

Geoestatística Baseada-em-Modelo

Raquel Menezes

Departamento de Matemática
Universidade do Minho

Novembro de 2023



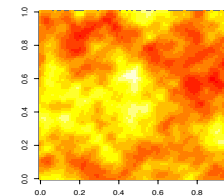
Model-based geostatistics (Diggle et al., 1998)

“...aplicação de princípios gerais de inferência e modelação estatística sobre problemas geoestatísticos, recorrendo a métodos baseados na verosimilhança.”

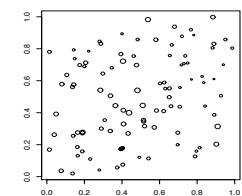
Um modelo **hierárquico de dois níveis**:

- 1 Campo latente $S(\mathbf{x}) : \mathbf{x} \in D \subset \mathbb{R}^2$
- 2 Processo de medição $Y \equiv Y(\mathbf{x}_i)$ (versão com ruído de S)

dados não-observados s



dados observados (y, \mathbf{x})



2 / 11

Um modelo geoestatístico Gaussiano

$$Y(\mathbf{x}_i) = \mu(\mathbf{x}_i) + S(\mathbf{x}_i) + Z_i, \quad i = 1 \dots n$$

- 1 Como modelar a tendência?
 - ▷ $\mu(\mathbf{x}) = E[Y(\mathbf{x})]$
 - ▷ Pode envolver covariáveis geo-referenciadas $X_1(\mathbf{x}), \dots, X_p(\mathbf{x})$
- 2 Como modelar a correlação espacial?
 - ▷ $S(\mathbf{x})$ **campo gaussiano estacionário latente (GF)** de **média zero**
 - ▷ $\text{Var}[S(\mathbf{x})] = \sigma^2$
 - ▷ $\text{Corr}[S(\mathbf{x}), S(\mathbf{x}')] = \rho(h; \phi) \quad \forall \mathbf{x}, \mathbf{x}' \text{ e } h = \|\mathbf{x} - \mathbf{x}'\|$
- 3 O que sabemos sobre as observações?
 - ▷ $Y(\mathbf{x}_i)|S(\mathbf{x}_i) \sim \text{Normal}(\mu(\mathbf{x}_i), \tau^2)$ i.i.d. e $\mu(\mathbf{x}_i) \sim \text{covariáveis} + S(\mathbf{x}_i)$
 - ▷ $Z_i \sim N(0, \tau^2)$ i.i.d. (erro de medição)

3 / 11

Dados Gaussianos – Modelo Geoestatístico Linear

Considere o conjunto de dados (Y_i, \mathbf{x}_i) , $i = 1, \dots, n$

- $S(\mathbf{x})$ campo latente Gaussiano
- $Y_i|S(\mathbf{x}_i) \sim \text{Normal}(\mu(\mathbf{x}_i), \tau^2)$
- $\mu(\mathbf{x}_i) = \beta_0 + \beta_1 X_1(\mathbf{x}_i) + \dots + \beta_p X_p(\mathbf{x}_i) + S(\mathbf{x}_i)$

Denominado na terminologia inglesa por *Mixed-Effects Model*

Efeitos-Mistos = Efeitos Fixos + Efeitos Aleatórios

- ▷ Efeitos Fixos \rightarrow covariáveis observáveis $X_1(\mathbf{x}_i), \dots, X_p(\mathbf{x}_i)$
- ▷ **Efeito Aleatório** \rightarrow efeito espacial não-observável $S(\mathbf{x}_i)$

4 / 11

Dados Não-Gaussianos – Modelo Geoestatístico Linear Generalizado

Considere agora o conjunto de dados (Y_i, \mathbf{x}_i) , onde as observações Y_i têm uma **distribuição diferente da Gaussiana** (por exemplo, Y_i são contagens ou presença/ausência em \mathbf{x}_i).

- $S(\mathbf{x})$ campo latente Gaussiano (**este pressuposto mantém-se**)
 - ▷ $\text{Cov}[S(\mathbf{x}), S(\mathbf{x} + \mathbf{h})] = \sigma^2 \rho(\mathbf{h}; \phi, \theta)$
- $Y_i | S(\mathbf{x}_i) \sim F(\mu(\mathbf{x}_i), \nu)$
 - ▷ $F(\cdot)$ é **outra distribuição** (por exemplo, Poisson ou Binomial)
 - ▷ ν parâmetro adicional (por ex. τ^2) consuante a distribuição vai variar o n° de parametros
- $g(\mu(\mathbf{x}_i)) = \beta_0 + \beta_1 X_1(\mathbf{x}_i) + \dots + \beta_p X_p(\mathbf{x}_i) + S(\mathbf{x}_i)$
 - ▷ $g(\cdot)$ é chamada **função de ligação** (por ex. $\log(\cdot)$ ou $\text{logit}(\cdot)$)

5 / 11

Dificuldades e soluções

Dados não-gaussianos e variável latente não observada
↓
Função de verossimilhança intratável (difícil calcular MLEs¹)

Estimação de parâmetros em Modelos Geoestatísticos Generalizados

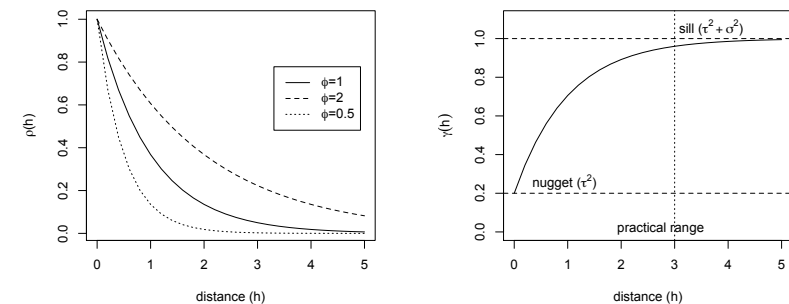
- Máxima Verossimilhança por Monte Carlo (Geyer e Thompson, 1992; Geyer, 1994; Zhang, 2002; Christensen, 2004)
- Verossimilhança Hierárquica (Lee e Nelder, 1996)
- Equações de estimação generalizadas, GEE (Liang e Zeger, 1986)
- Métodos numéricos de integração baseados na aproximação de Laplace

¹Maximum Likelihood Estimates

7 / 11

A função de correlação e o variograma

Processo isotrópico e estacionário de 2ª ordem



Esquerda: A função de **correlação exponencial** para diferentes valores do raio de influência ϕ

Direita: Representação esquemática de um **variograma típico** com seus parâmetros estruturais, sendo $\gamma(h) = \tau^2 + \sigma^2(1 - \rho(h)) = \text{Var}[Y(\mathbf{x}) - Y(\mathbf{x}')] = \sigma^2(1 - \rho(h))$ e $h = \|\mathbf{x} - \mathbf{x}'\|$

6 / 11

Dificuldades e soluções

Adicionalmente

- Modelos geoestatísticos tipicamente associados a matrizes de covariância Σ esparsas
- Quando o número de localizações N é grande, a inversão da matriz Σ pode ser demorada ou numericamente instável
- Problema conhecido na literatura como *big n problem* (Baddeley et al., 2015)

Lindgren, Rue and Lindstrom (2011) propõe abordagem baseada em **equações diferenciais parciais estocásticas** (SPDE), considerando funções de correlação de Matérn.

A ideia é **aproximar o Campo Gaussiano (GF) por um Campo Aleatório de Markov Gaussiano (GMRF)**, uma versão discretizada.

8 / 11

Abordagens alternativas

- Abordagem Bayesiana
 - ▷ Inferência Bayesiana usando algoritmos MCMC
Vantagem: considera diferentes fontes de variabilidade
Desvantagem: peso computacional, problemas de convergência, ...
 - ▷ Inferência Bayesiana recorrendo aos métodos **INLA**² (Rue, Martino, Chopin, 2009) + **SPDE** (Lindgren et al., 2011)
Vantagem: baixo peso computacional
- **Abordagem Frequencista** (considerada nesta UC)
 - ▷ Aproximação, via **SPDE**, do GF por um GMRF's sobre uma grelha regular/irregular
Vantagem: baixo peso computacional e *simplicidade*

²*Integrated Nested Laplace Approximation*, assume existência de modelos Gaussianos latentes
9 / 11

Estatística espacial e o R

- Package **sp** (Bivand, Pebesma e Gomez-Rubio 2008) e package **raster** (Hijmans, 2014) fornecem recursos para armazenar, manipular e visualizar dados espaciais
 - ▷ **sp** – objetos `SpatialPointsDataFrame` e `SpatialPolygonsDataFrame` para armazenar pontos e polígonos, sendo compatíveis com os mais comuns formatos SIG
 - ▷ **raster** – ferramentas semelhantes para dados *bitmap*/matriciais
- Package **geoR** (Diggle e Ribeiro, 2006) oferece ferramentas para inferência baseada em verossimilhança, desde 2000, incluindo as opções:
 - ▷ função de Matérn; inclusão de covariáveis; MLEs; anisotropia geométrica; transformação Box-Cox
- Para inferência Bayesiana, o package **INLA** (Rue et al., 2013) implementa a abordagem INLA para uma ampla variedade de modelos, incluindo modelos de campo aleatório de Markov Gaussiano
- Packages **glmmTMB**, **spaMM** e **spmodel**

spmodel: *spatial statistical modeling and prediction in R*

O package **spmodel** é usado para **inferência** e **predição** em modelos **geoestatísticos** e modelos para **dados agregados por áreas**.

- **Parâmetros estimados** por vários métodos, incluindo otimização baseada em verossimilhança e mínimos quadrados ponderados com base em variogramas.
- Outras funcionalidades de modelação incluem **anisotropia**, **efeitos aleatórios não espaciais** e abordagens para grandes volumes de dados.
- Estatísticas de ajuste do modelo são usadas para resumir, visualizar e **comparar modelos**.
- As previsões em locais não observados são facilmente obtidas.