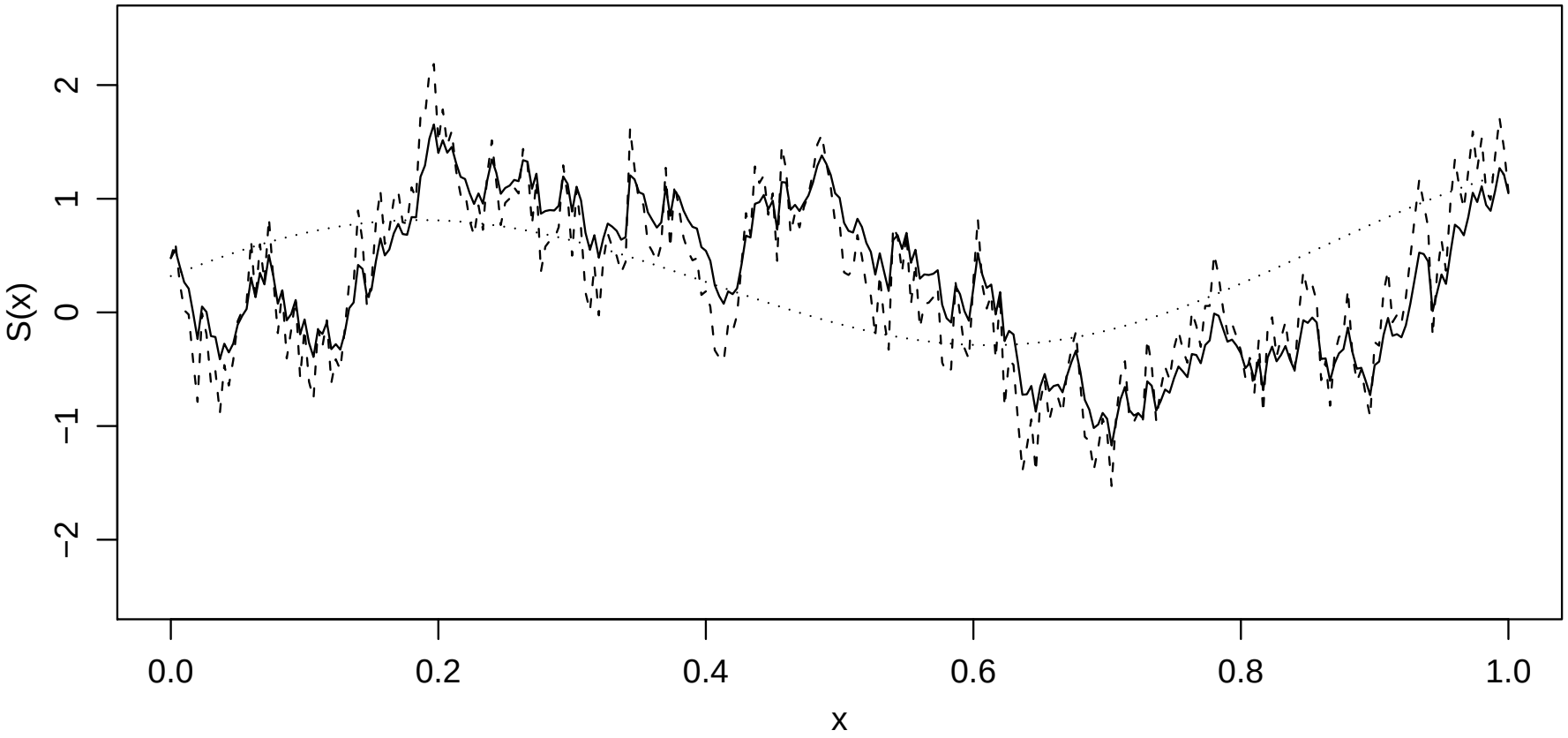


Other families (I): powered exponential

$$\rho(u) = \exp\{-(u/\phi)^{\kappa}\}$$

- scale parameter ϕ and shape parameter κ
- non-orthogonal parameters
- $0 < \kappa \leq 2$
- non-differentiable for $\kappa < 2$ e infinitely dif. for $\kappa = 2$
- asymptotically behaviour (pratical range)


κ = 0.7 , ϕ = 0.16
κ = 1 , ϕ = 0.25
κ = 2 , ϕ = 0.43
ρ(u)
u
0.0
0.2
0.4
0.6
0.8
1.0


S(x)
x
−2
−1
0
1
2
0.0
0.2
0.4
0.6
0.8
1.0

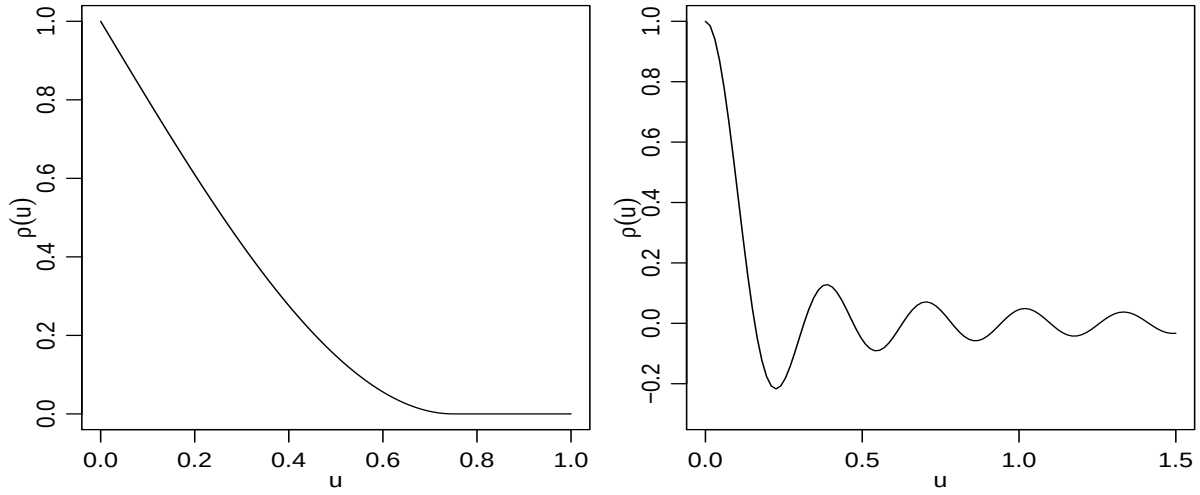
Other families (II): spherical and wave

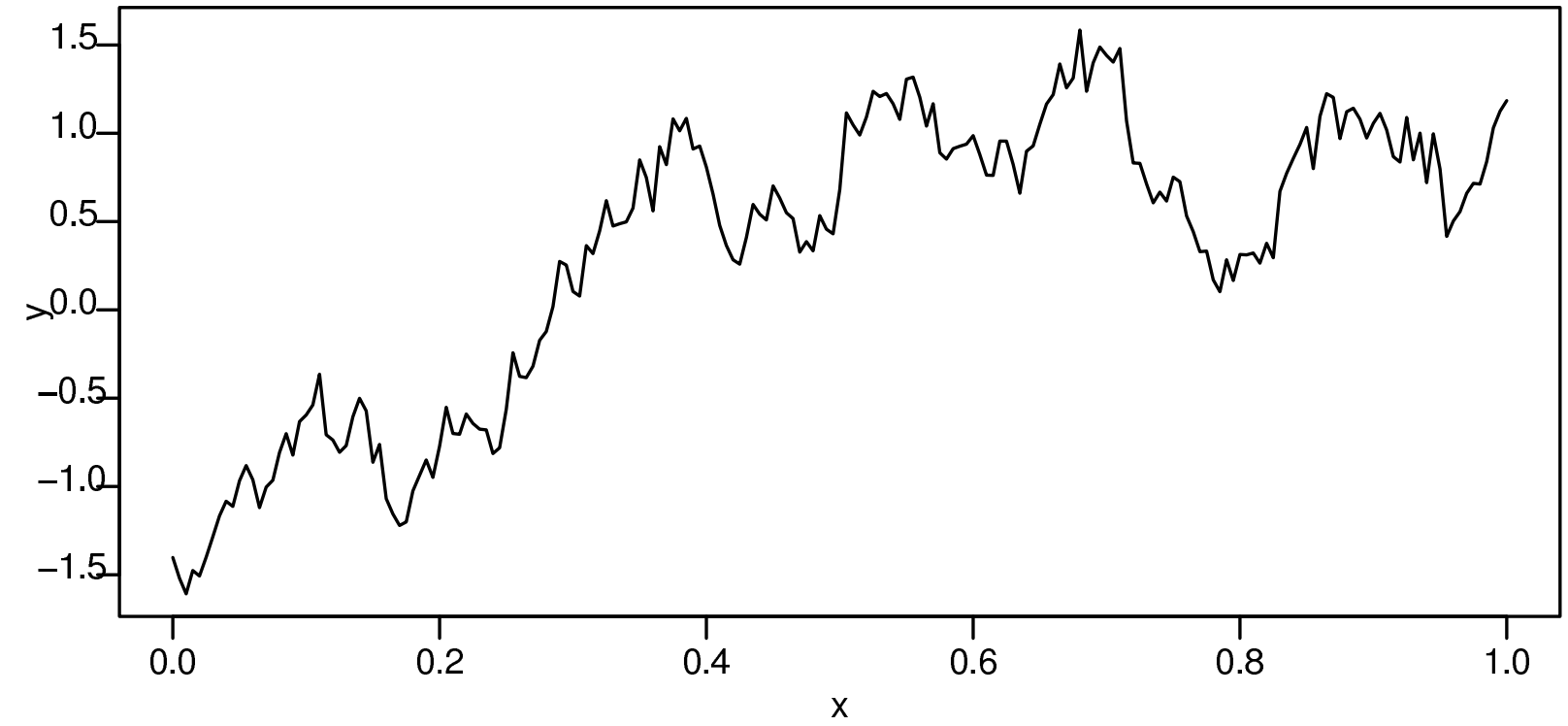
$$\rho(u) = [1 - 1.5(u/\phi) + 0.5(u/\phi)^3] \quad \mathbf{I}_{[0,\phi]}(u)$$

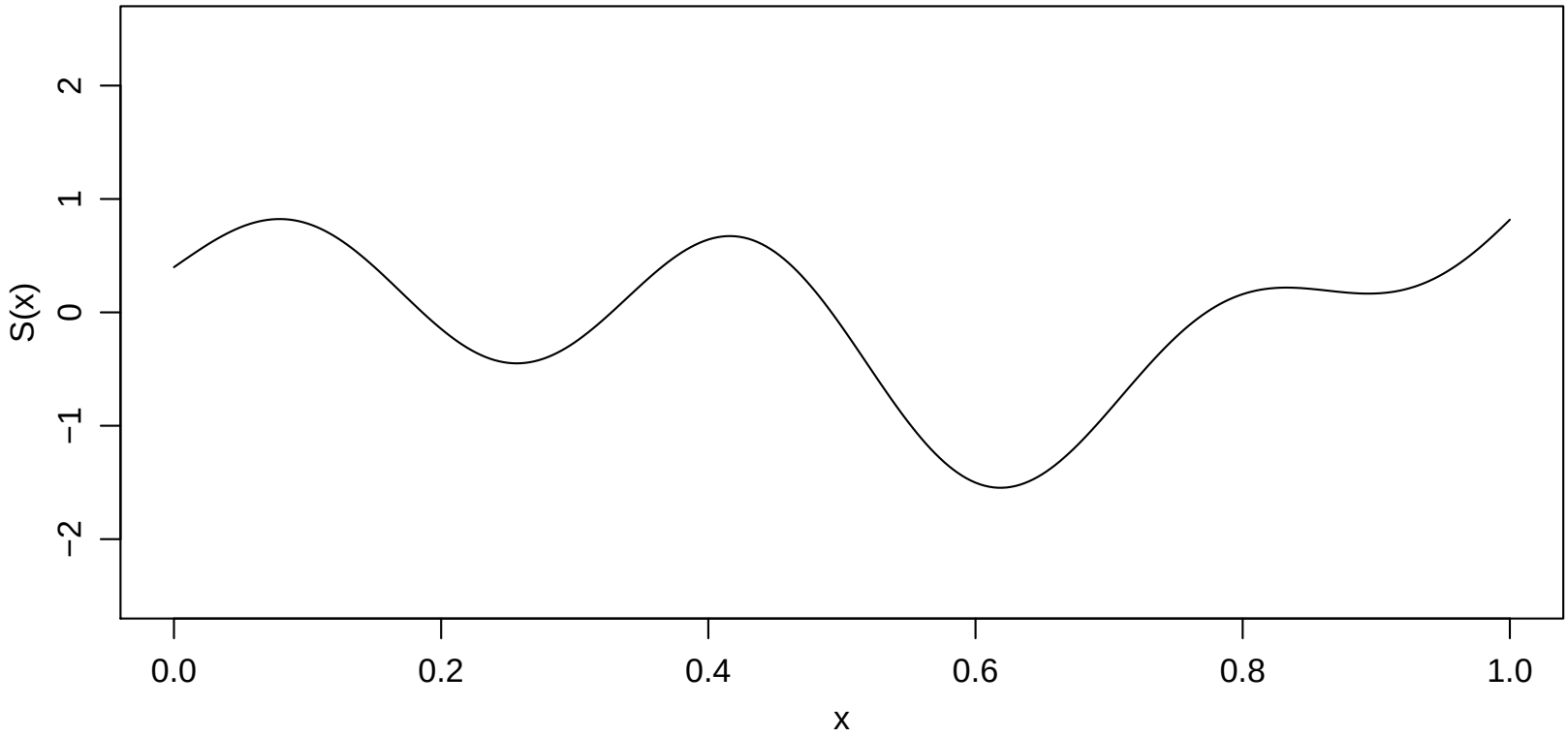
- finite range ϕ (overlapping volume between two spheres)
- non-differentiable at origin
- only once differentiable at $u = \phi$
- potential difficulties for MLE

$$\rho(u) = (u/\phi)^{-1}\sin(u/\phi)$$

- non-monotone (realisations reflect oscilatory behaviour)




1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
y
0.0
0.2
0.4
0.6
0.8
1.0
x


2
1
0
−1
−2
S(x)
0.0
0.2
0.4
0.6
0.8
1.0
x

Spatial case: the *nugget* effect

- discontinuity at the origin
- interpretations
  - measurement error ($\mathrm{Var}\,[\boldsymbol{Y}|\boldsymbol{S}]$)
  - micro-scale variation ($\mathrm{Var}\,[\boldsymbol{S}]$)
  - combination of both
- usually indistinguishable (linear model)
- except repeated measurements at coincident locations
- impact on predictions and their variance
- importance for sampling design

Notes on covariance functions I

- typically, but not necessarily, decreasing functions
- for monotonic models, the pratical range is defined as the distance where the correlation is 0 or 0.05 (for non-finite)
- assuming punctual support. Different supports (mis-aligned data) requires regularization (change of support)
- Variogram representations (wider class of processes)
  - Theoretical variogram (for cte mean)

    $$2V(u) = \mathrm{Var}\{Y(x_i) - Y(x_j)\} = \mathrm{E}\{[Y(x_i) - Y(x_j)]^2\}$$

  - *intrinsic* stationarity (intrinsic random functions, Matheron,1973)
  - validity (Gneiting, Sasvári and Schlather, 2001)

Covariance functions - some extensions

- **Compactly supported covarianve functions**
  - **spatial: Gneiting (2002)**
  - **spatio-temporal (Zastavnyi & Porcu, 2009)**
- **Spatio-temporal covariance functions (Gneiting, 2002)**
  - **stationarity, separability and symmetry**
  - **review: Gneiting, Genton and Guttorp (2007)**
- **Multivariate extension of Matérn model (Gneiting, Kleiber & Schlather, 2009)**

Terminology for variograms

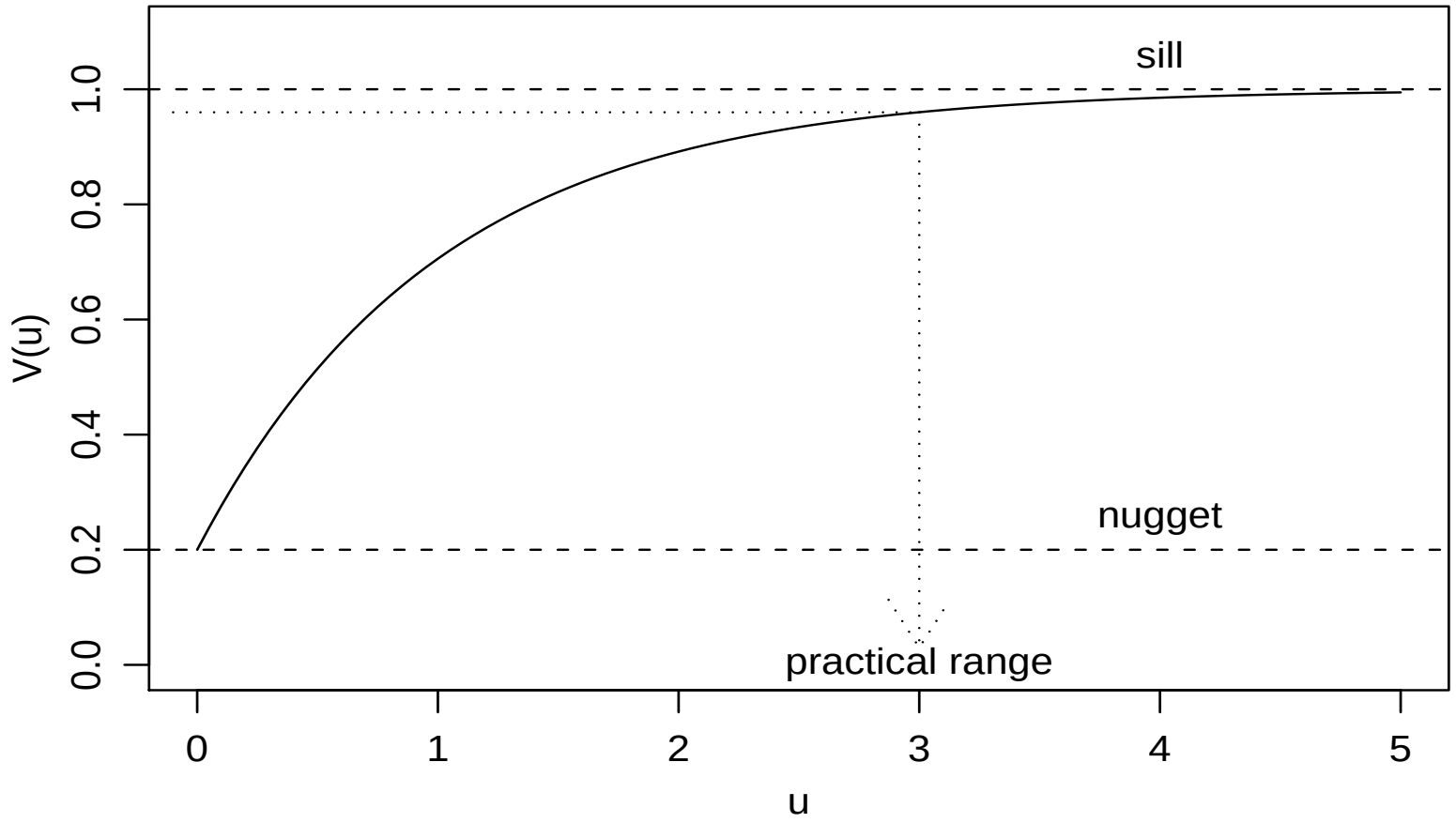
Under stationary Gaussian model:

$$V(u) = \tau^2 + \sigma^2\{1 - \rho(u;\phi)\}$$

- *the nugget variance*: τ^2
- *the sill*: $\sigma^2 = \mathrm{Var}\{S(x)\}$
- *the total sill*: $\tau^2 + \sigma^2 = \mathrm{Var}\{Y(x)\}$
- *the range*: ϕ, such $\rho_0(u) = \rho(u/\phi)$
- *the practical (effective) range*: u_0, such
  - $\rho(u) = 0$ (finite range correlation models)
  - $\rho(u) = 0.95\sigma^2$ (correlation functions approaching zero asymptotically)
  - or, in terms of variogram $\mathrm{V}(u) = \tau^2 + 0.95\sigma^2$
  - ***this is just a practical convention!***

Schematic representation

The theoretical variogram is a function which sumarises all the second order properties of the process

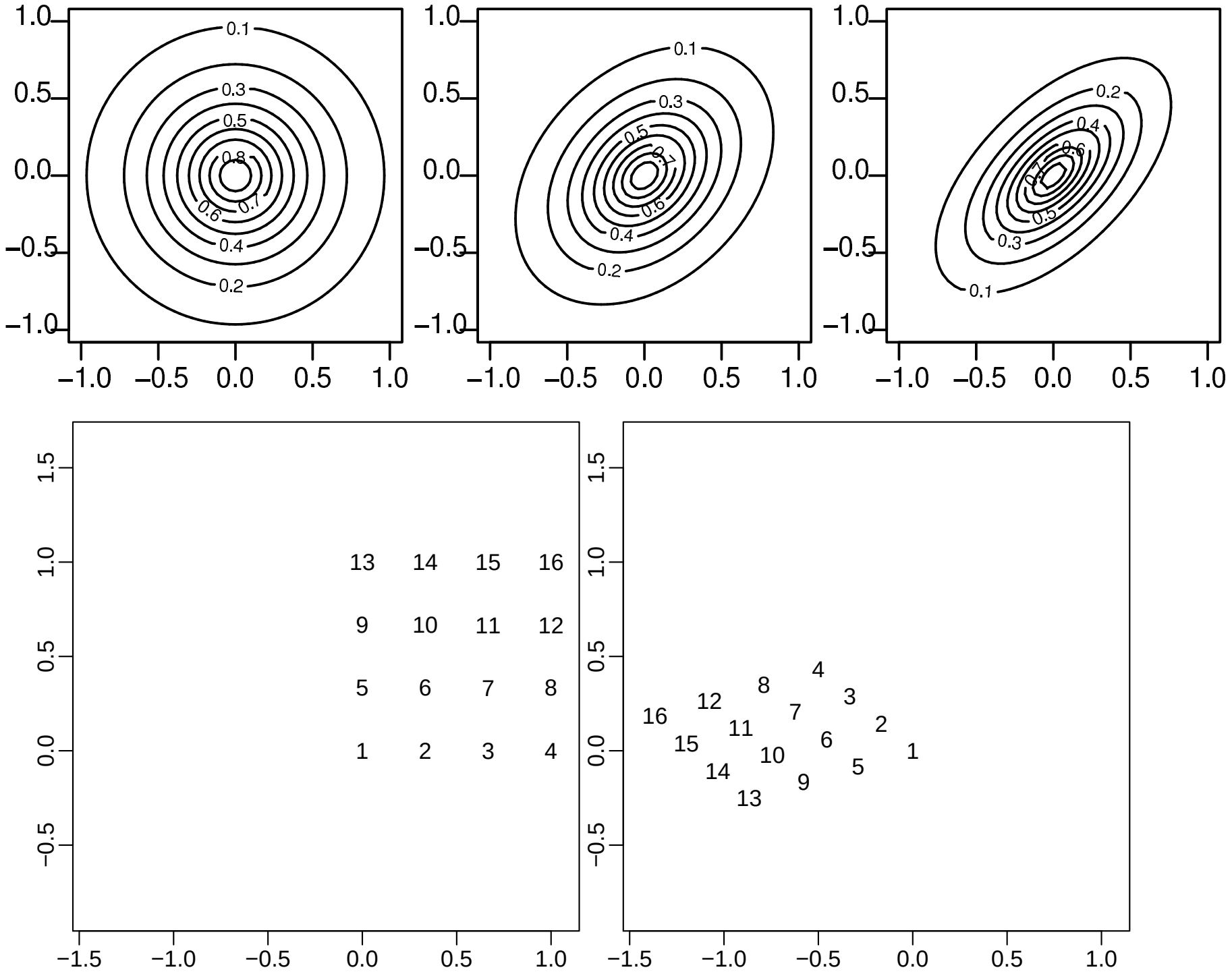


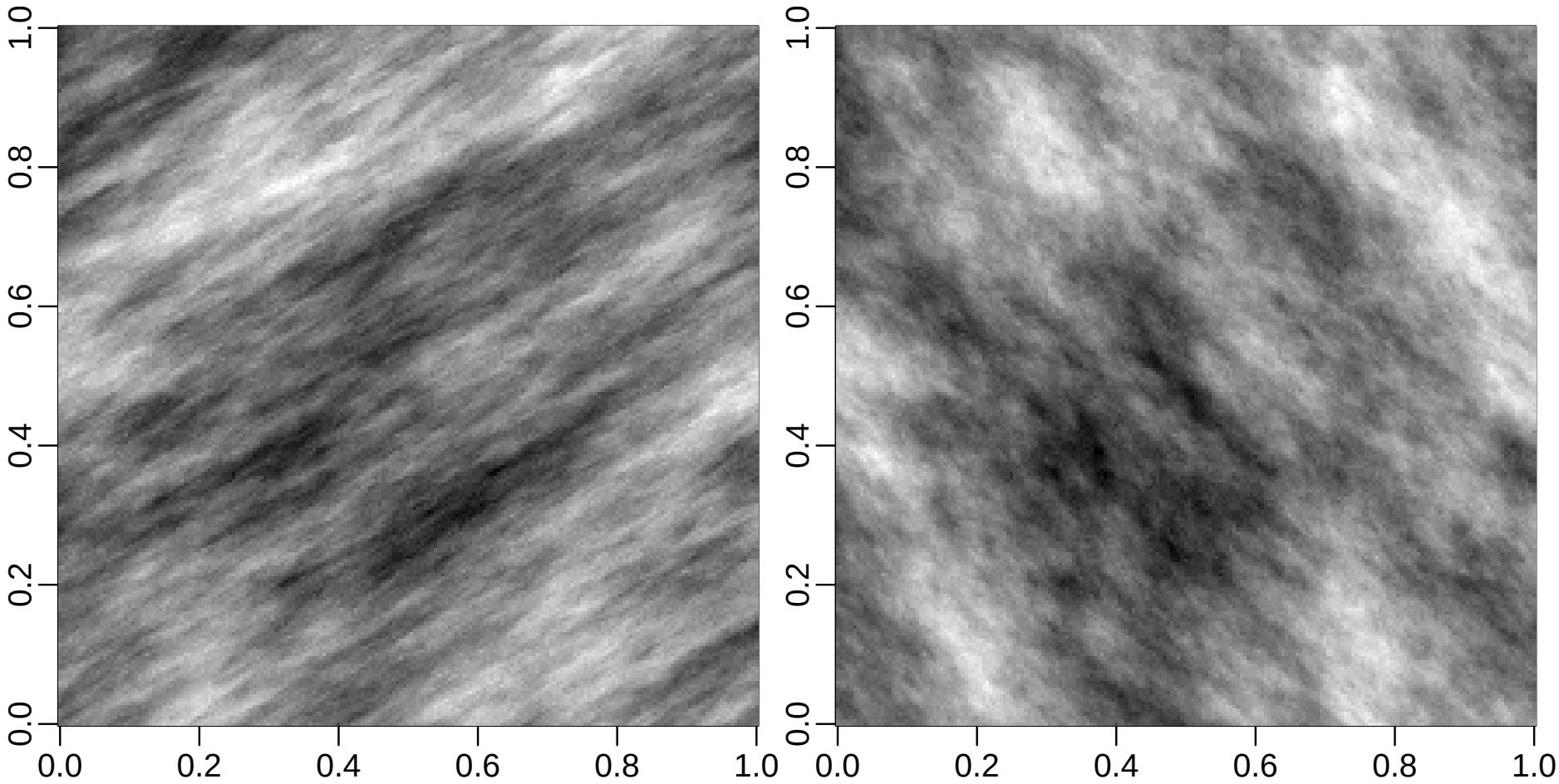
Directional effects

- environmental conditions wind, flow, soil formation, etc) can induce directional effects
- non-invariant properties of the cov. function under *rotation*
- simplest model: geometric anisotropy
- new coordinates by rotation and stretching of the original coordinates:

$$(x_1', x_2') = (x_1, x_2) \begin{pmatrix} \cos(\psi_A) & -\sin(\psi_A) \\ \sin(\psi_A) & \cos(\psi_A) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\psi_R} \end{pmatrix}$$

- add two parameters to the covariance function
- (ψ_A, ψ_R) anisotropy angle and ratio parameters


1.0
0.5
0.0
−0.5
−1.0
−1.0 −0.5 0.0 0.5 1.0
0.1
0.3
0.5
0.8
0.7
0.6
0.4
0.2
0.1
0.3
0.5
0.7
0.6
0.4
0.2
0.2
0.4
0.6
0.7
0.5
0.3
0.1
1.5
1.0
0.5
0.0
−0.5
−1.5 −1.0 −0.5 0.0 0.5 1.0
13 14 15 16
9 10 11 12
5 6 7 8
1 2 3 4
4
8
3
12
7
16
2
11
6
15
10
1
5
14
9
13



Realisations of a geometrically anisotropic Gaussian spatial processes whose principal axis runs diagonally across the square region with anisotropy parameters $(\pi/3, 4)$ *for the left-hand panel and* $(3\pi/4, 2)$ *for the right-hand panel.*

Simulating from the model

- For a finite set of locations x, $S(x)$ is multivariate Gaussian.

- A "standard" way for obtaining (unconditional) simulations of $S(x)$ is:
  - define the locations
  - define values for model parameters
  - compute Σ using the correlation function
  - obtain $\Sigma^{1/2}$, e.g. by Cholesky factorization of singular value decomposition
  - obtain simulations $S = \Sigma^{1/2} Z$ where Z is a vector of normal scores.

Simulating from the model (cont.)

- **Large simulations are often need in practice and require other methods, e.g.:**
  - **Wood and Chan (1994) – fast fourier transforms**
  - **Lantuéjoul (2002) – models and algorithms**
  - **Schlather (2001) – R package *RandomFields* : implements a diversity of algorithms (circulant embedding, turning bands, . . .)**
  - **Rue and Tjelmeland (2002) – approximation by Markov Gaussian Random Fields**
    Gibbs scheme using approximated sparse $(n-1) \times (n-1)$ full conditionals (GMRFlib)

Constructing multivariate models

One example: A common-component model

- assume independent processes $S_0^*(\cdot)$, $S_1^*(\cdot)$ and $S_2^*(\cdot)$
- Define a bivariate process $S(\cdot) = \{S_1(\cdot), S_2(\cdot)\}$
- $S_j(x) = S_0^*(x) + S_j^*(x) : j = 1, 2.$
- $S(\cdot)$ is a valid bivariate process with covariance structure

$$\text{Cov}\{S_j(x), S_{j'}(x-u)\} = \text{Cov}_0(u) + I(j = j')\,\text{Cov}_j(u)$$

- for different units it requires an additional scaling parameters so that $S_{0j}^*(x) = \sigma_{0j} R(x)$ where $R(x)$ has unit variance.

Approaches for multivariate models

- **CCM: Diggle and Ribeiro (2007); Bognola & Ribeiro (2008); Fanshawe & Diggle (2010)**
- **LMC: linear model of corregionalization**
- **Books: Chilès and Delfiner, (1999); Wackernagel (2003)**

Some recent developments:

- **Schmidt & Gelfand (2003); Gelfand, Schmidt, Banerjee & Sirmans (2004) – triangular structure and Bayesian inference**
- **Reich & Fuentes (2007) – semiparametric**
- **Majundar et. al. (2009) – convolution**
- **Apanasovich & Genton (2010) – latent dimention**
- **Jun (2009) – multivariate processes on a globe**
- **Gneiting, Klieber & Schlather (2009) – multivariate Matérn**

Non-stationary models

Stationarity is a convenient working assumption, which can be relaxed in various ways.

- Functional relationship between mean and variance: sometimes handled by a data transformation
- Non-constant mean:
  - replace constant μ by

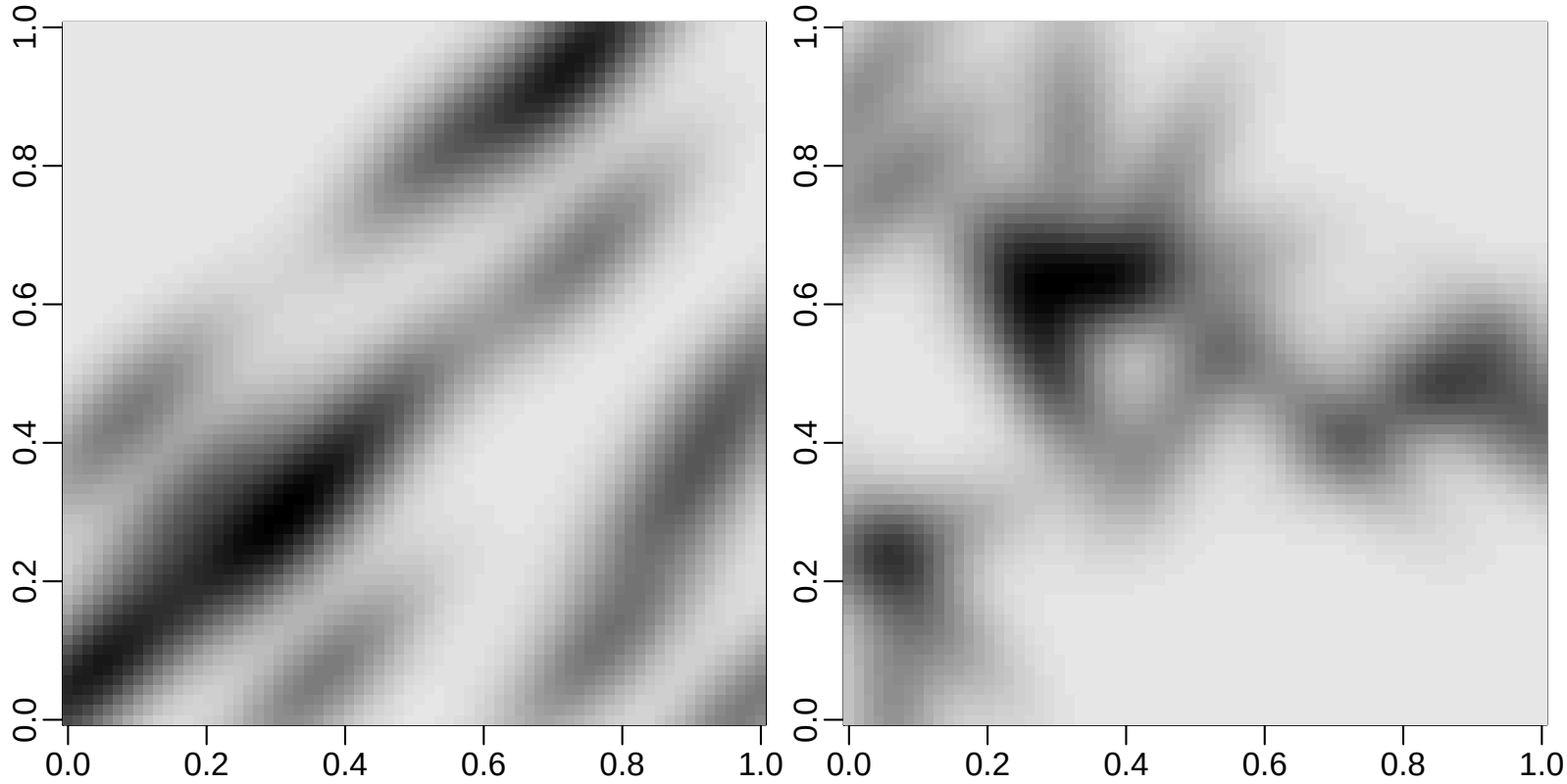
$$\mu(x) = F\beta = \sum_{j=1}^{k} \beta_j f_j(x)$$

  - trend surface and covariates
  - deterministic *vs* stochastic: interpretation of the process
  - exploratory analysis: possible non-linear relations

- **Non-stationary random variation:**
  - ***intrinsic*** **variation a weaker hypothesis (process has stationary increments, cf random walk model in time series), widely used as default model for discrete spatial variation (Besag, York and Molié, 1991).**
  - ***Spatial deformation*** **methods (Sampson and Guttorp, 1992) seek to achieve stationarity by complex transformations of the geographical space, x.**
  - **spatial convolutions (Higdon, 1998, 2002; Fuentes e Smith)**
  - **low-rank models (Hastie, 1996)**
  - **non-Euclidean distances (Rathburn, 1998)**
  - **locally directional effects**

Need to balance:
· increased flexibility of general modelling assumptions *against*
· over-modelling of sparse data,
leading to poor identifiability of model parameters.

An illustration



Globally (left) and locally (right) directional models

GIS integration

An example with the aRT package

- (aRT API R - Terralib)
- http://www.leg.ufpr.br/aRT

PART 2

Parameter Estimation

and

Spatial Prediction

Opening remarks

- The canonical problem is spatial prediction of the form

$$\hat{S}(x) = \mu(x) + \sum_{i=1}^{n} w_i(x)(y_i - \mu(x))$$

- The prediction problem can be tackled by adopting some criteria (e.g. minimise MSPE)

$$MSPE(\hat{T}) = \mathrm{E}\left[(T - \hat{T})^2\right] \text{ e.g. above } T = S(x)$$

- However this requires knowledge about:
  - the assumed model
  - the model parameters
- need to infer first and second-moment properties of the process from the available data

Inference (linear model)

- parameter estimation: likelihood based methods (other approaches are also used)

- spatial prediction: *simple kriging*

$$\hat{S}(x) = \mu + \sum_{i=1}^{n} w_i(x)(y_i - \mu)$$

- straightforward extension for $\mu(x)$

- Parameter uncertainty?
  usually ignored in traditional geostatistics
  (plug-in prediction)

An aside:

Distinguishing parameter estimation and spatial prediction

- assume a set of locations $x_i : i = 1, \ldots, n$ on a lattice covering the area
- interest: average level of pollution over the region
- consider the sample mean:

$$\bar{S} = n^{-1} \sum_{i=1}^{n} S_i$$

...

- **within a parameter estimation problem**
  - **estimator of the constant mean parameter** $\mu = \mathrm{E}[S(x)]$
  - **precision given by the M.S.E.** $\mathrm{E}\{[(\bar{S} - \mu)^2]\}$
  - $\mathrm{Var}[\bar{S}] = n^{-2} \sum_{i=1}^{n} \sum_{i=1}^{n} \mathrm{Cov}(S_i, S_j) \geq \sigma^2/n$
- **within a prediction problem**
  - **predictor of the spatial average** $S_A = |A|^{-1} \int_A S(x)dx$
  - **precision by the M.S.E.** $\mathrm{E}[(\bar{S} - S_A)^2]$, S_A **is r.v**
  - **precision (can even approach zero) given by**

$$\begin{aligned} \mathrm{E}[(\bar{S} - S_A)^2] &= n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}(S_i, S_j) \\ &+ |A|^{-2} \int_A \int_A \mathrm{Cov}\{S(x), S(x')\} dx dx' \\ &- 2(n|A|)^{-1} \sum_{i=1}^{n} \int_A \mathrm{Cov}\{S(x), S(x_i)\} dx. \end{aligned}$$

Exploratory Data Analysis (EDA)

- **Non-spatial**
  - **outliers**
  - **non-normality**
  - **arbitrary mean model: choice of potential covariates**
- **Spatial**
  - **spatial outliers**
  - **trend surfaces (scatterplots against covariates)**
  - **other potential spatial covariates**
  - **GIS tools**

First moment properties (trend)

The OLS estimator

$$\tilde{\beta} = (D'D)^{-1}D'Y$$

is unbiased irrespective the covariance structure (assuming the model is correct)

A more efficient GLS estimator:

$$\hat{\beta} = (D'V^{-1}D)^{-1}D'V^{-1}Y$$

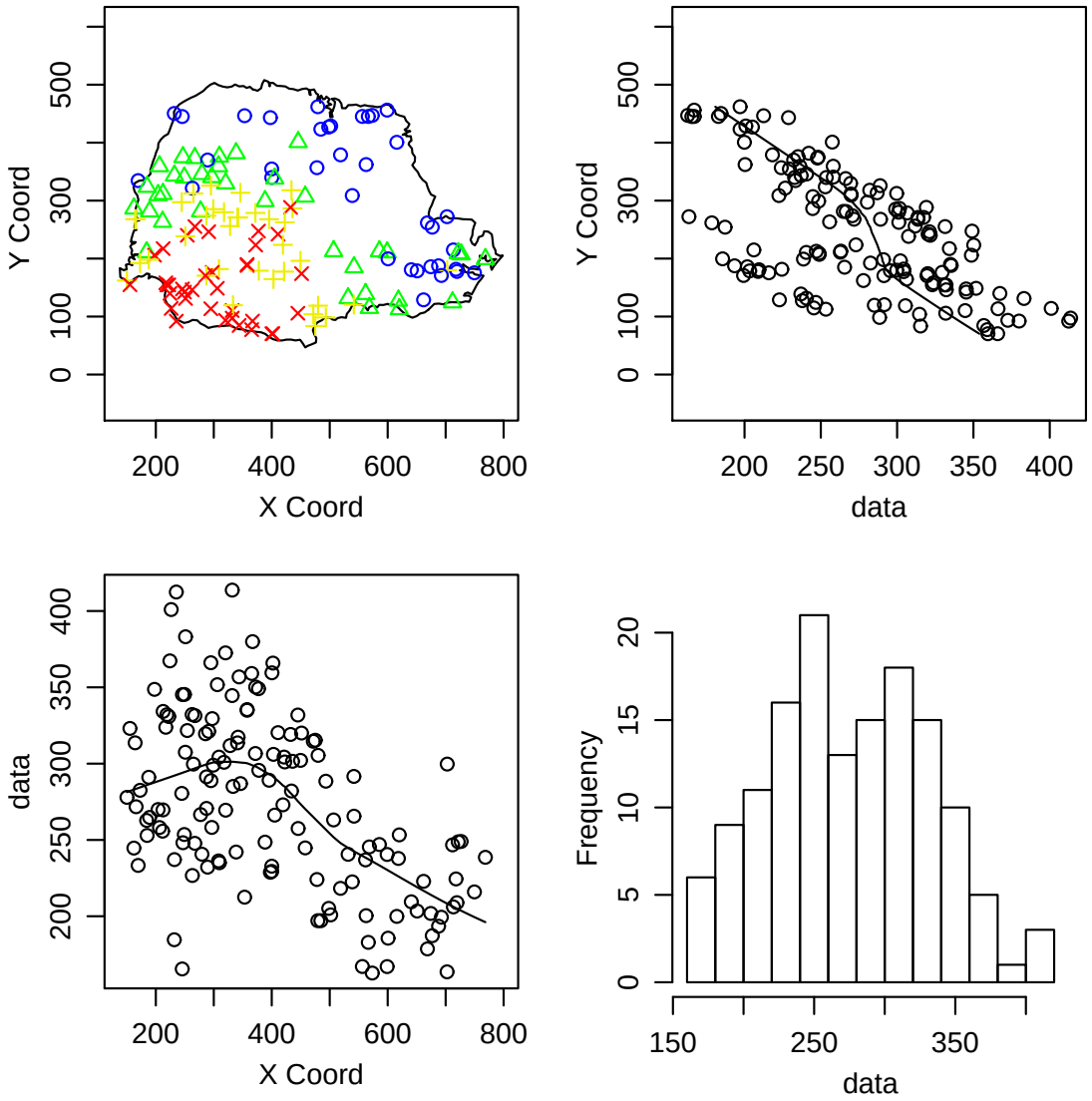
- **unbiased and smaller variance**
- **MLE**
- **requires knowledge about covariance parameters**

For non-cte mean, OLS residuals can inform about covariance structure

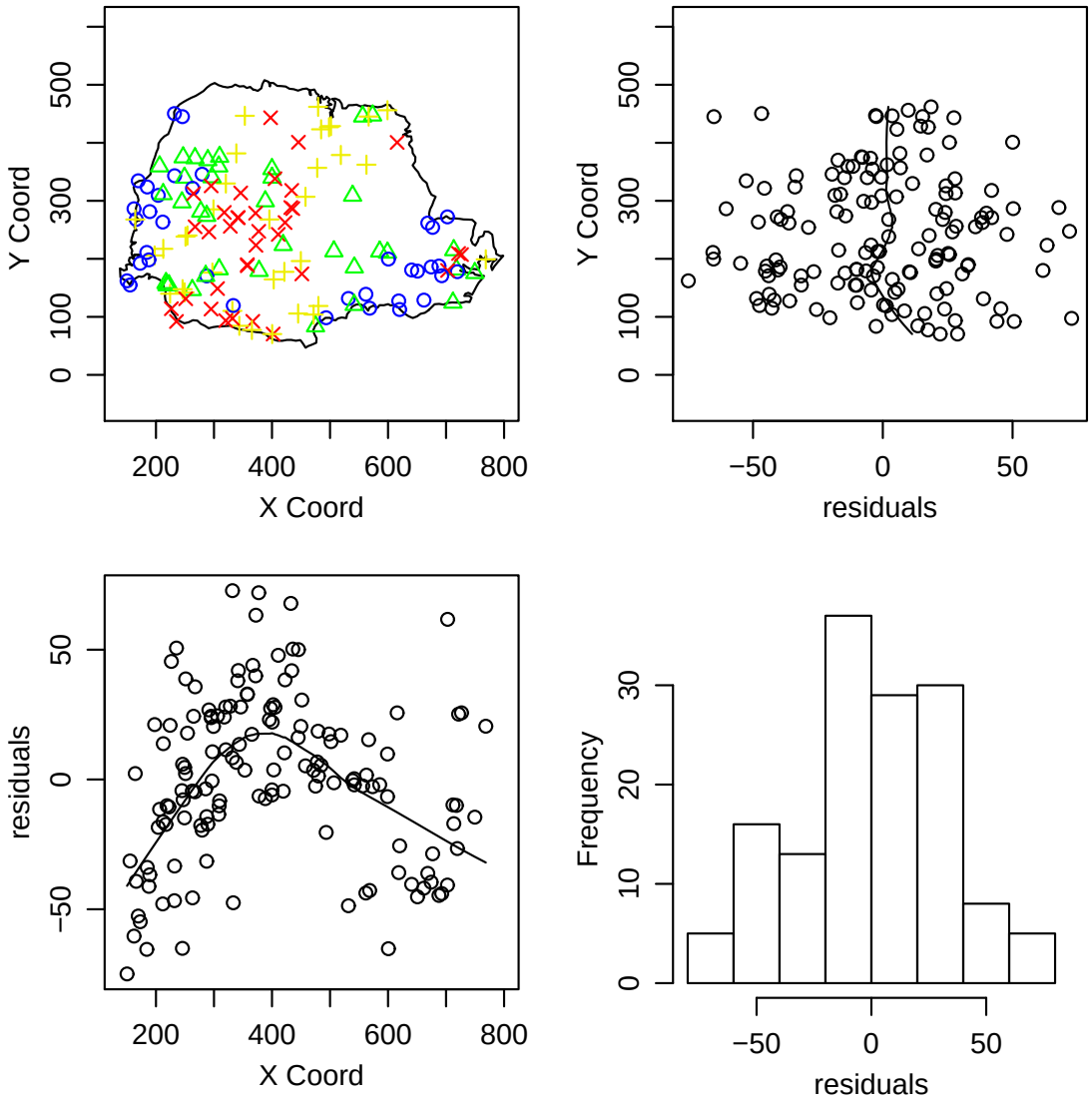
$$R = Y - D\tilde{\beta}$$

Strategies: two stages (which can be interactive) or joint estimation

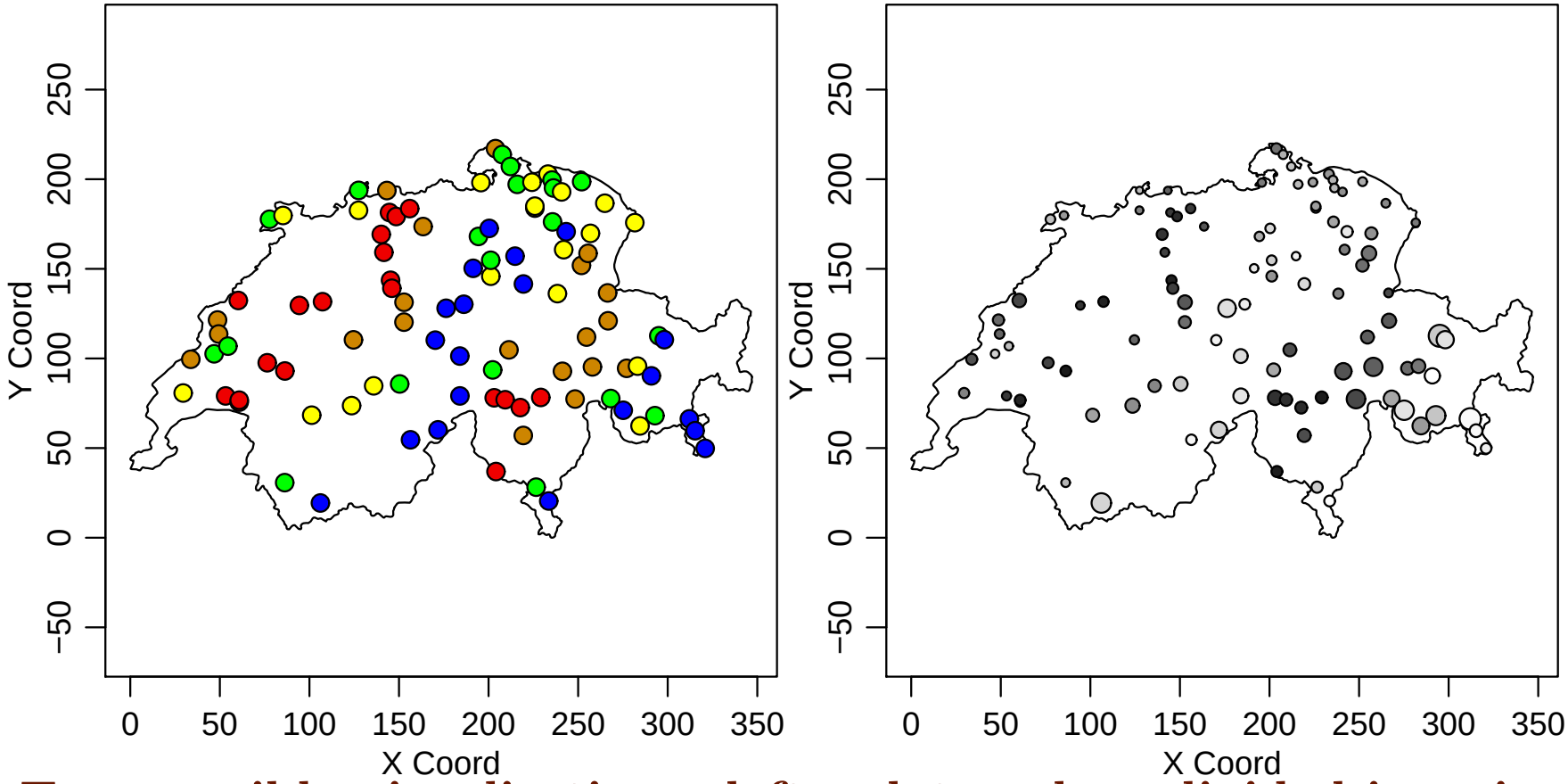
EDA: A quick exploratory display



EDA: Residual plots



EDA: Circle plot



Two possible visualisations: left – data values divided in quintiles, right – gray shade proportional to data, circle sizes proportional to a covariate value (elevation).

Second order properties
EDA: Empirical Variograms

- **Theoretical variogram (for cte mean)**

$$2V(u) = \text{Var}\{Y(x_i) - Y(x_j)\} = \text{E}\{[Y(x_i) - Y(x_j)]^2\}$$

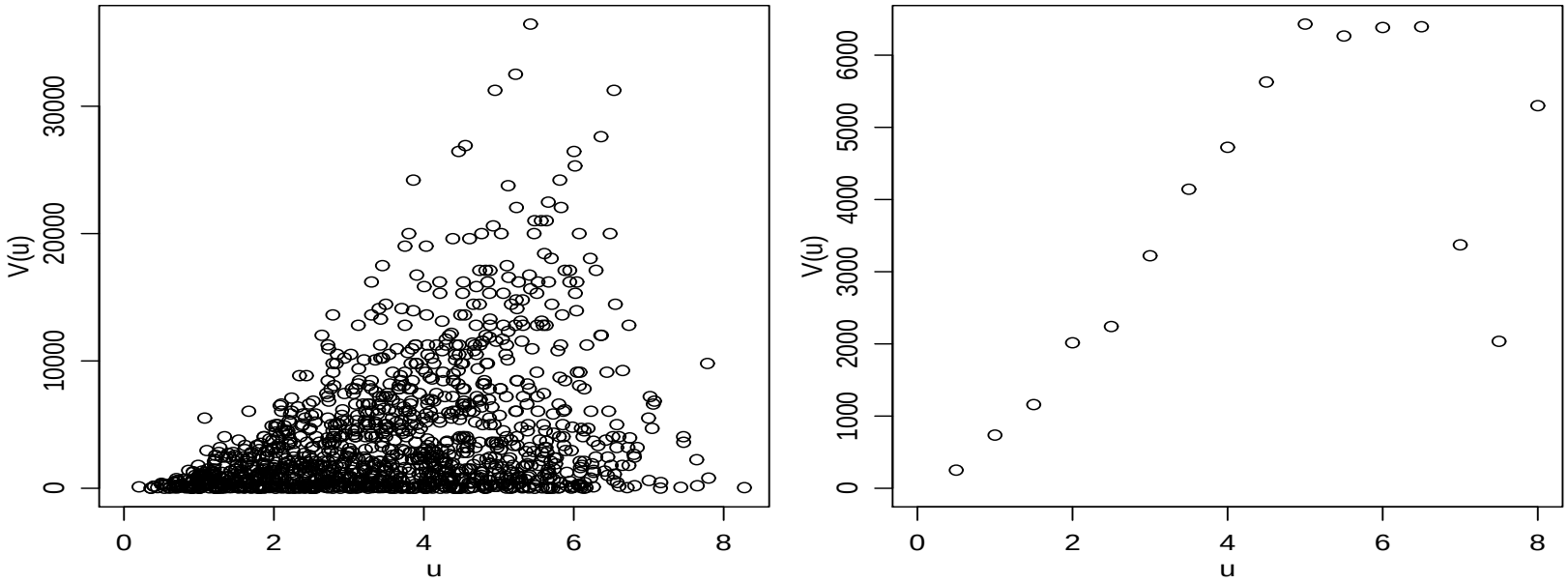
- **under the assumed Gaussian model:**

$$V(u) = \tau^2 + \sigma^2\{1 - \rho(u;\phi)\}$$

- **Empirical (semi-)variogram:** $\hat{V}(u)$

$$\hat{V}(u_{ij}) = \text{average}\{0.5[y(x_i) - y(x_j)]^2\} = \text{average}\{v_{ij}\}$$

where each average is taken over all pairs $[y(x_i), y(x_j)]$ **such that** $||x_i - x_j|| \approx u$

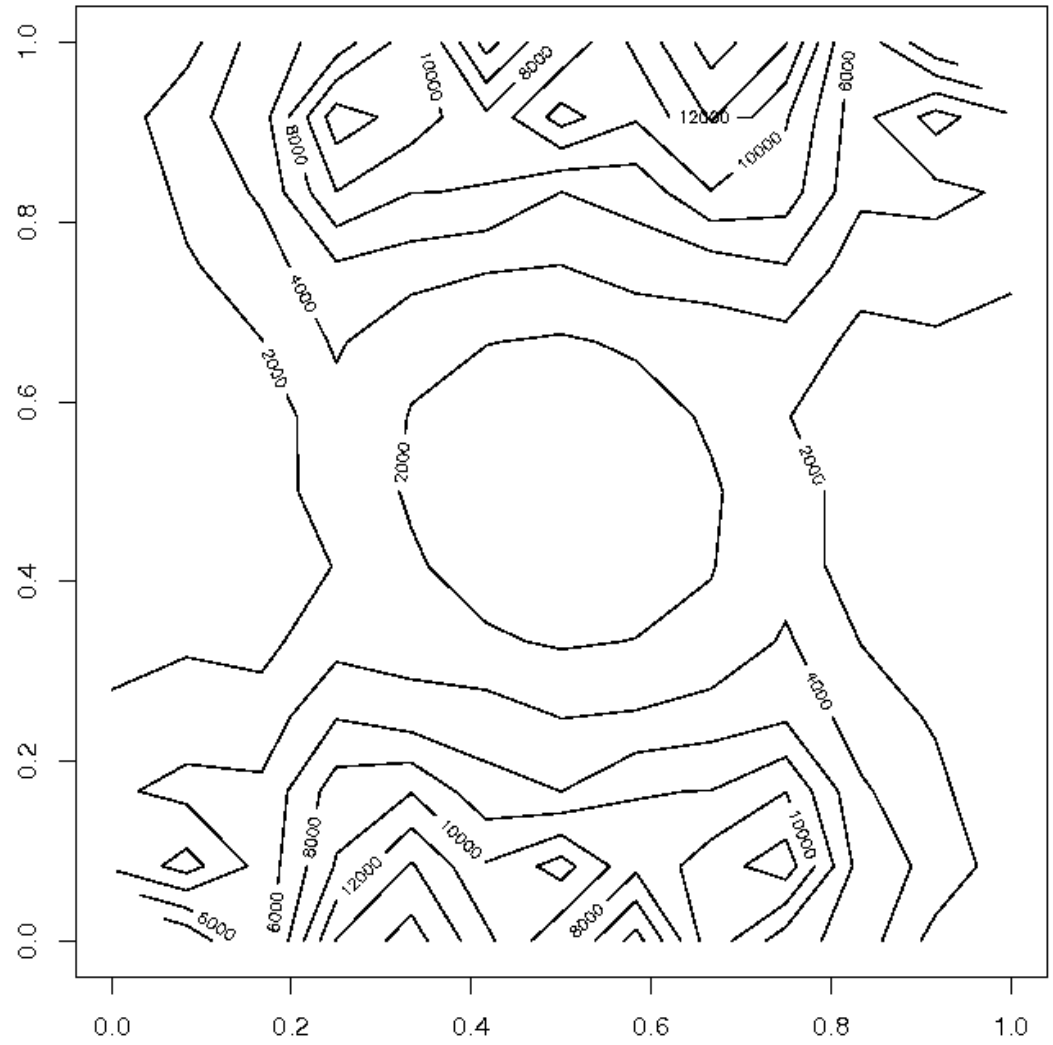


- variogram cloud: scatterplot of (u_{ij}, v_{ij})
- the empirical variogram is derived from the variogram cloud by averaging within bins: $u - h/2 \leq u_{ij} < u + h/2$
- *sample variogram* ordinates V_k; $(k-1)h < u_{ij} < kh$
- convention $u_k = (k - 0.5)h$ (interval mid-point)
- may adopt distinct h_k
- excludes zero from the smallest bin (deliberate)
- typically limited at a distance $u < u_{max}$

Some topics on empirical variograms

- biased for non-constant mean
- higher order polynomials *vs* spatial correlation
- for a process with non-constant mean (covariates) replace $y(x_i)$ by residuals $r(x_i) = y(x_i) - \hat{\mu}(x_i)$ from a trend removal
- usage of kernel or spline smoothers however
  notice $\frac{1}{2}n(n-1)$ points are not independent
  bandwidth issues, considering exploratory purposes
- a diversity of alternative estimators is available
  e.g. robust estimators (Genton, 1998a)
- Monte Carlo envelopes for empirical variograms

Exploring directional effects



Paradigms for parameter estimation

- Ad hoc (variogram based) methods
  - compute an empirical variogram
  - fit a theoretical covariance model
- Likelihood-based methods
  - typically under Gaussian assumptions
  - more generally needs MCMC or approximations
  - Optimal under stated assumptions, robustness issues
  - full likelihood not feasible for large data-sets
  - variations on the likelihood function (*pseudo-likelihoods*)
- Bayesian paradigma, combines estimation and prediction

Variogram model fitting

- fitting a typically non-linear variogram function (as e.g. the Matérn) to the empirical variogram provides a way to estimate the models parameters.

- e.g. a weighted least squares criteria minimises

$$W(\theta) = \sum_k n_k \{[\bar{V}_k - V(u_k; \theta)]\}^2$$

  where θ denotes the vector of covariance parameters and $\bar{V}_k$ is average of n_k variogram ordinates v_{ij}.

- in practice u is usually limited to a certain distance

- variations includes:
  – fitting models to the variogram cloud
  – other estimators for the empirical variogram
  – different proposals for weights
  – explicity account of covariance structure (Genton, 1998b)

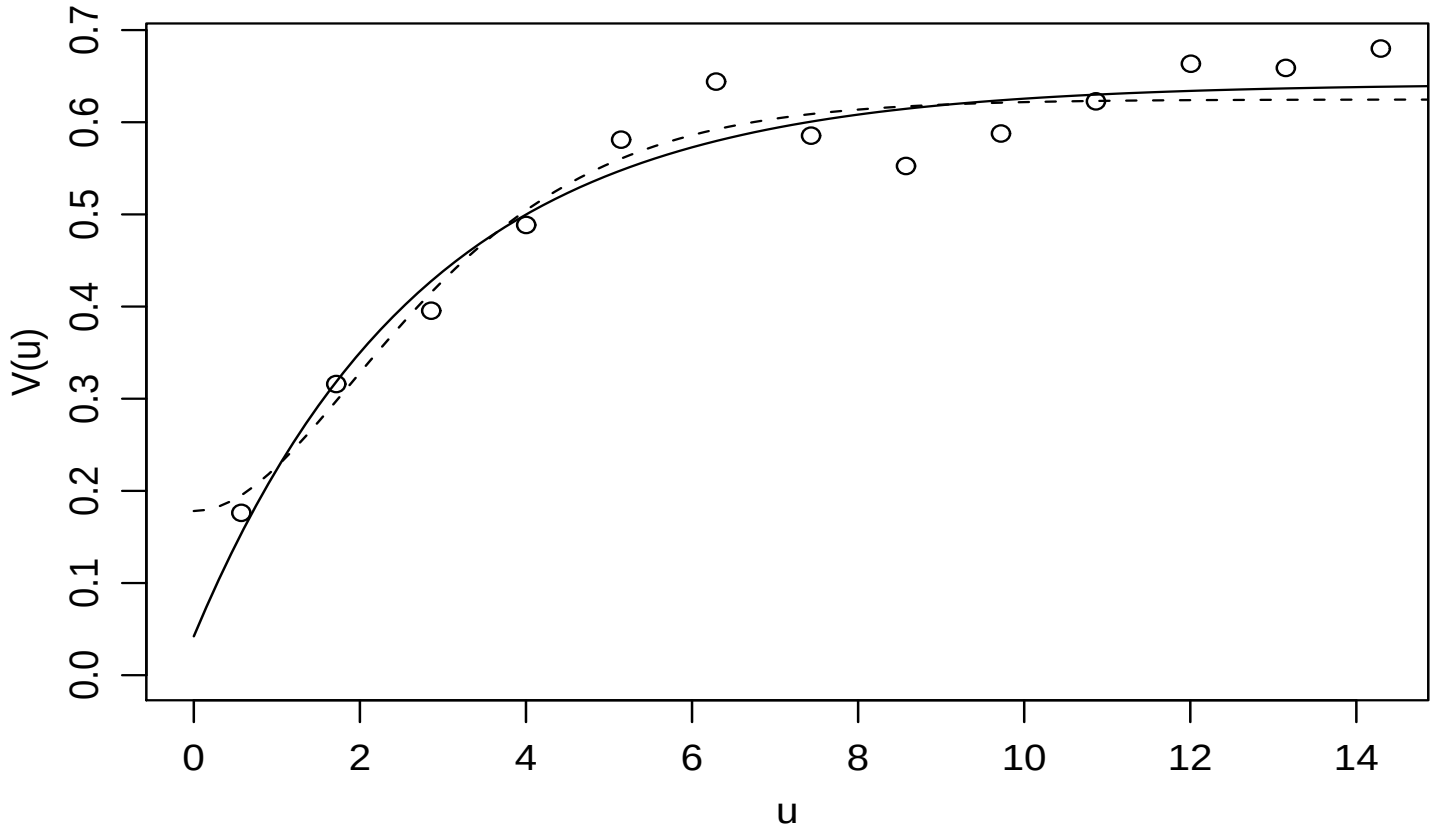
Comments on variograms - I

Difficulties with empirical variograms

- $v_{ij} \sim V(u_{ij})\chi^2_1$
- the v_{ij} are correlated
- the variogram cloud is therefore unstable, both pointwise and in its overall shape
- binning removes the first objection to the variogram cloud, but not the second
- is sensitive to mis-specification of $\mu(x)$

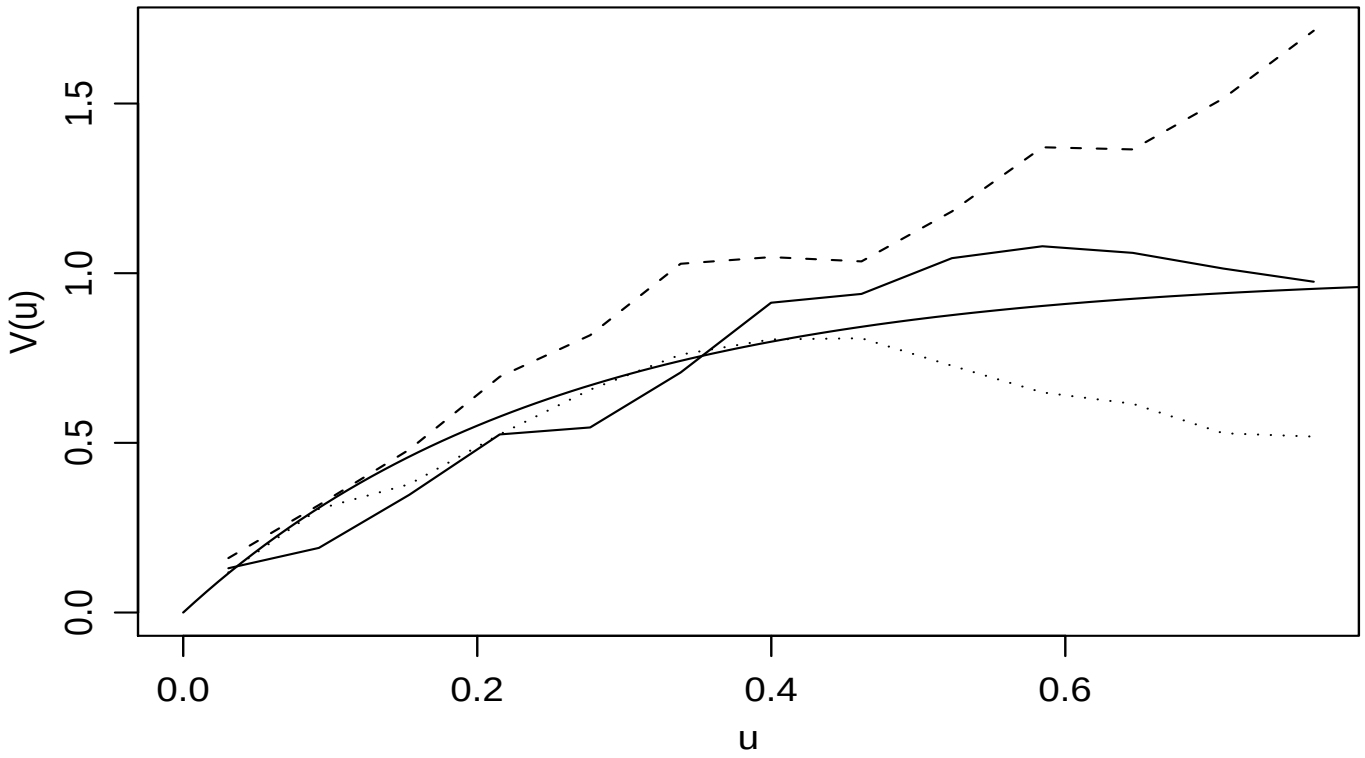
Comments on variograms - II

- **equally good fits for different "extrapolations" at origin**



Comments on variograms - III

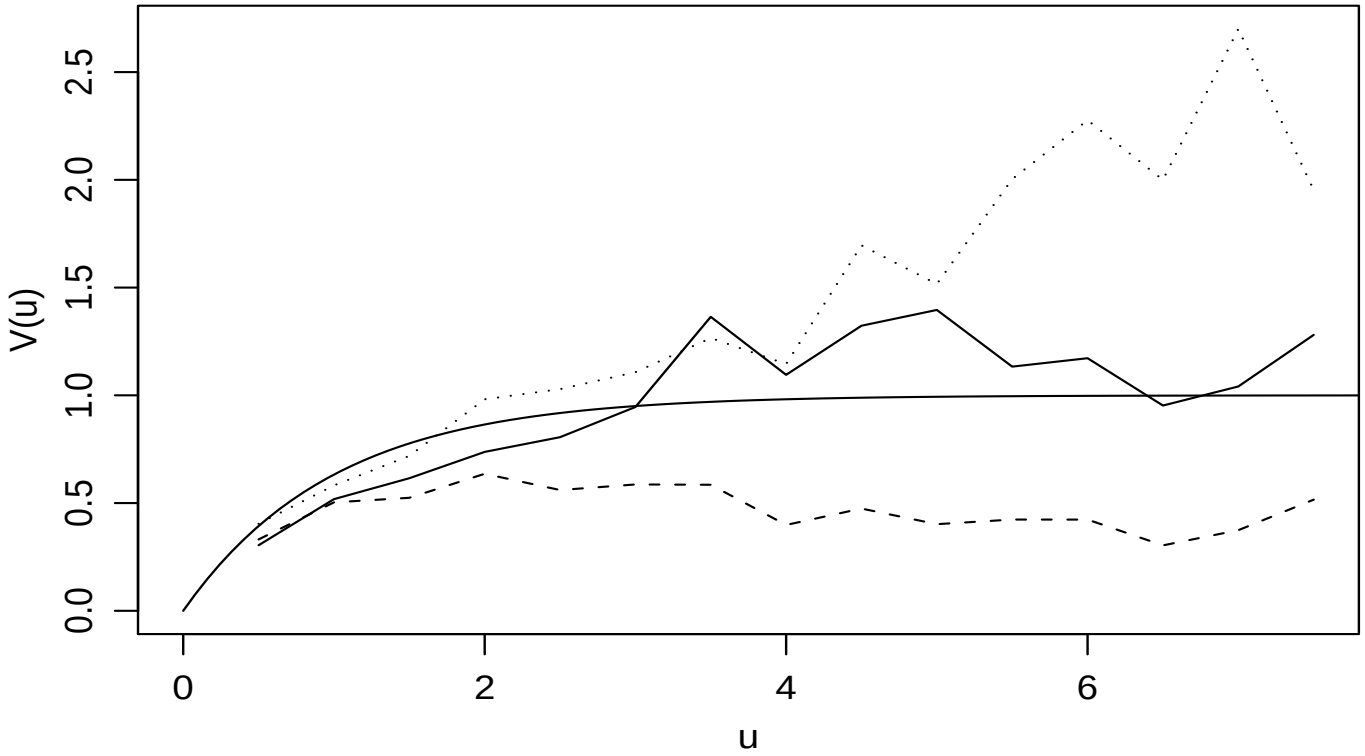
- **correlation between variogram points points**



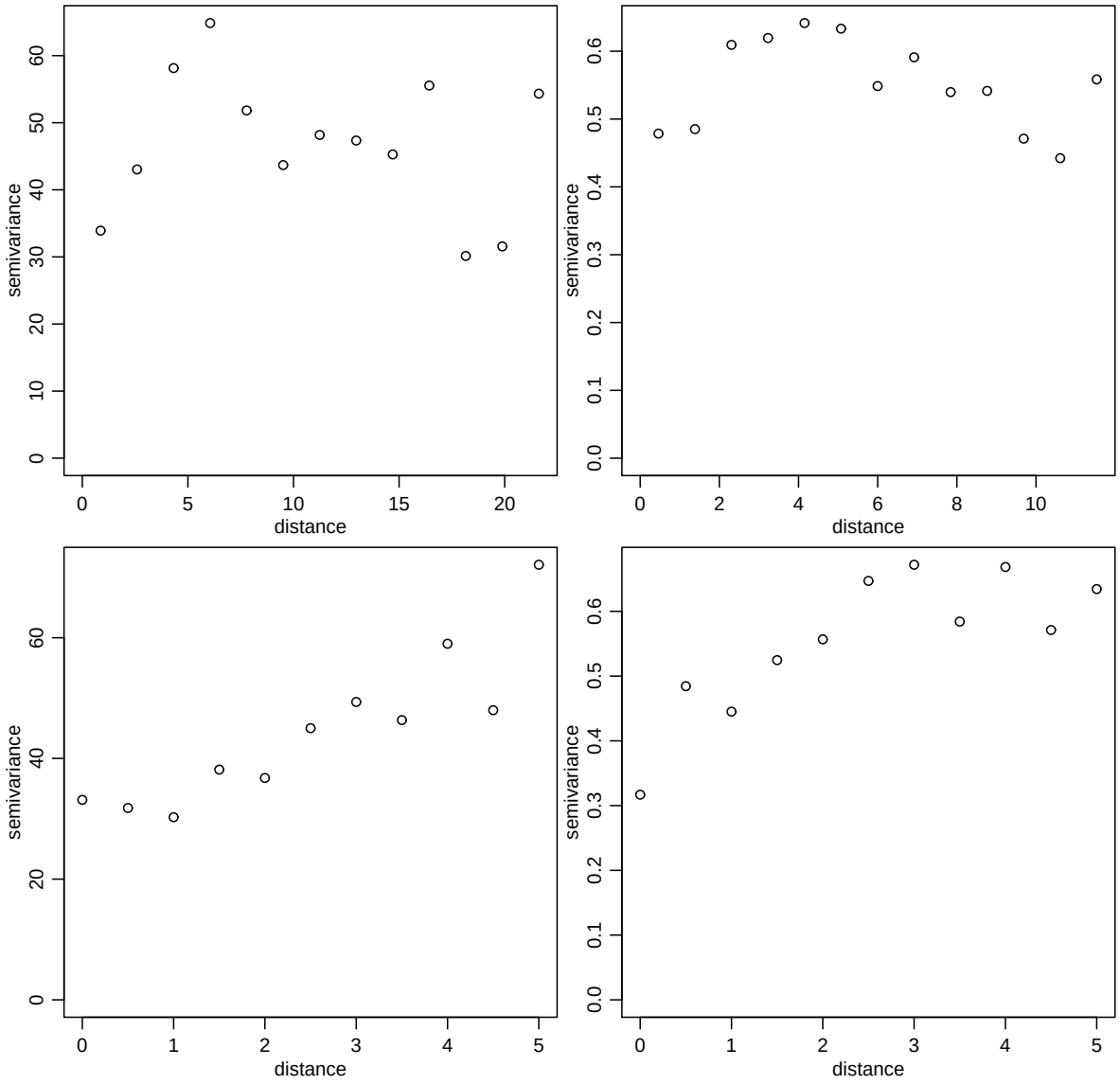
Empirical variograms for three simulations from the same model.

Comments on variograms - IV

- sensitivity to the specification of the mean
- solid smooth line: true model, dotted: empirical variogram, solide: empirical variogram from true residuals, dashed: empirical variogram from estimated residuals.



Comments on variograms - V



Parameter estimation: maximum likelihood

For the basic geostatistical model

$$Y \sim \text{MVN}(\mu 1, \sigma^2 R + \tau^2 I)$$

1 denotes an n-element vector of ones,

I is the $n \times n$ identity matrix

R is the $n \times n$ matrix with $(i,j)^{th}$ element $\rho(u_{ij})$ where $u_{ij} = ||x_i - x_j||$, the Euclidean distance between x_i and x_j.

Or more generally for

$$\begin{aligned} S(x_i) &= \mu(x_i) + S_c(x_i) \\ \mu(x_i) &= D\beta = \sum_{j=1}^{k} f_k(x_i)\beta_k \end{aligned}$$

where $d_k(x_i)$ is a vector of covariates at location x_i

$$Y \sim \text{MVN}(D\beta, \sigma^2 R + \tau^2 I)$$

The likelihood function is

$$L(\beta, \tau, \sigma, \phi, \kappa) \propto -0.5\{\log|(\sigma^2 R + \tau^2 I)| + (y - D\beta)'(\sigma^2 R + \tau^2 I)^{-1}(y - D\beta)\}.$$

- reparametrise $\nu^2 = \tau^2/\sigma^2$ and denote $\sigma^2 V = \sigma^2(R + \nu^2 I)$
- the log-likelihood function is maximised for

$$\hat{\beta}(V) = (D'V^{-1}D)^{-1}D'V^{-1}y$$
$$\hat{\sigma}^2(V) = n^{-1}(y - D\hat{\beta})'V^{-1}(y - D\hat{\beta})$$

- concentrated likelihood: substitute (β, σ^2) by $(\hat{\beta}, \hat{\sigma^2})$ and the maximisation reduces to

$$L(\tau_r, \phi, \kappa) \propto -0.5\{n \log|\hat{\sigma^2}| + \log|(R + \nu^2 I)|\}$$

Some technical issues

- poor quadratic approximations, unreliable Hessian matrices
- identifiability issues for more than two parameters in the correlation function
- for models such as *Matérn* and *powered exponential* ϕ and κ are not orthogonal
- For the Matérn correlation function we suggest to take κ in a discrete set $\{0.5, 1, 2, 3, \ldots, N\}$ ("profiling")
- other possible approach is reparametrization such as replacing ϕ by $\alpha = 2\sqrt{\kappa}\phi$ (Handcock and Wallis)
- stability: e.g. Zhang's remarks on σ^2/ϕ
- reparametrisations and asymptotics, e.g. $\theta_1 = \log(\sigma^2/\phi^{2\kappa})$ and $\theta_2 = \log(\phi^{2\kappa})$

Note: variations on the likelihood

- we strongly favor likelihood based methods.
- examining profile likelihoods can be reavealing on model identifiability and parameter uncertainty.
- restricted maximum likelihood is widely recommended leading to less biased estimators but is sensitive to mis-specification of the mean model. In spatial models distinction between $\mu(x)$ and $S(x)$ is not sharp.
- approximate likelihoods are useful for large data-sets.
- composite likelihood uses independent contributions for the likelihood function for each pair of points.
- Markov Random Fields can be used to approximate geostatistical models.
- ...

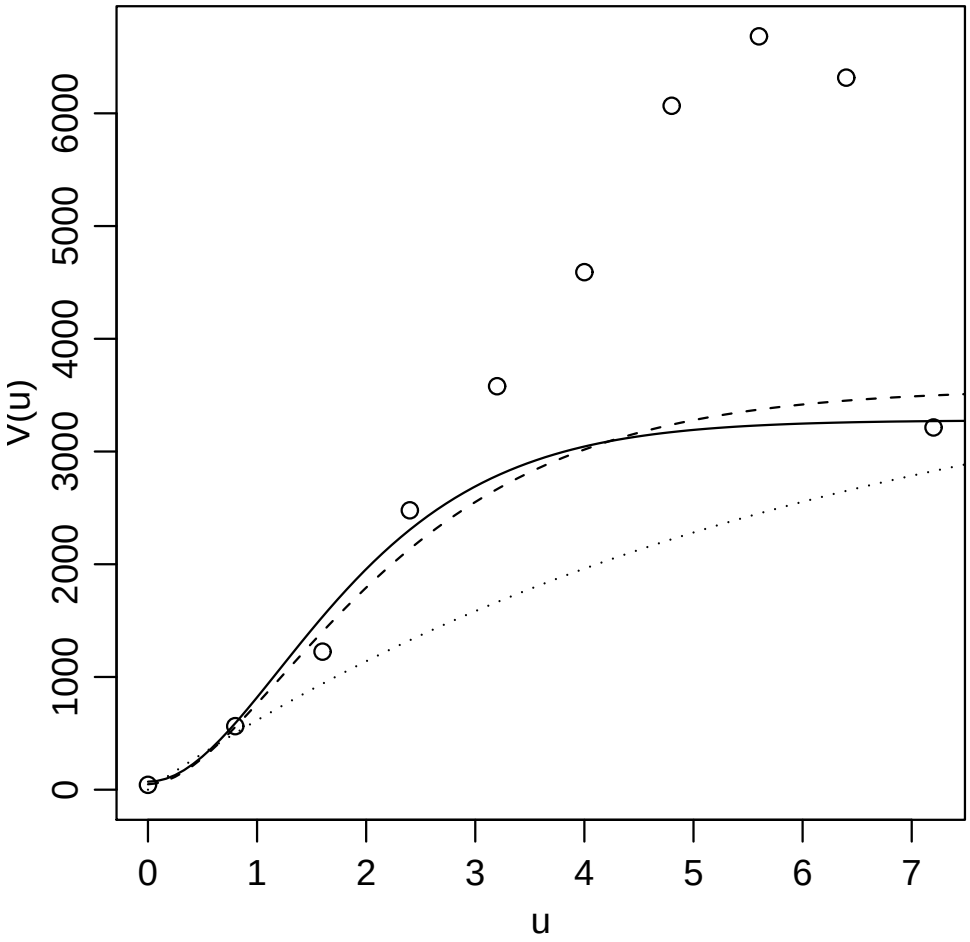
Example: Surface elevation data

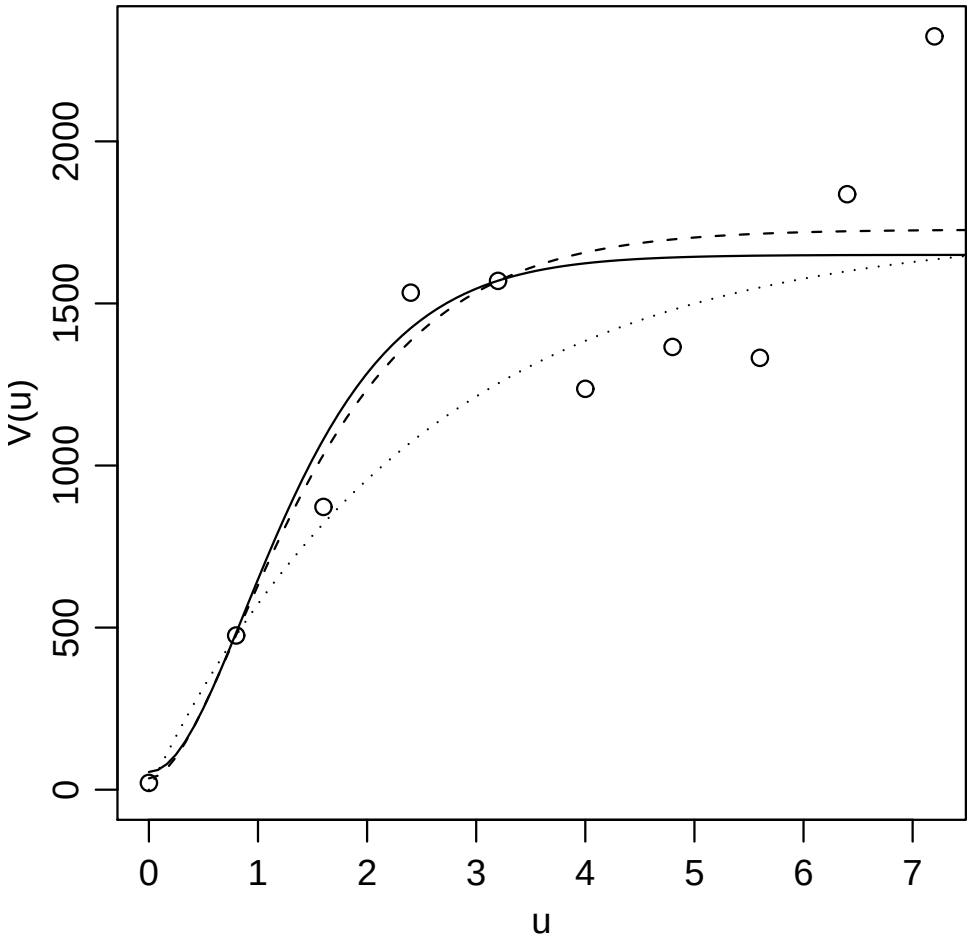
model with constant mean

model	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\phi}$	$\hat{\tau}^2$	logL
$\kappa = 0.5$	863.71	4087.6	6.12	0	-244.6
$\kappa = 1.5$	848.32	3510.1	1.2	48.16	-242.1
$\kappa = 2.5$	844.63	3206.9	0.74	70.82	-242.33

model with linear trend

model	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$	$\hat{\phi}$	$\hat{\tau}^2$	logL
$\kappa = 0.5$	919.1	-5.58	-15.52	1731.8	2.49	0	-242.71
$\kappa = 1.5$	912.49	-4.99	-16.46	1693.1	0.81	34.9	-240.08
$\kappa = 2.5$	912.14	-4.81	-17.11	1595.1	0.54	54.72	-239.75



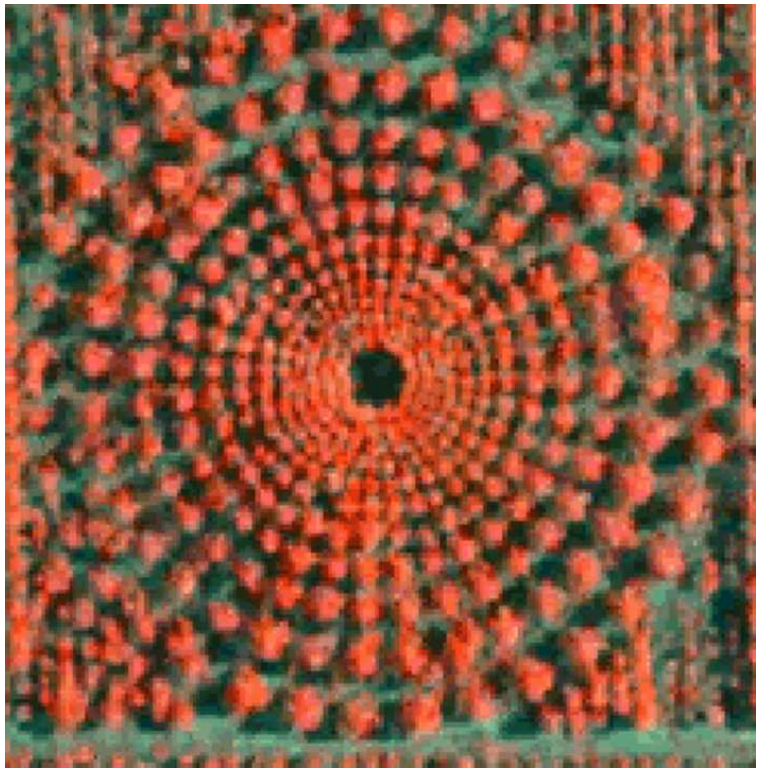


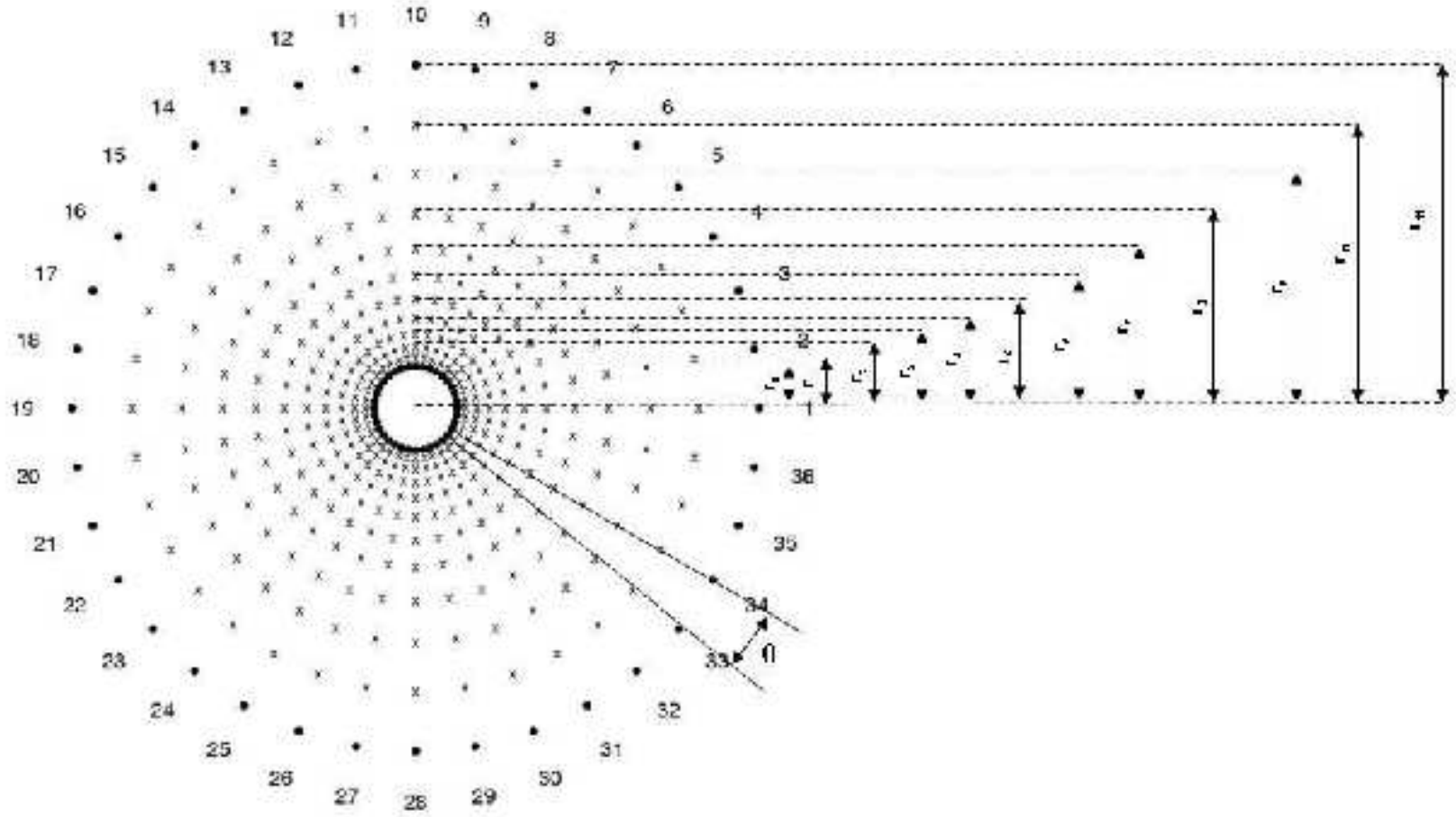
Example: experiment on systematic design

Example: experiment on systematic design

Motivation

- missing data
- reliable inference





LEGENDA

×- plantas úteis; • –bordadura; 1:36 - número dos raios; θ= 10°. r_0 = 5,30 m, r_1 = 6,42 m, r_2 = 7,78 m, r_3 = 9,42 m, r_4 = 11,41 m, r_5 = 13,82 m, r_6= 16,74 m, r_7=20,27 m, r_8 = 24,55 m, r_9 = 29,73, r_D =36,00 m, r_{II} = 43,60 m, calculado pela por: $m = r_0 a^n$, em que a=1,21 e r_0 = 5,30 m a distância radial do primeiro raio.

Prediction – general results

goal: predict the realised value of a (scalar) r.v. T, using data y a realisation of a (vector) r.v. Y.

predictor: of T is any function of Y, $\hat{T} = t(Y)$

a criterion – MMSPE: the *best* predictor minimises

$$MSPE(\hat{T}) = \mathrm{E}\left[(T - \hat{T})^2\right]$$

The MMSEP of T is $\hat{T} = \mathrm{E}(T|Y)$

The prediction mean square error of $\hat{T}$ is

$$\mathrm{E}[(T - \hat{T})^2] = \mathrm{E}_Y[\mathrm{Var}(T|Y)],$$

(the prediction variance is an estimate of $\mathrm{MSPE}(\hat{T})$).

$\mathrm{E}[(T - \hat{T})^2] \leq \mathrm{Var}(T)$, with equality if T and Y are independent random variables.

Prediction – general results (cont.)

- We call $\hat{T}$ the least squares predictor for T, and $\mathrm{Var}(T|Y)$ its prediction variance
- $\mathrm{Var}(T) - \mathrm{Var}(T|Y)$ measures the contribution of the data (exploiting dependence between T and Y)
- point prediction, prediction variance are summaries
- complete answer is the distribution $[T|Y]$ (analytically or a sample from it)
- not transformation invariant:
  $\hat{T}$ the best predictor for T does NOT necessarily imply that $g(\hat{T})$ is the best predictor for $g(T)$.

Prediction – Linear Gaussian model

Suppose the target for prediction is $T = S(x)$

The MMSEP is $\hat{T} = \mathrm{E}[S(x)|Y]$

- **$[S(x), Y]$ are jointly multivariate Gaussian. with mean vector $\mu 1$ and variance matrix**

$$\begin{bmatrix} \sigma^2 & \sigma^2 \mathrm{r}' \\ \sigma^2 \mathrm{r} & \tau^2 I + \sigma^2 R \end{bmatrix}$$

  where r is a vector with elements $r_i = \rho(||x - x_i||) : i = 1, \ldots, n$.

- $$\hat{T} = \mathrm{E}[S(x)|Y] = \mu + \sigma^2 \mathrm{r}'(\tau^2 I + \sigma^2 R)^{-1}(Y - \mu 1) \qquad (1)$$

- $$\mathrm{Var}[S(x)|Y] = \sigma^2 - \sigma^2 \mathrm{r}'(\tau^2 I + \sigma^2 R)^{-1}\sigma^2 \mathrm{r}$$

Prediction – Linear Gaussian model (cont.)

- for the Gaussian model $\hat{T}$ is linear in Y, so that

$$\hat{T} = w_0(x) + \sum_{i=1}^{n} w_i(x) Y_i$$

- equivalent to a least squares problem to find w_i which minimise $MSPE(\hat{T})$ within the class of linear predictors.
- Because the conditional variance does not depend on Y, the prediction MSE is equal to the prediction variance.
- Equality of prediction MSE and prediction variance is a special property of the multivariate Gaussian distribution, not a general result.

Prediction – Linear Gaussian model (cont.)

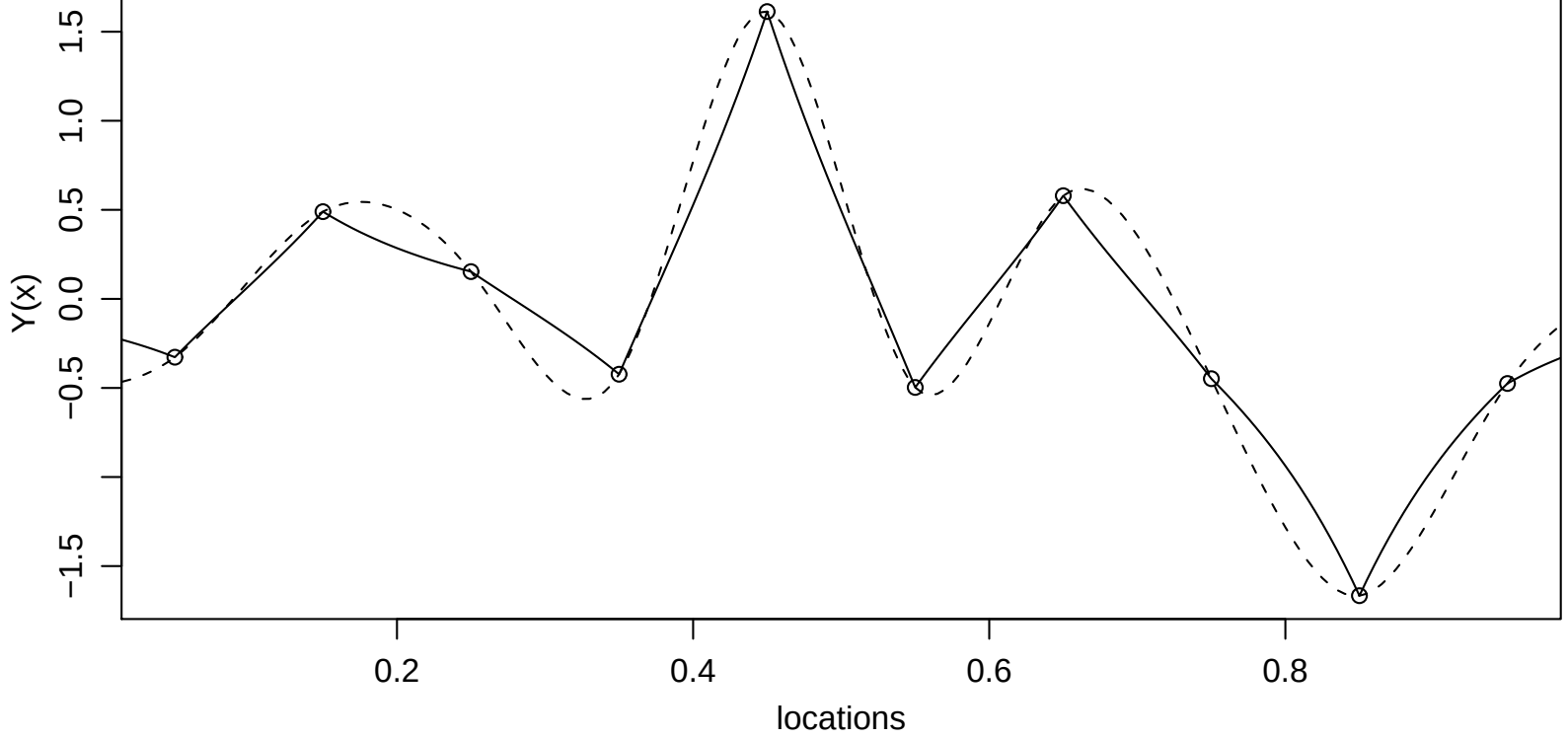
- Construction of the surface $\hat{S}(x)$, where $\hat{T} = \hat{S}(x)$ is given by (1), is called simple kriging.
- Assumes known model parameters.
- This name is a reference to D.G. Krige, who pioneered the use of statistical methods in the South African mining industry (Krige, 1951).

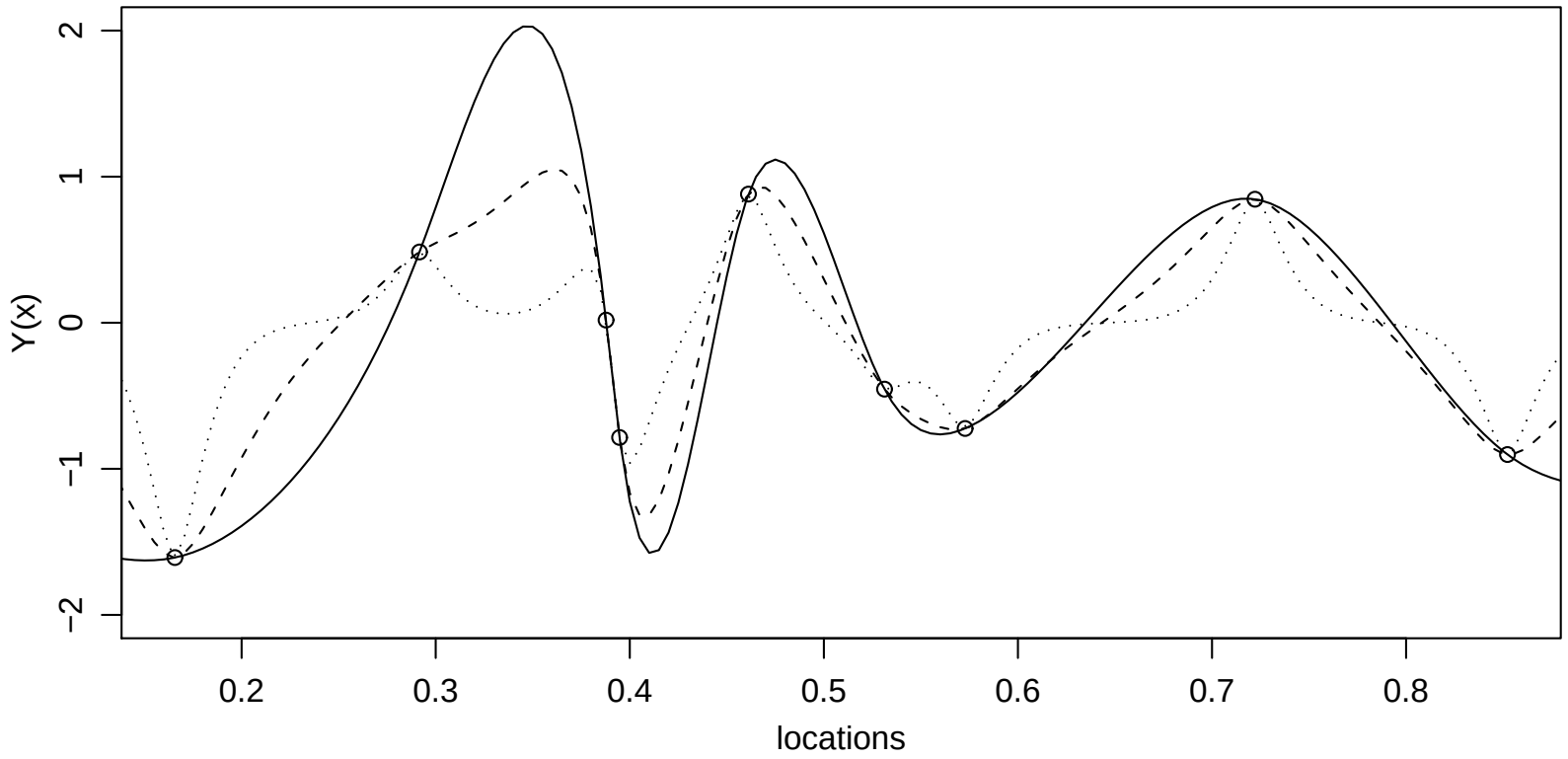
Features of spatial prediction

The minimum mean square error predictor for $S(x)$ is given by

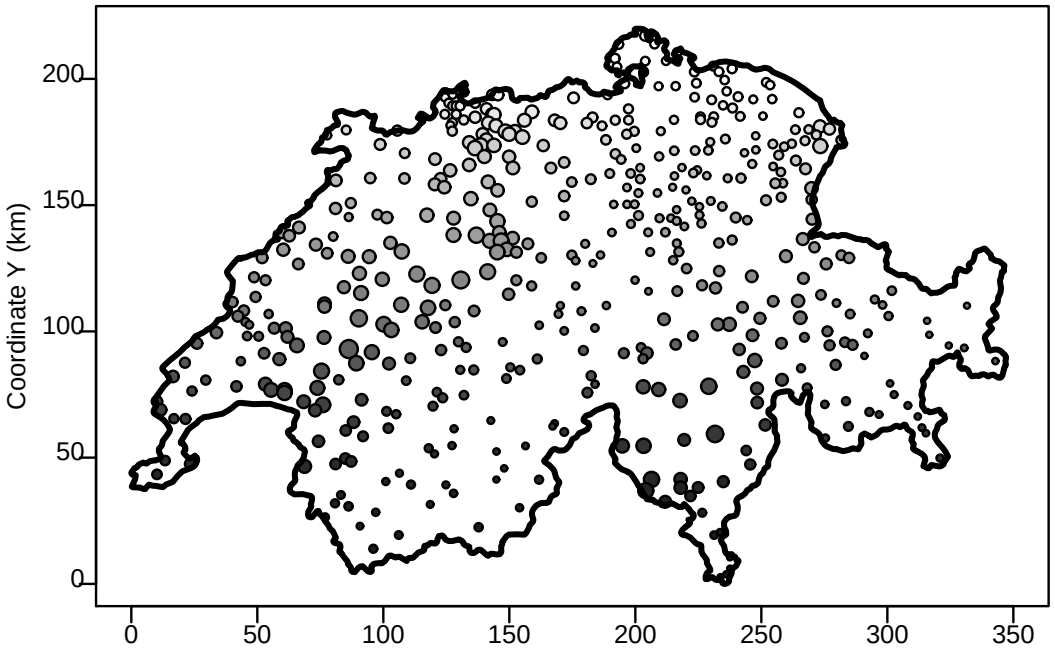
$$\begin{aligned} \hat{T} = \hat{S}(x) &= \mu + \sum_{i=1}^{n} w_i(x)(Y_i - \mu) \\ &= \{1 - \sum_{i=1}^{n} w_i(x)\}\mu + \sum_{i=1}^{n} w_i(x) Y_i \end{aligned}$$

- shows the predictor $\hat{S}(x)$ compromises between its unconditional mean μ and the observed data Y,
- the nature of the compromise depends on the target location x, the data-locations x_i and the values of the model parameters,
- $w_i(x)$ are the prediction weights.


Y(x)
1.5
1.0
0.5
0.0
−0.5
−1.5
0.2
0.4
0.6
0.8
locations


Y(x)
2
1
0
−1
−2
0.2
0.3
0.4
0.5
0.6
0.7
0.8
locations

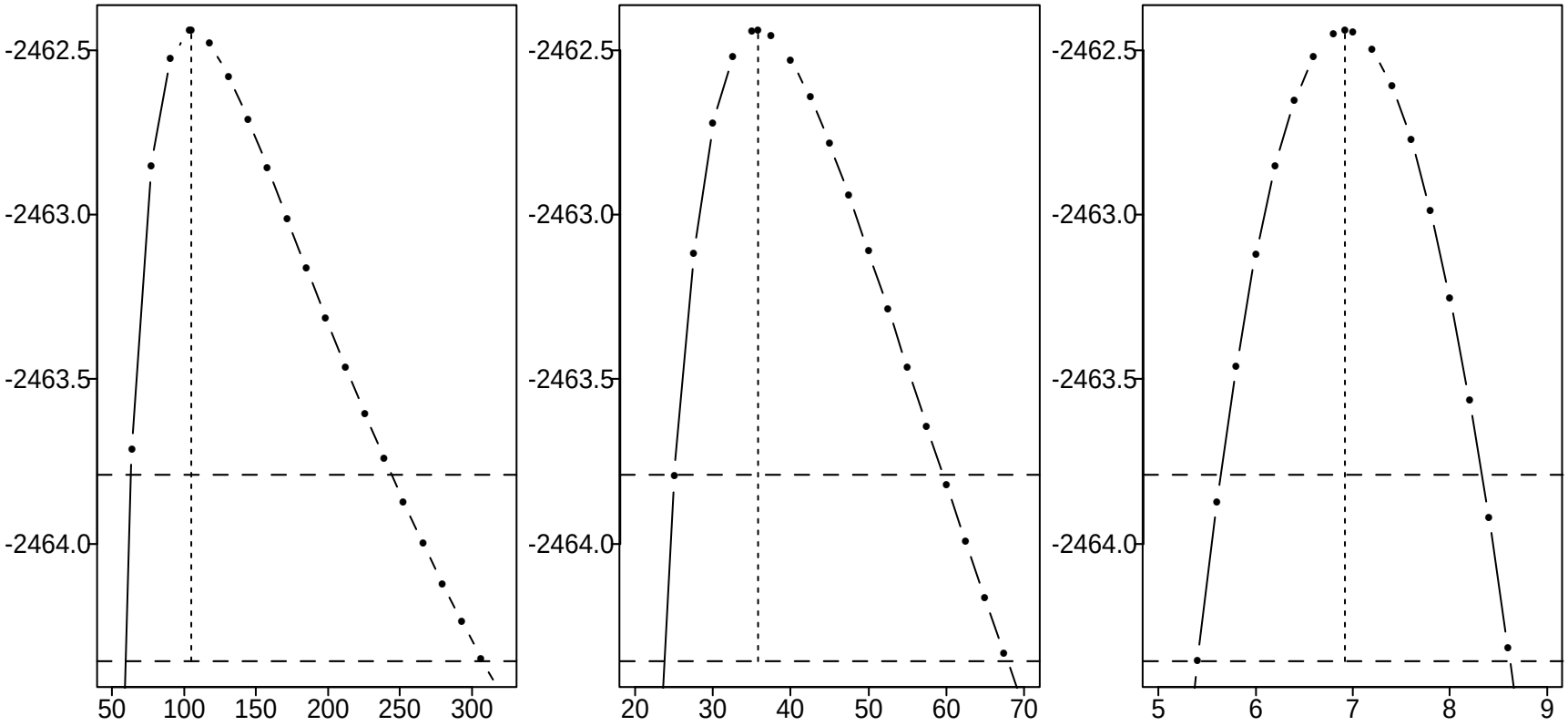
Swiss rainfall data – trans-Gaussian model

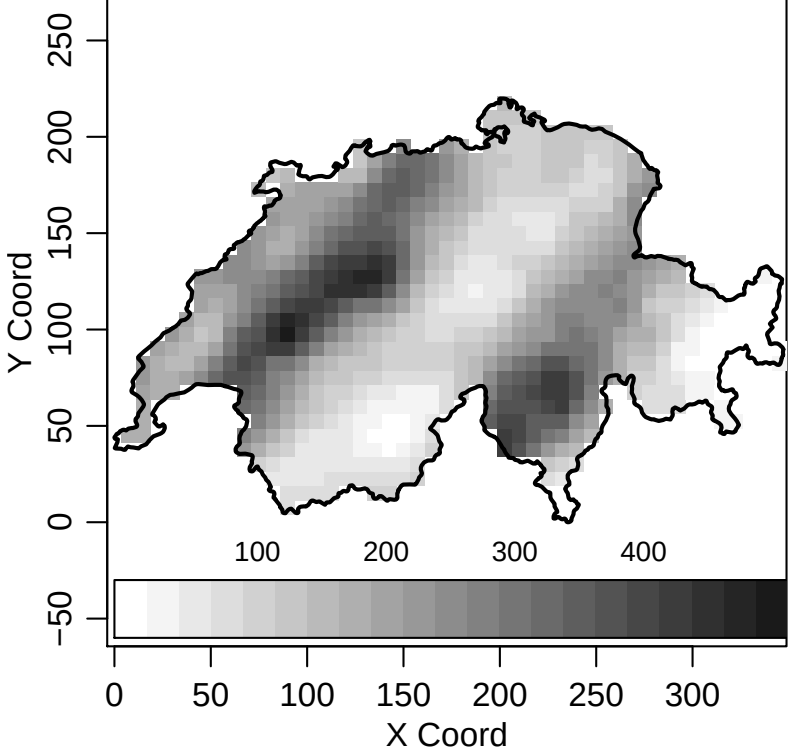


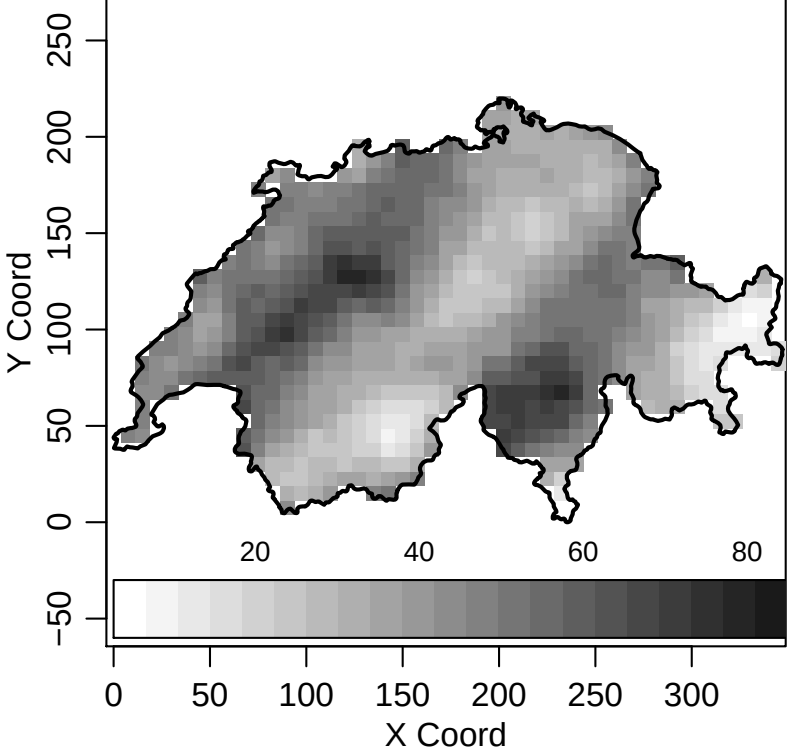
$$Y_i^* = h_\lambda(Y) = \begin{cases} \frac{(y_i)^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(y_i) & \text{if } \lambda = 0 \end{cases}$$

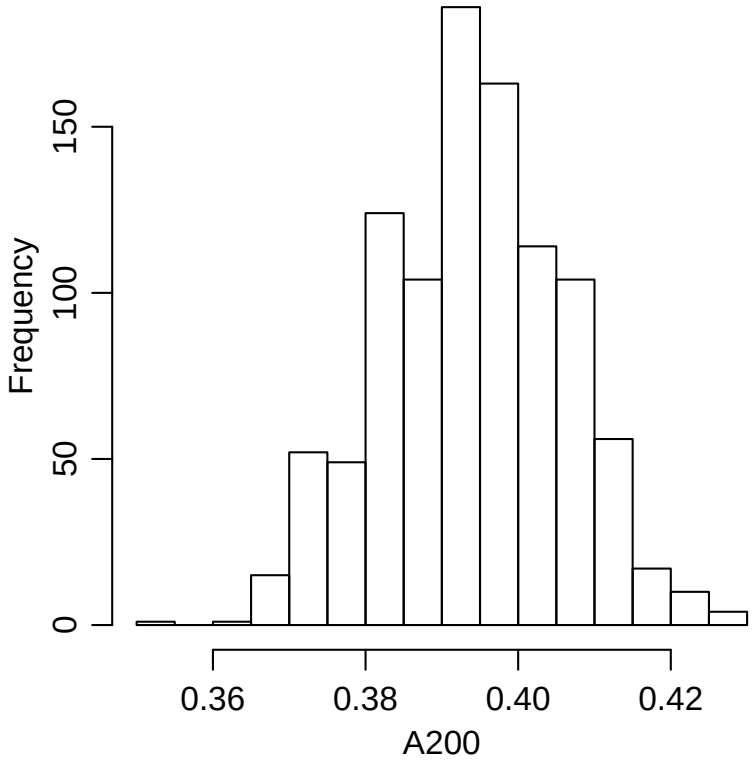
$$\ell(\beta, \theta, \lambda) = -\frac{1}{2}\{\log|\sigma^2 V| + (h_\lambda(y) - D\beta)'\{\sigma^2 V\}^{-1}(h_\lambda(y) - D\beta)\} + \sum_{i=1}^{n} \log\left((y_i)^{\lambda-1}\right)$$

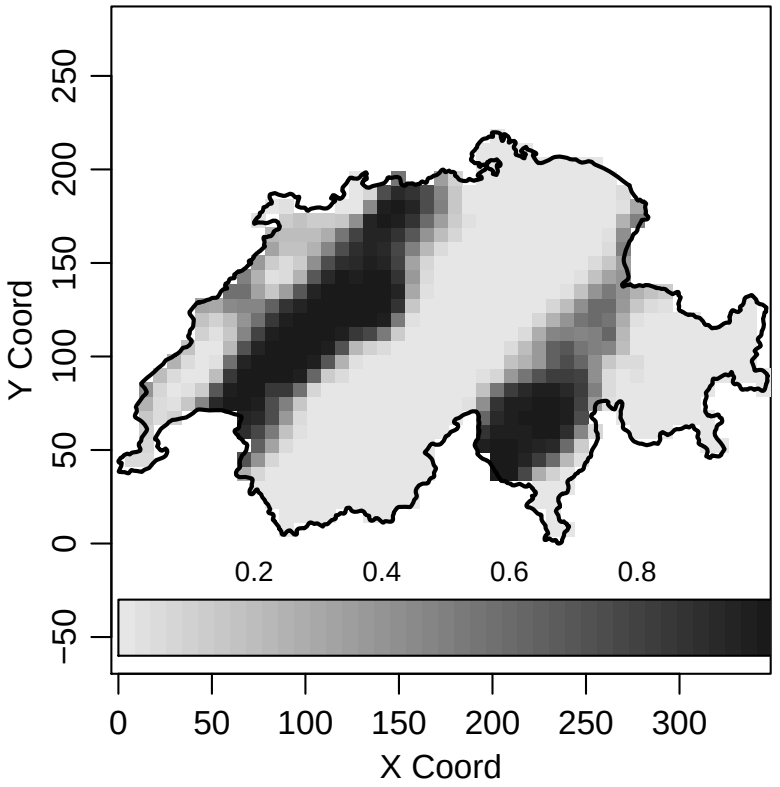
κ	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\phi}$	$\hat{\tau}^2$	$\log \hat{L}$
0.5	18.36	118.82	87.97	2.48	-2464.315
1	20.13	105.06	35.79	6.92	-2462.438
2	21.36	88.58	17.73	8.72	-2464.185




Y Coord
X Coord
100
200
300
400


Y Coord
X Coord
20
40
60
80


Frequency
A200
0.36
0.38
0.40
0.42


Y Coord
X Coord
0.2
0.4
0.6
0.8

PART 3

Bayesian Inference

Bayesian Basics

Bayesian inference deals with parameter uncertainty by treating parameters as random variables, and expressing inferences about parameters in terms of their conditional distributions, given all observed data.

- **model specification includes model parameters:**

$$[Y, \theta] = [\theta][Y|\theta]$$

- **inference using Bayes' Theorem:**

$$[Y, \theta] = [Y|\theta][\theta] = [Y][\theta|Y]$$

- **to derive the posterior distribution**

$$[\theta|Y] = [Y|\theta][\theta]/[Y] \propto [Y|\theta][\theta]$$

- **The prior distribution $[\theta]$ express the uncertainty about the model parameters**

- The posterior distribution $[\theta|Y]$ express the *revised* uncertainty after observing Y
- conjugacy is achieved in particular models where convenient choices of $[\theta]$ produces $[\theta|Y]$ within the same family
- more generally $[\theta|Y]$ may be an unknown and $[Y] = \int [Y|\theta][\theta]d\theta$ may need to be evaluated numerically.
- probability statements and estimates are based on the posterior density obtained through

$$p(\theta|y) = \frac{\ell(\theta; y)\pi(\theta)}{\int \ell(\theta; y)\pi(\theta)d\theta}$$

  are usually expressed as summary statistics (mean, median, mode) and/or Bayesian credibility intervals
- credible intervals are not uniquely defined (e.g. quantile based, highest density interval, etc)

Prediction

For Bayesian prediction expand the Bayes' theorem to include the prediction target, allowing for uncertainty on model parameters to be accounted for.

- **and for prediction**

$$[Y, T, \theta] = [Y, T|\theta][\theta]$$

- **derive the predictive distribution**

$$[T|Y] = \int [T, \theta|Y] d\theta = \int [T|Y, \theta][\theta|Y] d\theta$$

- **can be interpreted as a weighted prediction over possible values of $[\theta|Y]$**

- **in general, as data becomes more abundant $[\theta|Y]$ concentrates around $\hat{\theta}$**

Bayesian inference for the geostatistical model

Bayesian inference for the geostatistcal model expands the previous results acknowledging for Y and S as specified by the adopted model.

- **model specification:**

$$[Y, S, \theta] = [\theta][Y, S|\theta] = [\theta][S|\theta][Y|S, \theta]$$

- **inference using Bayes' Theorem:**

$$[Y, S, \theta] = [Y, S|\theta][\theta] = [Y][\theta, S|Y]$$

- **to derive the posterior distribution**

$$[\theta|Y] = \int [\theta, S|Y] dS = \int \frac{[Y|S, \theta][S|\theta][\theta]}{[Y]} dS$$

- **where $[Y] = \int\int [Y|\theta][S|\theta][\theta] dS d\theta$ is typically difficult to evaluate**

- **For prediction**

$$[Y, T, S, \theta] = [Y, T|S, \theta][S|\theta][\theta]$$

- **derive the predictive distribution**

$$[T|Y] = \int\int [T, S, \theta|Y] dS d\theta = \int\int [T|Y, S, \theta][S, \theta|Y] dS d\theta$$

- **and explore the conditional independence structure of the model to simplify the calculations**

Notes I

- likelihood function occupies a central role in both classical and Bayesian inference
- plug-in prediction corresponds to inferences about $[T|Y, \hat{\theta}]$
- Bayesian prediction is a weighted average of plug-in predictions, with different plug-in values of θ weighted according to their conditional probabilities given the observed data.
- Bayesian prediction is usually more cautious than plug-in prediction.
  Allowance for parameter uncertainty usually results in wider prediction intervals

Notes II

1. **The need to evaluate the integral which defines $[Y]$ represented a major obstacle to practical application,**

2. **development of Markov Chain Monte Carlo (MCMC) methods has transformed the situation.**

3. **BUT, for geostatistical problems, reliable implementation of MCMC is not straightforward. Geostatistical models don't have a natural Markovian structure for the algorithms work well.**

4. **in particular for the Gaussian model other algorithms can be implemented.**

Results for the Gaussian models - I

- fixing covariance parameters and assuming a (conjugate) prior for β

$$\beta \sim \mathrm{N}\left(m_\beta \,;\, \sigma^2 V_\beta\right)$$

- The posterior is given by

$$\begin{aligned}
[\beta|Y] &\sim \mathrm{N}((V_\beta^{-1} + D'R^{-1}D)^{-1}(V_\beta^{-1}m_\beta + D'R^{-1}y) \,; \\
&\qquad \sigma^2\,(V_\beta^{-1} + D'R^{-1}D)^{-1}) \\
&\sim \mathrm{N}\left(\hat{\beta} \,;\, \sigma^2\, V_{\hat{\beta}}\right)
\end{aligned}$$

- and the predictive distribution is

$$p(S^*|Y, \sigma^2, \phi) = \int p(S^*|Y, \beta, \sigma^2, \phi)\, p(\beta|Y, \sigma^2, \phi)\, d\beta.$$

- with mean and variance given by

$$
\begin{aligned}
E[S^*|Y] &= (D_0 - r'V^{-1}D)(V_\beta^{-1} + D'V^{-1}D)^{-1}V_\beta^{-1}m_\beta + \\
&\quad \left[r'V^{-1} + (D_0 - r'V^{-1}D)(V_\beta^{-1} + D'V^{-1}D)^{-1}D'V^{-1}\right]Y \\
\mathrm{Var}[S^*|Y] &= \sigma^2\left[V_0 - r'V^{-1}r + \right. \\
&\quad \left. (D_0 - r'V^{-1}D)(V_\beta^{-1} + D'V^{-1}D)^{-1}(D_0 - r'V^{-1}D)'\right].
\end{aligned}
$$

- predicted mean balances between prior and weighted average of the data
- The predictive variance has three interpretable components: a priori variance, the reduction due to the data and the uncertainty in the mean.
- $V_\beta \to \infty$ results can be related to REML and universal (or ordinary) kriging.

Results for the Gaussian models - II

- fixing correlation parameters and assuming a (conjugate) prior for $[\beta, \sigma^2] \sim N\chi^2_{ScI}\left(m_b, V_b, n_\sigma, S^2_\sigma\right)$ given by:

$$[\beta|\sigma^2] \sim \mathrm{N}\left(m_\beta\ ;\ \sigma^2 V_\beta\right) \text{ and } [\sigma^2] \sim \chi^2_{ScI}(n_\sigma, S^2_\sigma)$$

- The posterior is $[\beta, \sigma^2|y, \phi] \sim N\chi^2_{ScI}\left(\tilde{\beta}, V_{\tilde{\beta}}, n_\sigma + n, S^2\right)$

$$\tilde{\beta} = V_{\tilde{\beta}}(V_b^{-1} m_b + D' R^{-1} y)$$

$$V_{\tilde{\beta}} = (V_b^{-1} + D' R^{-1} D)^{-1}$$

$$S^2 = \frac{n_\sigma S^2_\sigma + m'_b V_b^{-1} m_b + y' R^{-1} y - \tilde{\beta}' V_{\tilde{\beta}}^{-1} \tilde{\beta}}{n_\sigma + n}$$

- The predictive distribution $[S^*|y] \sim t_{n_\sigma+n}\left(\mu^*, S^2\Sigma^*\right)$
- with mean and variance given by

$$\begin{aligned} \mathbf{E}[S^*|y] &= \mu^*, \\ \mathbf{Var}[S^*|y] &= \frac{n_\sigma + n}{n_\sigma + n - 2} S^2\Sigma^*, \end{aligned}$$

$$\begin{aligned} \mu^* &= (D^* - r'V^{-1}D)V_{\tilde{\beta}}V_b^{-1}m_b \\ &\quad + \left[r'V^{-1} + (D^* - r'V^{-1}D)V_{\tilde{\beta}}D'V^{-1}\right]y, \\ \Sigma^* &= V^0 - r'V^{-1}r + (D^* - r'V^{-1}D)(V_b^{-1} + V_{\hat{\beta}}^{-1})^{-1}(D^* - r'V^{-1}D)'. \end{aligned}$$

- valid if $\tau^2 = 0$
- for $\tau^2 > 0$, $\nu^2 = \tau^2/\sigma^2$ can regarded as a correlation parameter

Results for the Gaussian models - III

Assume a prior $p(\beta, \sigma^2, \phi) \propto \frac{1}{\sigma^2} p(\phi)$.

- **The posterior distribution for the parameters is:**

$$p(\beta, \sigma^2, \phi|y) = p(\beta, \sigma^2|y, \phi)\ p(\phi|y)$$

- **where $p(\beta, \sigma^2|y, \phi)$ can be obtained analytically and**

$$pr(\phi|y) \propto pr(\phi)\ |V_{\hat{\beta}}|^{\frac{1}{2}}\ |R_y|^{-\frac{1}{2}}\ (S^2)^{-\frac{n-p}{2}}$$

- **analogous results for more general prior:**

  $[\beta|\sigma^2, \phi] \sim N\left(m_b, \sigma^2 V_b\right)$ **and** $[\sigma^2|\phi] \sim \chi^2_{ScI}\left(n_\sigma, S^2_\sigma\right)$,

- **choice of prior for ϕ can be critical. (Berger, De Oliveira & Sansó, 2001)**

Algorithm 1:

1. **Discretise the distribution $[\phi|y]$, i.e. choose a range of values for ϕ which is sensible for the particular application, and assign a discrete uniform prior for ϕ on a set of values spanning the chosen range.**

2. **Compute the posterior probabilities on this discrete support set, defining a discrete posterior distribution with probability mass function $\tilde{pr}(\phi|y)$, say.**

3. **Sample a value of ϕ from the discrete distribution $\tilde{pr}(\phi|y)$.**

4. **Attach the sampled value of ϕ to the distribution $[\beta, \sigma^2|y, \phi]$ and sample from this distribution.**

5. **Repeat steps (3) and (4) as many times as required; the resulting sample of triplets (β, σ^2, ϕ) is a sample from the joint posterior distribution.**

The predictive distribution is given by:

$$\begin{aligned} p(S^*|Y) &= \int\int\int p(S^*, \beta, \sigma^2, \phi|Y)\, d\beta\, d\sigma^2\, d\phi \\ &= \int\int\int p(s^*, \beta, \sigma^2|y, \phi)\ d\beta\, d\sigma^2\, pr(\phi|y)\ d\phi \\ &= \int p(S^*|Y, \phi)\ p(\phi|y)\, d\phi. \end{aligned}$$

Algorithm 2:

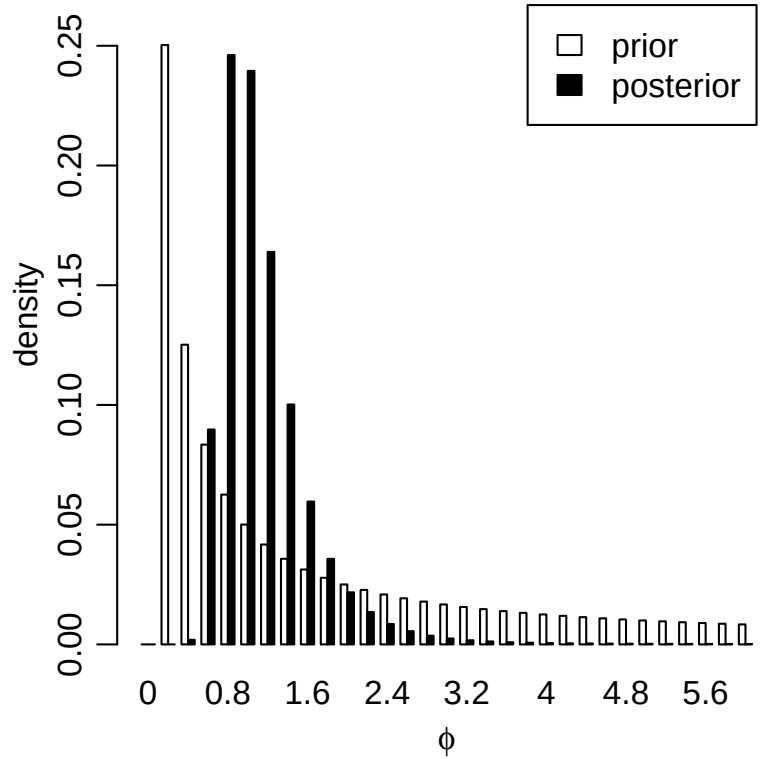
1. **Discretise $[\phi|Y]$, as in Algorithm 1.**
2. **Compute the posterior probabilities on the discrete support set. Denote the resulting distribution $\tilde{pr}(\phi|y)$.**
3. **Sample a value of ϕ from $\tilde{pr}(\phi|y)$.**
4. **Attach the sampled value of ϕ to $[s^*|y, \phi]$ and sample from it obtaining realisations of the predictive distribution.**
5. **Repeat steps (3) and (4) to generate a sample from the required predictive distribution.**

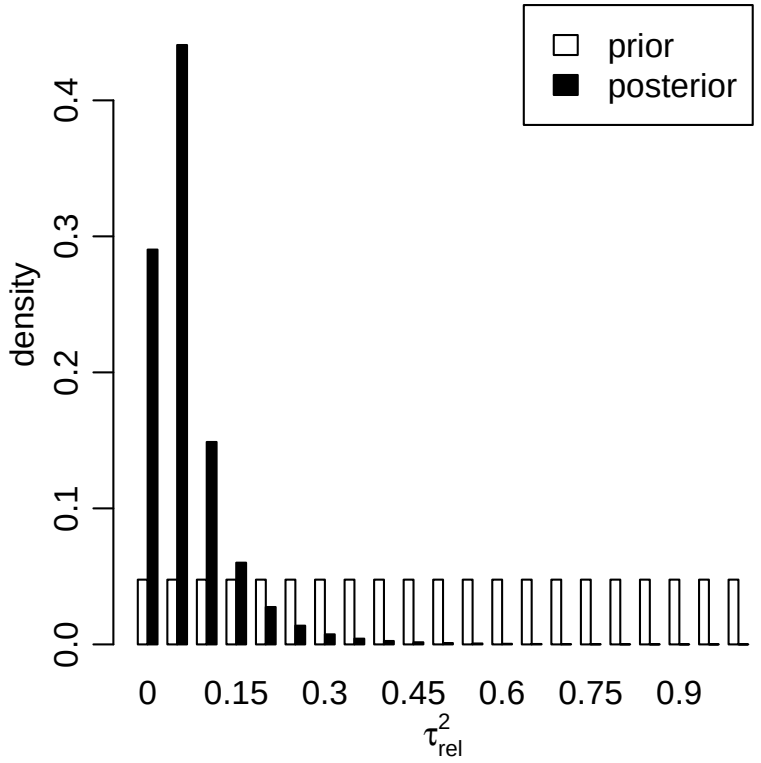
Notes

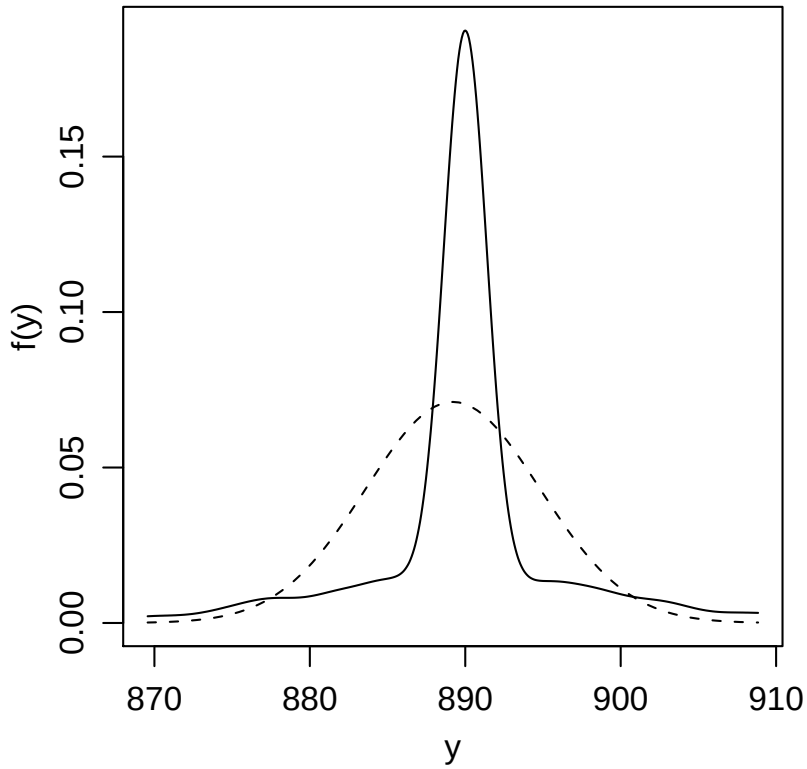
1. **The algorithms are of the same kind to treat τ and/or κ as unknown parameters.**

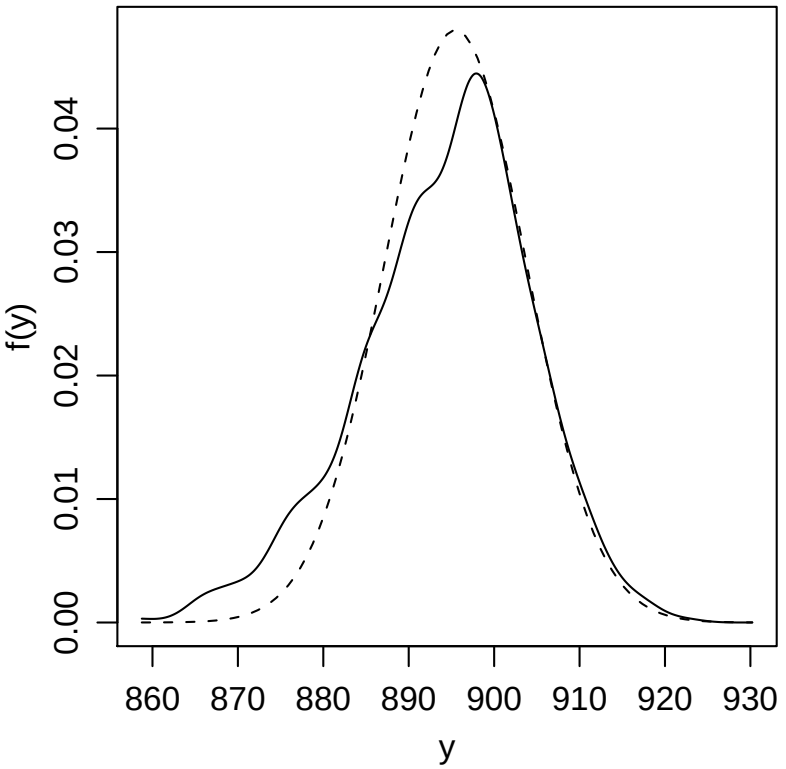
2. **We specify a discrete prior distribution on a multi-dimensional grid of values.**

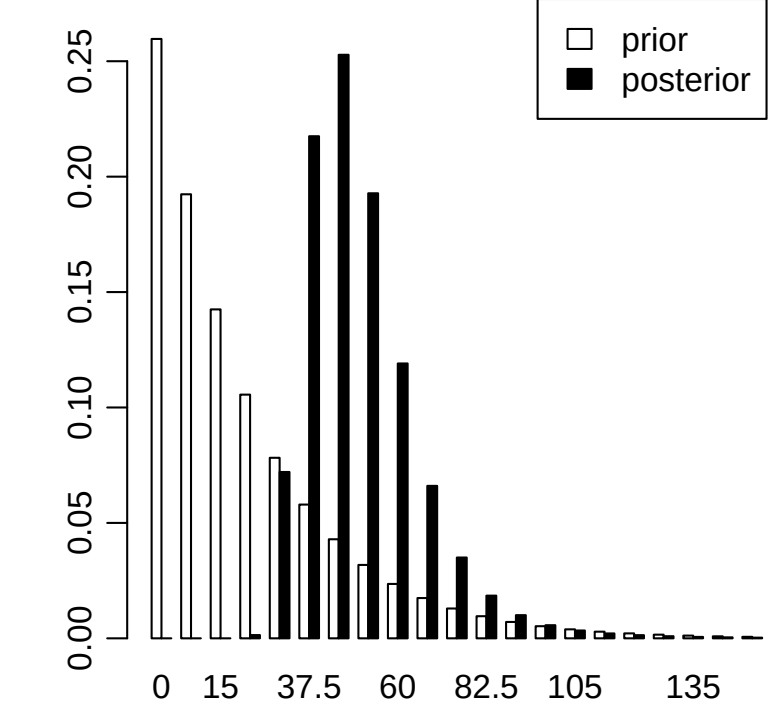
3. **This implies extra computational load (but no new principles)**

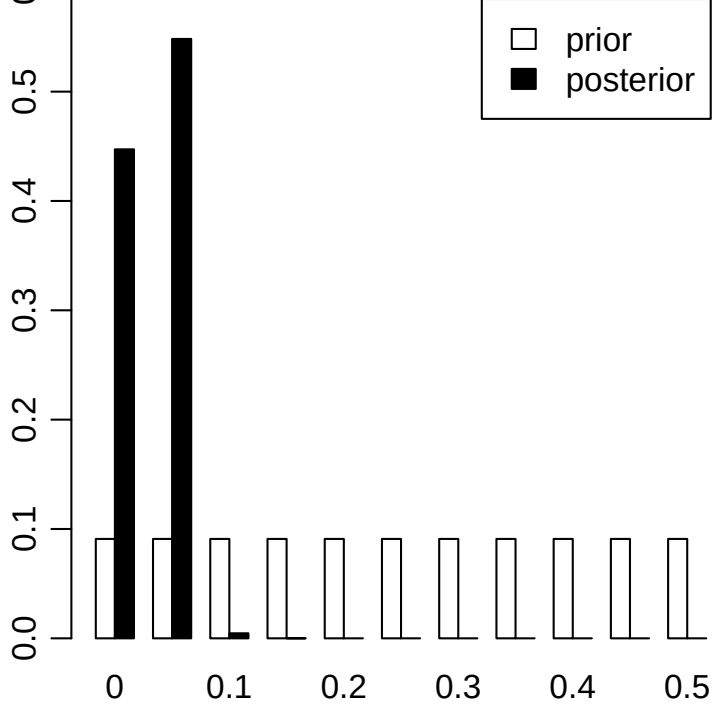

prior
posterior
density
0.00
0.05
0.10
0.15
0.20
0.25
0
0.8
1.6
2.4
3.2
4
4.8
5.6
ϕ

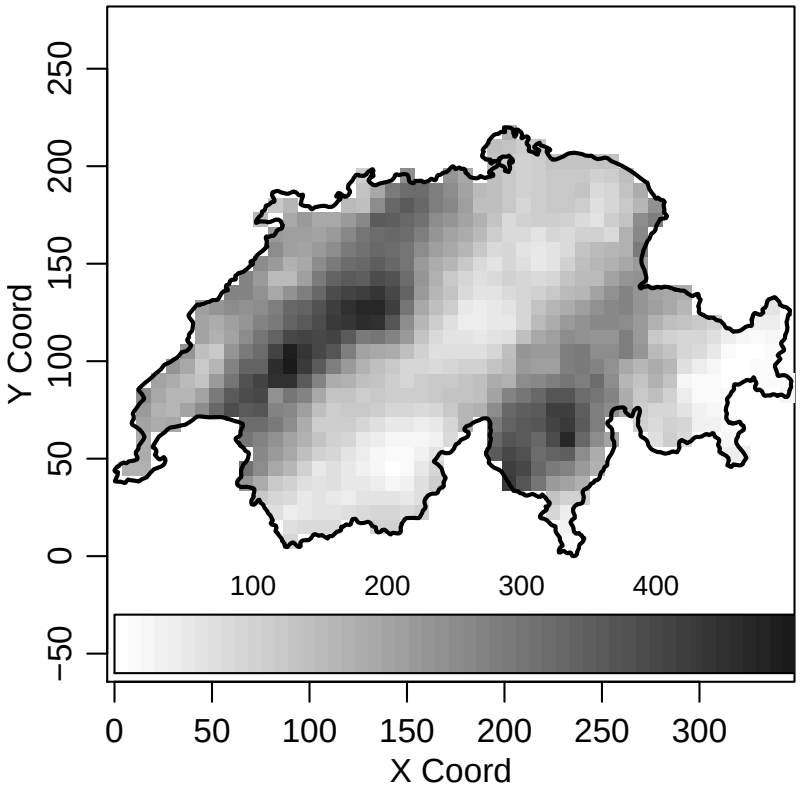

prior
posterior
density
0.0
0.1
0.2
0.3
0.4
0
0.15
0.3
0.45
0.6
0.75
0.9
τ^2_{rel}


f(y)
0.00
0.05
0.10
0.15
870
880
890
900
910
y


f(y)
0.00
0.01
0.02
0.03
0.04
860
870
880
890
900
910
920
930
y


prior
posterior
0.00
0.05
0.10
0.15
0.20
0.25
0
15
37.5
60
82.5
105
135


prior
posterior
0.0
0.1
0.2
0.3
0.4
0.5
0
0.1
0.2
0.3
0.4
0.5


Y Coord
−50
0
50
100
150
200
250
100
200
300
400
0
50
100
150
200
250
300
X Coord

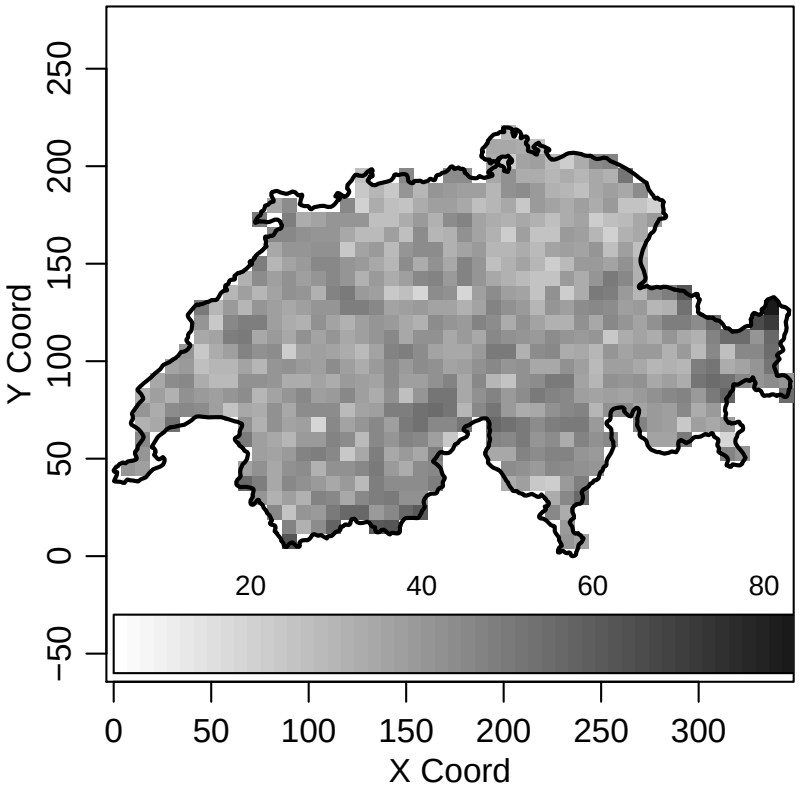
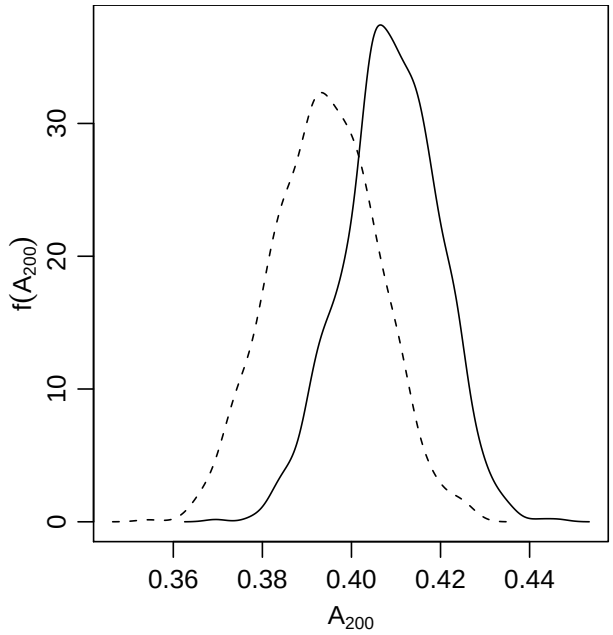
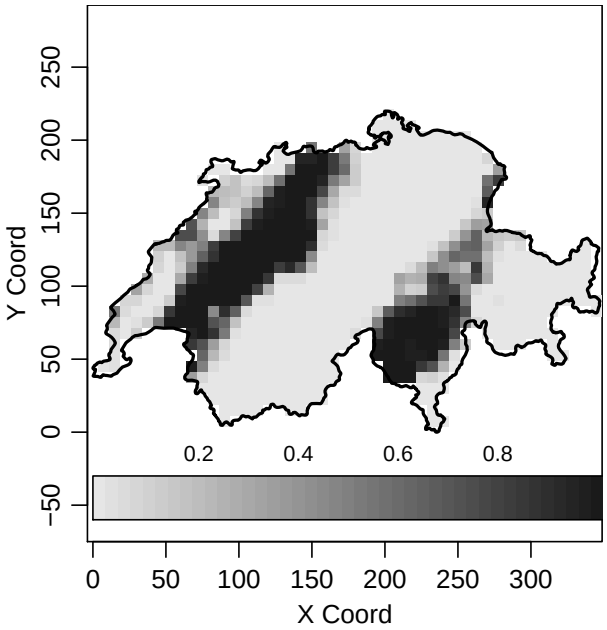

Y Coord
−50
0
50
100
150
200
250
20
40
60
80
0
50
100
150
200
250
300
X Coord

Table 1: Swiss rainfall data: posterior means and 95% central quantile-based credible intervals for the model parameters.

parameter	estimate	95% interval
β	**144.35**	**[53.08 , 224.28]**
σ^2	**13662.15**	**[8713.18 , 27116.35]**
ϕ	**49.97**	**[30 , 82.5]**
ν^2	**0.03**	**[0 , 0.05]**





Generalized linear geostatistical model

- Preserving the assumption of a zero mean, stationary Gaussian process $S(\cdot)$,
- our basic model can be generalized replacing the assumption of mutually independent $Y_i|S(\cdot) \sim \mathrm{N}(S(x), \tau^2)$ by assuming
  $Y_i|S(\cdot)$ are mutually independent within the class of generalized linear models (GLM)
- with a link function $h(\mu_i) = \sum_{j=1}^{p} d_{ij}\beta_j + S(x_i)$
- this defines a generalized linear mixed model (GLMM) with correlated random effects
- which provides a way to adapt classical GLM for geostatistical applications.

GLGM

- usually just a single realisation is available, in contrast with GLMM for longitudinal data analysis
- The GLM approach is most appealing when follows an natural sampling mecanism such as Poisson model for counts and logist-linear models for binary/binomial responses
- in principle transformed models can be considered for skewed distributions
- variograms for such processes can be obtained although providing a less obvious summary statistics
- empirical variograms of GLM residuals can be used for exploratory analysis

An example: a Poisson model

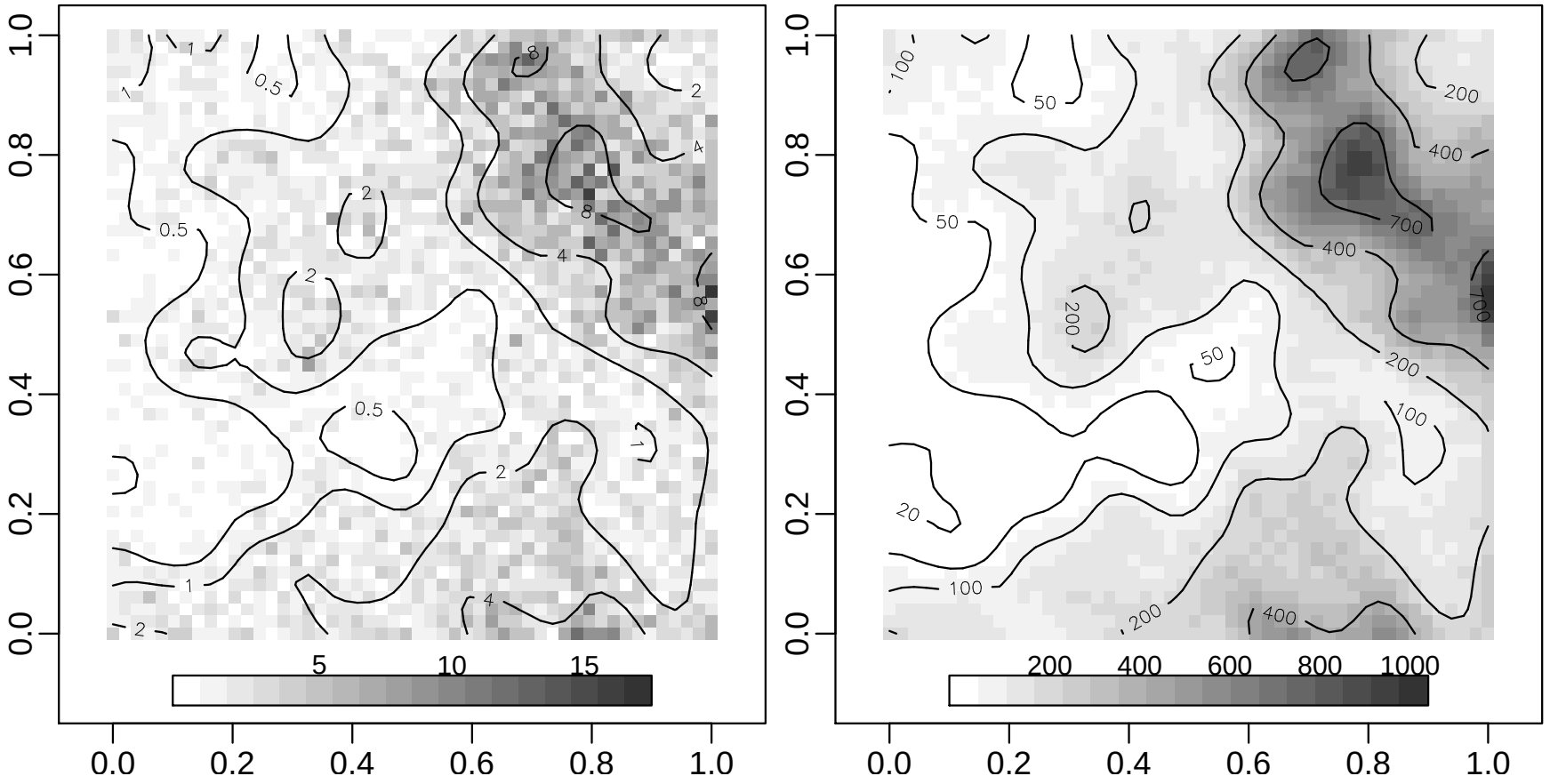
- $[Y(x_i) \mid S(x_i)]$ is Poisson with density

$$f(y_i; \zeta_i) = \exp(-\zeta_i)\zeta_i^{y_i}/y_i! \quad y_i = 0, 1, 2, \ldots$$

- link: $E[Y(x_i) \mid S(x_i)] = \zeta_i = h(\mu_i) = h(\mu + S(x_i))$
- log-link $h(\cdot) = \exp(\cdot)$
- more generaly the models can be expanded allowing for covariates and/or uncorrelated random effects

$$h(\mu_i) = \sum_{j=1}^{p} d_{ij}\beta_j + S(x_i) + Z_i$$

which, differently from Gaussian models, distinguish between the terms of the nugget effect: Poisson variation accounts for the anologue of measurement error and spatially uncorrelated component to the short scale variation



Simulations from the Poisson model; grey-scale shading represents the data values on a regular grid of sampling locations and contours represents the conditional expectation surface, with $\mu = 0.5$ on the left panel and $\mu = 5$ on the right panel.

Another example: a Binomial logistic model

- $[Y(x_i) \mid S(x_i)]$ is Binomial with density

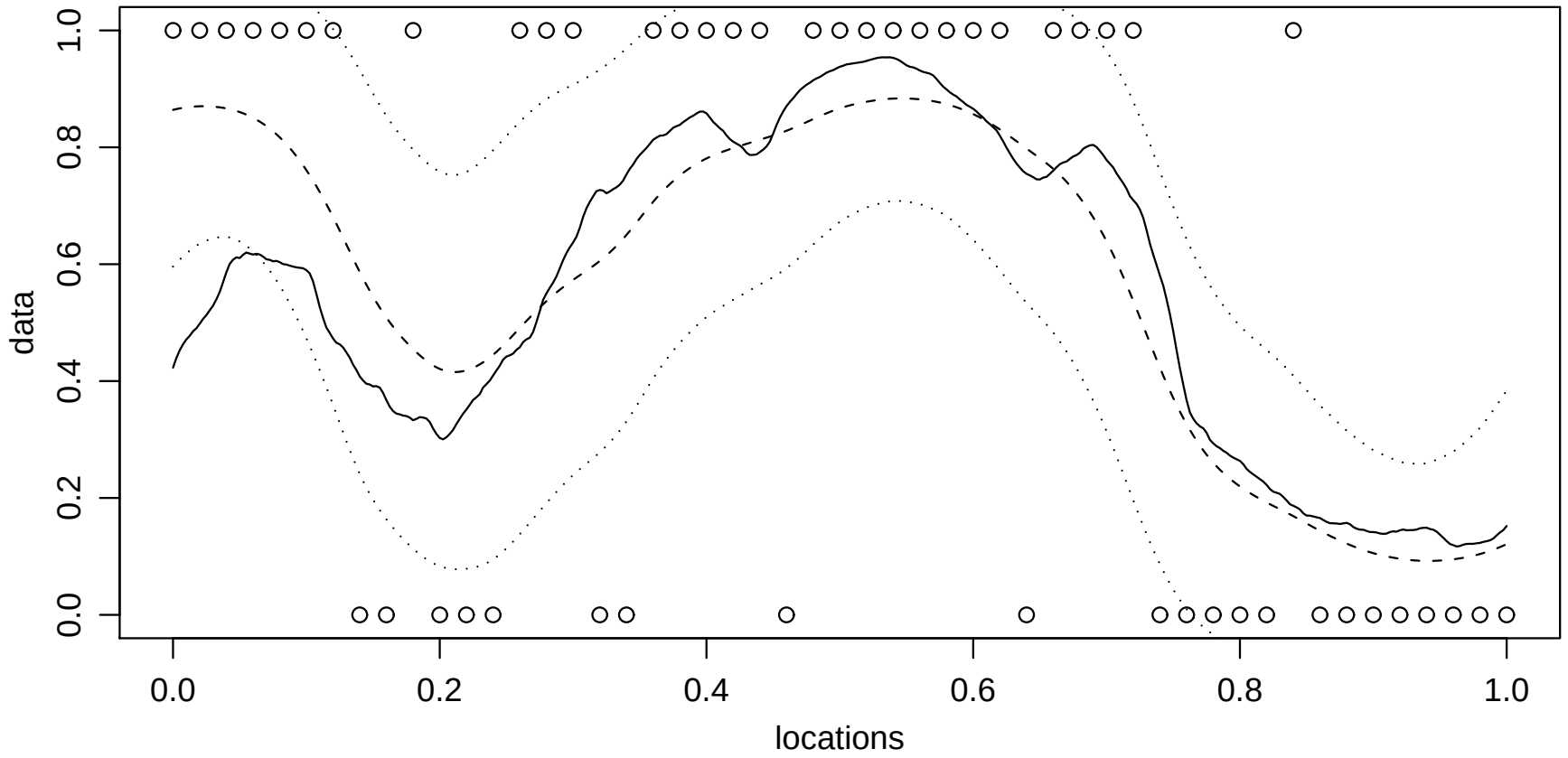
$$f(y_i;\zeta_i) = \binom{n_i}{y_i} \zeta_i^{y_i}(1-\zeta_i)^{(n_i-y_i)} \quad y_i = 0, 1, \ldots, n_i$$

- logistic link: $E[Y(x_i) \mid S(x_i)] = n_i\zeta_i = \frac{\exp\{\mu_i\}}{1+\exp\{\mu_i\}}$
- mean: $\mu_i = \mu + S(x_i)$
- again can be expanded as

$$h(\mu_i) = \sum_{j=1}^{p} d_{ij}\beta_j + S(x_i) + Z_i$$

- typically more informative with larger values of n_i

An simulated example from binary model



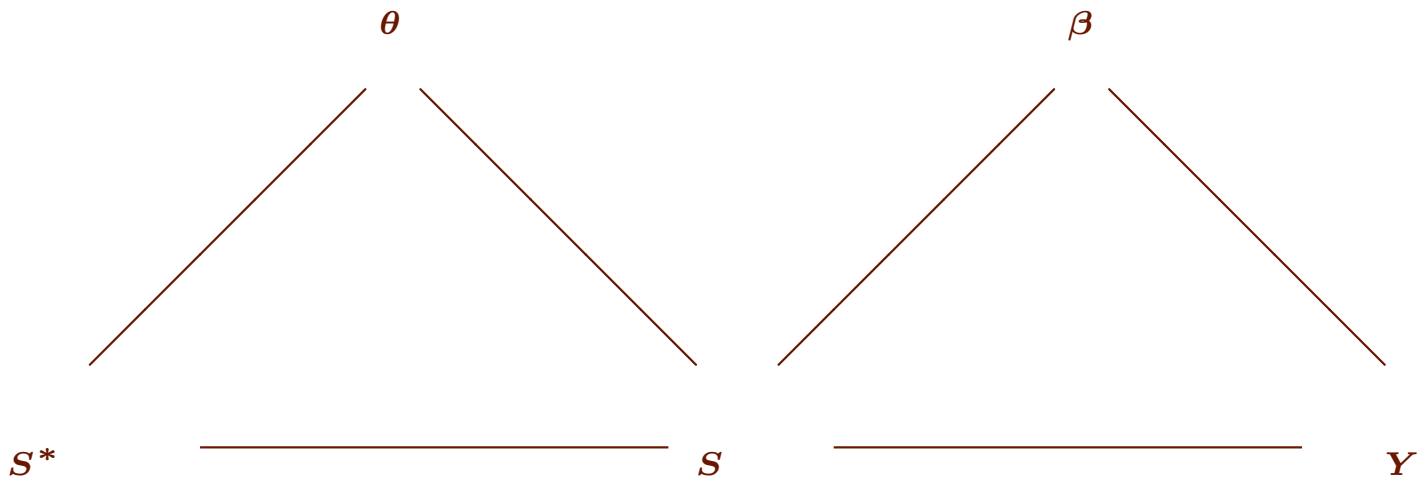
- **in this example the binary sequence is not much informative on $S(x)$**
- **wide intervals compared to the prior mean of $p(x)$**

Inference

- **Likelihood function**

$$L(\theta) = \int_{\mathbb{R}^n} \prod_i^n f(y_i; h^{-1}(s_i)) f(s \mid \theta) ds_1, \ldots, ds_n$$

- **Involves a high-dimensional (numerical) integration**
- **MCMC algorithms can exploit the conditional independence scructure of the model**



Prediction with known parameters

- Simulate $s(1), \ldots, s(m)$ from $[S|y]$ (using MCMC).
- Simulate $s^*(j)$ from $[S^*|s(j)]$, $j = 1, \ldots, m$ (multivariate Gaussian)
- Approximate $\mathrm{E}[T(S^*)|y]$ by $\frac{1}{m}\sum_{j=0}^{m} T(s^*(j))$
- if possible reduce Monte Carlo error by
  - calculating $\mathrm{E}[T(S^*)|s(j)]$ directly
  - estimate $\mathrm{E}[T(S^*)|y]$ by $\frac{1}{m}\sum_{j=0}^{m} \mathrm{E}[T(S^*)|s(j)]$

MCMC for conditional simulation

- Let $S = D'\beta + \Sigma^{1/2}\Gamma$, $\Gamma \sim N_n(0, I)$.
- Conditional density of $[\Gamma \,|Y = y]$

$$f(\gamma|y) \propto f(y|\gamma)f(\gamma)$$

  Langevin-Hastings algorithm

- Proposal: γ' from a $N_n(\xi(\gamma), hI)$ where $\xi(\gamma) = \gamma + \frac{h}{2}\nabla \log f(\gamma \mid y)$.
- E.g for the Poisson-log Spatial model: $\nabla \log f(\gamma|y) = -\gamma + (\Sigma^{1/2})'(y - \exp(s))$ where $s = \Sigma^{1/2}\gamma$.
- Expression generalises to other generalised linear spatial models.
- MCMC output $\gamma_1, \ldots, \gamma_m$. Multiply by $\Sigma^{1/2}$ and obtain: $s(1), \ldots, s(m)$ from $[S|y]$.

MCMC for Bayesian inference

Posterior:

- **Update Γ from $[\Gamma|y, \beta, \log(\sigma), \log(\phi)]$ (Langevin-Hasting described earlier)**
- **Update β from $[\beta|\Gamma, \log(\sigma), \log(\phi)]$ (RW-Metropolis)**
- **Update $\log(\sigma)$ from $[\log(\sigma)|\Gamma, \beta, \log(\phi)]$ (RW-Metropolis)**
- **Update $\log(\phi)$ from $[\log(\phi)|\Gamma, \beta, \log(\sigma)]$ (RW-Metropolis)**

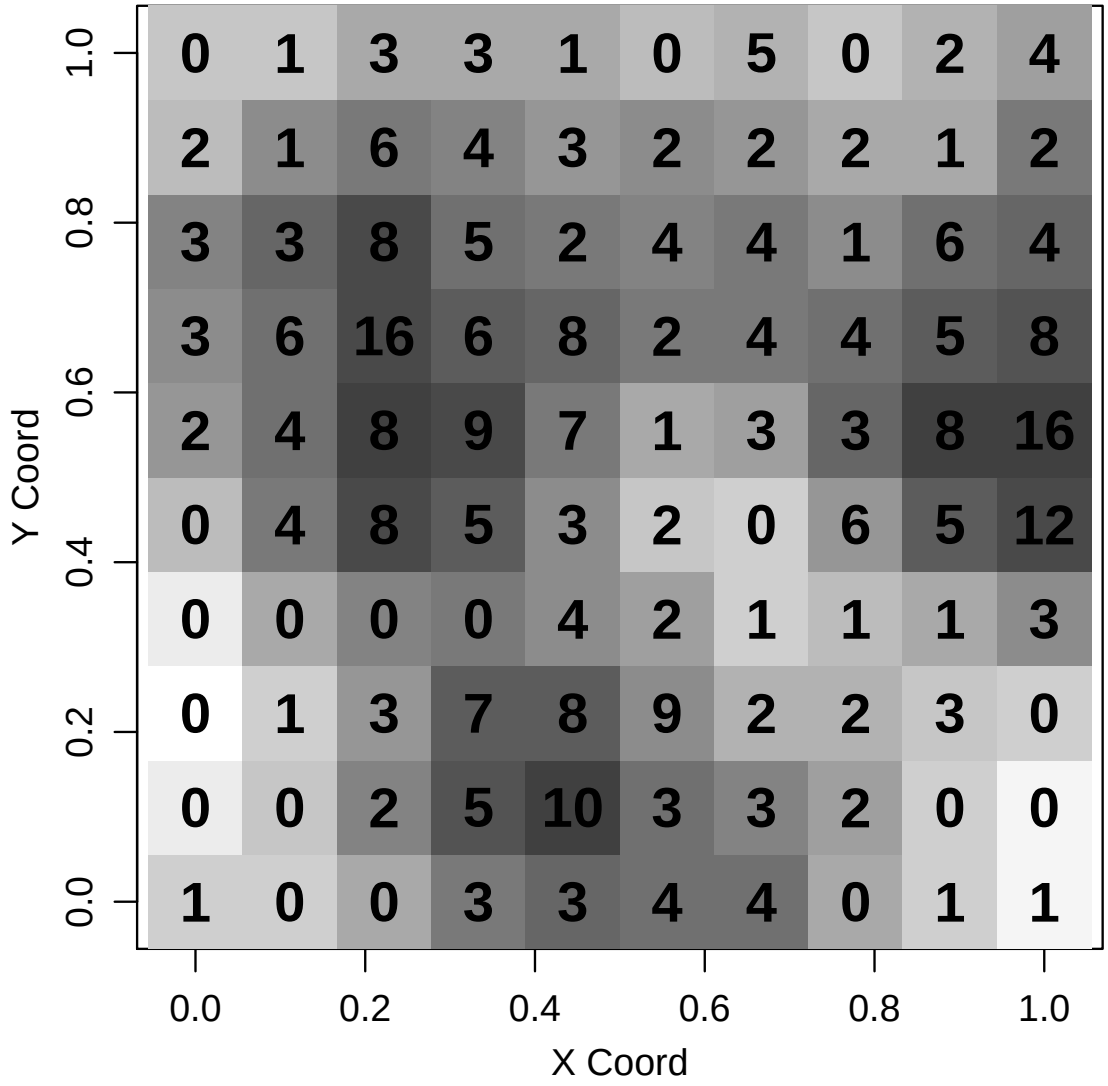
Predictive:

- **Simulate $(s(j), \beta(j), \sigma^2(j), \phi(j))$, $j = 1, \ldots, m$ (using MCMC)**
- **Simulate $s^*(j)$ from $[S^*|s(j), \beta(j), \sigma^2(j), \phi(j)]$, $j = 1, \ldots, m$ (multivariate Gaussian)**

Comments

- Marginalisation w.r.t β and σ^2 is possible using conjugate priors
- Discrete prior for ϕ is an advantage (reduced computing time).
- thinning: not to store a large sample of high-dimensional quantities.
- similar algorithms for MCMC maximum likelihood estimation

A simulated Poisson data



R code for simulation

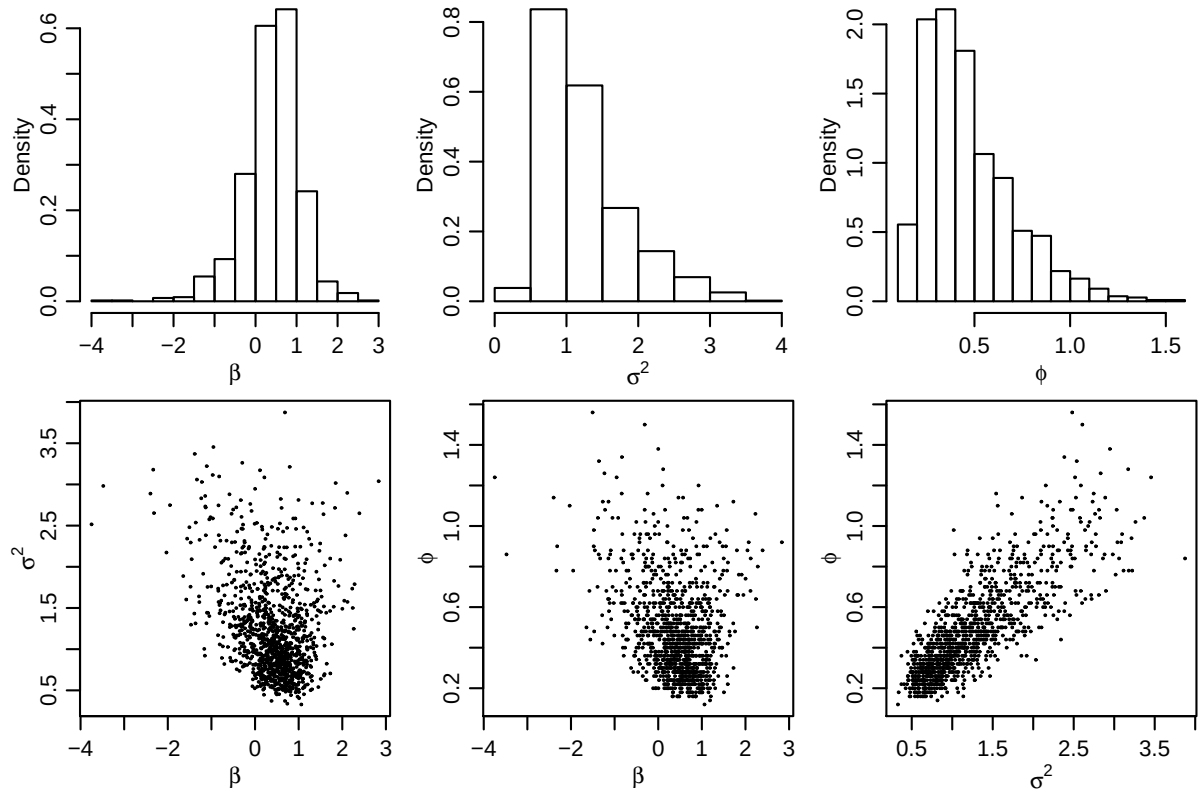
```
## setting the seed
> set.seed(371)
## defining the data locations on a grid
> cp <- expand.grid(seq(0, 1, l = 10), seq(0, 1, l = 10))
## simulating from the S process
> s <- grf(grid = cp, cov.pars = c(2, 0.2), cov.model = "mat",
  +     kappa = 1.5)
## visualising the S process
> image(s, col = gray(seq(1, 0.25, l = 21)))
## inverse link function
> lambda <- exp(0.5 + s$data)
## simulating the data
> y <- rpois(length(s$data), lambda = lambda)
## visualising the data
> text(cp[, 1], cp[, 2], y, cex = 1.5, font = 2)
```

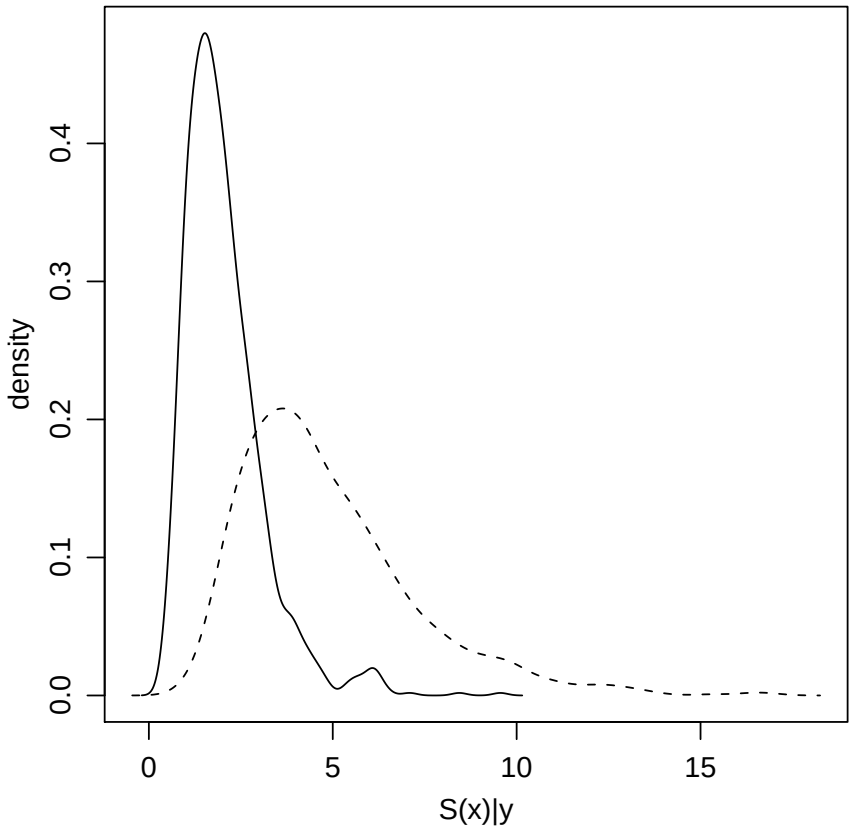
R code for the data analysis

```
set.seed(371)
## calibracao do algoritmo MCMC
MCc <- mcmc.control(S.scale=0.025, phi.sc=0.1, n.iter=110000,
                    burn.in=10000, thin=100, phi.start=0.2)
## especificacao de priors
PGC <- prior.glm.control(phi.prior="exponential", phi=0.2,
                   phi.discrete=seq(0,2,by=0.02),tausq.rel=0)
## opo de saida
OC <- output.glm.control(sim.pred=T)
## escolhendo 2 localizacoes para predicao
locs <- cbind(c(0.75, 0.15), c(0.25, 0.5))
##
pkb <- pois.krige.bayes(dt, loc=locs, prior=PGC, mcmc=MCc, out=OC)
```

Summaries of the posterior for the simulated Poisson data: posterior means and 95% central quantile-based intervals.

parameters	true values	posterior mean	95% interval
β	0.5	0.4	[0.08 , 1.58]
σ^2	2.0	1.24	[0.8 , 2.76]
ϕ	0.2	0.48	[0.3 , 1.05]




density
0.0
0.1
0.2
0.3
0.4
0
5
10
15
S(x)|y

Rongelap Island

— see other set of slides —

The Gambia malaria

— see other set of slides —

Covariance functions and variograms

- **In non-Gaussian settings, the variogram is a less natural summary statistic but can still be useful as a diagnostic tool**
- **for GLGM the model with constant mean:**

$$\mathrm{E}\left[Y(x_i)|S(x_i)\right] = \mu_i = g(\alpha + S_i) \quad v_i = v(\mu_i)$$

$$\begin{aligned}
\gamma_Y(u) &= \mathbf{E}[\frac{1}{2}(Y_i - Y_j)^2] \\
&= \frac{1}{2}\mathbf{E}_S[\mathbf{E}_Y[(Y_i - Y_j)^2|S(\cdot)]] \\
&= \frac{1}{2}\left(\mathbf{E}_S[\{g(\alpha + S_i) - g(\alpha + S_j)\}^2] + 2\mathbf{E}_S[v(g(\alpha + S_i))\right. \\
&\approx g'(\alpha)^2\gamma_S(u) + \bar{\tau}^2
\end{aligned}$$

- the variogram on the Y-scale is approximately proportional to the variogram of $S(\cdot)$ plus an intercept
- the intercept represents an average nugget effect induced by the variance of the error distribution of the model
- however it relies on a linear approximation to the inverse link function
- it may be inadequate for diagnostic analysis since the essence of the generalized linear model family is its explicit incorporation of a non-linear relationship between Y and $S(x)$.
- The exact variogram depends on higher moments of $S(\cdot)$
- explicit results are available only in special cases.

Spatial survival analysis

- specified through hazard function $h(t) = f(t)/\{1 - F(t)\}$,
- $h(t)\delta t$ is the conditional probability event will occour in the interval $(t, t + \delta t)$, given it has not occour until time t
- *proportional hazards model* with $\lambda_0(t)$, an unspecified baseline hazard function

$$h_i(t) = \lambda_0(t) \exp(d_i'\beta)$$

- $h_i(t)/h_j(t)$ does not change over time
- alternativelly, fully specified models are proposed
- *frailty* corresponds to *random effects* can be introduced by

$$h_i(t) = \lambda_0(t) \exp(z_i'\beta + U_i) = \lambda_0(t) W_i \exp(d_i'\beta)$$

- **e.g.** *log-Gaussian frailty model* **and** *gamma frailty model*
- **replacing U_i by $S(x_i)$ introduces *spatial frailties* (Li & Ryan, 2002; Banerjee, Wall & Carlin, 2003)**
- **$\mathrm{E}[S(x)] = -0.5\,\mathrm{Var}[S(x)]$ preserves interpretation of $\exp\{S(x)\}$ as a frailty process**
- **other possible approaches, e.g. Henderson, Shimakura and Gorst (2002) extends the gamma-frailty model**

PART 3

Geostatistical design

Geostatistical models for point process

- **Two possible connections between point process and geostatistics:**
  1. **measurement process replaced by a point process**
  2. **choice of data locations for $Y(x_i)$**

Cox point processes

Definition:

A Cox process is a point process in which there is an unobserved, non-negative-valued stochastic process $S = \{S(x) : x \in \mathbb{R}^2\}$ such that, conditional on S, the observed point process is an inhomogeneous Poisson process with spatially varying intensity $S(x)$.

- **fits into the general geostatistical framework**
- **derived as limiting form of a geostatistical model as $\delta \to 0$ for locations on lattice-spacing δ**
- ***log-Gaussian Cox process* is a tractable form of Cox process (e.g. Möller, Syversveen and Waagepetersen,1998; Brix & Diggle, 2001)**
- **inference generally requires computationally intensive Monte Carlo methods, implementation involves careful tuning**
- **moment-based method provides an analogue of the variogram, for exploratory analysis and *preliminary* estimation of model parameters**

Cox point processes

- intensity surface $\Lambda(x) = \exp\{S(x)\}$
- has mean and variance $\lambda = \exp\{\mu + 0.5\gamma(0)\}$
- also represents the expected number of points per unit area in the Cox process, and $\phi(u) = \exp\{\gamma(u)\} - 1$.
- $K(s)$: reduced second moment measure of a stationary point process
- $\lambda K(s)$: expected number of further points within distance s of an arbitrary point of the process
- For the log-Gaussian Cox process

$$K(s) = \pi s^2 + 2\pi\lambda^{-2} \int_0^s \phi(u) u du u$$

- **A non-parametric estimator:**

$$\hat{K}(s) = \frac{|A|}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} w_{ij}^{-1} I(u_{ij} \leq s)$$

- w_{ij} **allows for** *edge correction*

- ***preliminary estimates*** **of model parameters can then be obtained by minimising a measure of the discrepancy between theoretical and empirical** K**-functions**

Geostatistics and marked point processes

locations X signal S measurements Y

- Usually write geostatistical model as

$$[S, Y] = [S][Y|S]$$

- What if X is stochastic? Usual implicit assumption is

$$[X, S, Y] = [X][S][Y|S],$$

  hence can ignore $[X]$ for inference about $[S, Y]$.

- Resulting likelihood:

$$L(\theta) = \int [S][Y|S]dS$$

Marked point processes

locations X marks Y

- X is a point process
- Y need only be defined at points of X
- natural factorisation of $[X, Y]$?

Example 1. Spatial distribution of disease

$$X: \quad \text{population at risk}$$
$$Y: \quad \text{case or non-case}$$

- Natural factorisation is $[X, Y] = [X][Y|X]$
- Usual scientific focus is $[Y|X]$
- Hence, can ignore $[X]$

Example 2. Growth of natural forests

$$X: \quad \text{location of tree}$$
$$Y: \quad \text{size of tree}$$

- Two candidate models:
  - competitive interactions $\Rightarrow [X, Y] = [X][Y|X]$
  - environmental heterogeneity $\Rightarrow [X, Y] = [Y][X|Y]$?
- focus of scientific interest?

Preferential sampling

locations X signal S measurements Y

- **Conventional model:**

$$[X, S, Y] = [S][X][Y|S] \quad (1)$$

- **Preferential sampling model:**

$$[X, S, Y] = [S][X|S][Y|S, X] \quad (2)$$

- **Key point for inference:** even if $[Y|S, X]$ in (2) and $[Y|S]$ in (1) are algebraically the same, the term $[X|S]$ in (1) cannot be ignored for inference about $[S, Y]$, because of the shared dependence on the unobserved process S

A model for preferential sampling

$$[X, S, Y] = [S][X|S][Y|S, X]$$

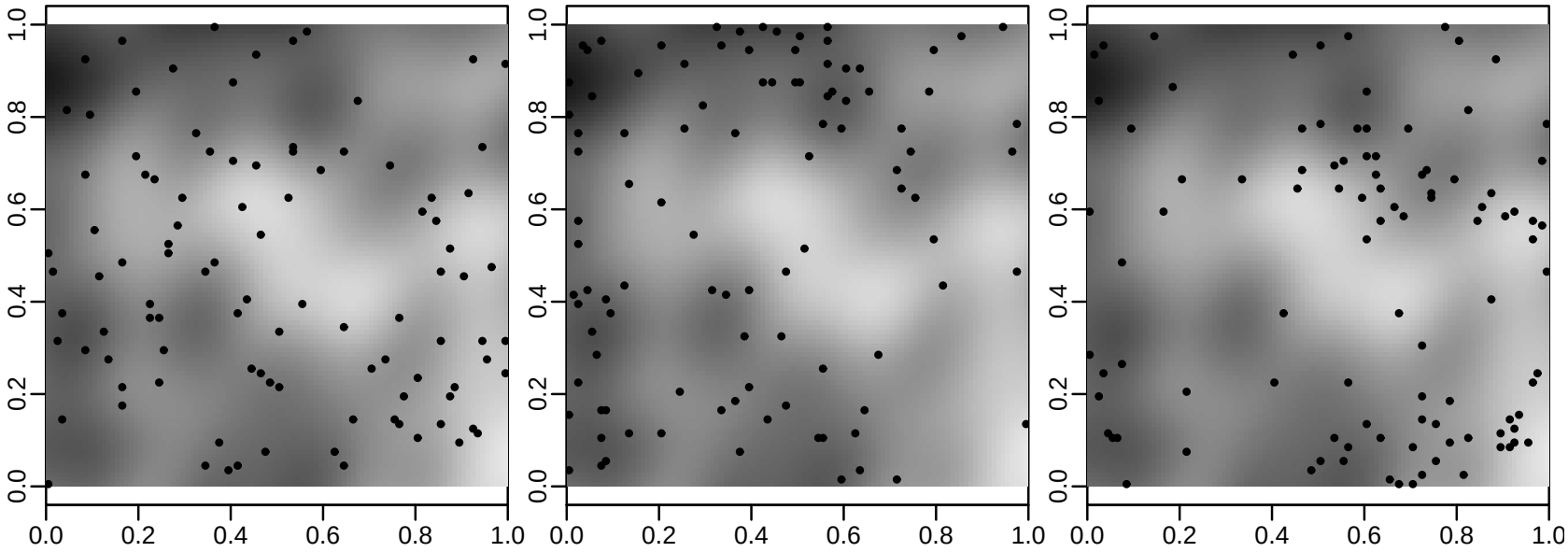
- $[S] = \mathrm{SGP}(0, \sigma^2, \rho)$ (stationary Gaussian process)
- $[X|S] =$ inhomogenous Poisson process with intensity

$$\lambda(x) = \exp\{\alpha + \beta S(x)\}$$

- $[Y|S, X] = \mathrm{N}\{\mu + S(x), \tau^2\}$ (independent Gaussian)

Diggle, Menezes & Su (2009)

Simulation of preferential sampling model



Locations (dots) and underlying signal process (grey-scale):

- **left-hand panel: uniform non-preferential**
- **centre-panel: clustered preferential**
- **right-hand panel: clustered non-preferential**

Likelihood inference

$$[X, S, Y] = [Y|S, X][X|S][S]$$

data are X and Y, hence likelihood is

$$L(\theta) = \int [X, S, Y] dS = \mathbf{E}_S \left[[Y|S, X][X|S] \right]$$

$S = \{S_X, S_{-X}\}$: $[S|Y] = [S_X|Y][X_{-X}|S_X]$; $[Y|X, S] = [Y|S_X]$

$$L(\theta) = \int [X|S] \frac{[Y|S_X]}{[S_X|Y]} [S_X][Y|S] dS = \mathbf{E}_{S|Y} \left[[X|S] \frac{[Y|S_X]}{[S_X|Y]} [S_X] \right]$$

evaluate expectation by Monte Carlo (with S on a lattice)

$$L_{MC}(\theta) = m^{-1} \sum_{j=1}^{m} [X|S_j] \frac{[Y|S_X]}{[S_X|Y]} [S_X]$$

requires (efficient) conditional simulations $[S|Y]$

Geostatistical design

- **Retrospective: add to, or delete from, an existing set of measurement locations $x_i \in A : i = 1, \ldots, n$.**

- **Prospective: choose optimal positions for a new set of measurement locations $x_i \in A : i = 1, \ldots, n$.**

For a compreensive account see Müller (2007)

Naïve design folklore

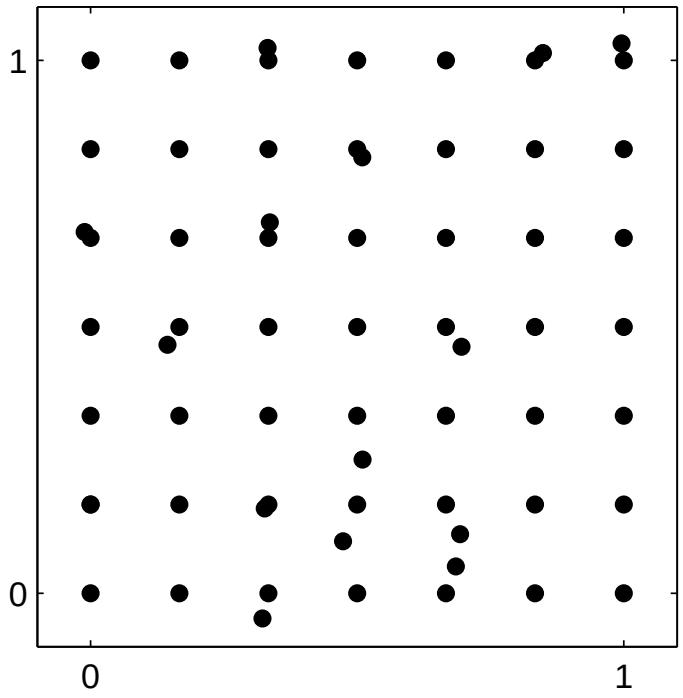
- **Spatial correlation decreases with increasing distance.**
- **Therefore, close pairs of points are wasteful.**
- **Therefore, spatially regular designs are a good thing.**

Less naïve design folklore

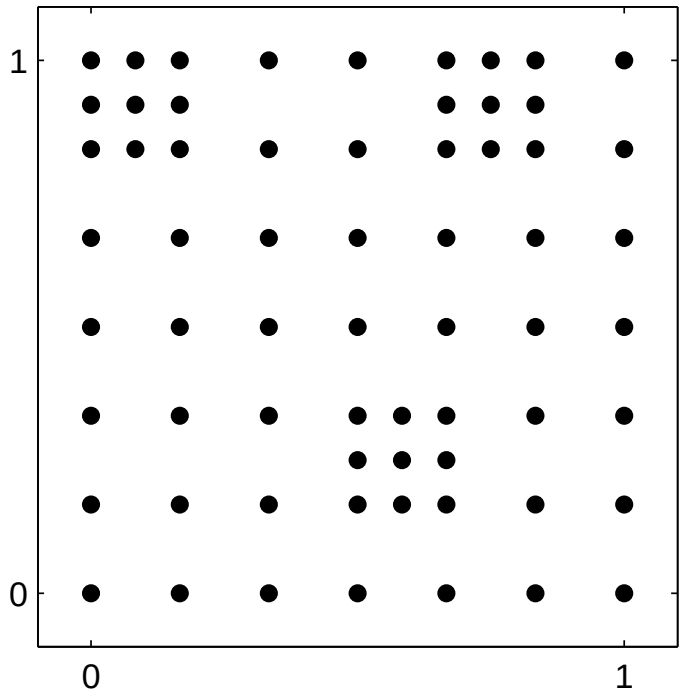
- Spatial correlation decreases with increasing distance.
- Therefore, close pairs of points are wasteful if you know the correct model.
- But in practice, at best, you need to estimate unknown model parameters.
- And to estimate model parameters, you need your design to include a wide range of inter-point distances.
- Therefore, spatially regular designs should be tempered by the inclusion of some close pairs of points.

Examples of compromise designs

A) Lattice plus close pairs design



B) Lattice plus in−fill design



Designs for parameter estimation

Comparison of random and square lattice designs, each with $n = 100$ *sample locations, with respect to three design criteria: spatial maximum of mean square prediction error* $M(x)$*; spatial average of mean square prediction error* $M(x)$*; scaled mean square error,* $100 \times MSE(T)$*, for* $T = \int S(x)dx$*. The simulation model is a stationary Gaussian process with parameters* $\mu = 0$*,* $\sigma^2 + \tau^2 = 1$*, correlation function* $\rho(u) = \exp(-u/\phi)$ *and nugget variance* τ^2*. The tabulated figures are averages of each design criterion over* $N = 500$ *replicate simulations.*

Model parameters		max $M(x)$		average $M(x)$		$MSE(T)$	
		Random	Lattice	Random	Lattice	Random	Lattice
$\tau^2 = 0$	$\phi = 0.05$	9.28	8.20	0.77	0.71	0.53	0.40
	$\phi = 0.15$	5.41	3.61	0.40	0.30	0.49	0.18
	$\phi = 0.25$	3.67	2.17	0.26	0.19	0.34	0.10
$\tau^2 = 0.1$	$\phi = 0.05$	9.57	8.53	0.81	0.76	0.54	0.41
	$\phi = 0.15$	6.22	4.59	0.50	0.41	0.56	0.28
	$\phi = 0.25$	4.44	3.34	0.37	0.30	0.47	0.22
$\tau^2 = 0.3$	$\phi = 0.05$	10.10	9.62	0.88	0.86	0.51	0.40
	$\phi = 0.15$	7.45	6.63	0.65	0.60	0.68	0.43
	$\phi = 0.25$	6.23	5.70	0.55	0.51	0.58	0.38

A Bayesian design criterion

Assume goal is prediction of $S(x)$ for all $x \in A$.

$$[S|Y] = \int [S|Y,\theta][\theta|Y]d\theta$$

For retrospective design, minimise

$$\bar{v} = \int_A \mathrm{Var}\{S(x)|Y\}dx$$

For prospective design, minimise

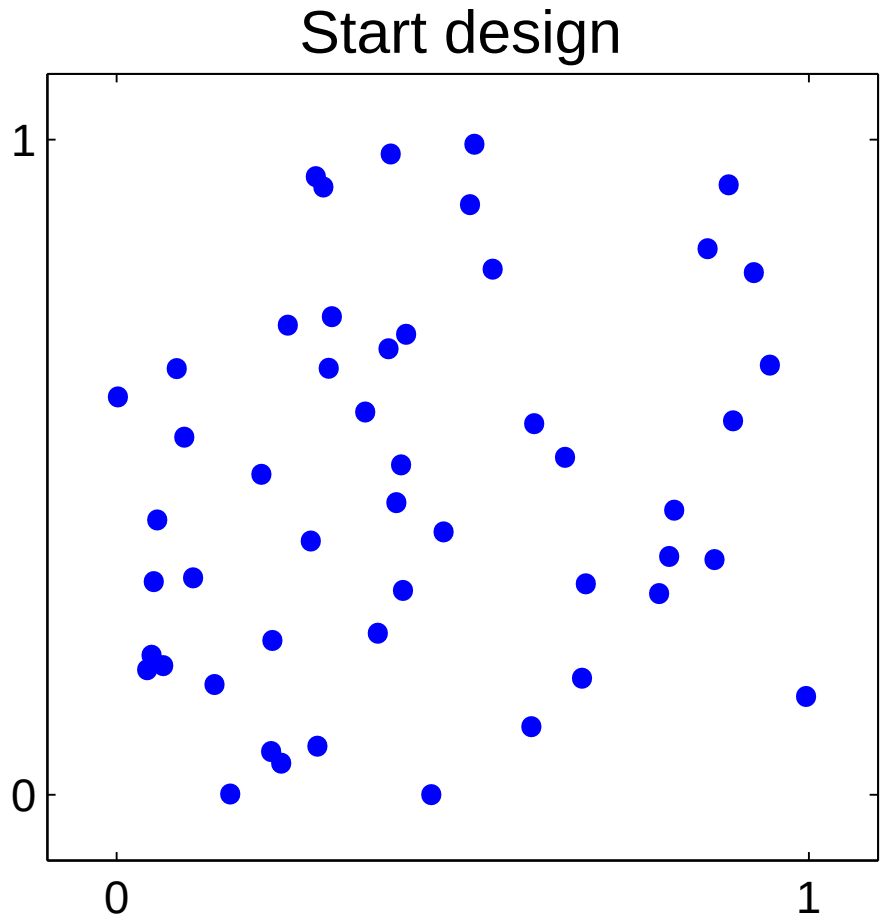
$$\mathrm{E}(\bar{v}) = \int_y \int_A \mathrm{Var}\{S(x)|y\}f(y)dy$$

where $f(y)$ corresponds to

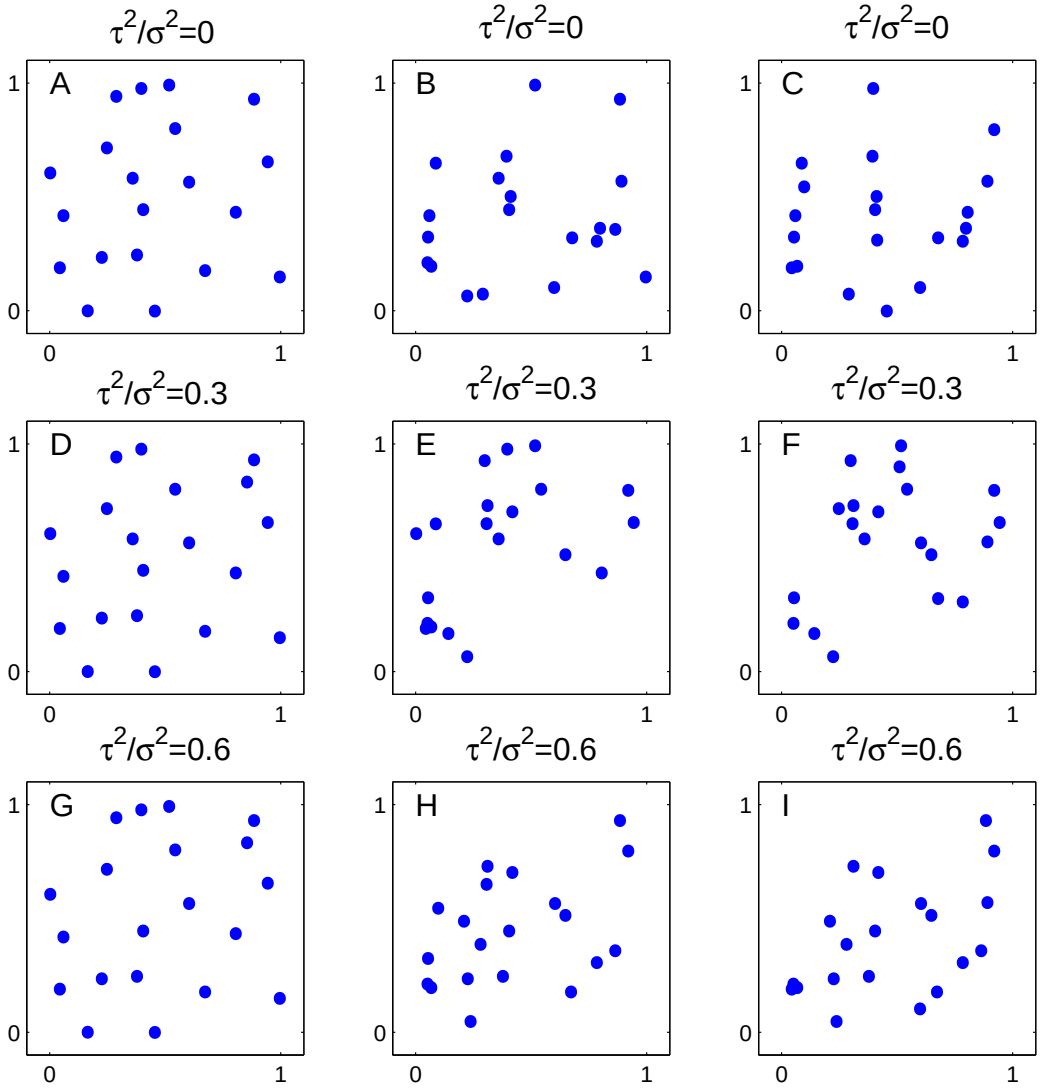
$$[Y] = \int [Y|\theta][\theta]d\theta$$

Results

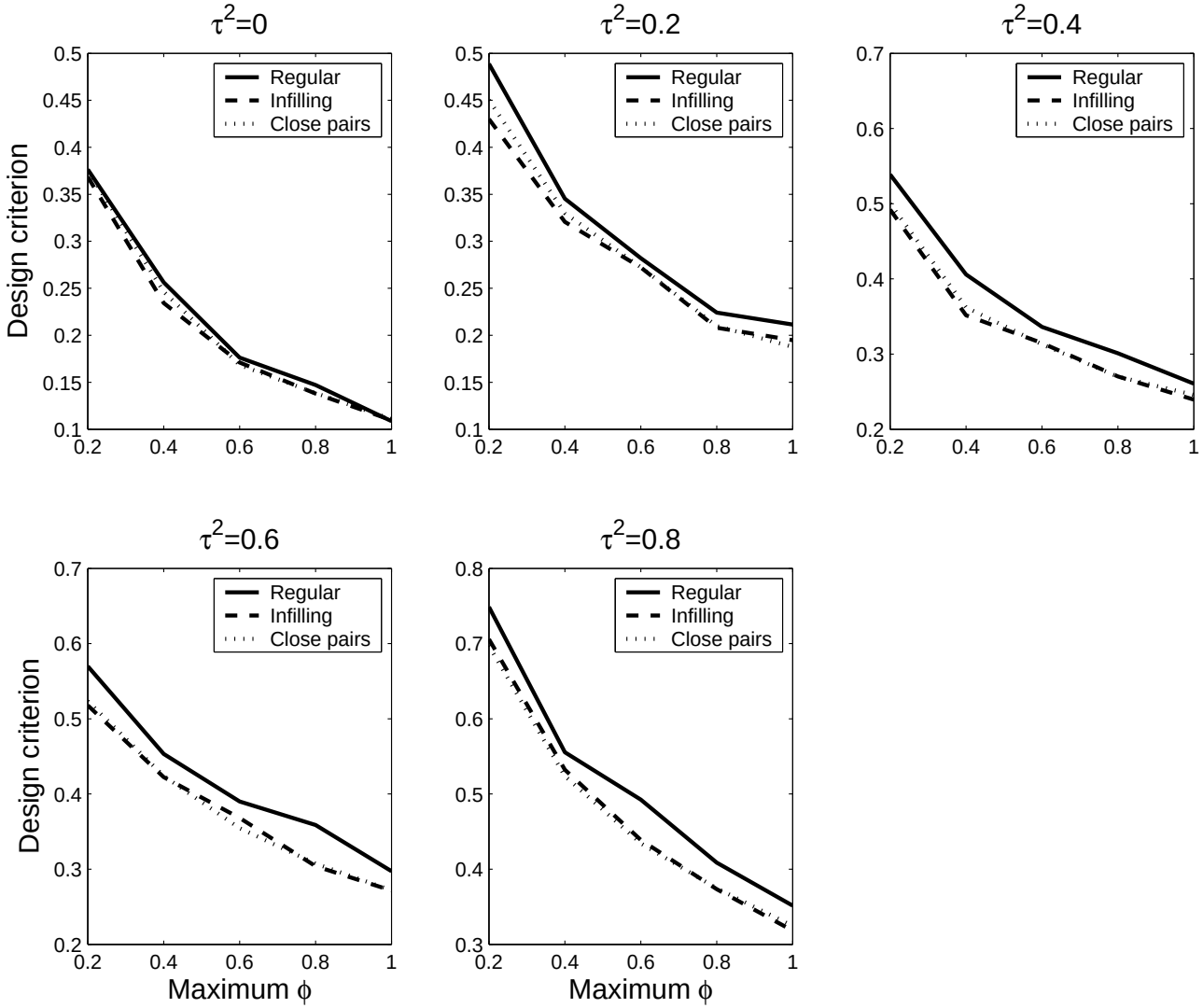
Retrospective: deletion of points from a monitoring network



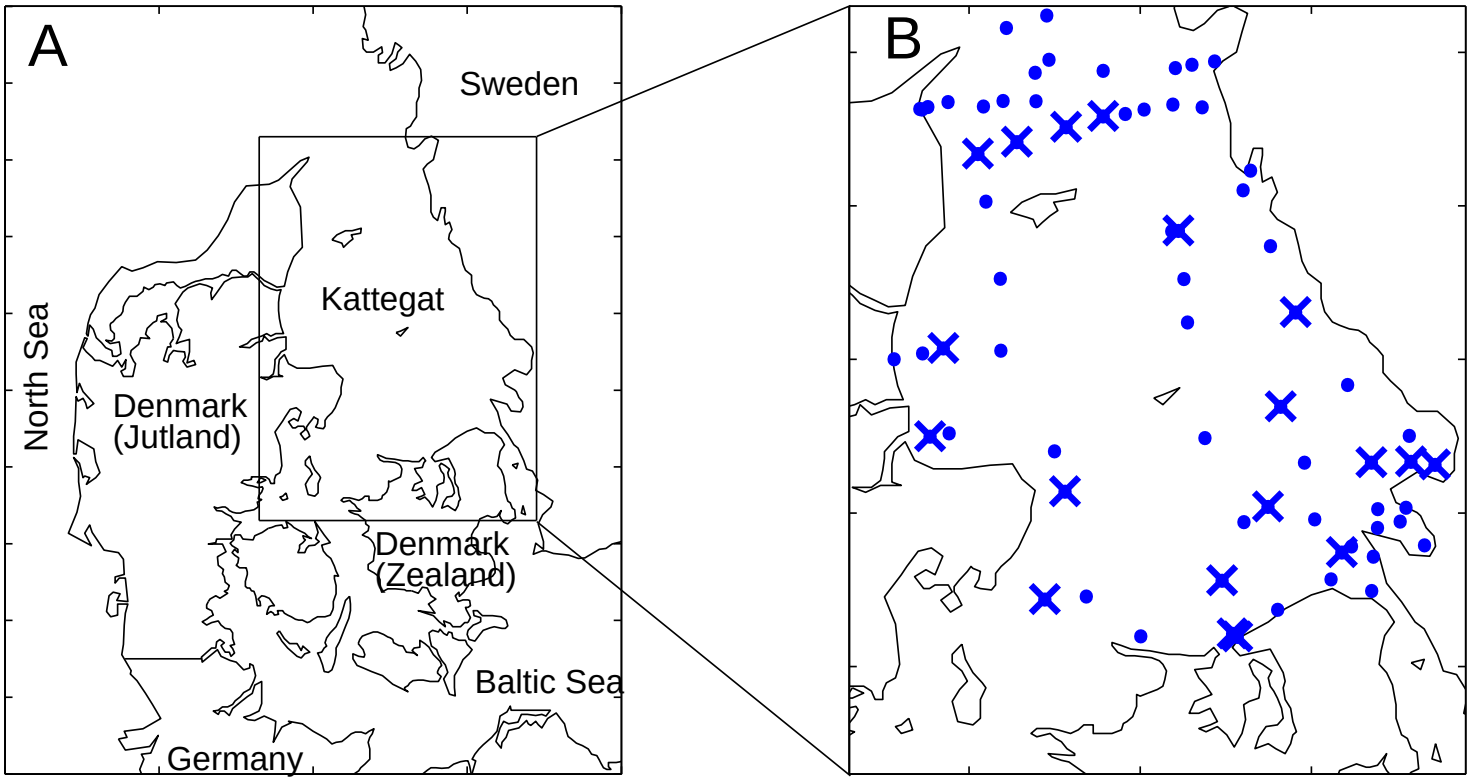
Selected final designs



Prospective: regular lattice vs compromise designs



Monitoring salinity in the Kattegat basin



Solid dots are locations deleted for reduced design.
Diggle and Lophaven (2004)

Further examples

1. **modelling fish stocks in the Portuguese coast**

2. **modelling geostatistical compositional data**

Further remarks on geostatistical design

1. Conceptually more complex problems include:
   (a) design when some sub-areas are more interesting than others;
   (b) design for best prediction of non-linear functionals of $S(\cdot)$;
   (c) multi-stage designs
   (d) spatio-temporal designs
2. Theoretically optimal designs may not be realistic
3. Goal here is NOT optimal design, but to suggest constructions for good, general-purpose designs.

Closing remarks

- **Geostatistical problems can be treated under statistical modelling approach**
- **Parameter uncertainty can have a material impact on prediction**
- **Bayesian paradigm deals naturally with parameter uncertainty**
- **Implementation through MCMC is not wholly satisfactory:**
  - **sensitivity to priors?**
  - **convergence of algorithms?**
  - **routine implementation on large data-sets?**

- **Model-based approach clarifies distinctions between:**
  - **the substantive problem;**
  - **formulation of an appropriate model;**
  - **inference within the chosen model;**
  - **diagnostic checking and re-formulation.**
- **Analyse problems, not data:**
  - **what is the scientific question?**
  - **what data will best allow us to answer the question?**
  - **what is a reasonable model to impose on the data?**
  - **inference: avoid *ad hoc* methods if possible**
  - **fit, reflect, re-formulate as necessary**
  - **answer the question.**

Some computational resources

- **R-project**:
  http://www.R-project.org
- **CRAN spatial task view**:
  http://cran.r-project.org/src/contrib/Views/Spatial.html
- **AI-Geostats web-site**:
  http://www.ai-geostats.org

Mostly used for this notes:

- **geoR package**:
  http://www.leg.ufpr.br/geoR
- **geoRglm package**:
  http://www.leg.ufpr.br/geoRglm

- **RandomFields package:**
  http://www.r-project.org