CE-056:
Fundamentos de Geoestatística

Paulo Justiniano Ribeiro Junior *

Departamento de Estatística
Universidade Federal do Paraná

Segundo semestre de 2004

Endereço para correspondência: Departamento de Estatística, Universidade Federal do Paraná, E-mail: Paulo.Ribeiro@est.ufpr.br

PARTE I:
INTRODUÇÃO

1. **Exemplos Básicos de dados espaciais**

2. **Terminologia para estatística espacial**

3. **Outros exemplos de dados geoestatísticos**

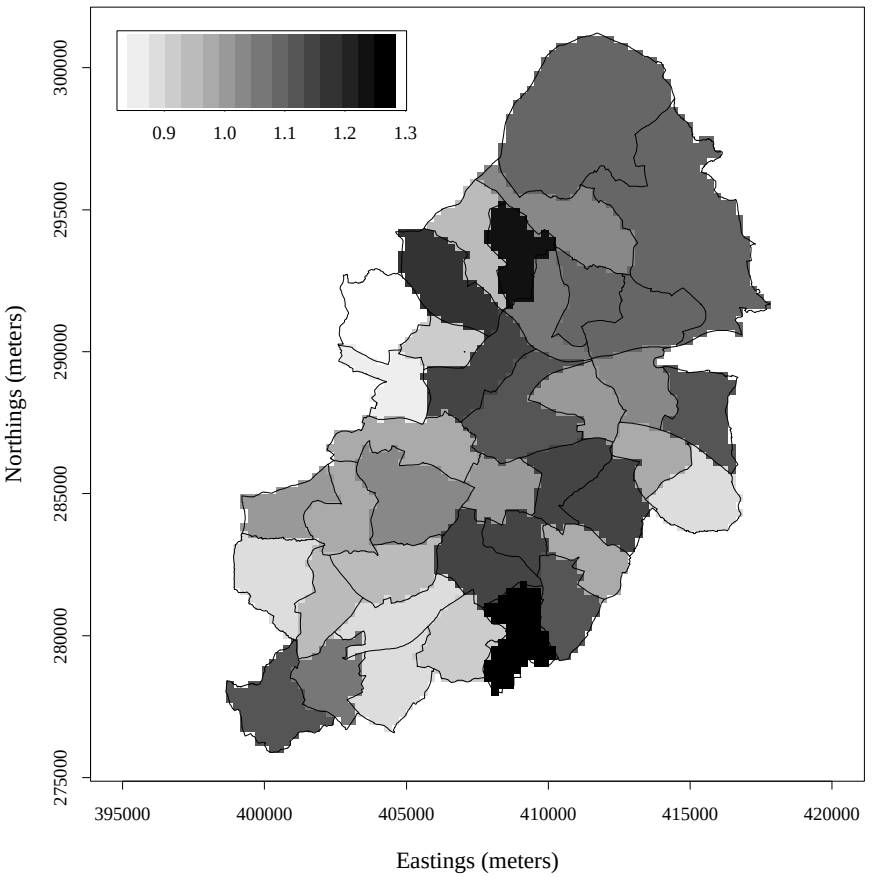
4. **Características de Problemas Geoestatísticos**

5. **Questões Centrais em Geoestatística**

1. Estatística Espacial: Exemplos Básicos

(a) **Taxas de câncer por regiões administrativas**

tons de cinza correspondem à variação estimada do risco relativo de câncer colorectal em 36 zonas eleitorais da cidade de Birmingham, UK.

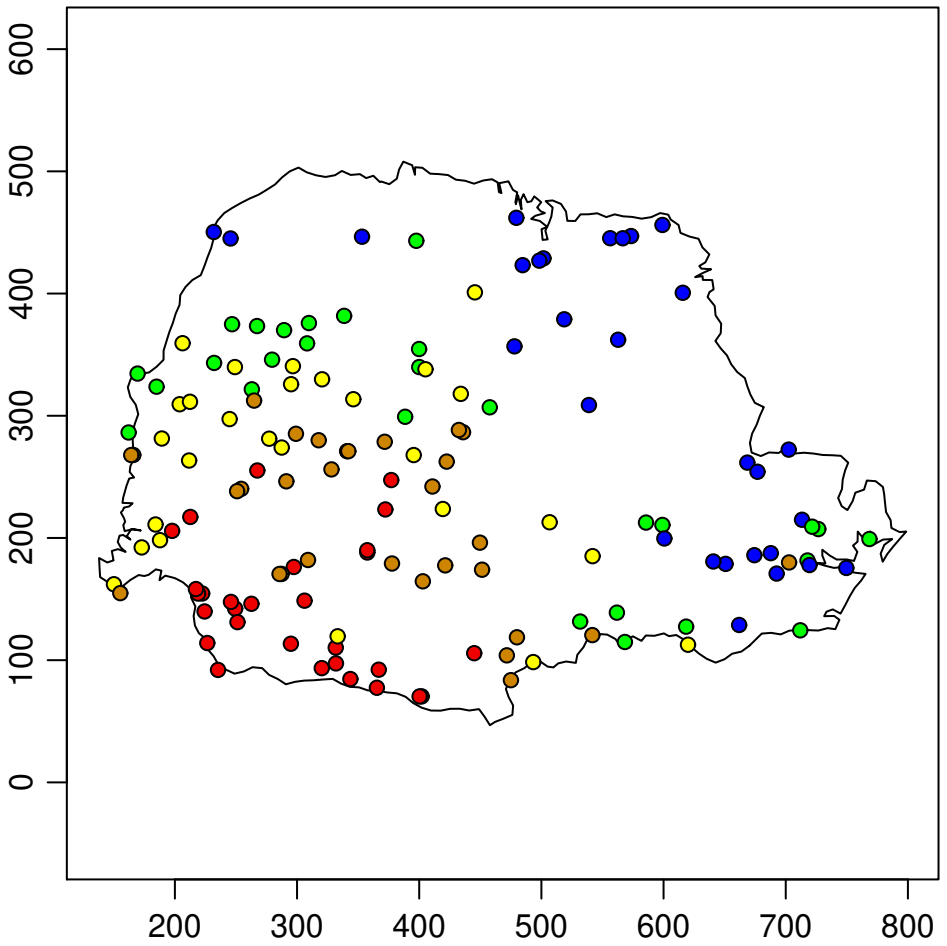


(b) **Precipitação no Estado do Paraná**

Medidas de chuva em 143 postos meteorológicos.

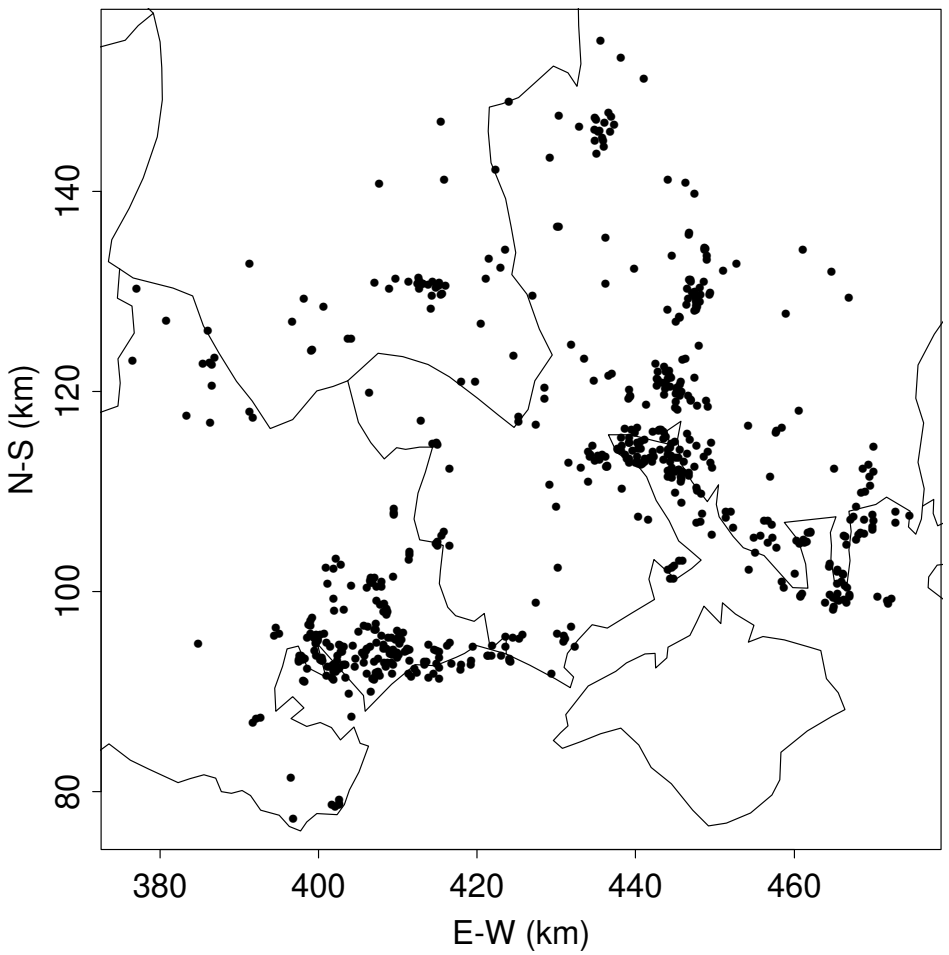
Médias históricas para o período de Maio-Junho (estação seca).

Maiores detalhes: tese de Jacinta L. Zamboti (2001).



(c) **Infecções bacterianas no sul da Inglaterra**

Localizações das residências de 651 casos notificados num per íodo de 1 ano na região central do sul da Inglaterra.



2. Terminologia para estatística espacial

(a) **Variação espacial discreta**
Estrutura básica. $Y_i : i = 1, ..., n$

- raramente ocorre naturalmente
- útil como estratégia pragmática
- modelos são tipicamente definidos indiretamente a partir de condicionais $[Y_i | Y_j, \forall j \neq i]$

(b) **Variação espacial contínua**
Estrutura básica. $Y(x) : x \in \mathbb{R}^2$

- dados $(y_i, x_i) : i = 1, ..., n$, localizações x_i podem ser:
  - não estocástica (ex. grade cobrindo a região em estudo A) ou estocástica, *porém independente do processo* $Y(x)$

(c) **Processo pontual espacial**
Estrutura básica. Conjunto contável de pontos $x_i \in \mathbb{R}^2$, generados estocásticamente.

- às vezes dados são agregados em regiões

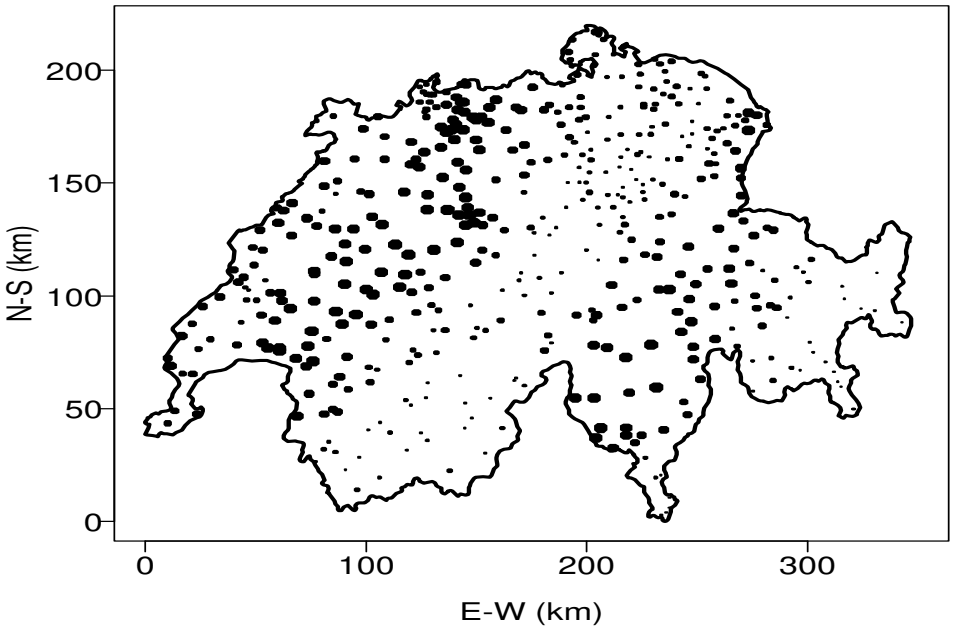
Estatística espacial é a seleção de métodos estatísticos nos quais a localização espacial tem papel explícito na análise dos dados.

Dois temas estratégicos

- não confundir *formato dos dados* com o *processo subjacente*.
- a escolha do modelo pode ser influenciada pelos objetivos científicos do estudo

3. Outros Exemplos de Problemas Geo-estatísticos

(a) Dados de chuva na Suíça

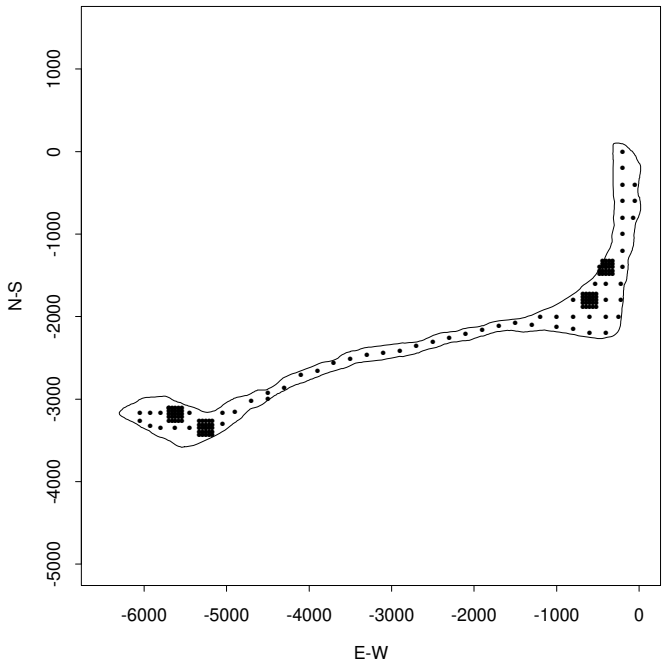


Localizações com tamanhos dos pontos proporcionais aos valores observados de precipitação

- 467 postos na Suíça
- medidas diárias de chuva em 8 de Maio de 1986
- dados do projeto:
  Spatial Interpolation Comparison 97
  `ftp://ftp.geog.uwo.ca/SIC97/`.

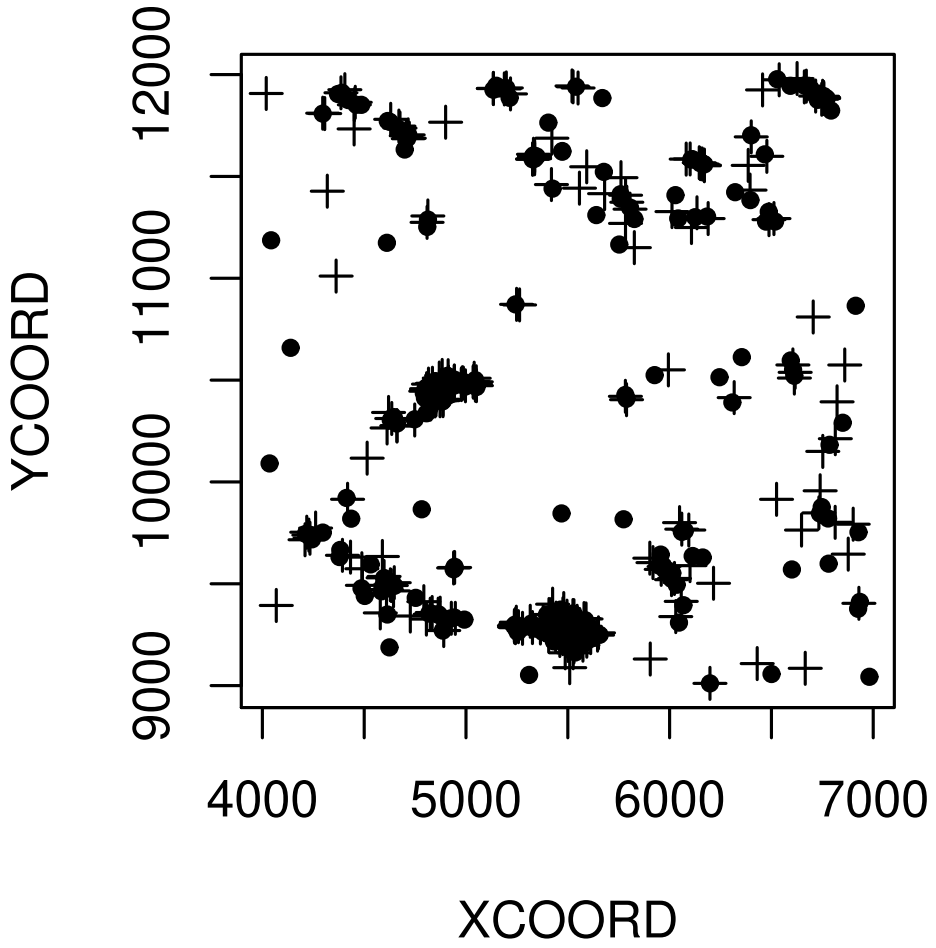
(b) **Ilha de Rongelap**

- estudo do resíduo de contaminação decorrente de testes de armas nucleares durante a década de 50
- ilha evacuada em 1985. Segura para reocupação
- pesquisa produz medidas com ruído Y_i de concentração de césio radioativo
- particular interesse em níveis máximos de concentração de césio



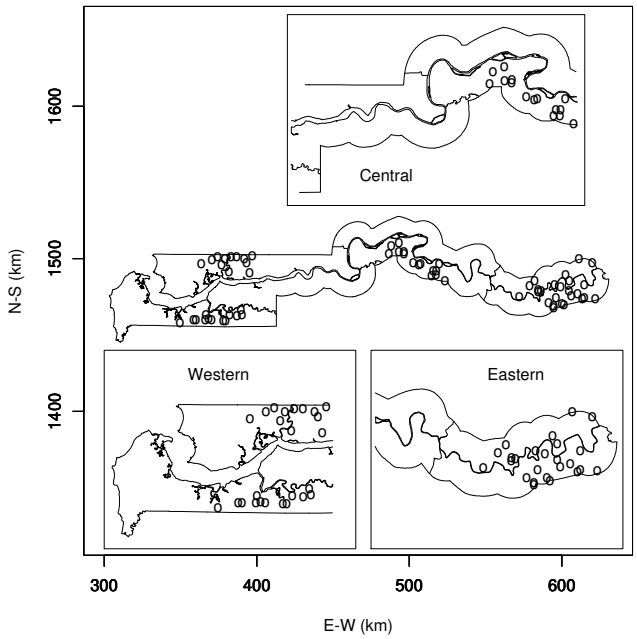
(c) **Espécies de líquens**

- fatores associados a distribuição espacial da presença de líquens em troncos de ávores
- resposta 0/1: presença ou ausência
- covariáveis: diâmetro, umidade, sombreamento, cobertura do tronco, viva



(d) **Malária em Gambia**

- na vila i, dado $Y_{ij} = 0/1$ denota ausência ou presença de malária no sangue da criança j
- covariáveis ao nível de vilas:
  - localização (coordenadas), presença de centro de saúde, índice de vegetação derivado de satélite
- covariáveis ao nível de crianças:
  - idade, uso e tratamento de mosquiteiro
- interesses: efeito das covariáveis e padrão espacial da variação residual



4. Características Principais dos Problemas Geoestatísticos

- dados consistem em **respostas** Y_i associadas com **localizações** x_i

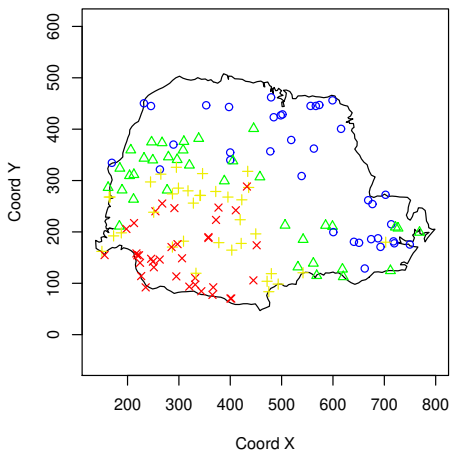
- em princípio, Y pode ser determinado em qualquer localização x dentro da região espacialmente contínua A

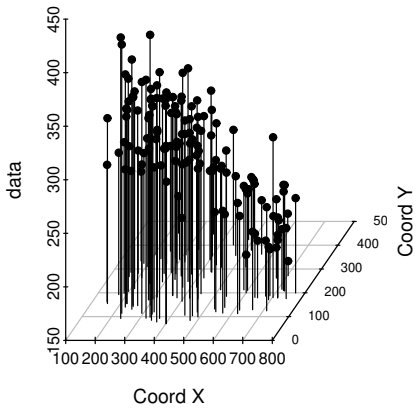
- assume-se que $\{Y(x) : x \in A\}$ é um processo estocástico

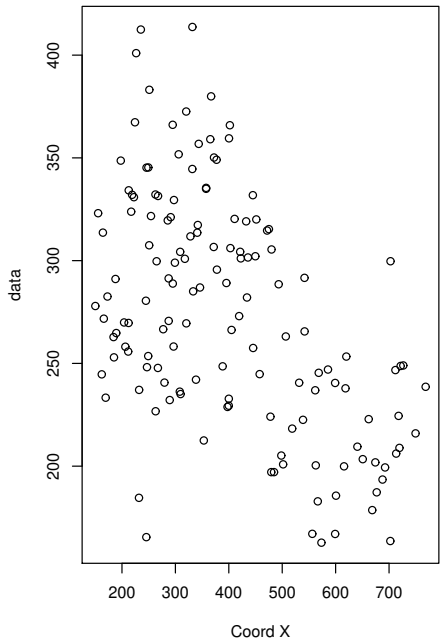
- x_i é tipicamente fixo. Se as localizações x_i são geradas por um processo estocástico pontual, assume-se que este processo é independente de $Y(x)$

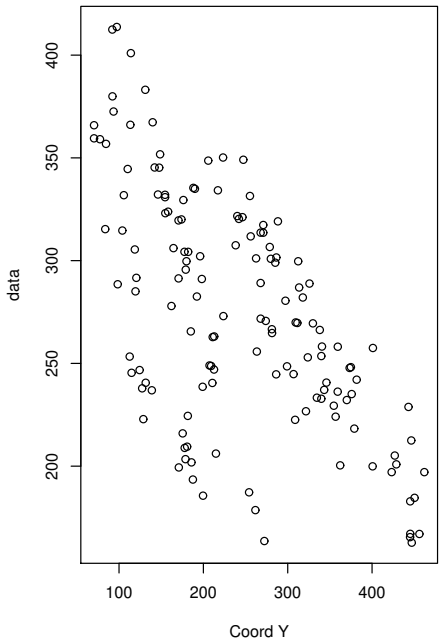
- objetivos científicos incluem a predição de um ou mais funcionais de processo (sem ruído) $\{S(x) : x \in A\}$

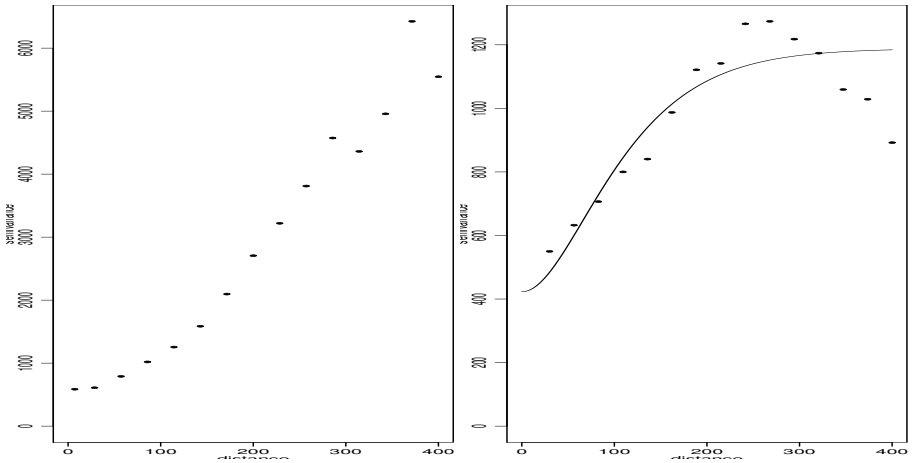
Exemplo básico: chuva no Paraná



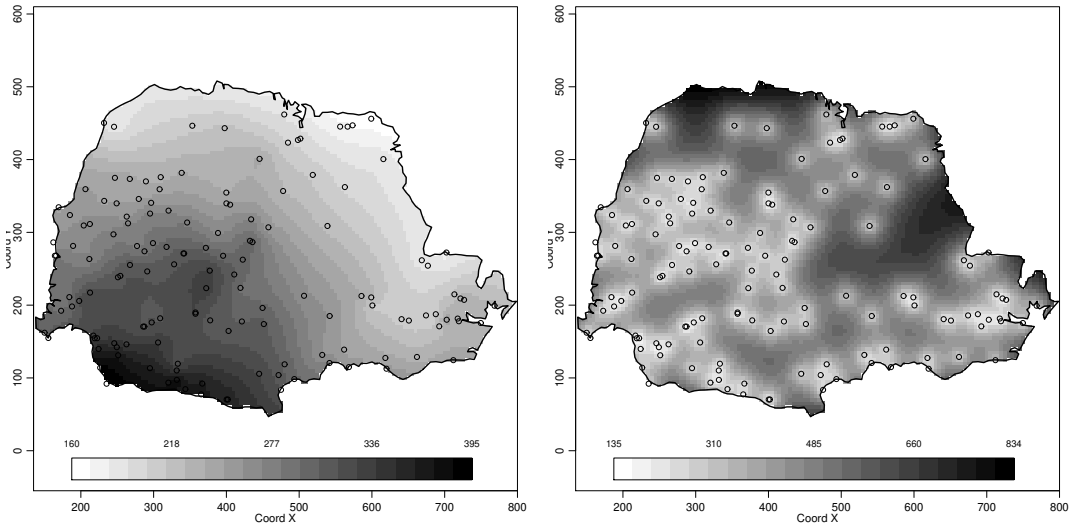








variogramas para dados originais (esquerda) e após retirada de tendência, com modelo ajustado (direita).



Krigagem: mapas de valores preditos (esquerda) e variâncias de predição (direita).

5. Questões Centrais

- **Delineamento**
  - quantas localizações?
  - quantas medidas?
  - configuração das localizações?
  - o que deve-se medir em cada localização?

- **Modelagem**
  - modelo probabilístico para o sinal $[S]$
  - modelo de probabilidade condicional para as medidas, $[Y|S]$

- **Estimação**
  - valores para parâmetros desconhecidos do modelo
  - inferências sobre os parâmetros ou funções destes

- **Predição**
  - avalia-se $[T|Y]$, a distribuição condicional aos dados do objetivo de predição

Geostatística Tradicional:

- evita referência explícita à especificação paramétrica dos modelos

- variogramas como instrumento de inferência (Matheron: "estimação e escolha")

- em geral usa-se estruturas complexas de variogramas

- concentra-se em estimadores lineares

- métodos e paradigmas específicos para:
  - predição pontual (SK, OK, KTE, UK)
  - predição de funcionais não lineares (IK, DK, ...)
  - estimação de densidades preditivas (IK, DK)
  - simulações das preditivas (SGSIM, SISIM, ...)

- *"kriging menu"*

PARTE II:

ESPECIFICAÇÃO DO MODELO GEOESTATÍSTICO

1. **“Model based geostatistics”**

2. **A Caminho de um Modelo Espacial**

3. **O Modelo Gaussiano em Detalhes**

4. **Função de Correlação**

5. **Efeitos Direcionais**

6. **Modelos Não-Estacionários**

1. “Model based geostatistics”

“Model based geostatistics means that we adopt a model-based approach to this class of problems, by which we mean that we start with an explicit stochastic model and derive associated methods of parameter estimation, interpolation and smoothing by the application of general statistical principles”.

Notação

$$(Y_i, x_i) : i = 1, ..., n$$

- $\{x_i : i = 1, ..., n\}$ é o **plano amostral**
- $\{Y(x) : x \in A\}$ é o **processo de medida**
- $\{S(x) : x \in A\}$ é o **processo do sinal**
- $T = \mathcal{F}(S)$ é o **objetivo de predição**
- $[S, Y] = [S][Y|S]$ é o **modelo geoestatístico**

2. A caminho da especificação de um modelo especial

Perspectiva histórica - paradigmas para inferência

(a) Modelos estatísticos:

- redução de dados
- escolha, estimação e predição

(b) Gauss e Legendre

- estudos de astronomia
- erros normais
- discrepância dados e modelo: min. quadrados
- 1° e 2° momentos

(c) Fisher e verossimilhança

- uso e interpretação da verossimilhança
- relação com min. quad.:
  $-2l = \frac{1}{\sigma^2}(y_i - \mu_i)^2$
- máximo, curvaturas, inferência, etc
- Royall, 1997
- pragmatismo e delineamentos

(d) Inferência: "Model-based" *vs* "design-based"

Perspectiva histórica - Modelos Lineares Generalizados

- Modelo linear

$$Y = X\beta + \varepsilon$$

- pode ser escrito como:

  $Y \sim N(\mu, \sigma^2)$
  $\mu = X\beta$

- e generalizado de 2 formas

  $Y \sim Q(\mu, ...)$
  $\eta = g(\mu) = X\beta$

- não mais requer
  - normalidade
  - variância constante
  - preocupação com escala

- verossimilança em destaque

- deviance: $D(\theta) = l(y, y) - l(y, \theta)$

- extensões
  - modelagem de superdispersão
  - modelos mixtos
  - modelos hierárquicos (multinível)
  - inferência Bayesiana

Modelo linear generalizado linear clássico

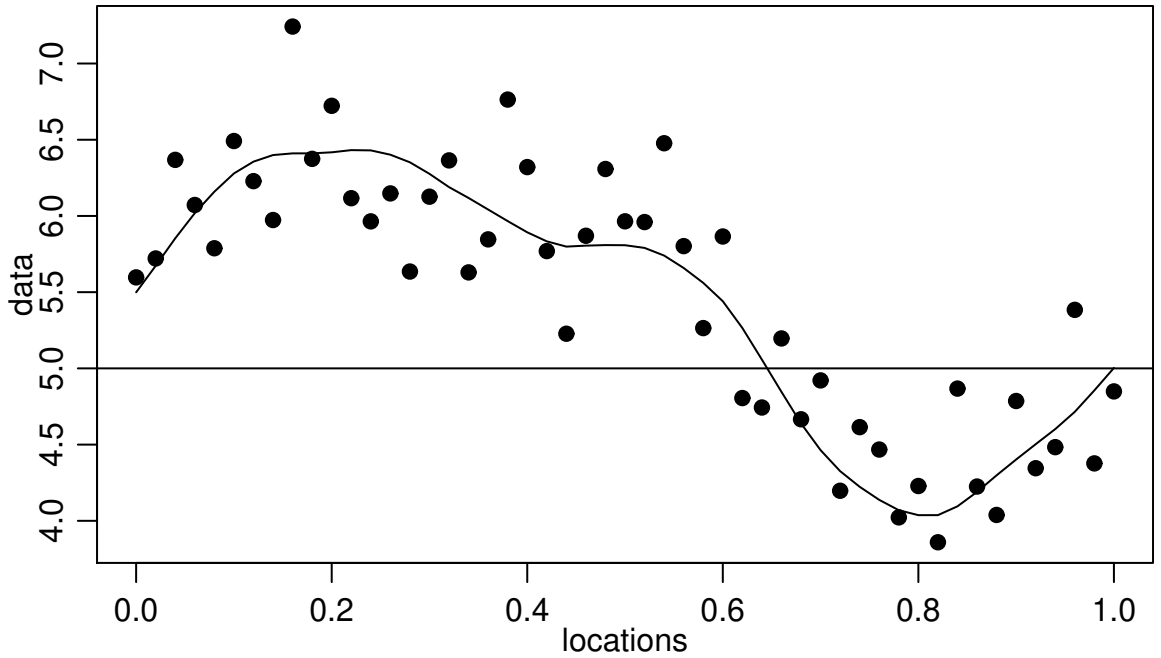
- $Y_i : i = 1, ..., n$
  mutuamente independentes, com $\mu_i = \mathrm{E}[Y_i]$
- $h(\mu_i) = \sum_{j=1}^{k} f_{ij}\beta_j$, com função de ligação conhecida $h(\cdot)$.

Modelo Linear Generalizado Mixto

- $Y_i : i = 1, ..., n$
  mutuamente independentes, com $\mu_i = \mathrm{E}[Y_i]$, conditional às ralizações de um conjunto de de variáveis aleatórias latentes U_i,
- $h(\mu_i) = U_i + \sum_{j=1}^{k} f_{ij}\beta_j$,
  para uma função de ligação conhecida $h(\cdot)$.

A modelo espacial (geoestatístico)

- $Y_i : i = 1, ..., n$
  mutuamente independentes, com $\mu_i = \mathrm{E}[Y_i]$, conditional às realizações de um conjunto de de variáveis aleatórias latentes U_i,
- $h(\mu_i) = U_i + \sum_{j=1}^{p} f_{ij}\beta_j$,
  para uma função de ligação conhecida $h(\cdot)$,
- $U_i = S(x_i)$
  onde $\{S(x) : x \in \mathbb{R}^2\}$ é um processo estocástico espacial.
- $h(\mu_i) = U_i + \sum_{j=1}^{k} f_{ij}\beta_j$,



simulação ilustrando os componentes do modelo: dados $Y(x_i)$ (pontos), sinal $S(x)$ (linha curva) e média μ (linha horizontal).

3. O Modelo Gaussiano

(a) $S(\cdot)$ é um processo Gaussiano estacionário com

i. $\mathrm{E}[S(x)] = 0$,

ii. $\mathrm{Var}\{S(x)\} = \sigma^2$

iii. $\rho(u) = \mathrm{Corr}\{S(x), S(x-u)\}$;

(b) a distribuição condicional de Y_i dado $S(\cdot)$ é Gaussiana com média $\mu + S(x_i)$ e variância τ^2;

(c) $Y_i : i = 1, ..., n$ são mutuamente independentes, condicional à $S(\cdot)$.

Uma formulação equivalente para o modelo Gaussiano:

$$Y_i = \mu + S(x_i) + Z_i : i = 1, ..., n.$$

onde $Z_i : i = 1, ..., n$ são mutuamente independentes e identicamente distribuídos com $Z_i \sim \mathrm{N}(0, \tau^2)$.

Desta forma a distribuição conjunta de Y é multivariada Normal,

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

onde:

1 denota um vetor de 1's com n elementos

I é matrix identidade $n \times n$

R é uma matrix $n \times n$ com $(i, j)^{th}$ elemento $\rho(u_{ij})$ onde

$u_{ij} = ||x_i - x_j||$, é distancia Euclideana entre x_i e x_j.

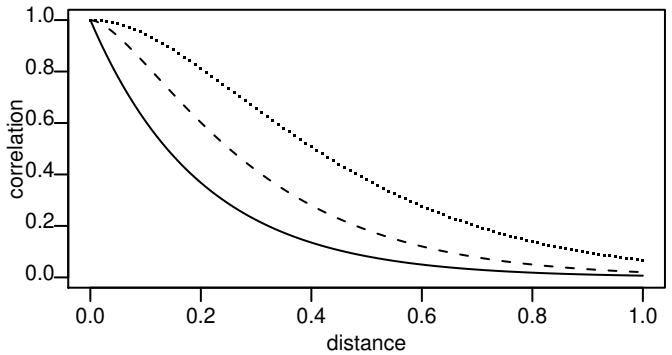
4. Especificação da função de correlação

A família de Matérn

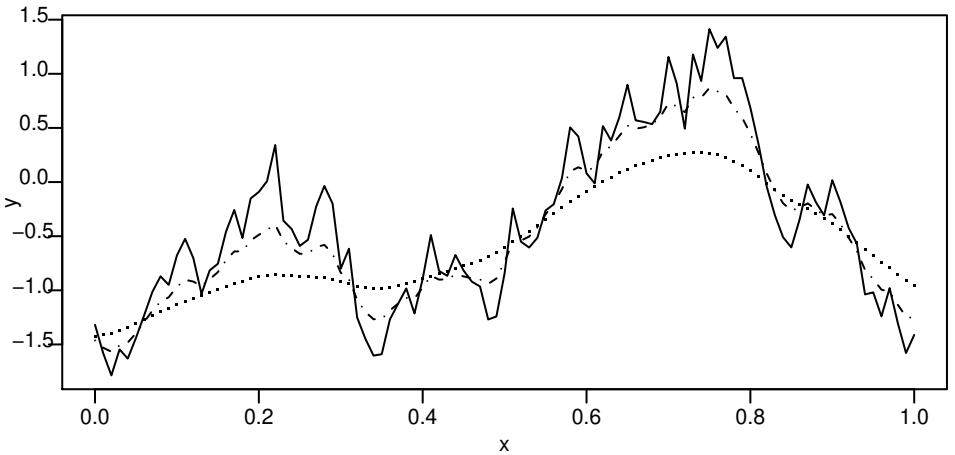
Função de correlação dada por

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

- κ e ϕ são parâmetros
- $K_{\kappa}(\cdot)$ denota função de Bessel de ordem κ
- válida para $\phi > 0$ e $\kappa > 0$.
- $\kappa = 0.5$: *modelo exponencial*
- $\kappa \to \infty$: *modelo Gaussiano*
- $S(x)$ é $\lceil \kappa - 1$ vezes diferenciável



Três exemplos de funções de Matérn com $\phi = 0.2$ and $\kappa = 1$ (linha sólida), $\kappa = 1.5$ (linha interrompida) and $\kappa = 2$ (pontos).



simulações de processos em 1-D com funções de correlação de de Matérn com $\phi = 0.2$ e $\kappa = 0.5$ (linha sólida), $\kappa = 1$ (linha interrompida) and $\kappa = 2$ (linha pontilhada).

VARIOGRAMAS

- o **variograma** de um processo $Y(x)$ é a função

$$V(x, x') = \frac{1}{2}\mathrm{Var}\{Y(x) - Y(x')\}$$

- para o modelo linear Gaussiano, com $u = ||x - x'||$,

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

- os parâmetros estruturais básicos são
  - *efeito pepita ("nugget")*: τ^2
  - *patamar ("sill")*: $\tau^2 + \sigma^2 = \mathrm{Var}\{Y(x)\}$
  - *o alcance ("range")*: ϕ, tal que $\rho(u) = \rho_0(u/\phi)$

- variogramas são definidos para uma classe mais ampla de processos, em comparação com correlogramas e covariogramas

- variogramas são largamente utilizados em geoestatística

5. Extensões do modelo básico

(a) Modelos Gaussianos transformados

- O modelo Gaussiano é claramente inapropriado para distribuições assimétricas.

- Certos dados podem indicar relações entre média e variância, que violam o modelo Gaussiano.

- Parâmetro extra λ da transformação Box-Cox introduz certa flexibilidade.

- O modelo fica então definido da forma:
  - assume-se $Y^* \sim MVN(F\beta, \sigma^2 V)$

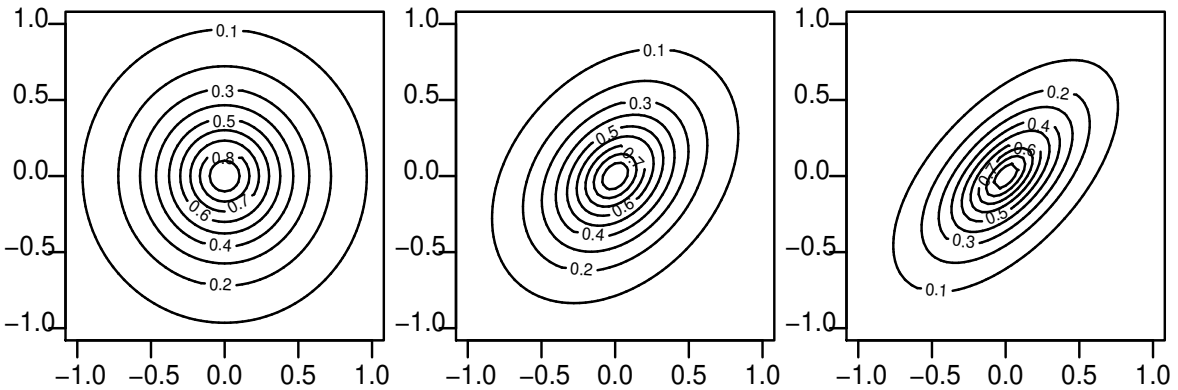
  - dados $y = (y_1, ..., y_n)$, são gerados por uma transformação do modelo linear Gaussiano $Y = h_\lambda^{-1}(Y^*)$ tal que:

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

(b) Efeitos Direcionais

- Condições ambientais podem induzir efeitos direcionais (vento, formação do solo, etc)

- como consequência a correlação espacial pode variar com a direção



contornos de correlação para modelo isotrópico (esq.) e dois modelos anisotrópicos (centro e dir.)

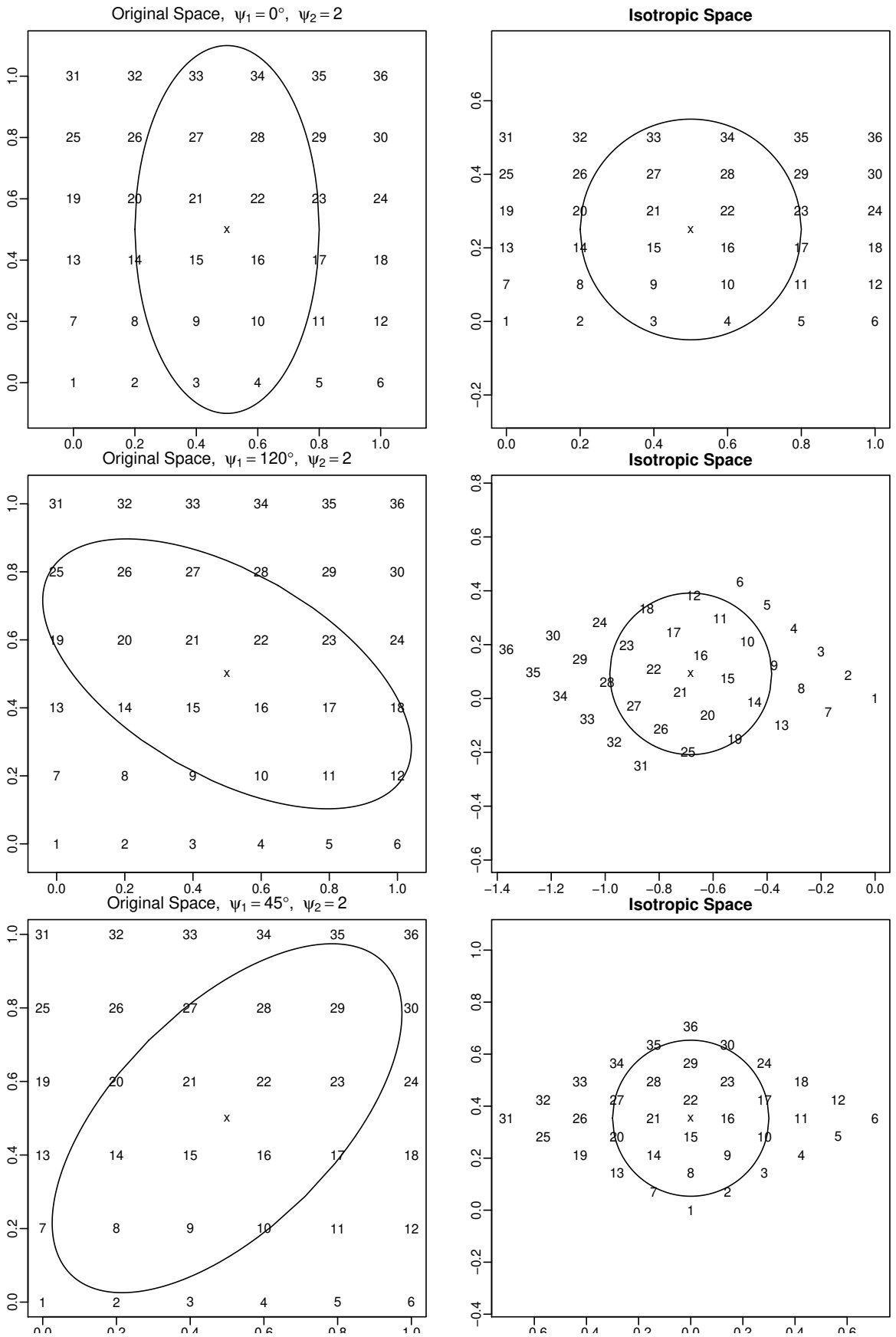
- *anisotropia geométrica*: possível (e simples) abordagem.

- dois parâmetros extra: *ângulo de anisotropia* ψ_A e *razão de anisotropia* ψ_R.

- rotação e contração/expansão das coordenadas originais:

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

“Correção”de anisotropia geometrica



(c) Modelos não estacionários

- *Modelos com médias não constantes (ou, incluindo covariáveis)*

  Substituir a média constante μ por

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

  para medidas $f_j(x)$ das covariáveis (lineares ou não lineares).

  Nota: corresponde a **krigagem universal** e **krigagem com tendência externa**.

- *Variação aleatória não estacionária*

  Variabilidade **intrinsica**: pressuposto mais fraco de estacionaridade (processo com incrementos estacionários, como passeios aleatórios em séries temporais), largamente utilizados como modelo "default"para variação espacial discreta (Besag, York and Molié, 1991).

  Métodos de **deformação espacial** (Sampson and Guttorp, 1992) buscam estacionaridade por transformações (complexas) do espaço geográfico, x.
  É preciso ter em mente o balanço entre a o aumento da flexibilidade de modelos mais gerais contra a sobre-modelagem de dados esparsos, que leva a pobre identificação dos parâmetros.

PARTE III:

PREDIÇÃO ESPACIAL

1. **Predição em processos estocásticos**

2. **Predição Geostatística**

3. **Predição no Modelo Gaussiano**

4. **O que a krigagem faz com os dados?**

5. **Predição de Funcionais**

1. Predição em processos estocásticos

General results for prediction

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, $\hat{T} = t(Y)$

best choice: needs a criterion

MMSPE: the *best* predictor minimises

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

Theorem 1.
The minimum mean square error predictor of T is

$$\hat{T} = \mathrm{E}(T|Y).$$

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

Comments

- 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 $[T|Y]$

- not transformation invariant:
  $\hat{T}$ the best predictor for T does NOT necessarily imply that $g(\hat{T})$ is the best predictor for $g(T)$.

2. Predição Geostatística

Suppose the **target** for prediction is $T = S(x)$
A **predictor** for T is a function $\hat{T} = \hat{T}(Y)$
The **mean square prediction error (MSPE)** is

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

The **the predictor** which minimises MSPE is

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

Two approaches:

- **Model-based geostatistics:**
  - specify a probability model for $[Y, T]$
  - choose $\hat{T}$ to minimise $MSPE(\hat{T})$ amongst all functions $\hat{T}(Y)$

- **Traditional (linear) geostatistics:**
  - Assume that $\hat{T}$ is linear in Y, so that

$$\hat{T} = b_0(x) + \sum_{i=1}^{n} b_i(x) Y_i$$

  - Choose b_i to minimise $MSPE(\hat{T})$ within the class of linear predictors

Coincident results under Gaussian assumptions

3. Predição sob o modelo Gaussiano

- assume that the target for prediction is $T = S(x)$
- $[T, Y]$ are jointly multivariate Gaussian.
- $\hat{T} = \mathrm{E}(T|Y)$, $\mathrm{Var}(T|Y)$ and $[T|Y]$ can be easily derived from a standard result:

Theorem 4. Let $X = (X_1, X_2)$ be jointly multivariate Gaussian, with mean vector $\mu = (\mu_1, \mu_2)$ and covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

ie $X \sim \mathrm{MVN}(\mu, \Sigma)$. Then, the conditional distribution of X_1 given X_2 is also multivariate Gaussian, $X_1|X_2 \sim \mathrm{MVN}(\mu_{1|2}, \Sigma_{1|2})$, where

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2)$$

and

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

For the geostatistical model:

$[T, Y]$ is multivariate Gaussian with mean vector $\mu\mathbf{1}$ and variance matrix

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

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

Hence, using Theorem 4 with $X_1 = T$ and $X_2 = Y$, we find that the minimum mean square error predictor for $T = S(x)$ is

$$\hat{T} = \mu + \sigma^2\mathbf{r}'(\tau^2 I + \sigma^2 R)^{-1}(Y - \mu\mathbf{1}) \tag{1}$$

with prediction variance

$$\mathrm{Var}(T|Y) = \sigma^2 - \sigma^2\mathbf{r}'(\tau^2 I + \sigma^2 R)^{-1}\sigma^2\mathbf{r}. \tag{2}$$

Notes

1. Because the conditional variance does not depend on Y, the prediction mean square error is equal to the prediction variance.

2. Equality of prediction mean square error and prediction variance is a special property of the multivariate Gaussian distribution, not a general result.

3. In conventional geostatistical terminology, construction of the surface $\hat{S}(x)$, where $\hat{T} = \hat{S}(x)$ is given by (1), is called *simple kriging*. This name is a reference to D.G. Krige, who pioneered the use of statistical methods in the South African mining industry (Krige, 1951).

4. O que a krigagem faz com os dados?

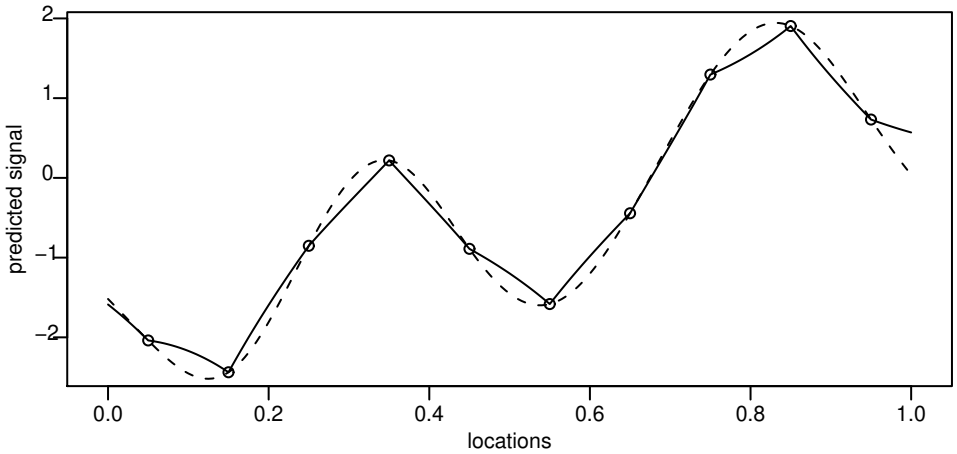
The minimum mean square error predictor for $S(x)$ is given by

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

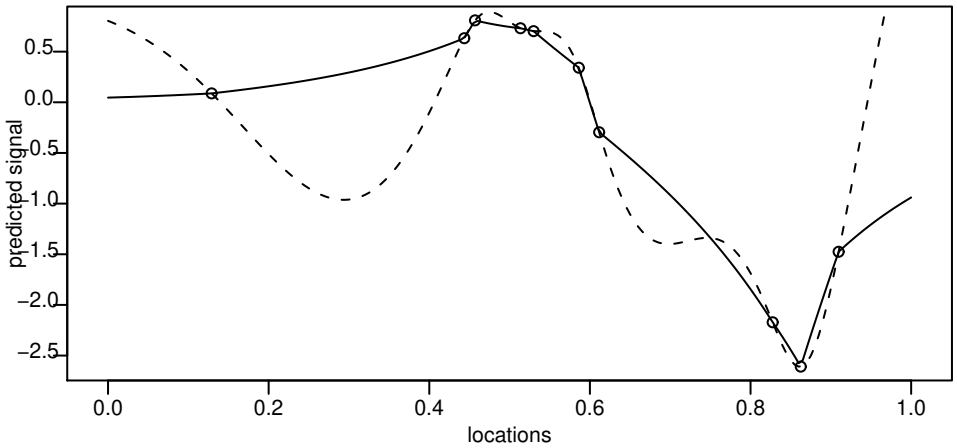
- the predictor $\hat{S}(x)$ compromises between its unconditional mean μ and the observed data Y
- the nature of the compromise depends on the target location x, the data-locations x_i and the values of the model parameters.
- call the $w_i(x)$ the *prediction weights*.

4.1 **Effects on predictions**

(a) **Varying the correlation function**

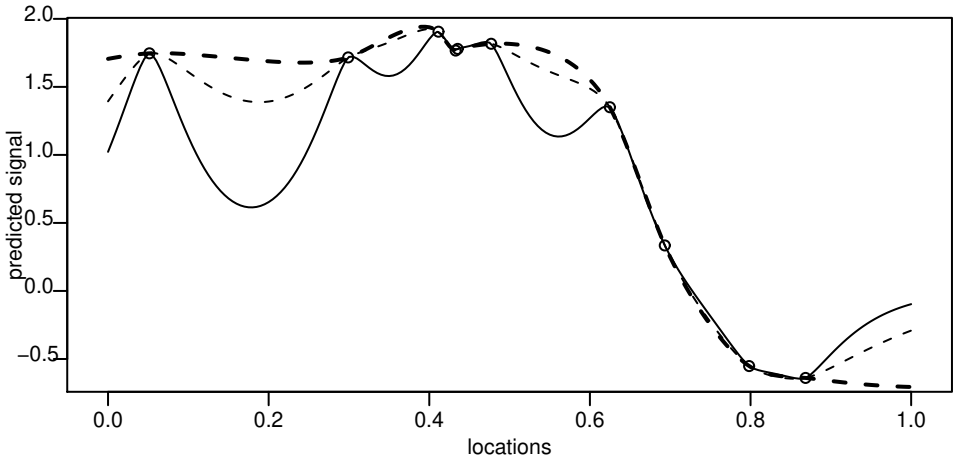


Predictions from 10 equally spaced data-points using exponential (solid line) or Matérn of order 2 (dashed line) correlation functions.



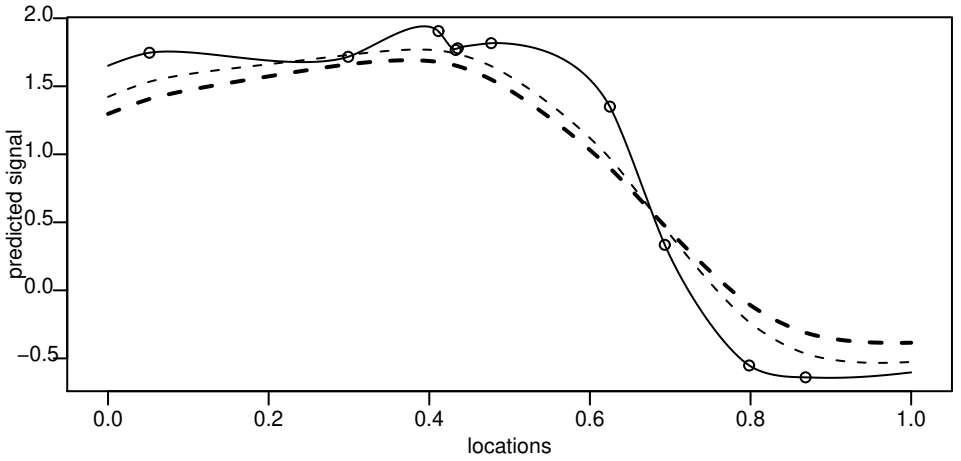
Predictions from 10 randomly spaced data-points using exponential (solid line) or Matérn of order 2 (dashed line) correlation functions.

(b) **Varying the correlation parameter**

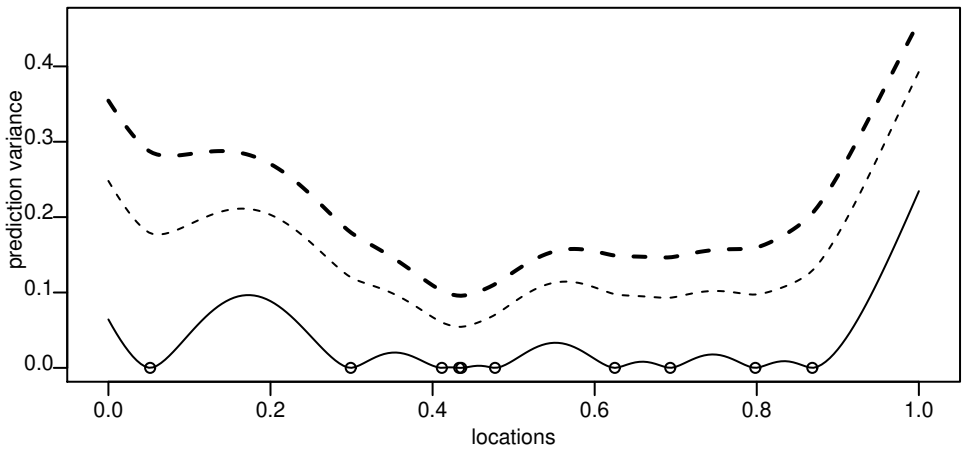


Predictions from 10 randomly spaced data-points using the Matérn ($\kappa = 2$) correlation function and different values of ϕ: 0.05 (solid line), 0.1 (dashed line) and 0.5 (thick dashed line).

(c) **Varying the noise-to-signal ratio**



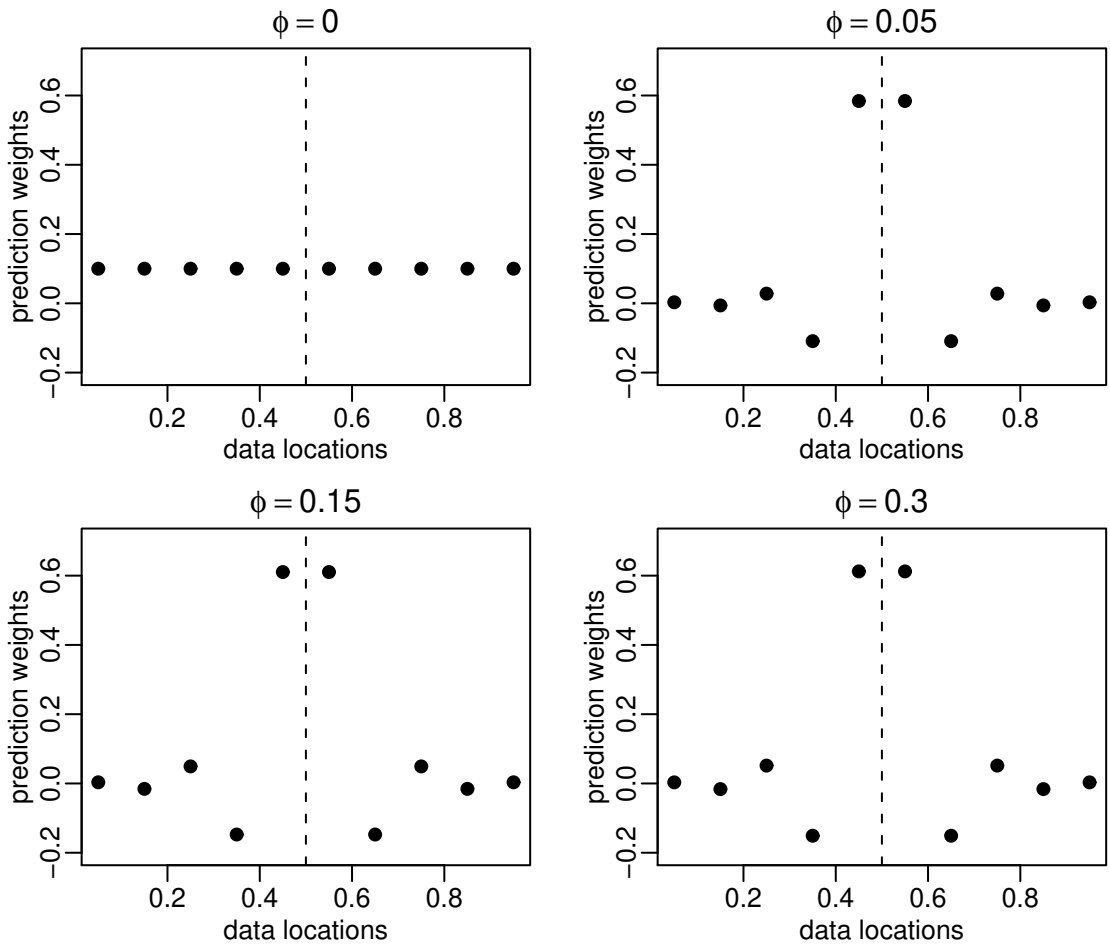
Predictions from 10 randomly spaced data-points using the Matérn correlation function and different values of τ^2: 0 (solid line), 0.25 (dashed line) and 0.5 (thick dashed line).



Prediction variances from 10 randomly spaced data-points using the Matérn correlation function and different values of τ^2: 0 (solid line), 0.25 (dashed line) and 0.5 (thick dashed line).

4.2 **Effects on kriging weights**

(a) **The prediction weights: varying** ϕ



Prediction weights for 10 equally spaced data-points with target location $x = 0.50$.

i. **varying parameter** $\phi = 0, 0.05, 0.15, 0.30$

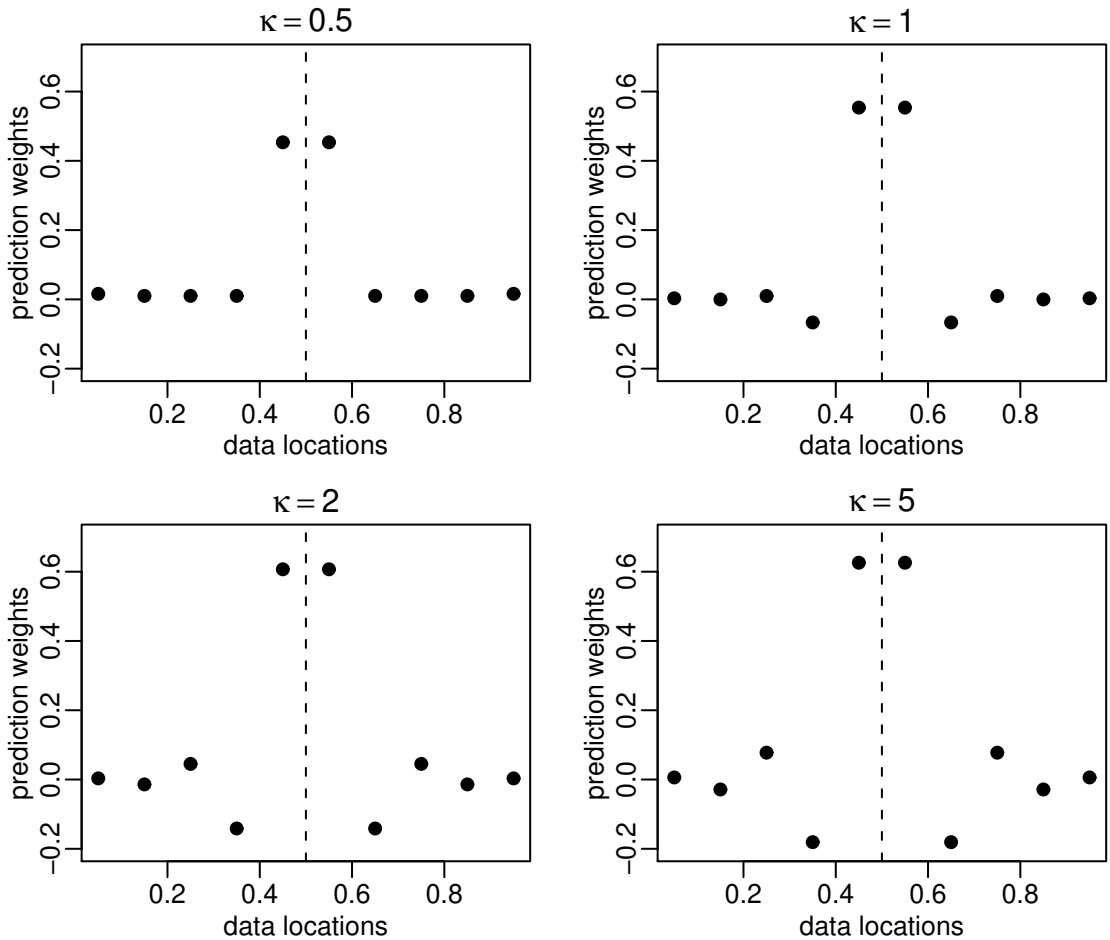
ii. locations: equally spaced $x_i = -0.05 + 0.1i$: $i = 1, ..., 10$

iii. prediction location: $x = 0.50$

iv. correlation function: Matérn with $\kappa = 2$

v. nugget: $\tau^2 = 0$

(b) **The prediction weights: varying** κ



Prediction weights for 10 equally spaced data-points with target location $x = 0.50$.

i. **varying parameter** $\kappa = 0.5, 1, 2, 5$

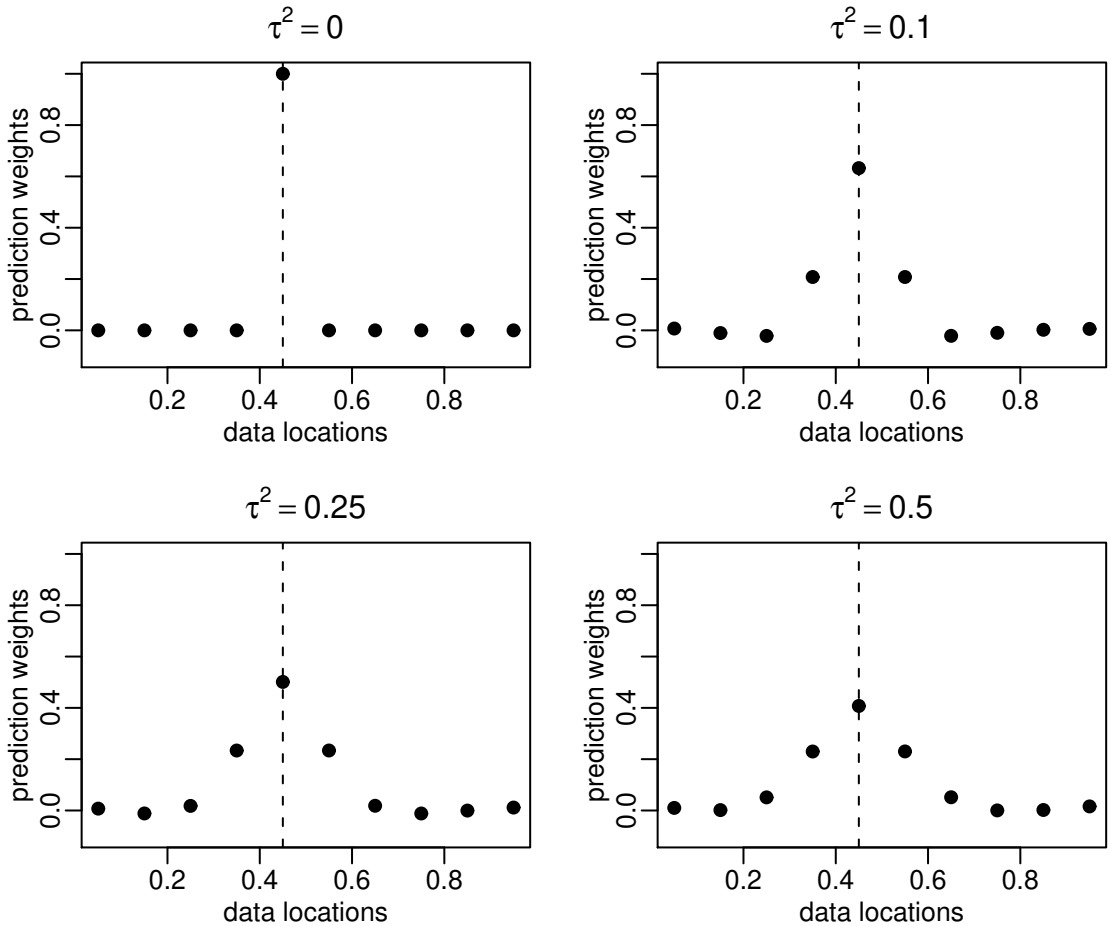
ii. locations: equally spaced $x_i = -0.05 + 0.1i$: $i = 1, ..., 10$

iii. prediction location: $x = 0.50$

iv. correlation function: Matérn with $\phi = 0.1$

v. Nugget: $\tau^2 = 0$

(c) **The prediction weights: varying** τ^2



Prediction weights for 10 equally spaced data-points with target location $x = 0.45$.

i. **varying parameter** $\tau^2 = 0, 0.1, 0.25, 0.5$

ii. locations: equally spaced $x_i = -0.05 + 0.1i$: $i = 1, ..., 10$

iii. prediction location: $x = 0.45$

iv. correlation function: Matérn with $\kappa = 2$ and $\phi = 0.1$

5. Predição de Funcionais

Let T be any *linear* functional of S,

$$T = \int_A w(x)S(x)dx$$

for some prescribed weighting function $w(x)$.

Under the Gaussian model:

- $[T, Y]$ is multivariate Gaussian;
- $[T|Y]$ is univariate Gaussian;
- the conditional mean and variance are:

$$\mathrm{E}[T|Y] = \int_A w(x)\mathrm{E}[S(x)|Y]dx$$

$$\mathrm{Var}[T|Y] = \int_A \int_A w(x)w(x')\mathrm{Cov}\{S(x), S(x')\}dxdx'$$

Note in particular that

$$\hat{T} = \int_A w(x)\hat{S}(x)dx$$

In words:

- given a predicted surface $\hat{S}(x)$, it is legitimate simply to calculate any linear property of this surface and to use the result as the predictor for the corresponding linear property of the true surface $S(x)$

- it is *NOT* legitimate to do this for prediction of non-linear properties

- for example, the maximum of $\hat{S}(x)$ is a very bad predictor for the maximum of $S(x)$ (this problem will be addressed later)

PARTE IV:

ESTIMAÇÃO DE PARÂMETROS

1. **Propriedades do Segundo Momento**

2. **Estimação usando Variogramas**

3. **Estimação por Verossimilhança**

4. **Predição "plug-in"**

5. **Estudo de Caso**

6. **Comentários e Extensões**

1. Propriedades do segundo momento

- o **variograma** é a função

$$V(x, x') = \frac{1}{2}\mathrm{Var}\{Y(x) - Y(x')\}$$

- para $u = ||x - x'||$,

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

- os paramêtros estruturais básicos são
  - *efeito pepita ("nugget")*: τ^2
  - *patamar ("sill")*: $\tau^2 + \sigma^2 = \mathrm{Var}\{Y(x)\}$
  - *o alcance ("range")*: ϕ, tal que $\rho(u) = \rho_0(u/\phi)$

- implicações práticas:
  - qualquer versão razoável do modelo linear Gaussiano tem pelo menos três parâmetros de covariância
  - um volume da dados substancial pode ser necessário para estimar maior número de parâmetros
  - a família **Matérn** possui um parâmetro extra para determinar a suavidade do processo $S(x)$

Paradigmas para estimação

- **Métodos "Ad-hoc" (baseados em variogramas)**
  - calcule o variograma empírico
  - ajuste um modelo teórico de variograma

- **Métodos baseados na verossimilhança**
  - tipicamente sob pressupostos de Gaussianidade
  - Ótimos sob as condições declaradas
  - maior demanda computacional
  - podem não ser robustos

- **Implementação Bayesiana**,
  - estimação e predição combinadas
  - cada vez mais aceitos

2. Estimação usando Variogramas

- O variograma é definido por

$$V(x, x') = \frac{1}{2}\mathrm{Var}\{Y(x) - Y(x')\}$$

- se $Y(x)$ é estacionário,

$$V(x, x') = V(u) = \frac{1}{2}\mathrm{E}[\{Y(x) - Y(x')\}^2]$$

  onde $u = ||x - x'||$

- sugere uma estimativa empírica para $V(u)$:

$$\hat{V}(u) = \mathrm{average}\{[y(x_i) - y(x_j)]^2\}$$

  onde cada média é tomada entre todos os pares $[y(x_i), y(x_j)]$ tal que $||x_i - x_j|| \approx u$

- para processo com média não constante a tendência pode ser removida:
  - defina $r_i = Y_i - \hat{\mu}(x_i)$
  - defina $\hat{V}(u) = \mathrm{average}\{(r_i - r_j)^2\}$,
    onde cada média é tomada entre todos os pares (r_i, r_j)

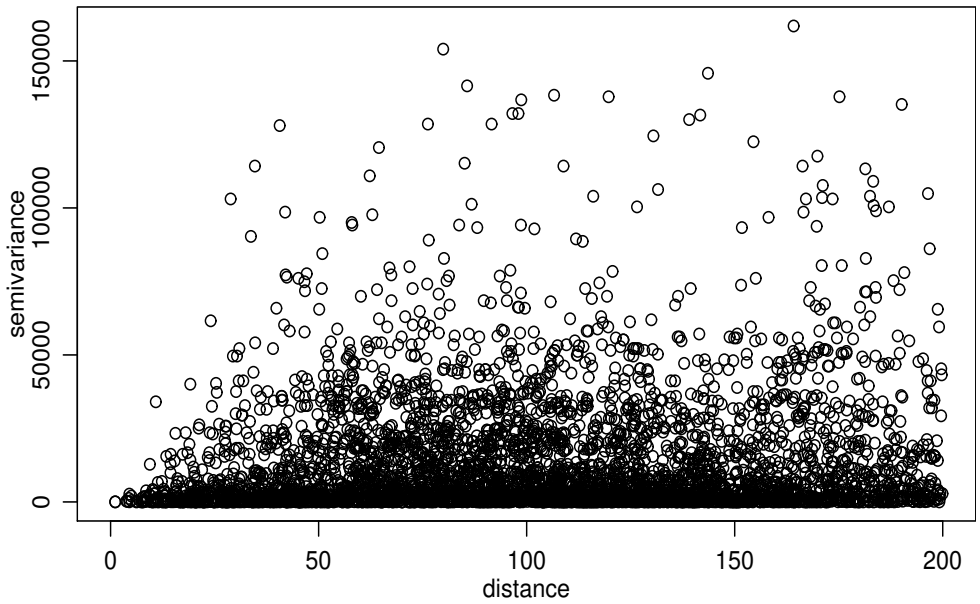
(a) **A nuvem variográfica**

- defina as quantidades:

$$
\begin{aligned}
r_i &= Y_i - \hat{\mu}(x_i) \\
u_{ij} &= ||x_i - x_j|| \\
v_{ij} &= \frac{(r_i - r_j)^2}{2}
\end{aligned}
$$

- a **nuvem de variograma** é um gráfico de pontos (u_{ij}, v_{ij})

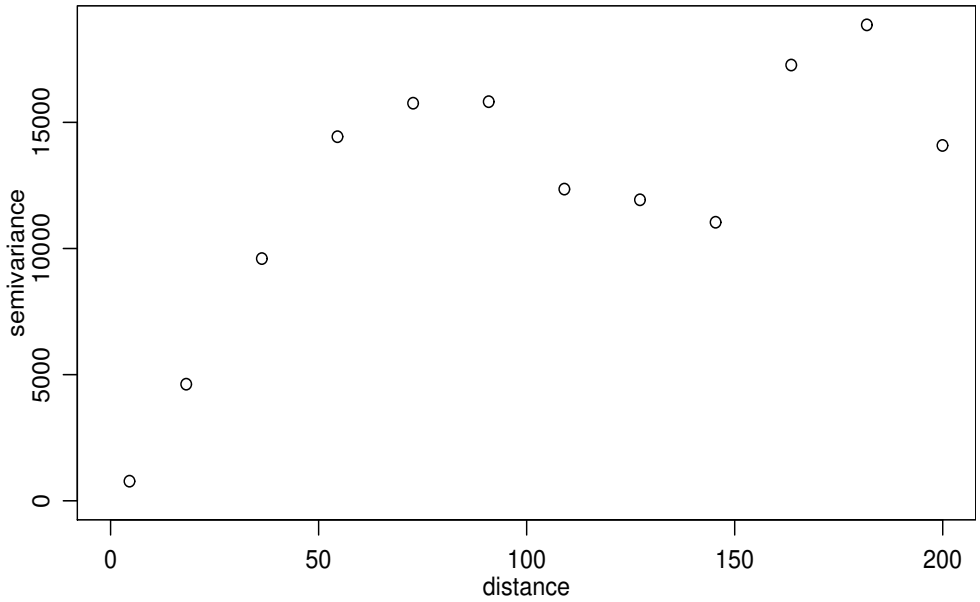
Exemplo: Dados de chuva na Suíça



(b) **O variograma empírico**

- obtido a partir da nuvem de variograma tomando médias de classes de distância: $u - h/2 \leq u_{ij} < u + h/2$
- forma k classes de distâncias, cada uma média de n_k pares,
- muito sensível à especificação de média do processo $\mu(x)$

Exemplo: Dados de chuva na Suíça



Variograma empírico

(c) **O ajuste de variogramas**

Estime os parâmetros $\tilde{\theta}$ minimizando um particular critério

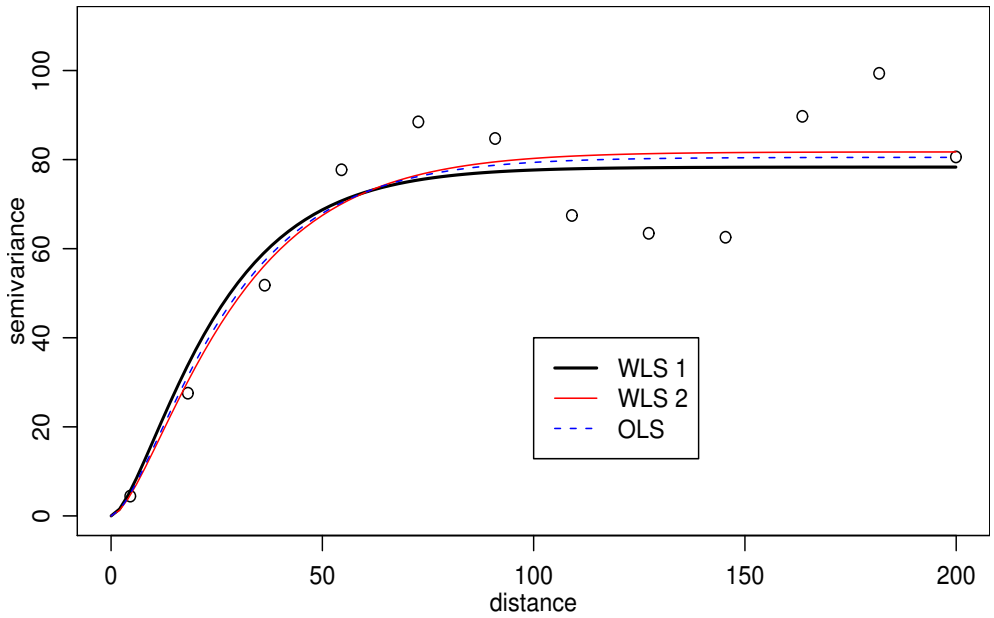
por exemplo, mínimos quadrados generalizados (Cressie, 1993)

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

onde $\bar{V}_k$ é a média das n_k ordenadas v_{ij} do variograma.

Outros critérios: OLS, WLS com diferentes pesos, GLS, quasi-verossimilhança.

Exemplo: Dados de chuva na Suíça

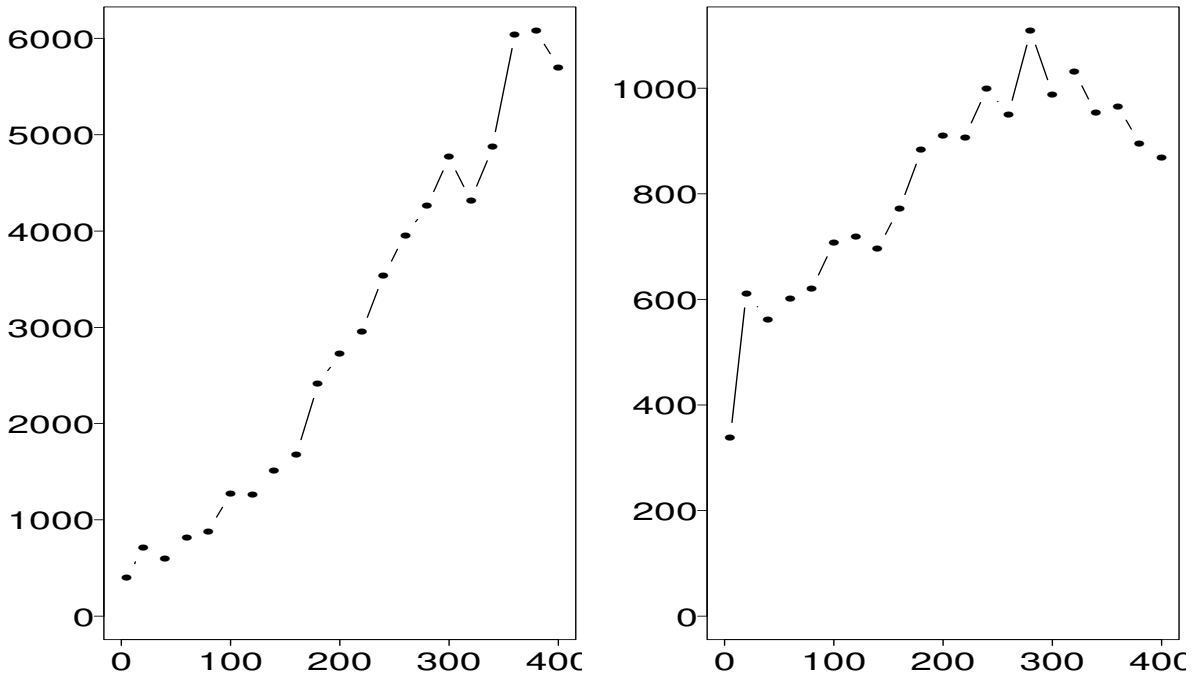


Ajuste de variograma empírico por WLS com pesos dados por n_k apenas (linha grossa), WLS com pesos sugeridos pro Cressie (linha cheia) o OLS (linha pontilhada)

(d) **Comentários sobre uso de variogramas para inferência**

i. Variogramas dos dados originais e resíduos podem ser muito diferentes

Exemplo: Dados do Paraná



variogramas empíricos dos dados originais (esq.) e resíduos após regressão em latitude, longitude e altitude (dir.)

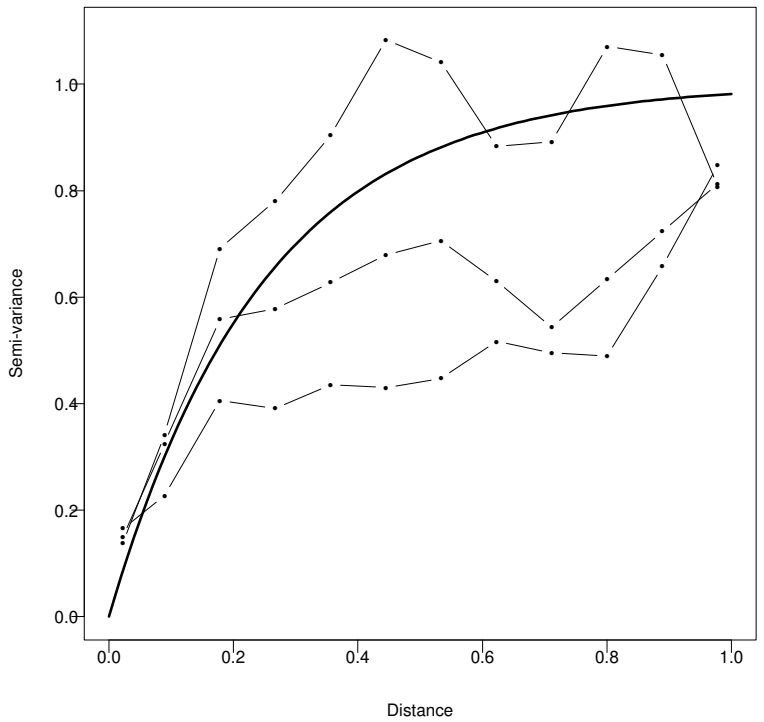
- variograma de dados originais reflete (inclui) variação de tendência geográfica de larga escala.

- variograma de resíduos elimina esta fonte de variação

ii. Quão instáveis são os variogramas empíricos?

- sob o modelo linear Gaussiano:
  - $v_{ij} \sim V(u_{ij})\chi^2_1$
  - the v_{ij} são correlacionados

iii. variogramas de diferentes realizações do mesmo processo podem ser bem diferentes



- linha sólida mostra o variograma verdadeiro do processo
- linha finas mostram variogramas empíricos de três realizações do mesmo processo (modelo)
- as altas correlações entre $\hat{V}(u)$ para sucessivos u conferem uma suavidade "enganadora"

iv. Ajuste de modelos por mínimos quadrados

- corresponde a um sistema de equações de estimação que produz estimativas viciadas de θ,
- mesmo assim é largamente utilizado na prática
- potencialmente "perigoso" devido as correlações inerentes aos sucessivos $\bar{V}_k$'s

v. estimação baseada em objeto que é por si estimado

vi. é possivel ajustar o modelo diretamente aos **dados** e não ao variograma.

3. Estimação por verossimilhança

O modelo Gaussiano é dado por:

$$Y_i|S \sim \mathrm{N}(S(x_i), \tau^2)$$

- $S(x_i) = \mu(x_i) + S_c(x_i)$
- $S_c(\cdot)$ é um processo estocástico Gaussiano com parâmetros de covariância (σ^2, ϕ, κ),
- $\mu(x_i) = F\beta = \sum_{j=1}^{k} f_k(x_i)\beta_k$, onde $f_k(x_i)$ é vetor de covariáveis na localização x_i

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

e a função de verossimilhaça é:

$$\begin{aligned}\ell(\beta, \tau, \sigma, \phi, \kappa) \propto \quad & -0.5\{\log|(\sigma^2 R + \tau^2 I)| + \\ & (y - F\beta)'(\sigma^2 R + \tau^2 I)^{-1}(y - F\beta)\}.\end{aligned}$$

para qual maximização (numérica) produz as estimativas de máxima verossimilhança

Modelos Gaussianos transformados

A log-verossimilhança é:

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

para $\theta = (\tau, \sigma, \phi, \kappa)$.

ML: tipicamente usada sob pressupostos de normalidade

- estimativas ótimas sob os pressupostos declarados

- porém computacionalmente caros e podem não ser robustos

- dificuldades computacionais para grande número de dados

- Implementação Bayesiana combinando estimação e predição tem sido cada vez mais aceita (ao menos entre estatísticos!).

- Para família de Matérn considere tomar κ em um conjunto discreto $\{0.5, 1, 2, 3, ..., N\}$

4. Predição "plug-in"

Em geral o interesse está em predizer

- o valor da realização do processo $S(\cdot)$ em um ponto
- ou a média de $S(\cdot)$ em uma região

$$T = |B|^{-1} \int_B S(x)dx$$

onde $|B|$ denota a área da região B.

Para o modelo Gaussiano o preditor de minímos quadrados de $T = S(x)$ é:

$$\hat{T} = \mu + \sigma^2 \mathbf{r}'(\tau^2 I + \sigma^2 R)^{-1}(Y - \mu \mathbf{1})$$

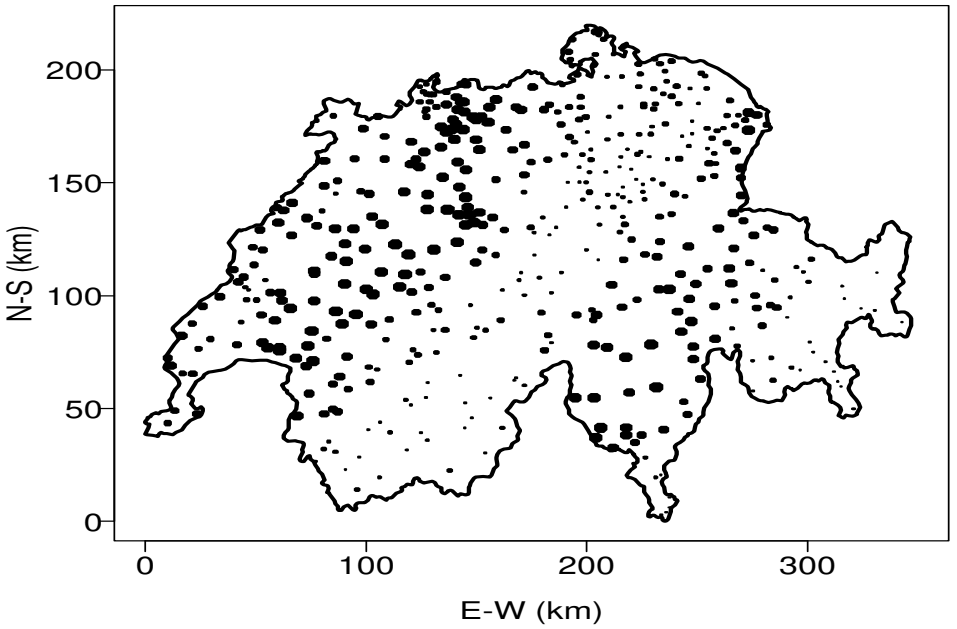
e a variância de predição

$$\mathrm{Var}(T|Y) = \sigma^2 - \sigma^2 \mathbf{r}'(\tau^2 I + \sigma^2 R)^{-1} \sigma^2 \mathbf{r}$$

onde os únicos termos desconhecidos são os parâmetros do modelo

A **predição "plug-in"** consiste em substituir os parâmetros por suas estimativas.

5. Estudo de caso: chuva na Suíça



Localizações com tamanho dos pontos proporcional aos valores observados. Distâncias em kilometros

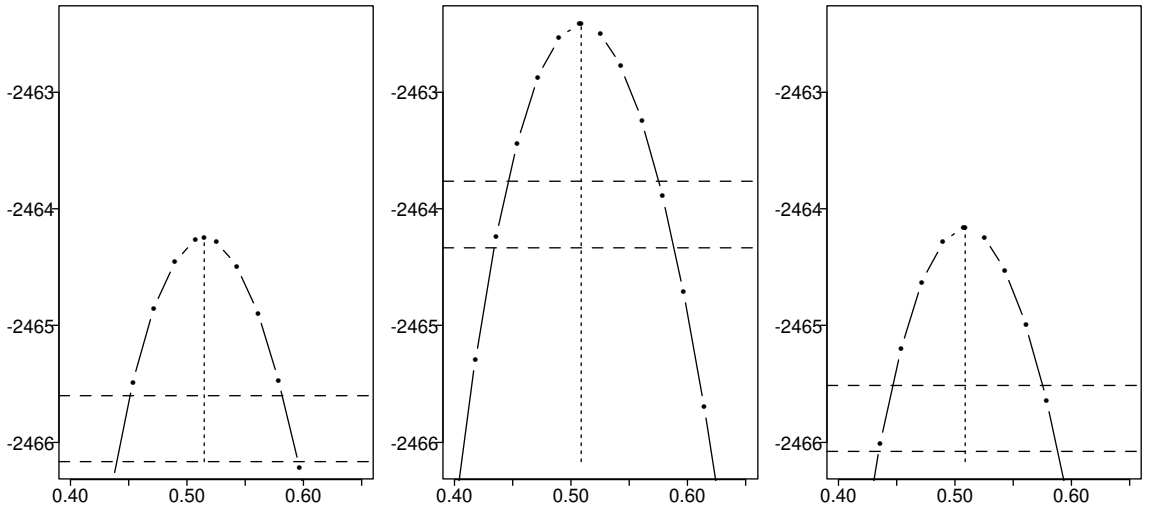
- 467 localizações
- medidas de precipitação em 8 de Maio 1986
- dados são valores inteiros com unidade de medida igual á $1/10$ mm
- 5 localizações com valores iguais à zero.

chuva na Suíça (cont.)

Estimação parâmetros de transformação e suavidade (modelo de Matérn)

κ	$\hat{\lambda}$	$\log \hat{L}$
0.5	0.514	-2464.246
1	0.508	-2462.413
2	0.508	-2464.160

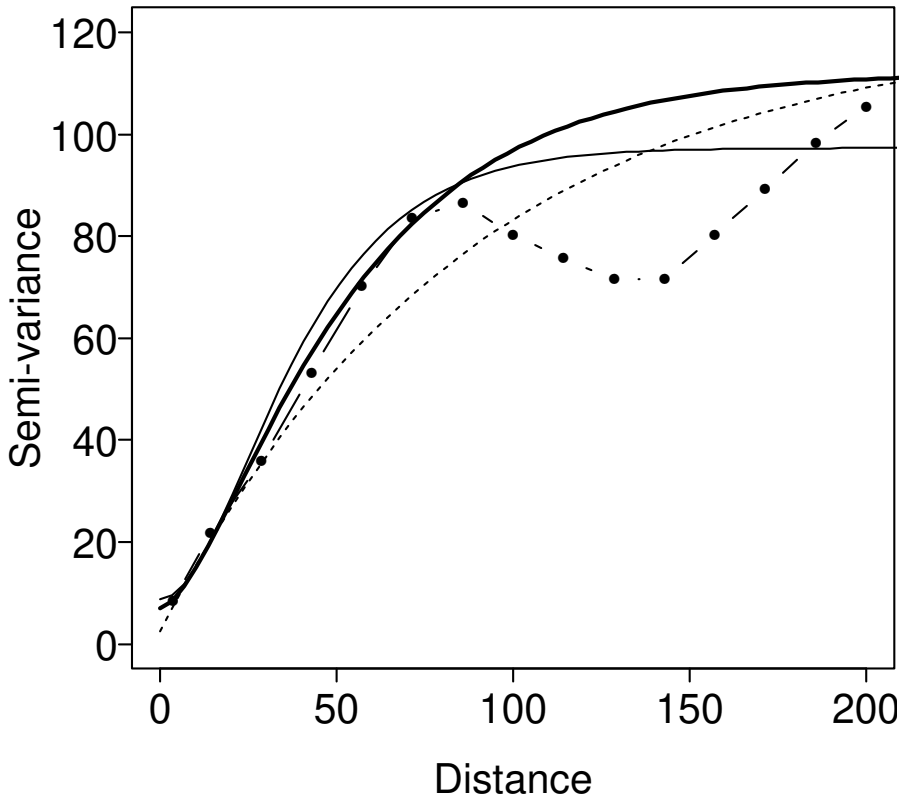
Estimativas de MV de $\hat{\lambda}$ e valores da log-verossimilhança $\log \hat{L}$ para diferentes valores de κ.



Verossimilhanças prefilhadas para λ. esquerda: $\kappa = 0.5$, meio: $\kappa = 1$, direita: $\kappa = 2$.

transformação logarítimica ou não-transformação são claramente NÃO indicadas!

chuva na Suíça (cont.)



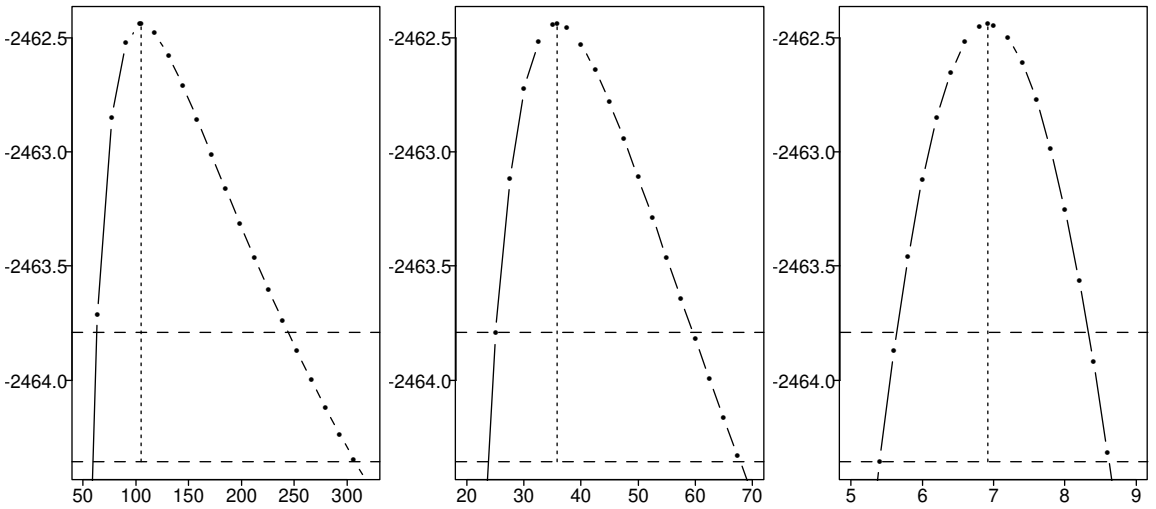
semivariograma empírico para dados transformados e variogramas teóricos com estimativas de MV para $\kappa = 0.5$ (linha interrompida), $\kappa = 1$ (linha grossa), $\kappa = 2$ (linha fina).

chuva na Suíça (cont.)

Estimativas para modelo com $\lambda = 0.5$

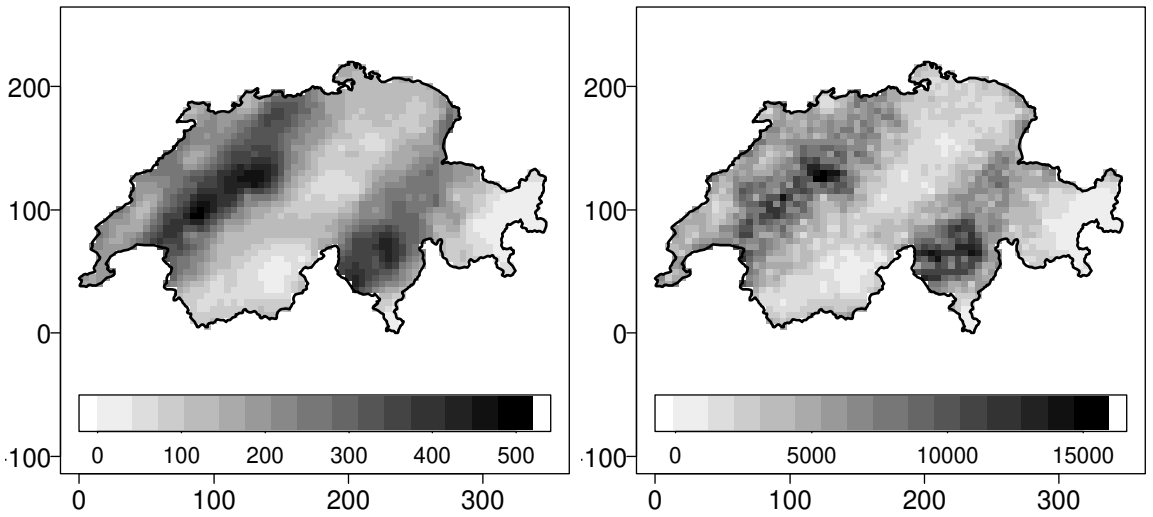
κ	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\phi}$	$\hat{\tau}^2$	$\log \hat{L}$
0.5	18.36	118.82	87.97	2.48	-2464.315
1	20.13	105.06	35.79	6.92	-2462.438
2	21.36	88.58	17.73	8.72	-2464.185

Maximum likelihood estimates $\hat{\beta}$, $\hat{\phi}$, $\hat{\sigma}$, $\hat{\tau}$ and the corresponding value of the likelihood function $\log \hat{L}$ for different values of the Matérn parameter κ, for $\lambda = 0.5$



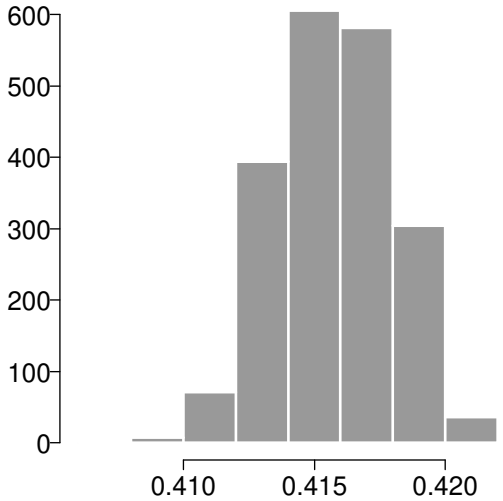
Verossimilhança perfilhada para parâmetros de covariância $\kappa = 1$ and $\lambda = 0.5$. esquerda: σ^2, meio: ϕ, direita: τ^2.

chuva na Suíça (cont.)



Mapas com predições (esquerda) e variâncias de predição (direita).

Predição da percentagem da área onde $Y(x) \geq 200$: $\tilde{A}_{200}$ é de 0.4157



Amostras da preditiva de $\tilde{A}_{200}$.

6. Outros tópicos e extensões

- Máxima verossimilhança restrita (REML)

- Verossimilhanças perfilhadas

- Estimação de modelos anisotrópicos

- Validação dos modelos

- Modelos não estacionários
  - tendências e/ou covariáveis
    (a) Predição *ad-hoc*:
    (b) estima β por OLS, $\tilde{\beta} = (F'F)^{-1}F'Y$, e calcula-se resíduos $Z = Y - F\tilde{\beta}$.
    (c) calcula-se o variograma empírico (dos resíduos) que é utilizado para formulação do modelo e estimação de parâmetros
    (d) reestima-se β por GLS e usa-se modelo ajustado para predição

    Nota: **krigagem universal** ou **krigagem com tendência externa**

- Relações funcionais entre médias e variâncias
- variação aleatória não estacionária
  Intrínseca Campos aleatórios de Markov (Besag, York and Molié, 1991).
  Deformações espaciais Sampson and Guttorp, 1992 tentam obter estacionaridade através de transformações não lineares do espaço geográfico x.
  Ver tese de Alexandra Smith (2001).
- flexibilidade *vs* identificabilidade

• papel dos variogramas empíricos
  - diagnóstico (abordagem "model-based")
  - ferramenta de inferência (abordagem tradicional)

• ambas abordagens anexam estimativas dos parametros ao modelo como se fosse valores verdadeiros.
  Predição "plug-in"
  - usualmente produz boas estimativas pontuais de $T = S(x)$
  - em geral sub-estima variância de predição
  - pode produzir estimativas inacuradas de outras quantidades objetivo T

PARTE V:

INFERÊNCIA BAYESIANA PARA O MODELO GAUSSIANO

1. **Inferência Bayesiana**

2. **Resultados para o modelo Gaussiano**

3. **Estudo de Caso: Dados da Suíça**

1. Análise Bayesiana - Conceitos Basicos

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.

For inference about model parameters, the full model specification now should include the model parameters:

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

Bayes' Theorem allows us to calculate:

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

Thus,

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

is the *posterior distribution* where

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

The Bayesian paradigm:

(a) **Model**

- the full model specification consists of $[Y, \theta] = [Y|\theta][\theta]$.
- formulate a model for the observable variable Y.
- this model defines $[Y|\theta]$ (and hence an expression for the log-likelihood $\ell(\theta; Y)$)

(b) **Prior**

- before we observe Y, the marginal $[\theta]$ expresses our uncertainty about θ
- call $[\theta]$ *prior distribution* for θ

(c) **Posterior**

- having observed Y, it is no longer an unknown (randomly varying) quantity
- therefore revise uncertainty about θ by conditioning on the observed value of Y
- call $[\theta|Y]$ *posterior distribution* for θ, and use it to make inferential statements

NOTE: the likelihood function occupies a central role in both classical and Bayesian inference

Prediction

Because Bayesian inference treats θ as a random variable, it makes no formal distinction between parameter estimation problems and prediction problems, and thereby provides a natural means of allowing for parameter uncertainty in predictive inference.

The general idea for prediction is to formulate a model for

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

and make inferences based on the conditional distribution

$$\begin{aligned} [T|Y] &= \int [T, \theta|Y] d\theta \\ &= \int [\theta|Y][T|Y, \theta] d\theta \end{aligned}$$

Comparing plug-in and Bayesian

- the 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, or in other words:

- allowance for parameter uncertainty usually results in wider prediction intervals

Notes:

(a) Until recently, the need to evaluate the integral which defines $[Y]$ represented a major obstacle to practical application.

(b) Development of Markov Chain Monte Carlo (MCMC) methods has transformed the situation.

(c) 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.

(d) in particular for the Gaussian model other algorithms can be implemented.

2. Results for the Gaussian Model

Uncertainty only in the mean parameter

Assume for now that only the mean parameter β is regarded as random with (conjugate) prior:

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

The posterior is given by

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

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

$$\begin{aligned}
E[S^*|Y] &= (F_0 - r'V^{-1}F)(V_\beta^{-1} + F'V^{-1}F)^{-1}V_\beta^{-1}m_\beta + \\
&\quad \left[r'V^{-1} + (F_0 - r'V^{-1}F)(V_\beta^{-1} + F'V^{-1}F)^{-1}F'V^{-1}\right] Y \\
\mathrm{Var}[S^*|Y] &= \sigma^2 \left[V_0 - r'V^{-1}r + \right. \\
&\quad \left. (F_0 - r'V^{-1}F)(V_\beta^{-1} + F'V^{-1}F)^{-1}(F_0 - r'V^{-1}F)'\right].
\end{aligned}$$

The predictive variance has three interpretable components: a priori variance, the reduction due to the data and the uncertainty in the mean.

$V_\beta \to \infty$ corresponds to universal (or ordinary) kriging.

Uncertainty for all model parameters

Assume (w.l.g.) a model without measurement error and the prior $p(\beta, \sigma^2, \phi) \propto \frac{1}{\sigma^2} p(\phi)$.

The posterior distribution:

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

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

Algorithm 1:

(a) 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.

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

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

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

(e) 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:

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

To sample from this distribution:

Algorithm 2:

(a) Discretise $[\phi|Y]$, as in Algorithm 1.

(b) Compute the posterior probabilities on the discrete support set. Denote the resulting distribution $\tilde{pr}(\phi|y)$.

(c) Sample a value of ϕ from $\tilde{pr}(\phi|y)$.

(d) Attach the sampled value of ϕ to $[s^*|y, \phi]$ and sample from it obtaining realisations of the predictive distribution.

(e) Repeat steps (3) and (4) as many times as required to generate a sample from the required predictive distribution.

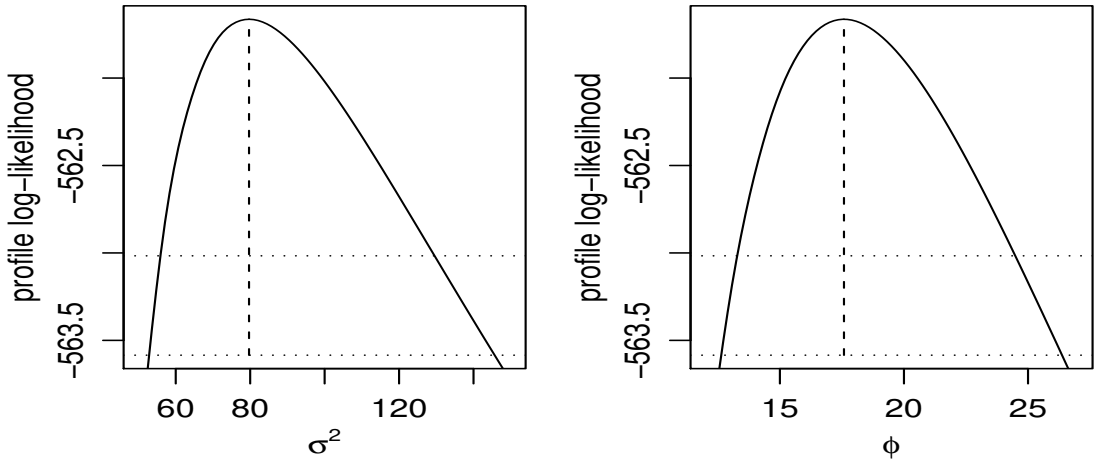
Note:

(a) The algorithms are of the same kind to treat τ and/or κ as unknown parameters.

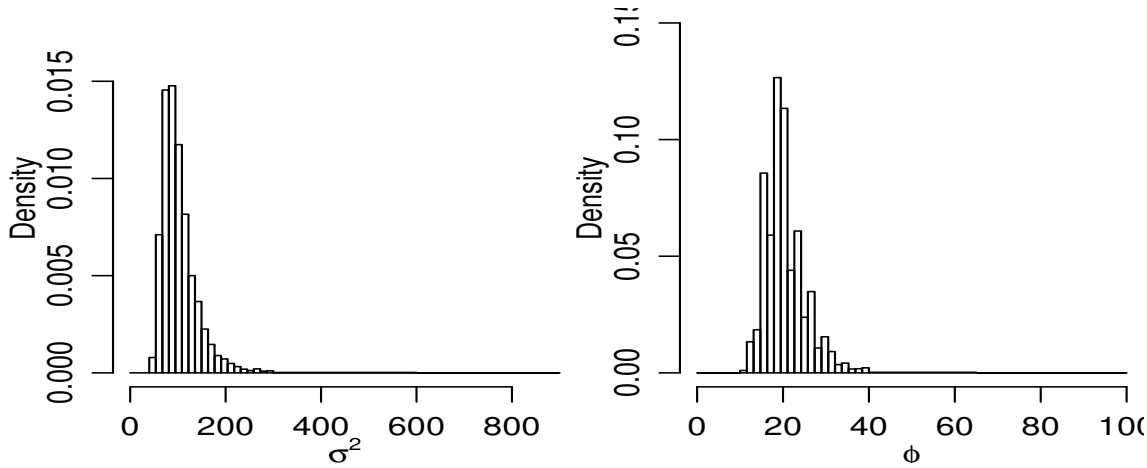
(b) We specify a discrete prior distribution on a multi-dimensional grid of values.

(c) This implies extra computational load (but no new principles)

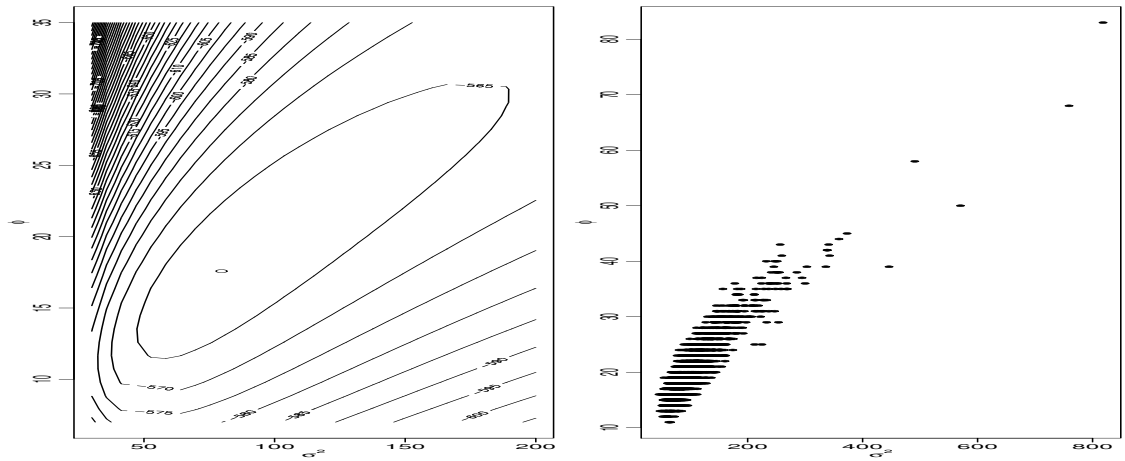
3. A Case Study: Swiss rainfall, 100 data



Profile likelihoods for covariance parameters: σ^2; ϕ

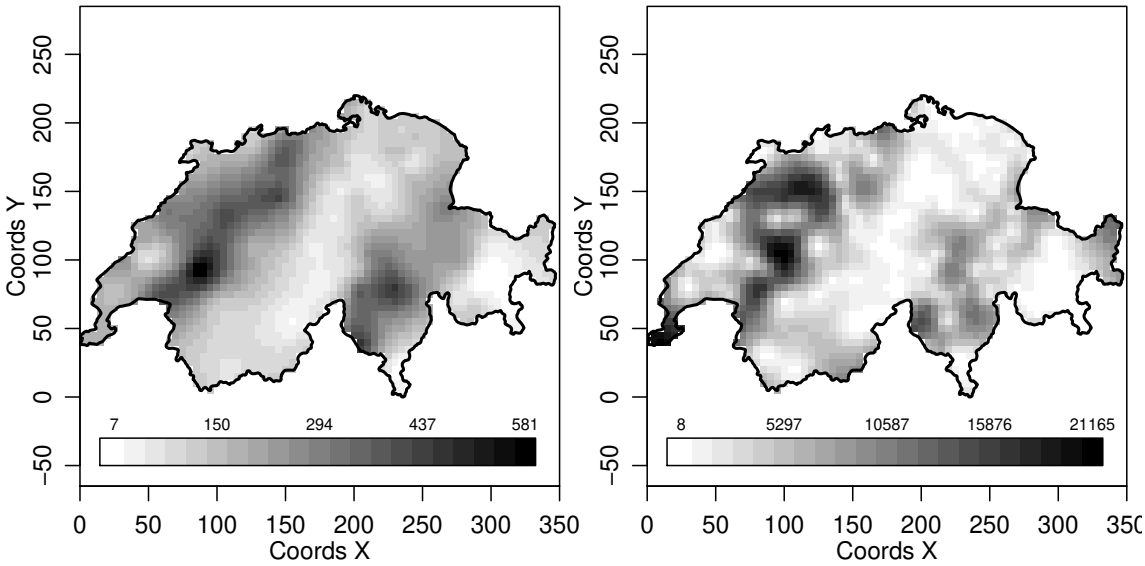


Posterior distributions for covariance parameters: σ^2; ϕ

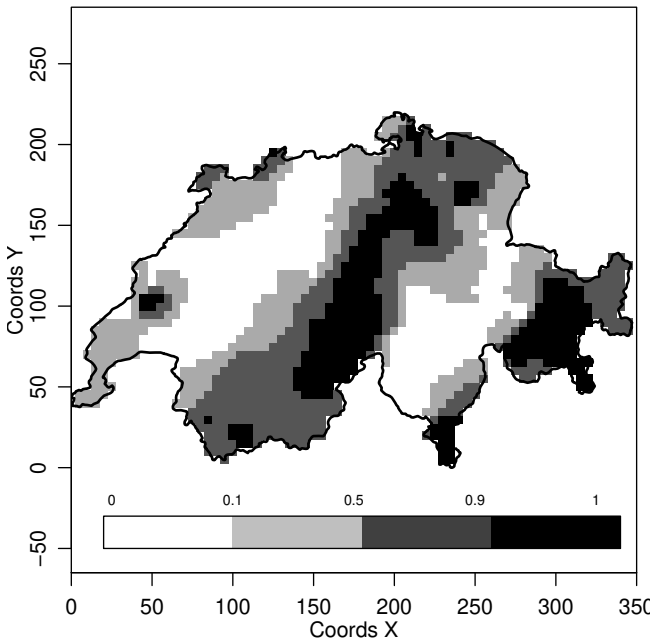


2D profile log-likelihood (left) and samples from posterior distributions (right) for parameters σ^2 and ϕ

Swiss rainfall: prediction results

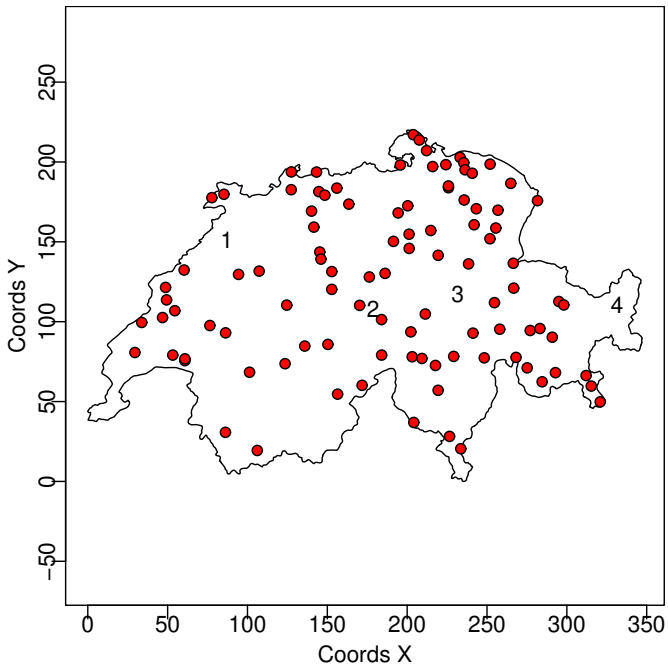


Predicted signal surfaces and associated measures of precision for the rainfall data: (a) posterior mean; (b) posterior variance

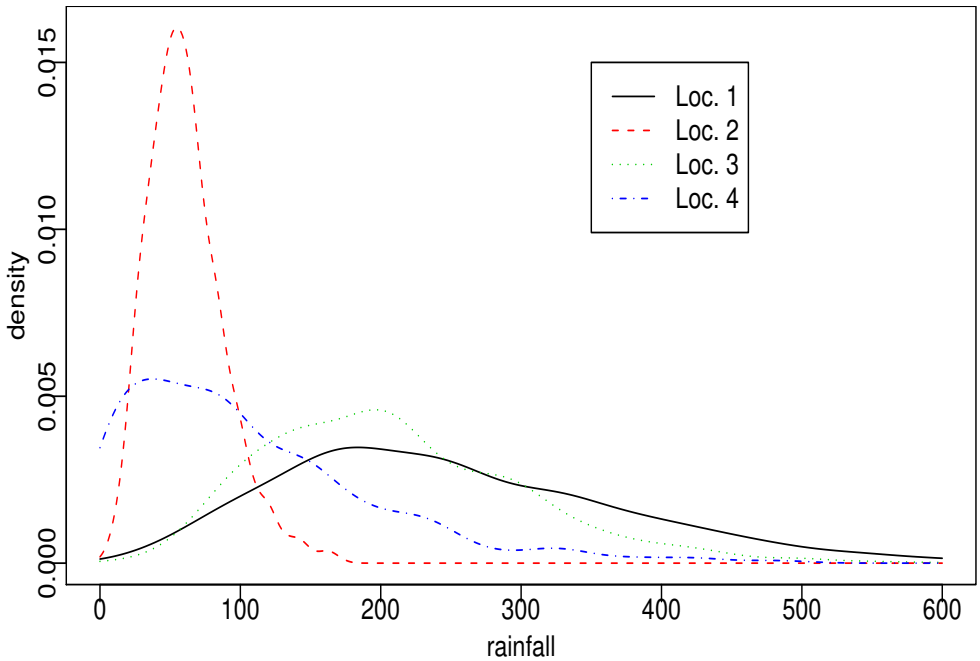


Posterior probability contours for levels 0.10, 0.50 and 0.90 for the random set $T = \{x : S(x) < 150\}$

Swiss rainfall: prediction results (cont.)



Recording stations and selected prediction locations (1 to 4)



Bayesian predictive distributions for average rainfall at selected locations.

PARTE VI:

MODELOS LINEARES GENERALIZADOS ESPACIAIS

1. **Modelos lineares Generalizados Espaciais**

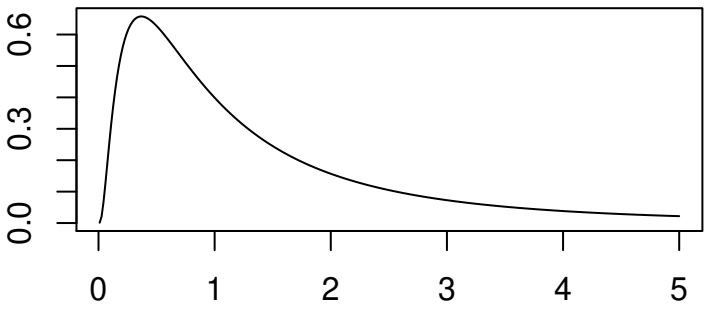
2. **Inferência via MCMC**

3. **estudo de caso: Ilha de Rongelap**

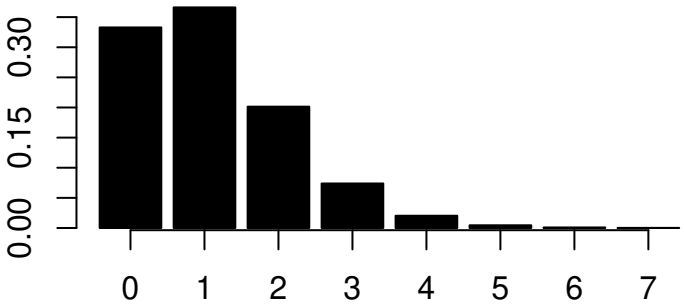
4. **estudo de caso: Malária em Gâmbia**

1. Modelos lineares Generalizados Espaciais

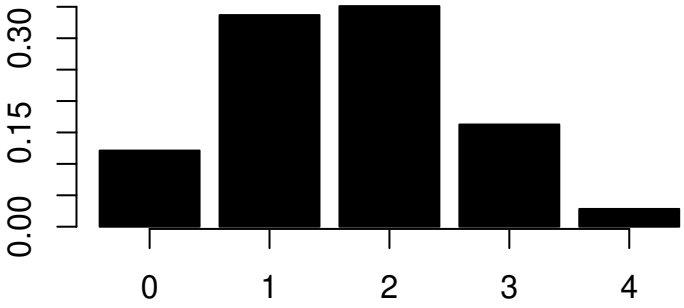
Dados Positivos:



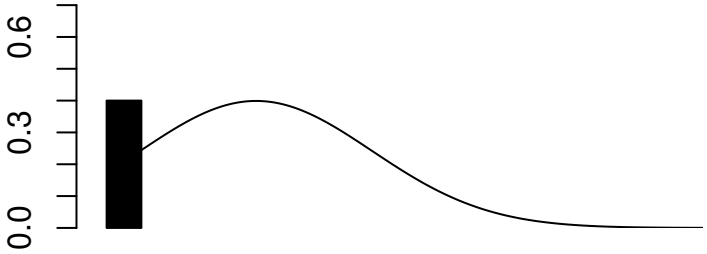
Dados Contagem:



Dados Binomial:



Dados Positivos com Zeros:



Examplos de Modelos

$x_1, \ldots, x_n$ posições com observações

Poisson-log

- $[Y(x_i) \mid S(x_i)]$ é Poisson com densidade

$$f(z; \mu) = \exp(-\mu)\mu^z/z! \quad z = 0, 1, 2, \ldots$$

- ligação: $E[Y(x_i) \mid S(x_i)] = \mu_i = \exp(S(x_i))$

Binomial-logit

- $[Y(x_i) \mid S(x_i)]$ é binomial com densidade

$$f(z; \mu) = \binom{r}{z}(\mu/r)^z(1 - \mu/r)^{r-\mu} \quad z = 0, 1, \ldots, r$$

- ligação: $\mu_i = E[Y(x_i) \mid S(x_i)]$, $S(x_i) = \log(\mu_i/(r - \mu_i))$

Função de Verossimilhança

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

Integral de alta dimensionalidade !!!

2. Inferência para o modelo geoestatístico linear generalizado

- avaliação da verossimilhança envolve integração numérica de multidimensional
- métodos aproximados (ex Breslow and Clayton, 1993) tem acurácia duvidosa
- MCMC é possível embora não rotineira

Esquemas para MCMC

- Ingredientes

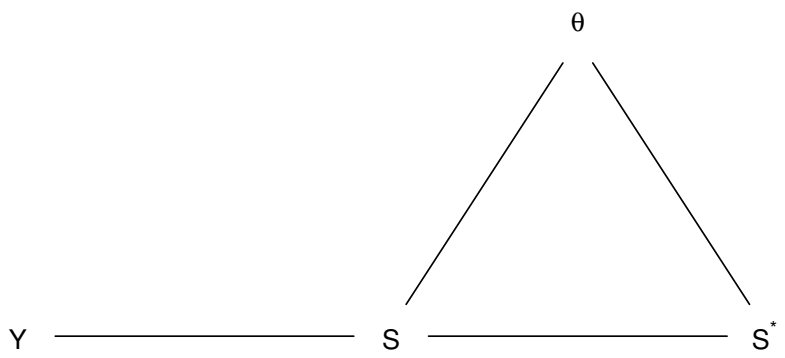
  · Prioris para os parâmetros de regressão β e de covariância θ

  · Dados: $Y = (Y_1, ..., Y_n)$

  · $S = (S(x_1), ..., S(x_n))$

  · $S^* =$ todos outros $S(x)$

- Estrutura de independência condicional



- use resultados das cadeias para contruir declarações à posteriori sobre $[T|Y]$, onde $T = \mathcal{F}(S^*)$

3. Estudo de caso: Ilha Rongelap

- **Ilha Rongelap**
  - aproximadamente a 2500 milhas sudoeste do Hawaii
  - contaminada por testes de armas nucleares em 1950's
  - evacuada em 1985
  - segura para re-assentamento?

- **Problemas estatísticos**
  - delineamento e medidas de campo de ^{137}Cs
  - estimar variação espacial da radiotividade de ^{137}Cs
  - comparação com padrões de segurança

O modelo Poisson

- Medidas básicas são contagens Y_i em intervalos de tempo t_i nas localizações x_i $(i = 1, ..., n)$

- estrutura dos dados sugere o modelo:

  · $S(x) : x \in R^2$ processo estacionário Gaussiano (radioatividade local)

  · $Y_i|\{S(\cdot)\} \sim$ Poisson(μ_i)

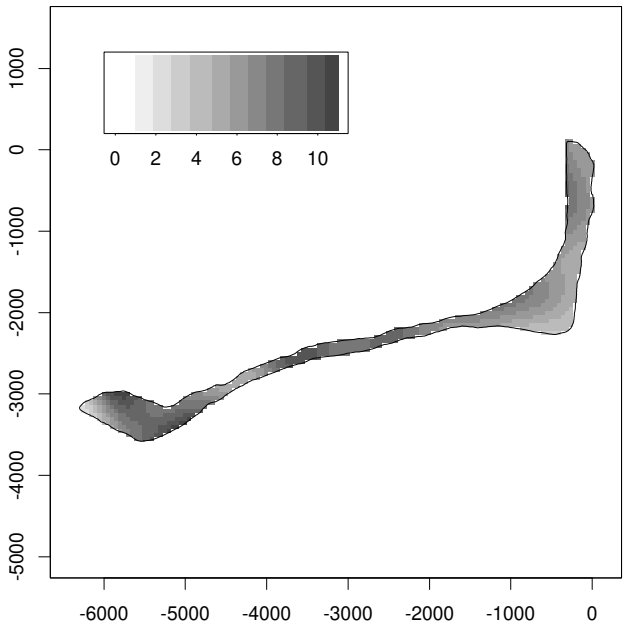
  · $\mu_i = t_i\lambda(x_i) = t_i \exp\{S(x_i)\}$.

- Objetivos:

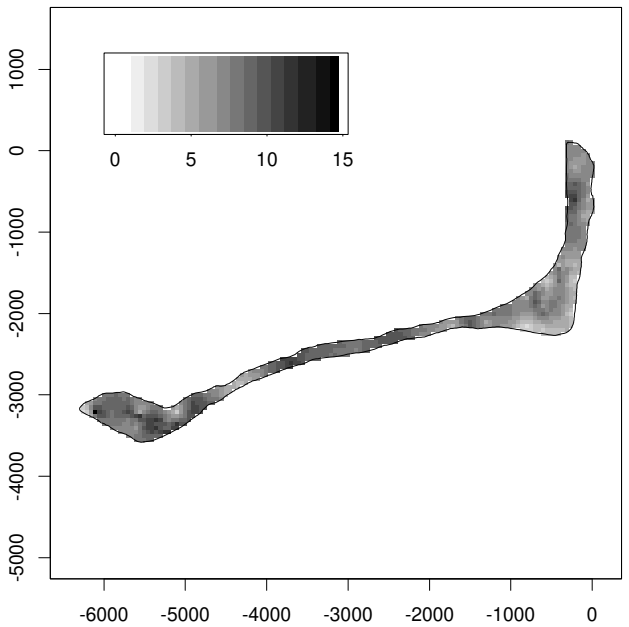
  · predizer $\lambda(x)$ sobre toda ilha

  · max $\lambda(x)$

  · arg(max $\lambda(x)$)

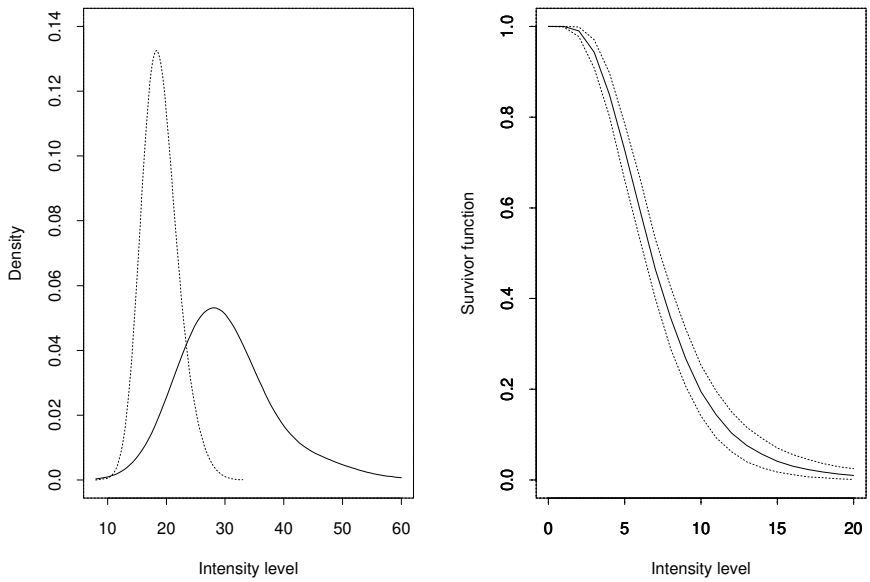


superfície de radiotividade predita utilizando krigagem logarítmica



superfície de radiotividade predita utilizando o modelo log-linear Poisson com processo latentre Gaussiano

Predição Bayesiana de funcionais não lineares da superfície de radiação



The left-hand panel shows the predictive distribution of maximum radioactivity, contrasting the effects of allowing for (solid line) or ignoring (dotted line) parameter uncertainty; the right-hand panel shows 95% pointwise credible intervals for the proportion of the island over which radioactivity exceeds a given threshold.

4. Estudo de caso: Malária em Gambia

- Neste exemplo a variação espacial é de interesse científico secundário.

- O objetivo primário e descrever a dependência entre a prevalência de parasitas de malária e as covariáveis medidas
  - em vilas
  - em indivíduos

- Particular interesse em saber se o índice de vegetação derivado de medidas de satélite pode ser utilizado como preditor da prevalência de malária.

  Isto ajudaria profissionais de saúde a alocar melhor os recursos que são escarsos.

Estrutura dos dados

- 2039 crianças em 65 vilas
- cada uma testada para presença de parasitas de malária no sangue

Covariáveis das crianças

- idade (dias)
- sexo (F/M)
- uso de mosquiteiro (nenhum, não tratado e tratado)

Covariáveis das vilas:

- localização
- indice de vegetação (satélite)
- presença de centro de saúde na vila

Modelo de regressão logística

- $Y_{ij} = 0/1$ presença ou ausência de parasitas de malária na jth criança da ith vila

- $f_{ij} =$ covariável da criança

- $w_i =$ covariável da vila

- $\text{logit}(P(Y_{ij} = 1|S(\cdot))) = f'_{ij}\beta_1 + w_i{}'\beta_2 + S(x_i)$

É razoável assumir infeções condicionalmente independentes na mesma vila?

Caso não, o modelo deve ser extendido para permitir variabilidade extra-binomial nãoespacial

- $U_i \sim \text{N}(0, \nu^2)$

- $\text{logit}P(Y_{ij} = 1|S(\cdot), U) = f'_{ij}\beta_1 + w_i{}'\beta_2 + U_i + S(x_i)$

Análise exploratória

- ajuste modelo logístico padrão sem $S(x)$ e/ou U

- calcule para cada vila:

  $N_i = \sum_{j=1}^{n_i} Y_{ij}$

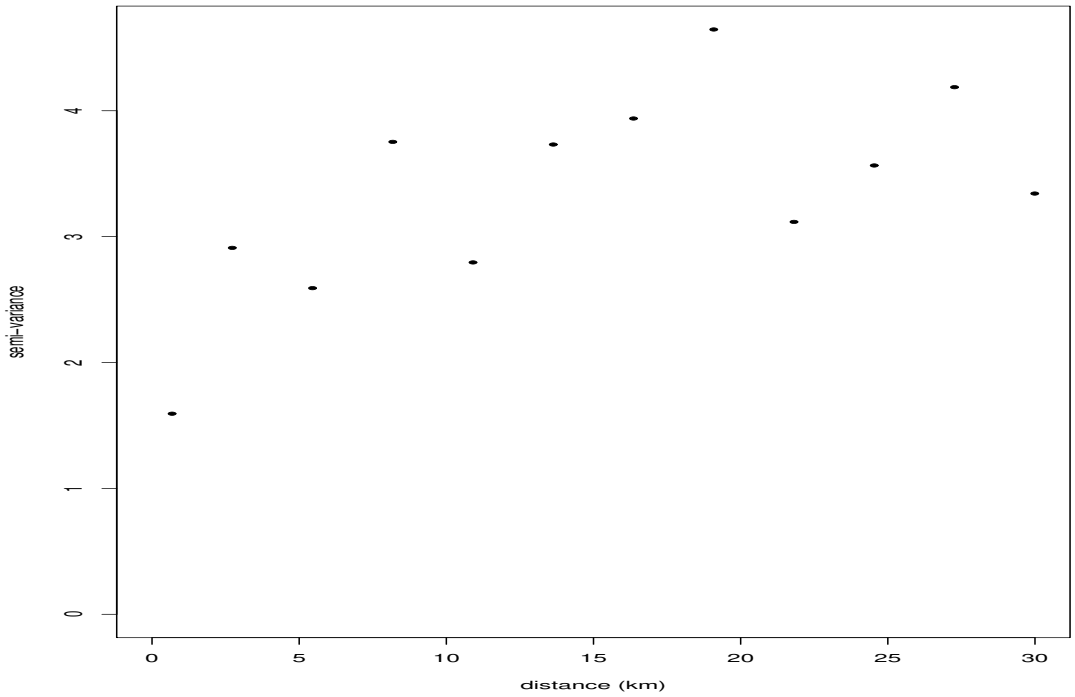
  $\mu_i = \sum_{j=1}^{n_i} \hat{P}_{ij}$

  $\sigma_i^2 = \sum_{j=1}^{n_i} \hat{P}_{ij}(1 - \hat{P}_{ij})$

- resíduos de vila, $r_i = (N_i - \mu_i)/\sigma_i$

- derivar dados r_i

- ajuste de parâmetros de covariância



Variograma do resíduos de vilas

Análise "model-based"

α = intercepto

β_1 = coeficiente para idade

β_2 = coeficiente uso de mosquiteiro

β_3 = coeficiente para mosquiteiro tratado

β_3 = coeficiente para indice de verde

β_4 = coeficiente para presença de centro de saúde

ν^2 = variância do efeito aleatório não espacial U_i

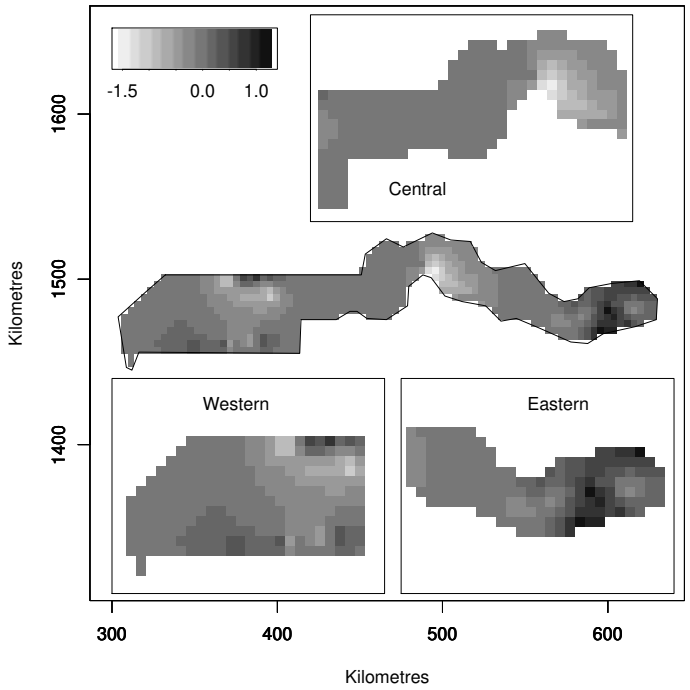
σ^2 = variância do preocesso espacial $S(x)$

ϕ = parâmetro de decaimento da correlação

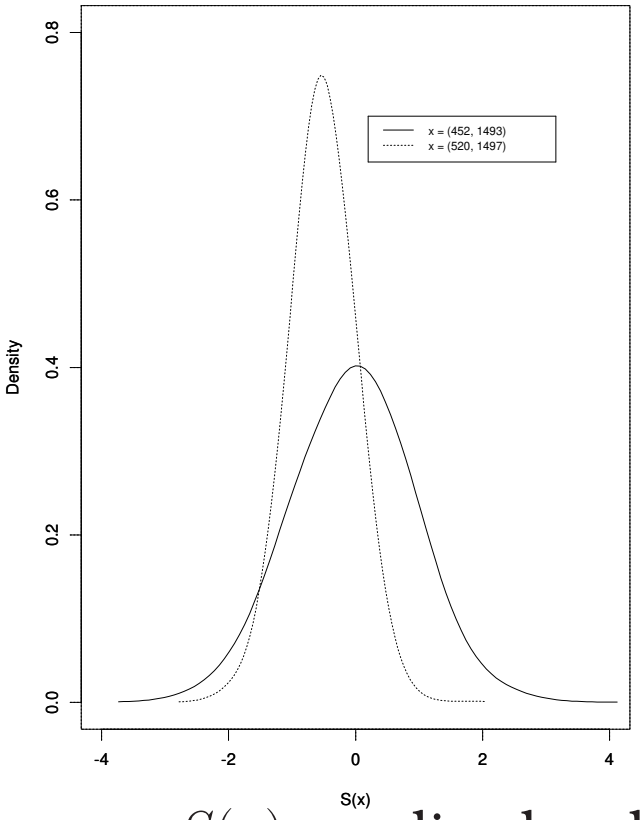
κ = parâmetro de suavidade

Param.	2.5% Qt.	97.5% Qt.	Mean	Median
α	-4.232073	1.114734	-1.664353	-1.696228
β_1	0.000442	0.000918	0.000677	0.000676
β_2	-0.684407	-0.083811	-0.383750	-0.385772
β_3	-0.778149	0.054543	-0.355655	-0.355632
β_4	-0.039706	0.071505	0.018833	0.020079
β_5	-0.791741	0.180737	-0.324738	-0.322760
ν^2	0.000002	0.515847	0.117876	0.018630
σ^2	0.240826	1.662284	0.793031	0.740790
ϕ	1.242164	53.351207	11.653717	7.032258
κ	0.150735	1.955524	0.935064	0.830548

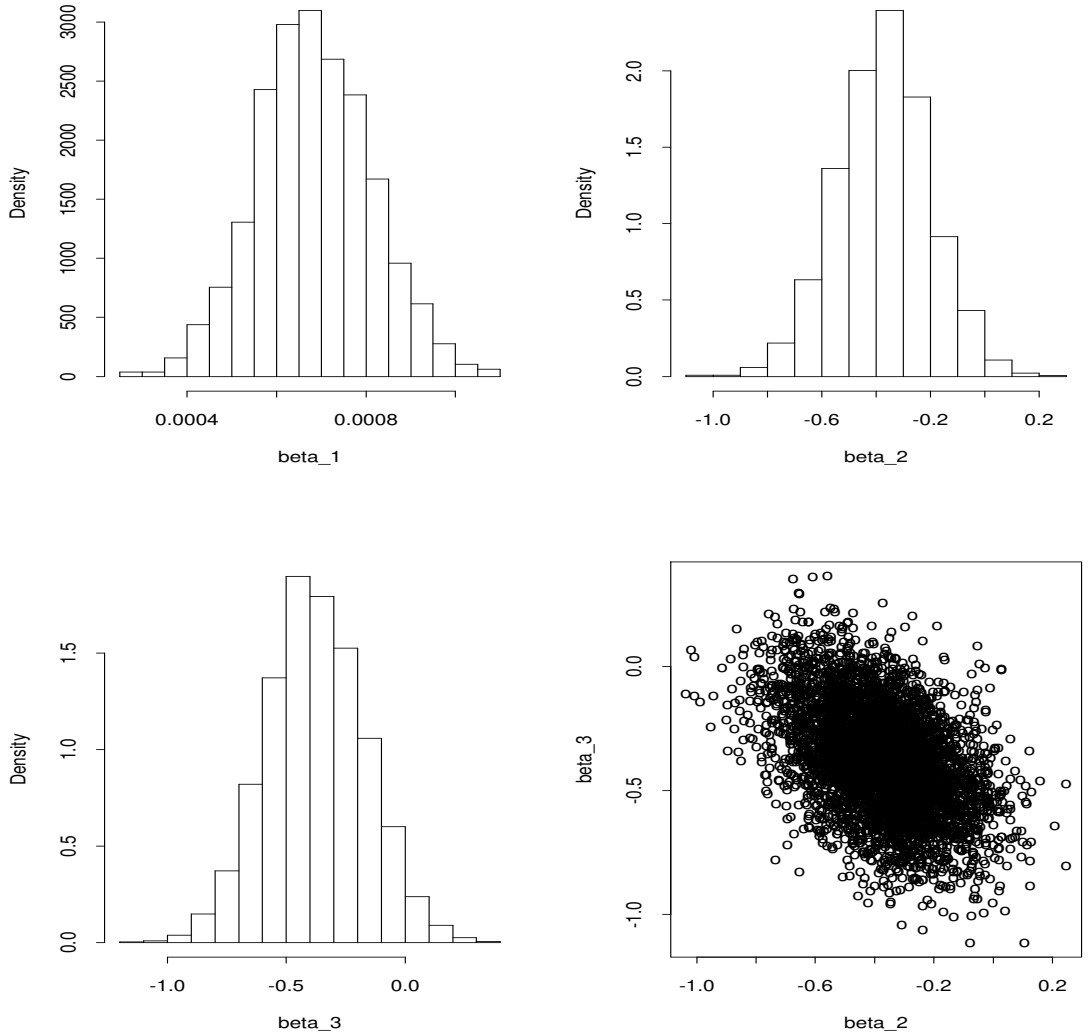
- ν^2 próximo de zero



superfície predita $\hat{S}(x)$ (média à posteriori)



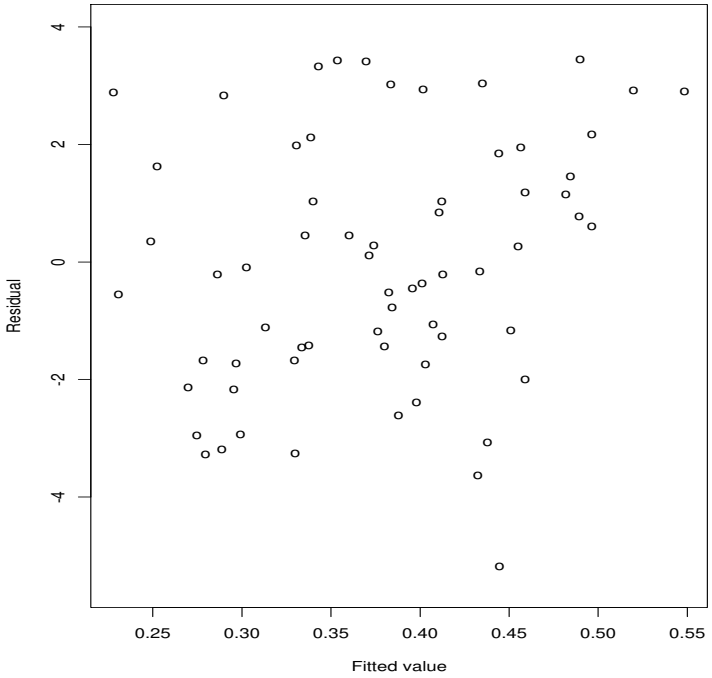
Posterioris para $S(x)$ em dias localizacoes, linha sólida – remota (452, 1493), linha interrompida – central (520, 1497)



posterioris para os parâmetros de regressão

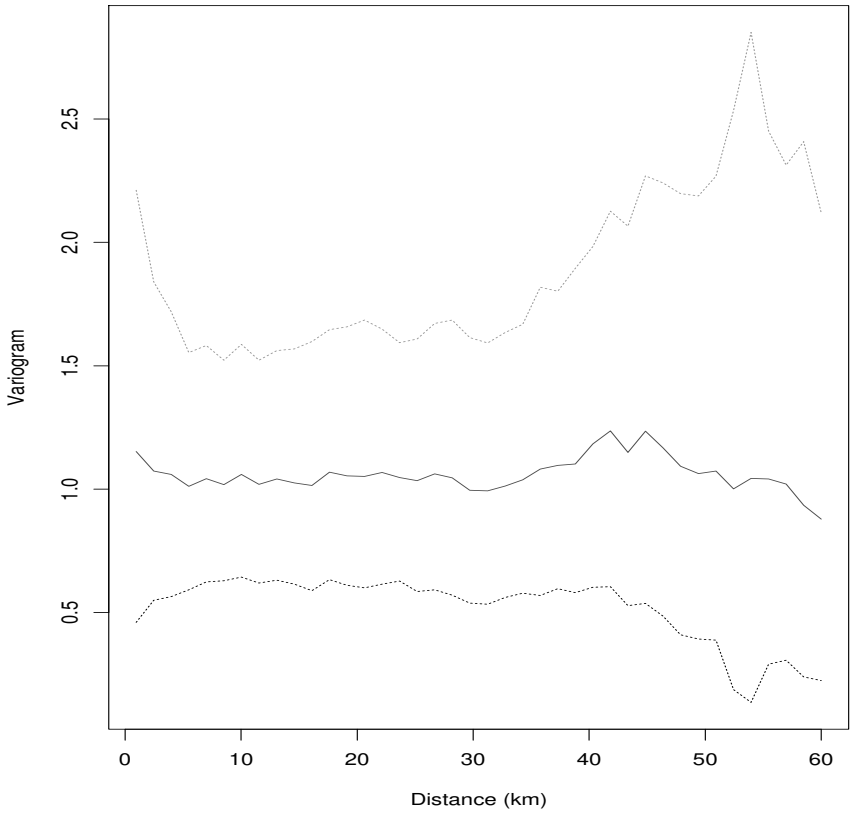
- β_1 = efeito de idade
- β_2 = efeito de mosquiteiro não tratado
- β_3 = efeito adicional de tratamento de mosquiteiro

Qualidade do ajusto do modelo



resíduos de vila *vs* valores ajustados

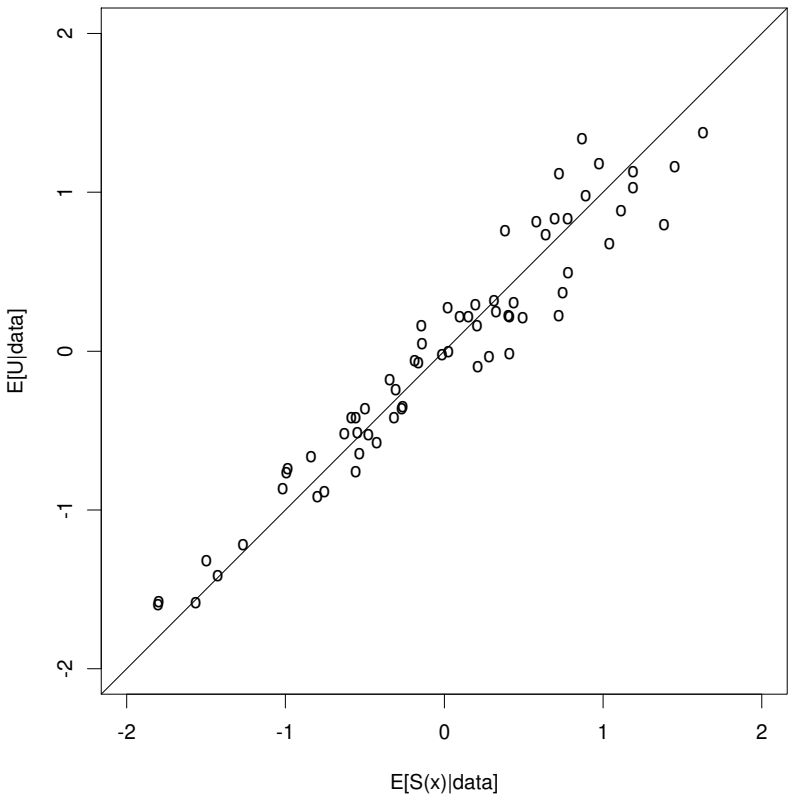
- $r_{ij} = (Y_{ij} - \hat{p}_{ij})/\sqrt{\{\hat{p}_{ij}(1 - \hat{p}_{ij})\}}$
- $r_i = \sum r_{ij}/\sqrt{n_i}$
- checa adequacidade do modelo para p_{ij}



variograma empírico de resíduos padronizados com intervalos de confiança (95%) construídos a partir de simulações do modelo ajustado

- $r_{ij} = (Y_{ij} - \hat{p}^*_{ij})/\sqrt{\{\hat{p}^*_{ij}(1 - \hat{p}^*_{ij})\}}$
- $r_i = \sum r_{ij}/\sqrt{n_i}$
- $\text{logit}(p^*_{ij}) = \hat{\alpha} + z'_{ij}\hat{\beta}\hat{S}(x_i)$
- checa adequacidade do modelo para $S(x)$

O modelo geostatistico é mesmo necessário?



média da posteriori para os efeitos aleatórios $\hat{U}_i$ de um GLMM não espacial contra médias a posteriori de $\hat{S}(x_i)$ nas localizações observadas no modelo geoestatístico

- alta correlação evidencia depêndencia espacial

GEE: uma alternativa para problemas onde a ênfase está nas covariáveis?