

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

Ana Beatriz Tozzo Martins

Orientador: Prof. PhD Paulo Justiniano Ribeiro Junior.

PPGMNE - UFPR

24 de setembro de 2008

Roteiro de Apresentação

1. Introdução
2. Objetivo
3. Revisão da literatura
 - 3.1 Geoestatística
 - 3.2 Dados Composicionais
4. Material e Métodos
5. Resultados Esperados
6. Cronograma.

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.
- Importância:
 - combate à erosão;
 - cultura a ser plantada;
 - preservação do meio ambiente;
 - manejo e uso do solo;
- 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

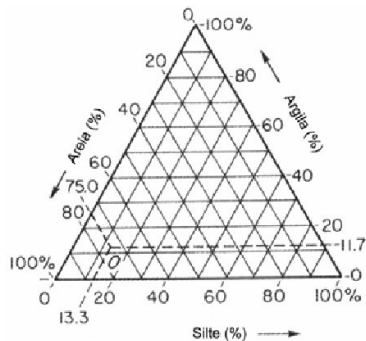


Figura: Triângulo de Feret.

Introdução

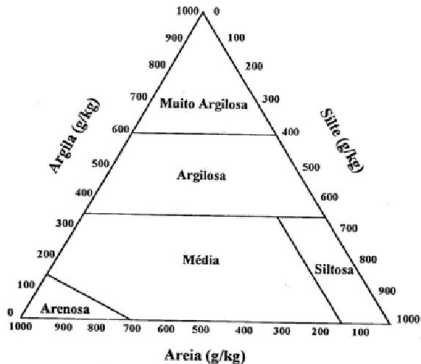


Figura: Diagrama triangular simplificado, utilizado pela EMBRAPA, para a classificação textural do solo.

Tratamento Estatístico

- **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).

Objetivo Geral

Obter um modelo geoestatístico para dados composicionais de solo que sejam espacializados e multivariados e que compatibilize a estrutura de covariância induzida pelo modelo multivariado com a estrutura de covariância induzida pelo modelo de dados composicionais representando o resultado através de mapas 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;
- Derivar métodos de inferência baseados em verossimilhança para a estimação dos parâmetros desconhecidos do modelo;
- Adequar métodos bayesianos aos parâmetros do modelo;
- Desenvolver recursos computacionais livres para análise de dados composicionais;
- Elaborar mapas temáticos de modelos composicionais de solo em estudo de caso.

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 - Modelo Geoestatístico Gaussiano

- $S(\underline{x}) \sim N(\mu; \sigma^2)$ - processo gaussiano estacionário;
- Função de correlação: $\rho(u_{ij}) = \text{Corr}[S(\underline{x}_i), S(\underline{x}_j)]$,
$$u_{ij} = \|\underline{x}_i - \underline{x}_j\|, i, j = 1, \dots, n$$

Diggle, Ribeiro Jr e Christensen (2003)

Revisão da Literatura - Modelo Geoestatístico Gaussiano

- $Y_i = S(\underline{x}_i) + Z_i \quad i = 1, \dots, n$
 $Z_i \sim N(0; \tau^2)$ erros aleatórios independentes de $S(\underline{x})$;

- $\underline{Y} \sim N_n(\mu \underline{1}; \sigma^2 \mathbf{R} + \tau^2 \mathbf{I})$

$\underline{1}$ vetor com n elementos iguais a 1,

\mathbf{R} matriz de ordem $n \times n$ com elementos $\rho(u_{ij})$ e

\mathbf{I} matriz identidade de ordem $n \times n$.

- Reparametrização $\nu^2 = \frac{\tau^2}{\sigma^2} \Rightarrow \underline{Y} \sim N_n(\mu \underline{1}; \sigma^2 \mathbf{V})$

Diggle, Ribeiro Jr (2007)

Revisão da Literatura - Componentes do Modelo

- **Estacionariedade:** distribuição espacial de $S(\underline{x})$ invariante sob translação das coordenadas, média e variância constantes em A e não dependem da localização \underline{x} .
- **Isotropia:** covariância depende somente da distância entre dois pontos e não da direção nos quais estão separados.
- $\rho(u) \Rightarrow$ suavidade do processo $S(\underline{x}) \Rightarrow$ grau de diferenciabilidade.

Revisão da Literatura - Componentes do Modelo

- Função de correlação de Matérn:

$$\rho(u, k, \phi) = \frac{1}{2^{k-1}\Gamma(k)} \left(\frac{u}{\phi}\right)^k K_k\left(\frac{u}{\phi}\right)$$

- $K_k(\cdot)$: função de Bessel modificada de ordem k ;
- $\phi > 0$ parâmetro de escala associado ao alcance;
- $k > 0$ medida da diferenciabilidade do processo $S(\underline{x})$
 $\{0, 5; 1, 5; 2, 5\}$;
- Amplitude prática: distância u_0 no qual $\rho(u_0) = \alpha$ e
 $\{3\phi; 4, 75\phi; 5, 92\phi\}$;

Revisão da Literatura - Componentes do Modelo

- Função de correlação de Matérn:
 - $$\begin{cases} k = 0,5 \Rightarrow \rho(u) = \exp(-u/\phi) \\ k \rightarrow \infty \Rightarrow \rho(u) = \exp\{-(u/\phi)^2\} . \end{cases}$$
- Não estacionariedade na média: $\mu(\underline{x}) = \beta_0 + \sum_{j=1}^p \beta_j d_j(\underline{x})$
 $d_j(\underline{x})$ variáveis explanatórias espaciais dependentes ou não das coordenadas;
- Efeito Pepita: $\tau^2 = \text{Var}(Z_i) = \text{EM} + \text{VME}$
EM : Erro de medida
VME: Variação de pequena escala

Revisão da Literatura - Estimação de parâmetros do Modelo

- Tendência: Método dos Mínimos Quadrados

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

- Método da Máxima Verossimilhança:

- $\underline{Y} \sim N_n(\mathbf{D}\underline{\beta}; \sigma^2\mathbf{R} + \tau^2\mathbf{I})$

\mathbf{D} matriz $n \times p$ de covariáveis;

$\underline{\beta}$ vetor de parâmetros da regressão;

\mathbf{R} depende de ϕ e k

Revisão da Literatura - Estimação de parâmetros do Modelo

- Método da Máxima Verossimilhança

Função de log-verossimilhança

$$l(\underline{\beta}, \sigma^2) = -\frac{1}{2} \{n \log(2\pi) + \log(|\sigma^2 \mathbf{V}|) + \frac{1}{\sigma^2} [\underline{Y}' \mathbf{V}^{-1} \underline{Y} - 2 \underline{Y}' \mathbf{V}^{-1} \mathbf{D} \underline{\beta} + \underline{\beta}' (\mathbf{D}' \mathbf{V}^{-1} \mathbf{D}) \underline{\beta}]\}$$

Revisão da Literatura - Estimação de parâmetros do Modelo

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

$$l_0(\nu^2, \phi, k) = -\frac{1}{2} \{ n \log(2\pi) + n \log(\hat{\sigma}^2) + \log|\mathbf{V}| + n \}$$

Revisão da Literatura - Predição Linear Espacial

- $\underline{S} \sim N_n(\mu \underline{1}; \sigma^2 \mathbf{R}); \underline{Y} \sim N_n(\mu \underline{1}; \sigma^2 \mathbf{V});$
- Preditor pontual de T : $\hat{T} = E(T|\underline{Y});$
- $T = S(\underline{x}) \Rightarrow [S(\underline{x}), \underline{Y}]$
 - $\hat{T} = \mu + \underline{r}'\mathbf{V}^{-1}(\underline{Y} - \mu \underline{1});$
 - $Var[T|\underline{Y}] = \sigma^2 + (1 - \underline{r}'\mathbf{V}^{-1}\underline{r}).$
- $\hat{S}(\underline{x}_0) = \{1 - \sum_{i=1}^n a_i(\underline{x}_0)\}\mu + \sum_{i=1}^n a_i(\underline{x}_0)Y_i$

Revisão da Literatura - Krigagem

- Krigagem ordinária:

- $\hat{T} = \hat{S}(\underline{x}) = \sum_{i=1}^n a_i(\underline{x}) Y_i; \quad \sum_{i=1}^n a_i(\underline{x}) = 1;$

- $\hat{\mu} = (\underline{1}' \mathbf{V}^{-1} \underline{1})^{-1} \underline{1}' \mathbf{V}^{-1} \underline{Y};$

- $\hat{T} = (\underline{1}' \mathbf{V}^{-1} \underline{1})^{-1} \underline{1}' \mathbf{V}^{-1} \underline{Y} + \underline{r}' \mathbf{V}^{-1} [\underline{Y} - (\underline{1}' \mathbf{V}^{-1} \underline{1})^{-1} \underline{1}' \mathbf{V}^{-1} \underline{Y}].$

Revisão da Literatura - Inferência Bayesiana para Predição Espacial

- $\underline{\theta} = [\underline{\beta}, \sigma^2, \phi, \tau^2]$ e $p(\underline{Y} = \underline{y} | \underline{\theta})$;
- $\underline{Y} \sim NM(\mathbf{D}\underline{\beta}; \sigma^2\mathbf{R} + \tau^2\mathbf{I})$;
- Função de Verossimilhança:

$$L(\underline{\theta} | \underline{Y}) \propto |\sigma^2\mathbf{R} + \tau^2\mathbf{I}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2} (\underline{Y} - \mathbf{D}\underline{\beta})' (\sigma^2\mathbf{R} + \tau^2\mathbf{I})^{-1} (\underline{Y} - \mathbf{D}\underline{\beta})\right\}.$$

Revisão da Literatura - Inferência Bayesiana para Predição Espacial

- Teorema de Bayes: $p(\underline{\theta}|\underline{Y}) \propto p(\underline{\theta}) p(\underline{Y}|\underline{\theta})$
- Distribuição *posteriori*:

$$p(\underline{\beta}, \sigma^2, \phi, \tau^2 | \underline{Y}) \propto p(\underline{\beta}, \sigma^2, \phi, \tau^2) |\sigma^2 \mathbf{R} + \tau^2 \mathbf{I}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\underline{Y} - \mathbf{D}\underline{\beta})' (\sigma^2 \mathbf{R} + \tau^2 \mathbf{I})^{-1} (\underline{Y} - \mathbf{D}\underline{\beta}) \right\}$$

- Distribuição preditiva dos parâmetros do modelo $p(\underline{\theta}|\underline{Y})$:

$$p(\underline{Y}_0, \underline{Y}) = \int p(\underline{Y}_0 | \underline{Y}, \underline{\theta}) p(\underline{\theta} | \underline{Y}) d\underline{\theta}$$

Revisão da Literatura - Matriz de covariância, de Correlação e Variograma Cruzado

$\underline{Y}(\underline{x})$ processo estacionário,

$$\Gamma(\underline{x}, \underline{x}') = \begin{bmatrix} \sigma_1^2 & \text{Cov}[Y_1(\underline{x}), Y_2(\underline{x}')] & \cdots & \text{Cov}[Y_1(\underline{x}), Y_d(\underline{x}')] \\ \text{Cov}[Y_2(\underline{x}), Y_1(\underline{x}')] & \sigma_2^2 & \cdots & \text{Cov}[Y_2(\underline{x}), Y_d(\underline{x}')] \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}[Y_d(\underline{x}), Y_1(\underline{x}')] & \text{Cov}[Y_d(\underline{x}), Y_2(\underline{x}')] & \cdots & \sigma_d^2 \end{bmatrix}$$

Revisão da Literatura - Matriz de covariância, de Correlação e Variograma Cruzado

Variograma Cruzado

- Variáveis nas mesmas localizações

$$V_{jk}^*(u) = \sigma_j \sigma_k \left\{ 1 - \frac{1}{2} [\rho_{jk}(u) + \rho_{jk}(-u)] \right\},$$

- Variáveis não necessariamente nas mesmas localizações

$$V_{jk}(u) = \frac{1}{2} (\sigma_j^2 + \sigma_k^2) - \sigma_j \sigma_k \rho_{jk}(u).$$

- Variáveis padronizadas

$$V_{jk}^*(u) = V_{jk}(u)$$

Revisão da Literatura - Modelo Bivariado

- $$\begin{cases} Y_{i1} &= \mu_1(\underline{x}_i) + S_1(\underline{x}_i) + Z_1(\underline{x}_i) \\ Y_{i'2} &= \mu_2(\underline{x}_{i'}) + S_2(\underline{x}_{i'}) + Z_2(\underline{x}_{i'}) \end{cases}$$
- $\underline{Y} = (\underline{Y}_1, \underline{Y}_2)' = (Y_{11}, Y_{21}, \dots, Y_{n_11}, Y_{12}, Y_{22}, \dots, Y_{n_22})'$
medidas nas localizações $\underline{x}_{ij}, \underline{x}_{i'j}$ $i, i' = 1, \dots, n_j, j = 1, 2;$
- $\underline{S}(\underline{x}) = (S_1(\underline{x}), S_2(\underline{x}))$ processo gaussiano estacionário com
 $\underline{\mu} = (0, 0)'$, $\sigma_j^2 = \text{Var}(S_j(\underline{x}))$
estrutura de correlação
$$\begin{bmatrix} \text{Corr}(S_1(\underline{x}), S_1(\underline{x} - u)) & \text{Corr}(S_1(\underline{x}), S_2(\underline{x} - u)) \\ \text{Corr}(S_1(\underline{x}), S_2(\underline{x} - u)) & \text{Corr}(S_2(\underline{x}), S_2(\underline{x} - u)) \end{bmatrix};$$
- $Z_j \sim N(0; \tau^2).$

Revisão da Literatura - Modelo Bivariado

- $$\begin{cases} Y_{i1} &= \mu_1(\underline{x}_i) + \sigma_{01}U_0(\underline{x}_i) + \sigma_1U_1(\underline{x}_i) + Z_1(\underline{x}_i) \\ Y_{i'2} &= \mu_2(\underline{x}_{i'}) + \sigma_{02}U_0(\underline{x}_{i'}) + \sigma_2U_2(\underline{x}_{i'}) + Z_2(\underline{x}_{i'}) \end{cases}$$

Revisão da Literatura - Modelo de Corregionalização

- Modelo de corregionalização linear

$$\underline{Y}(\underline{x}) = \mathbf{D}\underline{w}(\underline{x})$$

- \mathbf{D} é uma matriz de dimensão $p \times p$
- $w_j(\underline{x})$, $j = 1, \dots, p$, processos espaciais iid.
- $\mathbf{D}\mathbf{D}' = \mathbf{T} \rightarrow$ modelo de corregionalização linear intrínseca
- \mathbf{D} triangular inferior e $\underline{Y} = (\underline{Y}(\underline{x}_1), \underline{Y}(\underline{x}_2), \dots, \underline{Y}(\underline{x}_n))'$
- $\Sigma = \mathbf{R} \otimes \mathbf{T}$

Schmidt e Sansó(2006)



Revisão da Literatura - Modelo de Corregeionalização

- $w_j(\underline{x})$ independentes mas não identicamente distribuídos
- $w_j(\underline{x})$ processo gaussiano com média 0, variância 1, função de correlação estacionária $\rho_j(u)$

$$\Sigma = \sum_{j=1}^p (\mathbf{R}_j \otimes \mathbf{T}_j)$$

$\mathbf{T}_j = \underline{d}_j \underline{d}_j'$ com \underline{d}_j sendo a j -ésima coluna de \mathbf{D} .

Revisão da Literatura - Dados Composicionais

- Butler e Glasbey (2008): Registram informação sobre frequências relativas associadas a diferentes componentes de um sistema.
- Aitchison (1986):
 - a. Vetores cujos elementos são proporções de algum todo.
 - b. **Composição:** Vetor $\underline{Y} = (Y_1, Y_2, \dots, Y_B)'$ satisfazendo:
 - $Y_1 \geq 0, \dots, Y_B \geq 0$;
 - $Y_1 + Y_2 + \dots + Y_B = 1$.

Revisão da Literatura - Características de Um Conjunto De Dados Composicionais

Reyment e Savazzi (1999) apud Labus (2005):

- representado em forma matricial;
- linha: unidade observacional;
- coluna: parte ou variável;
- elementos não negativos;
- cada linha soma 1 (100%);
- coeficientes de correlação mudam se uma das variáveis é excluída da matriz e as linhas somam 1 (100%). O mesmo ocorre se um novo componente é adicionado.

Revisão da Literatura - Diagrama Ternário: Malha Quadrada

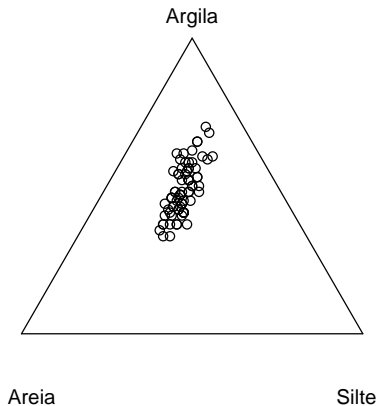


Figura: Diagrama ternário para dados da Malha Quadrada.

Revisão da Literatura - Composição Regionalizada

Seja $\Omega \subset \mathbb{R}^n$ domínio espacial, \mathbb{R}^n espaço real de dimensão n .

- **Composição Regionalizada:**

$\underline{Y}(\underline{x}) = [Y_1(\underline{x}), Y_2(\underline{x}), \dots, Y_B(\underline{x})]'$, $\underline{x} \in \Omega$, satisfazendo:

- 1. $Y_1(\underline{x}) \geq 0, \dots, Y_B(\underline{x}) \geq 0$;
- 2. $Y_1(\underline{x}) + Y_2(\underline{x}) + \dots + Y_B(\underline{x}) = 1$.

- **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 - Base Regionalizada

- Vetor $\underline{W}(\underline{x})$, $\underline{x} \in \Omega \subset \mathbb{R}^n$ cujos componentes são todos medidos na mesma escala e positivos.
- **Espaço Amostral:**

$$\mathbb{R}_+^B = \{\underline{W}(\underline{x}) \in \mathbb{R}^B; W_i(\underline{x}) > 0, i = 1, \dots, B\}$$

- **Operador fechamento:** Base \Rightarrow Composição

$$\begin{aligned} C: \mathbb{R}_+^B &\longrightarrow \mathbb{S}^B \\ \underline{W}(\underline{x}) &\longrightarrow C[\underline{W}(\underline{x})] = \frac{\underline{W}(\underline{x})}{\underline{j}'\underline{W}(\underline{x})}, \end{aligned}$$

\underline{j}' vetor de 1 's.

Revisão da Literatura - Subcomposição Regionalizada

- Seja $\underline{Y}(\underline{x})$ composição regionalizada de D partes;
- $s \subset \{1, 2, \dots, B\}$ tal que $\underline{Y}_s(\underline{x})$ é subvetor cujos elementos são os componentes de $\underline{Y}(\underline{x})$ correspondentes às partes em s ;
- **Subcomposição Regionalizada:**

$$C[\underline{Y}_s(\underline{x})] = \frac{\underline{Y}_s(\underline{x})}{\underline{j}'\underline{Y}_s(\underline{x})};$$

- Se $\underline{s}(\underline{x}) = C[\underline{Y}_s(\underline{x})] \Rightarrow \frac{s_i(\underline{x})}{s_j(\underline{x})} = \frac{Y_i(\underline{x})}{Y_j(\underline{x})}, \quad i, j \in s.$

Revisão da Literatura - Amalgamação

- $\underline{Y}(\underline{x}) = [Y_1(\underline{x}), Y_2(\underline{x}), \dots, Y_B(\underline{x})]'$ uma composição regionalizada de B partes;
- $\underline{Y}(\underline{x})$ dividida em C ($C \leq B$) subconjuntos mutuamente exclusivos;
- $\underline{Y}_i(\underline{x})$, $i = 1, \dots, C$, vetor cujos componentes são os elementos do i -ésimo subconjunto.
- **Amalgamação Regionalizada:** composição regionalizada de C partes

$$\underline{A}(\underline{x}) = [A_1(\underline{x}), A_2(\underline{x}), \dots, A_C(\underline{x})]'$$

em que $A_i(\underline{x}) = \underline{j}' \underline{Y}_i(\underline{x})$, \underline{j}' vetor de 1 's.

Revisão da Literatura - Partição

Partição Regionalizada: Amalgamação juntamente com subcomposições.

- $P_C[\underline{Y}(\underline{x})] = [\underline{s}_1(\underline{x}), \underline{s}_2(\underline{x}), \dots, \underline{s}_C(\underline{x}); \underline{A}(\underline{x})]$

em que $\underline{s}_i(\underline{x}) = C[Y_i(\underline{x})]$, $i = 1, \dots, C$.

- Preserva informações contidas nos subvetores.

Revisão da Literatura: Problema - Dados Composicionais

Efeito de **correlação espúria** (Pawlowsky e Olea, 2004):

- Covariâncias sujeitas à controles não estocásticos



interpretação errônea da estrutura de covariância espacial;

- Singularidade da matriz de covariância de uma composição.

Revisão da Literatura - Transformação

- **Graf (2006):** Soma constante \Rightarrow correlação negativa entre os componentes.
- **Aitchison (1986):**
 - Propõe transformação que generaliza a transformação logística $\ln \frac{Y}{1-Y}$ para um vetor composicional de 2 partes;
 - Magnitudes relativas ou razões \Rightarrow tratabilidade e interpretação estatística.
- **Transformações logísticas:** Aitchison (1982), Aitchison et al. (2000), Odeh et al. (2003).

Revisão da Literatura - Transformação

Transformação razão log-aditiva (ALR):

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

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

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

Revisão da Literatura - Transformação

Transformação logística generalizada aditiva (AGL):

$$\begin{aligned} \text{AGL} : \quad \mathbb{R}^{B-1} &\longrightarrow \mathbb{S}^B \\ \text{ALR}[\underline{Y}(\underline{x})] &\longrightarrow \text{AGL}\{\text{ALR}[\underline{Y}(\underline{x})]\} = \underline{Y}(\underline{x}) = \frac{\exp[\underline{Y}(\underline{x})]}{\sum_j \exp[\underline{Y}(\underline{x})]}; \end{aligned}$$

$$\underline{Y}(\underline{x}) = (Y_1(\underline{x}), Y_2(\underline{x}), \dots, 0)'.$$

$$\text{Se } c \neq 0 \quad \Rightarrow \quad \underline{Y}(\underline{x}) = c \cdot \text{AGL}\{\text{ALR}[\underline{Y}(\underline{x})]\}.$$

Revisão da Literatura - Transformação

Transformação Razão Log Centrada (CLR):

$$\begin{aligned} \text{CLR} : \mathbb{R}_+^B &\rightarrow \mathbb{R}^B \\ \underline{W}(\underline{x}) &\rightarrow \text{CLR}[\underline{W}(\underline{x})] = \ln \frac{\underline{W}(\underline{x})}{g[\underline{W}(\underline{x})]} \end{aligned}$$

ou

$$\begin{aligned} \text{CLR} : \mathbb{S}^B &\longrightarrow \mathbb{R}^B \\ \underline{Y}(\underline{x}) &\longrightarrow \text{CLR}[\underline{Y}(\underline{x})] = \ln \frac{\underline{Y}(\underline{x})}{g[\underline{Y}(\underline{x})]} \end{aligned}$$

g : média geométrica.

Revisão da Literatura - Perturbação e Potência

Aitchison (1986), Aitchison e Egozcue(2005), Tolosana-Delgado et al.(2005):

Operações que definem uma estrutura de espaço vetorial de dimensão $B - 1$ no simplex.

- 1. **Perturbação** (\oplus) : operação comutativa;
- 2. **Potência** (\odot) : produto externo.

Revisão da Literatura - Perturbação

Perturbação: Multiplicação de composições componente a componente e divisão de cada componente pela soma de todos.

$$\begin{aligned}\underline{Y}_1 \oplus \underline{Y}_2 &= (Y_{11}, Y_{12}, \dots, Y_{1B}) \oplus (Y_{21}, Y_{22}, \dots, Y_{2B}) \\ &= \mathcal{C}(Y_{11} Y_{21}, Y_{12} Y_{22}, \dots, Y_{1B} Y_{2B}).\end{aligned}$$

Revisão da Literatura - Potência

- **Potência**: análogo à multiplicação por um escalar no espaço real.

$$\alpha \odot (Y_{11}, Y_{12}, \dots, Y_{1B}) = \mathcal{C}(Y_{11}^\alpha, Y_{12}^\alpha, \dots, Y_{1B}^\alpha).$$

- Vetor de **diferenças** composicionais:

$$\underline{Y}_1 \ominus \underline{Y}_2 = \underline{Y}_1 \oplus (-1 \odot \underline{Y}_2).$$

- **Produto interno:**

$$\langle \underline{Y}_1, \underline{Y}_2 \rangle = \sum_{i=1}^B \ln \left(\frac{Y_{1i}}{g(\underline{Y}_1)} \right) \ln \left(\frac{Y_{2i}}{g(\underline{Y}_2)} \right).$$

- **Distância de Aitchison:**

$$d(\underline{Y}_1, \underline{Y}_2) = \sqrt{\sum_{i=1}^B \left(\ln \left(\frac{Y_{1i}}{g(\underline{Y}_1)} \right) - \ln \left(\frac{Y_{2i}}{g(\underline{Y}_2)} \right) \right)^2}$$

- **Centro:**

$$\text{cen}(\underline{Y}) = \frac{1}{g_s} (g(Y_1) \quad g(Y_2) \quad \dots \quad g(Y_B))'$$

- $g(Y_i)$ - média geométrica do i -ésimo componente
- $g_s = g(Y_1) + g(Y_2) + \dots + g(Y_B)$.

- **Medida de dispersão:**

$$\text{Totvar}(\underline{Y}) = \frac{1}{B} \sum_{i < j} \text{Var} \left(\ln \left(\frac{Y_i}{Y_j} \right) \right).$$

Revisão da Literatura - Domínio de Confiança

$$B_{1-\alpha}(\underline{Y}) = \left\{ \underline{Y} \in S^B \mid \left(\ln \left(\frac{\underline{Y}_{-B}}{\underline{Y}_B} \right) - \mu \right)' \Sigma^{-1} \left(\ln \left(\frac{\underline{Y}_{-B}}{\underline{Y}_B} \right) - \mu \right) \leq \chi_{b;1-\alpha}^2 \right\}$$

Revisão da Literatura - Diagrama Ternário: Malha Quadrada

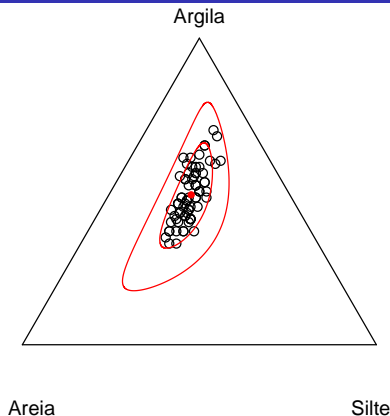


Figura: Diagrama ternário para dados da Malha Quadrada, incluindo o centro da distribuição e regiões 2 e 4σ

Revisão da Literatura - Estacionariedade

Segundo Pawlowsky-Glahn e Olea (2004)

$\underline{Y}(\underline{x})$ é **Estacionária de 2ª ordem** se $f(\underline{Y}(\underline{x}))$ satisfaz:

- $E(f(\underline{Y}(\underline{x}))) = \mu$ existe e não depende de \underline{x} ;
- $\text{Cov}(f(\underline{Y}(\underline{x}_2)), f(\underline{Y}(\underline{x}_1))) = \Sigma(\underline{x}_2 - \underline{x}_1)$ existe e não depende de $\underline{x}_1, \underline{x}_2$, mas somente da diferença $\underline{h} = \underline{x}_2 - \underline{x}_1$.

Revisão da Literatura - Função Intríntrica

$\underline{Y}(\underline{x})$ é **função intríntrica** se $f(\underline{Y}(\underline{x}))$ satisfaz:

- $E(f(\underline{Y}(\underline{x}))) = \mu$ existe e não depende de \underline{x} ;
- $\text{Cov}(f(\underline{Y}(\underline{x}_2)) - f(\underline{Y}(\underline{x}_1))) = \Gamma(\underline{x}_2 - \underline{x}_1)$

existe e não depende de \underline{x}_1 , \underline{x}_2 , mas somente da diferença $\underline{h} = \underline{x}_2 - \underline{x}_1$.

Revisão da Literatura - Estrutura de Covariância Espacial

Estrutura de Covariância Espacial: conjunto de funções B^4

$$\sigma_{ij \cdot kl}(\underline{h}) = \text{Cov} \left(\ln \frac{\underline{Y}_i(\underline{x})}{\underline{Y}_k(\underline{x})}, \ln \frac{\underline{Y}_j(\underline{x} + \underline{h})}{\underline{Y}_l(\underline{x} + \underline{h})} \right),$$

$i, j, k, l \in \{1, 2, \dots, B\}$, para $\underline{x}, \underline{x} + \underline{h} \in \Omega$.

Pawlowsky e Olea (2004).

Revisão da Literatura - Estrutura de Covariância Espacial

- Matriz de **LR autocovariâncias** (Matriz Variação)

$$\mathbf{T}(\underline{h})_{B \times B} = (\tau_{i,j}(\underline{h})) = (\sigma_{ii,jj}(\underline{h}))$$

- Matriz de **ALR covariâncias cruzadas**

$$\Sigma(\underline{h})_{B-1 \times B-1} = (\sigma_{ij}(\underline{h})) = (\sigma_{ij, BB}(\underline{h}))$$

- Matriz de **CLR covariâncias cruzadas**, $\Xi(\underline{h})$ com elementos

$$\xi_{ij}(\underline{h}) = \text{Cov} \left(\ln \frac{X_i(\underline{x})}{g(\underline{X}(\underline{x}))}, \ln \frac{Y_j(\underline{x} + \underline{h})}{g(\underline{Y}(\underline{x} + \underline{h}))} \right),$$

Revisão da Literatura - Estrutura de Covariância Espacial Intríntrica

Estrutura de Cov. Espacial Intríntrica: conjunto de funções B^4

$$V_{ij \cdot kl}(\underline{h}) = \frac{1}{2} \text{Cov} \left(\ln \frac{\underline{Y}_i(\underline{x})}{\underline{Y}_k(\underline{x})} - \ln \frac{\underline{Y}_i(\underline{x} + \underline{h})}{\underline{Y}_k(\underline{x} + \underline{h})}, \ln \frac{\underline{Y}_j(\underline{x})}{\underline{Y}_l(\underline{x})} - \ln \frac{\underline{Y}_j(\underline{x} + \underline{h})}{\underline{Y}_l(\underline{x} + \underline{h})} \right)$$

$i, j, k, l \in \{1, 2, \dots, B\}$ para $\underline{x}, \underline{x} + \underline{h} \in \Omega$.

Revisão da Literatura - Estrutura de Covariância Espacial Intríntrica

- Matriz de **LR variogramas** (Matriz Variação Intríntrica)

$$\Gamma(\underline{h})_{B \times B} = (V_{i,j}(\underline{h})) = (V_{ii,jj}(\underline{h}))$$

- Matriz de **ALR variograma cruzado**

$$\Psi(\underline{h})_{B-1 \times B-1} = (\psi_{ij}(\underline{h})) = (V_{ii,BB}(\underline{h}))$$

- Matriz de **CLR covariâncias cruzadas**, $\Delta(\underline{h})$ com elementos

$$\delta_{ij}(\underline{h}) = \text{Cov} \left(\ln \frac{Y_i(\underline{x})}{g(Y(\underline{x}))}, \ln \frac{Y_j(\underline{x} + \underline{h})}{g(Y(\underline{x} + \underline{h}))} \right)$$

Material e Métodos

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_{Agua}, pH_{KCl}, Ca, Mg, K, Al, H, C, N, CTC, S, V, M, NC, CEC, CN.
- Gonçalves (1997) - ESALQ/USP-UEM: $n = 412$

Material e Métodos - Método

- Banerjee, Carlin e Gelfand (2004), Diggle e Ribeiro Jr (2007), Schmidt e Sansó (2006): declaração explícita de modelo;
- Aitchison (1986): independência entre as localizações;
- Pawlowsky e Olea (2004): espacialização;
- Obage (2005): análise bayesiana sem espacialização.

Material e Métodos - 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 Métodos - Método

- Análise geoestatística
 - Análise descritiva: normalidade dos dados, tendência direcional;
 - Transformação Box & Cox ou, por exemplo, a arco seno;
 - Método dos mínimos quadrados ordinários: estimativas dos parâmetros;
 - Modelo geoestatístico: verossimilhança, função de correlação da família de Matérn ;
 - Métodos bayesianos: distribuição preditiva $p(\underline{Y}_0, \underline{Y})$;
 - Adaptar modelos bivariados aos dados composicionais incluindo a restrição soma dos componentes igual a 1.

Material e Métodos - Método

- Análise de dados composicionais
 - Diagrama ternário;
 - Estatísticas descritivas: centro da distribuição; variância total; matriz de LR autocovariâncias, CRL covariâncias cruzadas;
 - Composição modelada usando geoestatística;
 - Método de krigagem empregado na construção do mapa de classificação do solo.

Material e Métodos - Método

Modelos Iniciais:

$$1. \begin{cases} Y_{i1} &= \mu_1(x_i) + Z_1(x_i) \\ Y_{i'2} &= \mu_2(x_{i'}) + Z_2(x_{i'}) \end{cases}$$

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

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

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

Modelos Iniciais:

$$2. \begin{cases} Y_{i1} &= \mu_1(\underline{x}_i) + S_1(\underline{x}_i) + Z_1(\underline{x}_i) \\ Y_{i'2} &= \mu_2(\underline{x}_{i'}) + S_2(\underline{x}_{i'}) + Z_2(\underline{x}_{i'}) \end{cases}$$
$$\begin{cases} Y_{i1} &= \mu_1(\underline{x}_i) + \sigma_1 U(\underline{x}_i; \phi) + Z_1(\underline{x}_i) \\ Y_{i'2} &= \mu_2(\underline{x}_{i'}) + \sigma_2 U(\underline{x}_{i'}; \phi) + Z_2(\underline{x}_{i'}) \end{cases}$$

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

Resultados Preliminares

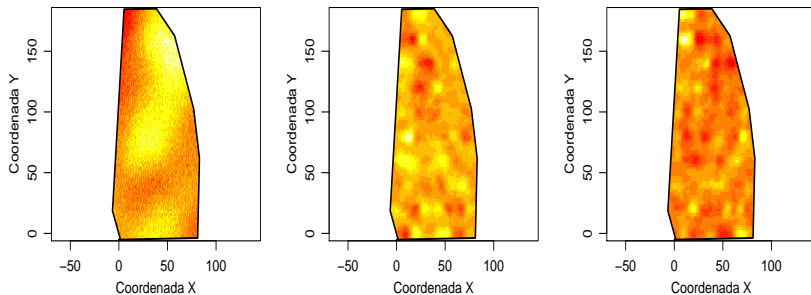


Figura: Mapas da porcentagem de areia (à esquerda), silte (centro) e argila (à direita).

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.

Cronograma

| Ano | 2008 | | | | 2009 | | | | | | | | | | | | 2010 | | |
|--|------|----|----|----|------|----|----|----|----|----|----|----|----|----|----|----|------|----|--|
| Mês | 09 | 10 | 11 | 12 | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 01 | 02 | |
| 1. Qualificar o projeto junto ao PPGMNE | X | | | | | | | | | | | | | | | | | | |
| 2. Completar a revisão da literatura e Elaborar Metodologia de análise de dados | X | X | X | X | X | | | | | | | | | | | | | | |
| 3. Aplicar metodologia aos conjuntos de dados | | | | | X | X | X | X | X | X | X | | | | | | | | |
| 4. Elaborar relatório preliminar com resultados e discussão | | | | | | | | X | X | X | | | | | | | | | |
| 5. Elaborar versão preliminar da tese | | | | | | | | | | | X | X | X | X | | | | | |
| 6. Fazer revisão ortográfica e gramatical da tese | | | | | | | | | | | | | | | | X | X | | |
| 7. Redação e submissão de artigos | | | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | | |
| 8. Confeção da tese | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | | |
| 9. Defesa da tese | | | | | | | | | | | | | | | | | X | X | |

Bibliografia

- AITCHISON, J. The statistical analysis of compositional data. **Journal of the Royal Statistical Society, Series B**, v. 44, n.2, p. 139-177, 1982.
- AITCHISON, J. **The statistical analysis of compositional data**. New Jersey: The Blackburn Press, 1986.
- AITCHISON, J.; BARCELÓ-VIDAL, C.; MARTÍN-FERNANDÉZ, J. A.; PAWLOWSKY-GLAHN, V. Logratios analysis and compositional distance. **Mathematical Geology**, v. 32, n.3, p. 563-580, 2000.
- AITCHISON, J.; EGOZCUE, J. J. Compositional data analysis: where are we and where should we be heading?. **Mathematical Geology**, v. 37, n.7, p. 829-850, oct. 2005.
- BANERJEE, S.; CARLIN B. P. e GELFAND , A. E. **Hierarchical modelling and analysis for spatial data**, Boca Raton: Chapman and Hall, 2004.

Bibliografia

- BOOGAART, K.G.v.d. Using the R package "compositions", v. 0.9, 2005. Disponível em: <<http://www.stat.boogaart.de/compositions/UsingCompositions.pdf>>. Acesso em: 09/02/08.
- BUTLER, A.; GLASBEY, C. A latent Gaussian model for compositional data with zeros. **Journal of the Royal Statistical Society, Series C** (in press), 2008. Disponível em: <http://www.bioss.ac.uk/staff/adam/documents/ButlerGlasbey_resubmit.pdf>. Acesso em: 13/02/08.
- DIGGLE, P. J. e RIBEIRO JR, P. J. **Model-based geostatistics**, USA: Springer Series in Statistics, 2007
- DIGGLE, P. J.; RIBEIRO JR, P. J. e CHRISTENSEN, O. F. An introduction to model-based geostatistics. **Spatial Statistics and Computational Methods**, New York: Møller, J. Springer, p. 43-86, 2003

Bibliografia

- 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.
- GRAF, M. Precision of compositional data in a stratified two-Stage cluster sample: comparison of the swiss earnings structure survey 2002 and 2004. **Survey Research Methods Section, ASA**, Session 415: Sample Survey Quality V, 3066-3072, 2006. Disponível em: <<http://www.amstat.org/sections/SRMS/proceedings/y2006/Files/JSM2006-000771.pdf>>. Acesso em: 18/05/08.
- LABUS, M. Compositional data analysis as a toll for interpretation of rock porosity parameters. **Geological Quarterly**, v. 49, n.3, p. 347-354, 2005
- ODEH, I. O. A.; TOOD, A. J.; TRIANTAFILIS, J. Spatial prediction of soil particle-size fractions as compositional data. **Soil Science**, v. 168, n.7, p. 501-515, July 2003.

Bibliografia

- 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.
- PAWLOWSKY-GLAHN, V.; OLEA, R. A. **Geostatistical analysis of compositional data**. New York: Oxford University Press, Inc., 2004.
- REYMENT, R. A. and SAVAZZI E. **Aspects of multivariate statistical analysis in geology**. Amsterdam: Elsevier, 1999.
- 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**.
- TOLOSANA-DELGADO, R.; OTERO, N.; PAWLOWSKY-GLAHN, V. Some Basic Concepts of Compositional Geometry. **Mathematical Geology**, v. 37, n.7, p. 563-580, oct 2005.
- 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
- Membros da banca: Ph.D. Alexandra Mello Schmidt e Dr. Antônio Carlos Gonçalves
- UEM/DES - Universidade Estadual de Maringá/Departamento de Estatística
- PPGMNE - Programa de Pós-Graduação em Métodos Numéricos em Engenharia
- LEG - Laboratório de Estatística e Geoinformação
- CNPQ - Conselho Nacional de Desenvolvimento Científico e Tecnológico
- Aos professores e colegas de curso.