

MÉTODOS GEOESTATÍSTICOS BASEADOS EM MODELOS, APLICADOS A DADOS COMPOSICIONAIS NA CLASSIFICAÇÃO DE SOLOS

Ana Beatriz Tozzo Martins

Orientador: Prof. PhD Paulo Justiniano Ribeiro Junior.

DES - UEM / PPGMNE - UFPR

6 de novembro de 2008

Motivação

- Ciências da Terra: dados expressos como frações ou porcentagens.
- Exemplos:
 - textura de solos;
 - composição química de uma rocha;
 - estruturas de dados resultantes de algoritmos de classificação.

Walvoort, D. J. J. e Gruijter, J.J. (2001).

Introdução

- Dados de solo: areia, silte e argila.
- Classificação do solo de acordo com a granulometria (NBR 6502/95 da ABNT):

Classificação	Diâmetro dos Grãos (mm)
Argila	(0; 0,002]
Silte	(0,002; 0,06]
Areia	(2,0; 0,06]
Pedregulho	(60,0; 2,0]

Introdução

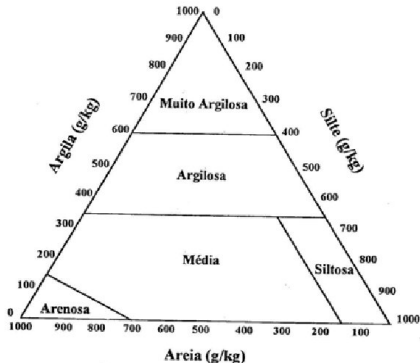


Figura: Diagrama triangular simplificado, utilizado pela EMBRAPA, para a classificação textural do solo.

Objetivo Geral

- Modelos e suas características:
 - são espaciais;
 - são multivariados;
- Desafio: compatibilizar estas estruturas de covariância;
- Aplicação: Representar o resultado através de mapa de classificação de solo.

Objetivos Específicos

- Investigar formas alternativas à proposta de Pawlowsky-Glahn e Olea (2004);
- Construir um modelo em que as dependências espacial e entre variáveis sejam consideradas na obtenção de uma função de covariância válida;
- Adequar métodos bayesianos aos parâmetros do modelo;
- Elaborar mapas temáticos de modelos composicionais de solo em estudo de caso.

Revisão da Literatura - Geoestatística

- Descreve variáveis com variação contínua no espaço;
- Usada para interpolar variável observada em alguns pontos em vários locais do espaço.

Revisão da Literatura - Modelo Geoestatístico Gaussiano

- Dados y_1, y_2, \dots, y_n nas localizações $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ em uma região $A \subset \mathbb{R}^2$ relacionados ao sinal $S(\underline{x})$; $\underline{x} \in \mathbb{R}^2$;
- y_i é uma realização de $Y_i = Y(\underline{x}_i)$ onde Y_i é uma versão ruído de $S(\underline{x}_i)$;
- O delineamento amostral será tal que $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ sejam fixos ou estocasticamente independentes de $Y(\underline{x}_1), Y(\underline{x}_2), \dots, Y(\underline{x}_n)$;
- Modelo geoestatístico:

$$[S(\underline{x}), Y(\underline{x})] = [S(\underline{x})][Y(\underline{x})|S(\underline{x})];$$

A distribuição do delineamento amostral $\{\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n\}$ é assumido independente de $S(\underline{x})$ e $Y(\underline{x})$.

Revisão da Literatura - Dados Composicionais

- Dados expressos em proporções analisados considerando a relação entre proporções;
- Aitchison (1986):
 - b. **Composição:** Vetor $\underline{Y} = (Y_1, Y_2, \dots, Y_B)'$ satisfazendo:
 - $Y_1 \geq 0, \dots, Y_B \geq 0$;
 - $Y_1 + Y_2 + \dots + Y_B = 1$.

- **Espaço Amostral:**

$$\mathbb{S}^B = \{\underline{Y}(\underline{x}) \in \mathbb{R}^B; Y_i(\underline{x}) > 0, i = 1, \dots, B; \underline{j}'\underline{Y}(\underline{x}) = 1\}$$

Revisão da Literatura - Diagrama Ternário: Malha Quadrada

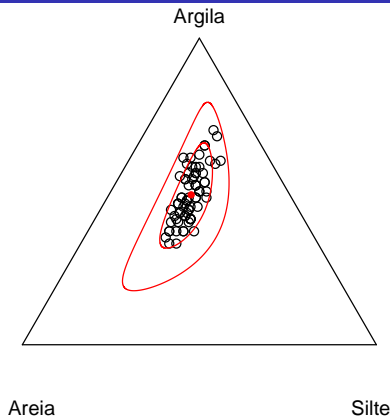


Figura: Diagrama ternário para dados da Malha Quadrada, incluindo o centro da distribuição e regiões 2 e 4σ

Revisão da Literatura - Transformação

Transformação razão log-aditiva (ALR):

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

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

Transformação logística generalizada aditiva (AGL):

$$\text{AGL} : \mathbb{R}^{B-1} \longrightarrow \mathbb{S}^B$$

$$\text{ALR}[\underline{Y}(\underline{x})] \longrightarrow \text{AGL}\{\text{ALR}[\underline{Y}(\underline{x})]\} = \underline{Y}(\underline{x}) = \frac{\exp[\underline{Y}(\underline{x})]}{\sum_j \exp[\underline{Y}(\underline{x})]}$$

Pawlowsky et al. (1995), Pawlowsky e Olea (2004).

Material e Método

Material: Dados de areia, silte e argila.

- Lago Ártico extraído de Aitchison (1986): $n = 39$
- Gener - ESALQ/USP: $n = 412$ Areia Muito Grossa, Areia Média, Areia Fina e Areia Muito Fina, além de Altimetria do terreno e Superfície,
- Bassoi - ESALQ/USP: $n = 250$ Cota, Areia Grossa, Silte, Argila, pH_{Água}, pH_{KCl}, Ca, Mg, K, Al, H, C, N, CTC, S, V, M, NC, CEC, CN.
- Gonçalves (1997) - ESALQ/USP-UEM: $n = 412$

Resultados Preliminares

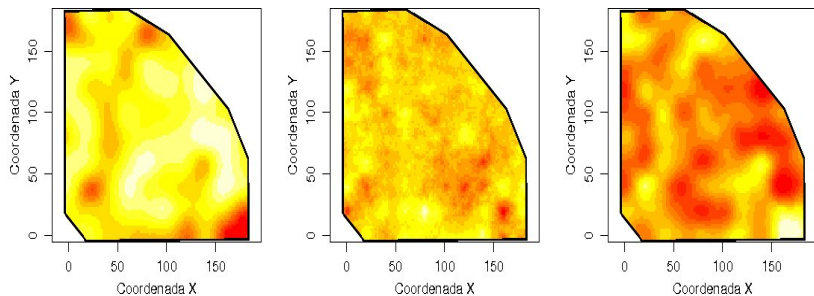


Figura: Mapas da porcentagem de areia (à esquerda), silte (centro) e argila (à direita) obtidos por análise univariada.

Material e Método - Método

- **Geoestatística:** Diggle, Tawn e Moyeed (1998), Banerjee, Carlin e Gelfand (2004), Schmidt e Sansó (2006), e Diggle e Ribeiro Jr.(2007);
- **Dados Composicionais:** Aitchison(1986);
- **Análise Geoestatística de Dados Composicionais:** Pawlowsky-Glahn e Olea(2004);
- **Análise Bayesiana de Dados Composicionais:** Obage (2007).

Material e Método - Método

- Integração das metodologias:
 - Imposição de restrição no espaço paramétrico;
 - Via priori utilizando-se inferência bayesiana.
- Como:
 - Geoestatística e teoria de dados composicionais;
 - Teoria de verossimilhança e inferência bayesiana;
 - Formas de implementar as restrições induzidas no espaço paramétrico.
- Performance das metodologias verificadas na análise de dados reais que motivam o trabalho.

Material e Método - Método

Modelos Iniciais - Independência entre as localizações:

$$1. \begin{cases} Y_{i1} &= \mu_1(\underline{x}_i) + Z_1(\underline{x}_i) \\ Y_{i'2} &= \mu_2(\underline{x}_{i'}) + Z_2(\underline{x}_{i'}) \end{cases}$$

com $i, i' = 1, \dots, n_j$, $j = 1, 2$ e $Z_i \sim N(0; \tau_i^2)$.

$$f(\underline{Y}) = \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi|\Sigma|}} \exp \left\{ -\frac{1}{2} (\underline{Y} - \underline{\mu}_{\underline{Y}})' \Sigma^{-1} (\underline{Y} - \underline{\mu}_{\underline{Y}}) \right\} \right]$$

- $\hat{\underline{\mu}} = \left(\sum_{i=1}^n X' \Sigma^{-1} X \right)' \sum_{i=1}^n (X' \Sigma^{-1} \underline{Y})$
- $\hat{\tau}_1^2 = \frac{\sum_{i=1}^n (\underline{Y} - \underline{\mu}_{\underline{Y}}) G^{-1} (\underline{Y} - \underline{\mu}_{\underline{Y}})}{n}$

Material e Método - Método

Modelos Iniciais - Dependência entre as localizações:

$$2. \begin{cases} Y_{i1} &= \mu_1(\underline{x}_i) + S_1(\underline{x}_i) + Z_1(\underline{x}_i) \\ Y_{i'2} &= \mu_2(\underline{x}_{i'}) + S_2(\underline{x}_{i'}) + Z_2(\underline{x}_{i'}) \end{cases}$$

$$\begin{cases} Y_{i1} &= \mu_1(\underline{x}_i) + \sigma_1 U(\underline{x}_i; \phi) + Z_1(\underline{x}_i) \\ Y_{i'2} &= \mu_2(\underline{x}_{i'}) + \sigma_2 U(\underline{x}_{i'}; \phi) + Z_2(\underline{x}_{i'}) \end{cases}$$

- $\hat{\underline{\beta}} = (\mathbf{D}'\mathbf{V}^{-1}\mathbf{D})^{-1}\mathbf{D}'\mathbf{V}^{-1}\underline{Y}$
- $\hat{\sigma}_1^2 = n^{-1} [(\underline{Y} - \mathbf{D}\hat{\underline{\beta}})'\mathbf{V}^{-1}(\underline{Y} - \mathbf{D}\hat{\underline{\beta}})]$

Resultados Esperados

- Espera-se que a imposição no modelo bivariado de uma restrição induzida pelo modelo de dados composicionais torne o modelo bivariado compatível com a estrutura de dados correlacionados;
- Avaliar o modelo proposto em comparação com o modelo considerando as variáveis separadamente;
- Desenvolver programas computacionais para a implementação e análise do modelo proposto;
- Aplicar a metodologia em dados reais que motivam este trabalho, por exemplo, na construção de mapa temático de classificação espacial do solo segundo concentrações de areia, silte e argila.

Resultados Preliminares

Parâmetros Estimados pelo modelo 1:

$$\hat{\mu} = \begin{bmatrix} -0.5625 \\ -0.7436 \end{bmatrix} \quad \hat{\Sigma} = 0.1986 \begin{bmatrix} 1 & 0.5180 \\ 0.5180 & 0.3973 \end{bmatrix}$$

$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\rho}$
0.4486	0.2827	0.8219

Bibliografia

- AITCHISON, J. **The statistical analysis of compositional data**. New Jersey: The Blackburn Press, 1986.
- BANERJEE, S.; CARLIN B. P. e GELFAND , A. E. **Hierachical modelling and analysis for spatial data**, Boca Raton: Chapman and Hall, 2004.
- DIGGLE, P. J. e RIBEIRO JR, P. J. **Model-based geostatistics**, USA: Springer Series in Statistics, 2007.
- DIGGLE, P. J.;TAWN, J. A.; MOYEED, R. A. Model-based geostatistics.**Applied Statistics**,v. 47, n.3, p. 299-350, 1998. Disponível <<http://em.citeseer.ist.psu.edu/diggle98modelbased.html>>, Acesso em:10/01/08.
- GONÇALVES, A. C. A. **Variabilidade espacial de propriedades físicas do solo para fins de manejo da irrigação**. 99 f. Tese (Tese de Doutorado em Agronomia) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1997.

Bibliografia

- OBAGE, S. C. **Uma análise bayesiana para dados composicionais**. 69 p. Dissertação (Mestrado em Estatística), Centro de Ciências Exatas e de Tecnologia, Universidade Federal de São Carlos, 2007.
- PAWLOWSKY-GLAHN, V.; OLEA, R. A. **Geostatistical analysis of compositional data**. New York: Oxford University Press, Inc., 2004.
- PAWLOWSKY-GLAHN, V.; OLEA, R. A.; DAVIS, J. C. Estimation of regionalized compositions: a comparison of three methods. **Mathematical Geology**, v.27, n.1, p. 105-127. 1995.
- SCHMIDT, A. M.; SANSÓ, B. Modelagem bayesiana da estrutura de covariância de processos espaciais e espaço temporais. *17 SINAPE e ABE-Associação Brasileira de Estatística*. Caxambu: Associação Brasileira de Estatística, 2006, **Minicurso**.
- WALVOORT, D. J. J.; GRUIJTER, J. J. Compositional kriging: A spatial interpolation method for compositional data. **Mathematical Geology**, v.33, n.8, p. 951-966, nov 2001.

OBRIGADA PELA ATENÇÃO!

Agradecimentos

- Orientador: Ph.D. Paulo Justiniano Ribeiro Junior
- PPGMNE - Programa de Pós-Graduação em Métodos Numéricos em Engenharia
- UEM/DES - Universidade Estadual de Maringá/ Dpto. de Estatística
- CNPQ - Conselho Nacional de Desenvolvimento Científico e Tecnológico
- LEG - Laboratório de Estatística e Geoinformação/UFPR