



## INFERÊNCIA NO MODELO GEOESTATÍSTICO PARA DADOS COMPOSICIONAIS

Ana Beatriz Tozzo Martins<sup>1</sup> Paulo Justiniano Ribeiro Jr<sup>2</sup> Wagner Hugo Bonat<sup>3</sup> Antônio Carlos Andrade Gonçalves<sup>4</sup>



<sup>1</sup>Msc. Eng. Produção e Sistemas, PPGMNE/UFPR-DES/UEM, abtmartins@uem.br;

<sup>2</sup>PhD Estatística, LEG/UFPR, paulojus@leg.ufpr.br;

<sup>3</sup>Estatístico, LEG/UFPR, wbonat@leg.ufpr.br;

<sup>4</sup>Dr. Agronomia, DAG/UEM, acagoncalves@uem.br



### 1. Resumo

O problema que motiva este trabalho é o estudo das frações granulométricas do solo. Considerou-se a análise de dados composicionais para tratar todas as frações conjuntamente, ao invés de tratá-las de forma individual como aparece muitas vezes na literatura, porém levando em consideração a espacialização das observações através da geoestatística. Dessa forma, o trabalho consistiu em propor um modelo e desenvolver algoritmos para se fazer inferência via verossimilhança dos parâmetros desse modelo avaliando, inclusive, diferentes métodos de otimização. Como exemplo, aplicou-se a metodologia proposta à análise de um conjunto de dados reais referentes às porcentagens de areia, silte e argila que formam as frações granulométricas. O trabalho é concluído com a interpretação dos valores obtidos para as estimativas dos parâmetros.

### 2. Introdução

#### 2.1 Dados composicionais

• Ciências da Terra: É comum os dados serem expressos como frações ou porcentagens. Aitchison (1986).

• Exemplo:  
- textura de solos ou granulometria que se refere a proporção de areia, silte e argila do solo;

• Literatura:  
Dados Composicionais: Aitchison (1986);  
Análise Geoestatística de Dados Composicionais: Pawlowsky-Glahn e Olea (2004);  
Inferência Bayesiana de Dados Composicionais Sem Efeito Espacial: Obage (2007);  
Inferência Bayesiana Espacial: Tjelmeland e Lund (2003).

• Composição: Vetor  $Y = (Y_1, Y_2, \dots, Y_B)'$  satisfazendo:

$$Y_1 \geq 0, \dots, Y_B \geq 0;$$

$$Y_1 + Y_2 + \dots + Y_B = 1.$$

$$\text{Espaço Amostral: } \mathbb{S}^B = \{Y \in \mathbb{R}^B; Y_i > 0, i = 1, \dots, B; \sum Y_i = 1\}$$

• Base: Vetor  $W(x)$ ,  $x \in \Omega \subset \mathbb{R}^n$  com componentes medidos na mesma escala e positivos.  
Espaço Amostral:  $\mathbb{R}_+^B = \{W(x) \in \mathbb{R}^B; W_i(x) > 0, i = 1, \dots, B\}$

• Operador fechamento: Base  $\Rightarrow$  Composição

$$C: \mathbb{R}_+^B \rightarrow \mathbb{S}^B$$

$$W(x) \rightarrow C[W(x)] = \frac{W(x)}{\sum_j W_j(x)}, \quad j' \text{ vetor de } 1^s.$$

• Operações que definem uma estrutura de espaço vetorial de dimensão  $B-1$  no simplex: Perturbação e Potência.

• Transformação razão log-aditiva (ALR):

$$\text{ALR: } \mathbb{S}^B \rightarrow \mathbb{R}^{B-1}$$

$$Y(x) \rightarrow \text{ALR}[Y(x)] = \left( \ln \frac{Y_1(x)}{Y_B(x)}, \dots, \ln \frac{Y_{B-1}(x)}{Y_B(x)} \right)'$$

Pawlowsky e Olea (2004).

#### 2.2 Objetivo

Combinar características espaciais e composicionais dos dados referentes à areia, silte e argila através da especificação de um modelo paramétrico composicional espacial e da inferência pela verossimilhança.

### 3. Material e Métodos

• Modelo geoestatístico bivariado composicional proposto:

$$\begin{cases} Y_1(x_i) = \mu_1(x_i) + S_1(x_i) + Z_1(x_i) \\ Y_2(x_i) = \mu_2(x_i) + S_2(x_i) + Z_2(x_i) \end{cases} \quad i, i' = 1, \dots, 76$$

★  $Y_1 = \ln(\text{areia}/\text{argila})$ ,  $Y_2 = \ln(\text{silte}/\text{argila})$ ,

★  $S_j(x) \sim N(0; \sigma_j^2)$ , efeitos espaciais com função de correlação comum  $\rho_U(x; \phi)$ ,  $j = 1, 2$ ,

★  $Z_j(x) \sim N(0; \tau_j^2)$ , efeitos composicionais no modelo.

Este pode ser reescrito como

$$\begin{cases} Y_1(x_i) = \mu_1(x_i) + \sigma_1 U(x_i; \phi) + Z_1(x_i) \\ Y_2(x_i) = \mu_2(x_i) + \sigma_2 U(x_i; \phi) + Z_2(x_i) \end{cases}$$

A inferência sobre  $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \tau_1, \tau_2, \phi, \rho)'$  é feita através da teoria de verossimilhança. Ao considerar  $\Sigma = \sigma_1^2 R + \tau_1^2 I_b = \sigma_1^2 V$ , em que  $R$  é a matriz de covariâncias relacionadas aos efeitos espaciais e  $I_b$  é a matriz bloco diagonal com elementos relacionados às covariâncias entre as composições tem-se a função de log verossimilhança dada por:

$$l(\theta) = -0,5[n \log(2\pi) + 2n \log(\sigma_1) + \log(|V|) + (1/\sigma_1^2) Qe]$$

em que  $Qe = (Y - \mu_Y)' \Sigma^{-1} (Y - \mu_Y)$ ,  $\mu_Y = X\mu$  e  $X$  é a matriz associada às médias das variáveis.

• Estimadores de máxima verossimilhança de  $\sigma_1$  e  $\mu$

$$\hat{\sigma}_1 = \sqrt{Qe/n} \quad e \quad \hat{\mu} = (X'V^{-1}X)^{-1}(X'V^{-1}Y)$$

Com a reparametrização  $\eta = \sigma_2/\sigma_1$ ,  $\nu_1 = \tau_1/\sigma_1$  e  $\nu_2 = \tau_2/\sigma_2$  (DIGGLE e RIBEIRO JR, 2007) obtém-se

$$l(\theta) = -0,5[\log(|V|) + n(\log(2\pi) + \log(Qe) - \log(n) + 1)]$$

em que  $\theta' = (\eta, \nu_2, \nu_1, \phi, \rho)'$  é maximizado numericamente pelos métodos "L-BFGS-B", "Nelder-Mead", "Gradiente Conjugado" e "BFGS", implementados no R (R development Core Team, 2008).

• Os erros padrão na escala original dos parâmetros são obtidos através do método delta em que

$$\text{Var}_{\theta}(g(\theta)) \geq \left( \frac{\partial g(\theta)}{\partial \theta} \right)' I_E(\theta)^{-1} \left( \frac{\partial g(\theta)}{\partial \theta} \right),$$

em que  $\partial g(\theta)/\partial \theta$  é matriz diagonal com elementos  $\partial l(\theta)/\partial \theta_i^*$ ,  $i^* = 1, \dots, 5$  e  $I_E(\theta)$  é a matriz informação esperada de Fisher para  $\theta$  baseada nos dados  $Y$  que é substituída pela matriz hessiana obtida no processo de maximização.

• Dados de 152 amostras de solo com percentuais de areia, silte e argila, provenientes de Gonçalves(1997).

### 4. Resultados

Tabela 1: Estimativas dos parâmetros de acordo com os métodos de otimização e intervalo de confiança para o método L-BFGS-B.

Parâmetro	Iniciais	Nelder Mead	Gradiente Conjugado	BFGS	L-BFGS-B	Erro Padrão	Límite Inferior	Límite Superior
$\mu_1$	0,5625	-0,7688	-0,7599	0,7686	-0,7599	0,4403	-1,2002	-0,3196
$\mu_2$	0,7436	-0,7963	-0,7941	-0,7963	-0,7941	0,2286	-0,0227	-0,5655
$\sigma_1$	0,1007	0,4581	0,4505	0,4580	0,4505	0,1208	0,3297	0,5713
$\sigma_2$	0,0399	0,1171	0,1153	0,1170	0,1153	0,0515	0,0638	0,1668
$\tau_1$	0,1007	0,2887	0,2870	0,2886	0,2870	0,0451	0,2419	0,3321
$\tau_2$	0,0399	0,2672	0,2670	0,2671	0,2670	0,0314	0,2356	0,2984
$\phi$	66,9117	71,0890	66,9113	71,0057	66,9085	42,5565	24,3520	109,4649
$\rho$	0,8219	0,9522	0,9544	0,9522	0,9544	0,0680	0,8864	1,0225
log.vero	42,5541	-0,8273	-0,8292	-0,8273				

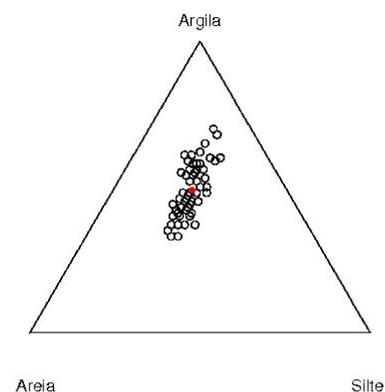


Figura 1: Diagrama ternário das porcentagens de areia, silte e argila.

### 5. Conclusão

A análise composicional revela que as amostras de solo apresentam maior conteúdo de argila do que de areia o que permite a classificação como solo argiloso. Os métodos de otimização considerados convergiram e as estimativas pelos quatro métodos são coincidentes com precisão de uma casa decimal com exceção do parâmetro  $\phi$  cujos valores são próximos. A análise do modelo indica que existe dependência espacial quanto a  $Y_1$  devido à razão  $(\sigma_1/\tau_1) > 1$ , mas não com relação a  $Y_2$ , onde  $(\sigma_2/\tau_2) < 1$ . Próximos passos incluem propostas para modelagem conjunta espacial com implementações de predição bayesianas permitindo obter incertezas associadas às classificações. A metodologia deve ser testada para a verificação da aplicabilidade e generalidade da proposta.

### Referências Bibliográficas

- AITCHISON, J. The statistical analysis of compositional data. New Jersey: The Blackburn Press, 1986, 416 p.
- DIGGLE, P. J.; RIBEIRO JR, P. J. Model-based geostatistics, USA: Springer Series in Statistics, 2007, 228 p.
- GONÇALVES, A. C. A. Variabilidade espacial de propriedades físicas do solo para fins de manejo da irrigação. 1997. 119p. Tese (Doutorado em Agronomia) - Escola Superior de Agricultura "Luiz de Queiroz". Universidade de São Paulo, Piracicaba.
- OBAGE, S. C. Uma análise bayesiana para dados composicionais. 2007. 69p. Dissertação (Mestrado em Estatística) - Universidade Federal de São Carlos, São Carlos.
- PAWLOWSKY-GLAHN, V.; OLEA, R. A. Geostatistical analysis of compositional data. New York: Oxford University Press, Inc., 2004.
- TJELMELAND, H.; LUND, K. V. Bayesian modelling of spatial compositional data. Journal of Applied Statistics, v.30, n.1, p.87-100, 2003.

### Agradecimento

Esse trabalho foi parcialmente financiado pela FINEP projeto CT-INFRA/UFPR.