

Um Modelo Geostatístico Bivariado Para Análise de Dados Composicionais

Ana Beatriz Tozzo Martins - PPGMNE-LEG/UFPR
DES/UEM

Paulo Justiniano Ribeiro Junior - LEG/UFPR

Wagner Hugo Bonat - PPGMNE-LEG/UFPR

Antônio Carlos Andrade Gonçalves - DAG/UEM

julho de 2009

Roteiro de Apresentação

1. Introdução: Motivação, literatura, objetivo;
2. Metodologia: Modelo Geoestatístico Bivariado;
3. Resultados: Estudo de caso de frações granulométricas de solo;
4. Conclusão.

Motivação

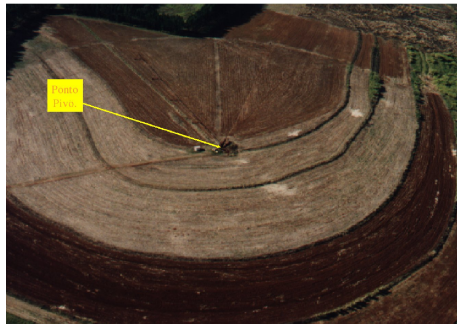


Figura: Quadrante irrigado por sistema pivô-central no campo experimental da ESALQ-USP.

SOLO

frações granulométricas
(composição)



propriedades físico-hídricas



prática agrícola



agricultura de precisão

Dados Composicionais - Aitchison (1986)

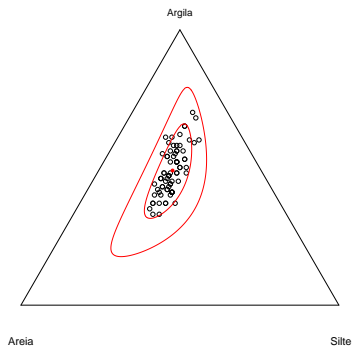


Figura: Diagrama Ternário das composições.

• **Composição:** $\underline{X} = (X_1, X_2, \dots, X_B)'$

$$X_1 \geq 0, \dots, X_B \geq 0$$

$$X_1 + X_2 + \dots + X_B = 1$$

$$\mathbb{S}^B = \{ \underline{X} \in \mathbb{R}^B; X_i > 0, i=1, \dots, B; \mathbf{1}'\underline{X}=1 \}$$

• **Base:** $\underline{W} = (W_1, W_2, \dots, W_B)'$

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

• **Operador fechamento:**

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

$$\underline{W} \longrightarrow \mathcal{C}[\underline{W}] = \frac{\underline{W}}{\mathbf{1}'\underline{W}}$$

Literatura

- Dados composicionais:
 - Aitchison (1986).
- Modelos espaciais uni e multivaridados:
 - Diggle e Ribeiro Jr (2007);
 - Schmidt e Gelfand (2003);
 - Banerjee, Carlin e Gelfand (2004);
 - Schmidt e Sansó (2006).
- Combinação de dados composicionais e espaciais:
 - Pawlowsky e Olea (2004);
 - Tjelmeland e Lund (2003).

Objetivo

- Estender o modelo geoestatístico bivariado para dados composicionais, derivando e implementando estimação baseada na verossimilhança;
- Obter preditores espaciais no simplex que permitam a construção de mapas de predição das frações do solo na área de estudo.

Fundamentos Para Análise de Dados Composicionais

Transformação razão log-aditiva (alr):

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

$$\underline{X}(\underline{x}) \longrightarrow \text{alr}(\underline{X}(\underline{x})) = \left(\ln \frac{X_1(\underline{x})}{X_B(\underline{x})}, \dots, \ln \frac{X_{B-1}(\underline{x})}{X_B(\underline{x})} \right)'$$

Composição: $\underline{X}(\underline{x}) = (X_1(\underline{x}), X_2(\underline{x}), X_3(\underline{x}))' = (\text{Areia}, \text{Silte}, \text{Argila})'$

$$\underline{Y}(\underline{x}) = (Y_1(\underline{x}), Y_2(\underline{x}))' = \left(\ln \frac{X_1(\underline{x})}{X_3(\underline{x})}, \ln \frac{X_2(\underline{x})}{X_3(\underline{x})} \right)'$$

Transformação logística generalizada aditiva (agl):

$$\text{agl} = \text{alr}^{-1}$$

Modelo Espacial

Modelo geoestatístico bivariado para dados composicionais:

$$\begin{cases} Y_1(\underline{x}_i) &= \mu_1(\underline{x}_i) + \sigma_1 U(\underline{x}_i; \phi) + \tau_1 Z(\underline{x}_i; \rho) \\ Y_2(\underline{x}_{i'}) &= \mu_2(\underline{x}_{i'}) + \sigma_2 U(\underline{x}_{i'}; \phi) + \tau_2 Z(\underline{x}_{i'}; \rho) \end{cases}$$

- $\underline{x}_i, \underline{x}_{i'} \in \mathbb{R}^2$; $i, i' = 1, \dots, n_1$, n_1 tamanho amostral;
- $\underline{Y}_{n \times 1}(\underline{x}) = (Y_1(\underline{x}_1), Y_2(\underline{x}_1), \dots, Y_1(\underline{x}_{n_1}), Y_2(\underline{x}_{n_1}))'$;
- $U \sim N(\underline{0}; \Sigma_U)$, Σ_U : variâncias unitárias e covariâncias em função de $\rho_U(\phi)$, exponencial;
- $Z \sim N(\underline{0}; \Sigma_Z)$, Σ_Z : variâncias unitárias e covariâncias em função de ρ induzida pela estrutura composicional.

Verossimilhança

- $\underline{\theta} = (\mu_1, \mu_2, \sigma_1, \sigma_2, \tau_1, \tau_2, \phi, \rho)'$ $\xrightarrow{\text{Reparam.}}$ $\underline{\theta}^* = (\eta, \nu_1, \nu_2, \phi, \rho)'$

- **Log-Verossimilhança** \rightarrow **Normal multivariada com EMV:**

$$\hat{\underline{\mu}} = (\mathbf{D}'\mathbf{V}^{-1}\mathbf{D})^{-1}(\mathbf{D}'\mathbf{V}^{-1}\underline{Y}) \quad \text{e} \quad \hat{\sigma}_1 = \sqrt{\hat{Q}e/n}$$

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

- **Log-verossimilhança concentrada** \rightarrow **solução numérica:**

$$l(\underline{\theta}^*, \underline{Y}) = -0,5 \left[\ln(|\mathbf{V}|) + n \left(\ln(2\pi) + \ln(\hat{Q}e) - \ln(n) + 1 \right) \right]$$

- **Método delta:**

- $\text{Var}(\underline{\hat{\mu}}) = \mathbf{I}_o(\underline{\hat{\mu}})^{-1} = \hat{\sigma}_1^2 (\mathbf{D}' \hat{\mathbf{V}}^{-1} \mathbf{D})^{-1}$

- $\text{Var}(\hat{\sigma}_1) = \mathbf{I}_o(\hat{\sigma}_1)^{-1} = \frac{\hat{\sigma}_1^3}{3\hat{Q}e - n\hat{\sigma}_1}$

- **Cokrigagem de \underline{Y}_0 em $\underline{x}_0 = (x_{10}, x_{20}, \dots, x_{n_20})$:**

- $\underline{\mu}_{\underline{Y}_0|\underline{Y}} = \underline{\mu}_{\underline{Y}_0} + \Sigma_{\underline{Y}_0\underline{Y}} \Sigma_{\underline{Y}\underline{Y}}^{-1} (\underline{Y} - \underline{\mu}_{\underline{Y}})$

- $\Sigma_{\underline{Y}_0|\underline{Y}} = \Sigma_{\underline{Y}_0\underline{Y}_0} - \Sigma_{\underline{Y}_0\underline{Y}} \Sigma_{\underline{Y}\underline{Y}}^{-1} \Sigma_{\underline{Y}\underline{Y}_0}$

Predição (cont.)

- **Integração de Gauss-Hermite multivariada** de ordem k

$$\bullet \int_{\mathbb{R}^{B-1}} g(\underline{Z}) f(-\underline{Z}'\underline{Z}) d\underline{Z} \approx$$

$$\sum_{i_1=1}^k \sum_{i_2=1}^k \cdots \sum_{i_{B-1}=1}^k \omega_{i_1} \omega_{i_2} \cdots \omega_{i_{B-1}} g(Z_{i_1}, Z_{i_2}, \dots, Z_{i_{B-1}})$$

- pesos $\omega_{i_1} \omega_{i_2} \cdots \omega_{i_{B-1}}$ e abscissas $Z_{i_1}, Z_{i_2}, \dots, Z_{i_{B-1}}$

Abramowitz e Stegun (1972)

- **Alternativa por simulação**

Predição (cont.)

Momentos no simplex:

- $\underline{\mu}_X = \int_{\mathbb{S}^B} \underline{X} f(\underline{X}) d\underline{X} \quad \Rightarrow \quad \underline{\mu}_X = \int_{\mathbb{R}^{B-1}} g_1(\underline{Z}) f(-\underline{Z}'\underline{Z}) d\underline{Z}$
- $\Sigma_X = \int_{\mathbb{S}^B} (\underline{X} - \underline{\mu}_X)(\underline{X} - \underline{\mu}_X)' f(\underline{X}) d\underline{X} \quad \Rightarrow \quad \Sigma_X = \int_{\mathbb{R}^{B-1}} g_2(\underline{Z}) f(-\underline{Z}'\underline{Z}) d\underline{Z}$
- $\underline{Z} = \frac{1}{\sqrt{2}} (R')^{-1} (\text{alr}(\underline{X}) - \underline{\mu}_Y)$; R cholesky de Σ_Y
- $g_1(\underline{Z}) = \pi^{-\frac{B-1}{2}} \text{agl}(\underline{\mu}_Y + \sqrt{2}R'\underline{Z})$
- $g_2(\underline{Z}) = \pi^{-\frac{B-1}{2}} \left(\text{agl}(\underline{\mu}_Y + \sqrt{2}R'\underline{Z}) - \underline{\mu}_X \right) \left(\text{agl}(\underline{\mu}_Y + \sqrt{2}R'\underline{Z}) - \underline{\mu}_X \right)'$

Análise de Frações Granulométricas de Um Solo

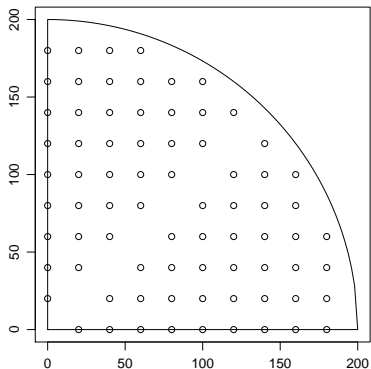


Figura: Distribuição das localizações na área de estudo.

- Gonçalves (1997):
Tese - ESALQ-USP;
 - Área: quadrante irrigado por um sistema pivô-central;
 - Grade de amostragem $20 \times 20\text{m}$
- ⇒ 76 amostras: $\left\{ \begin{array}{l} \text{Areia} \\ \text{Silte} \\ \text{Argila.} \end{array} \right.$

Análise de Frações Granulométricas de Um Solo

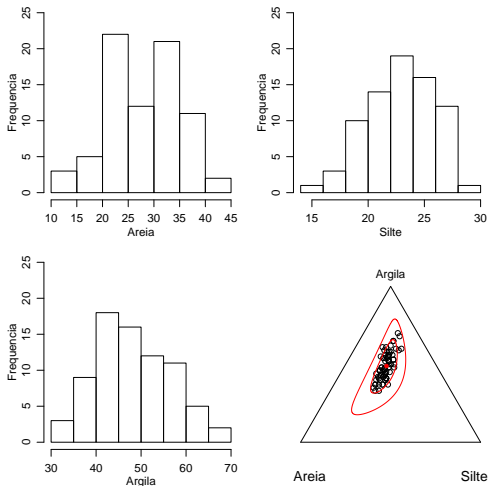


Figura: Distribuição de areia, silte e argila e diagrama ternário das composições.

Análise de Frações Granulométricas de Um Solo

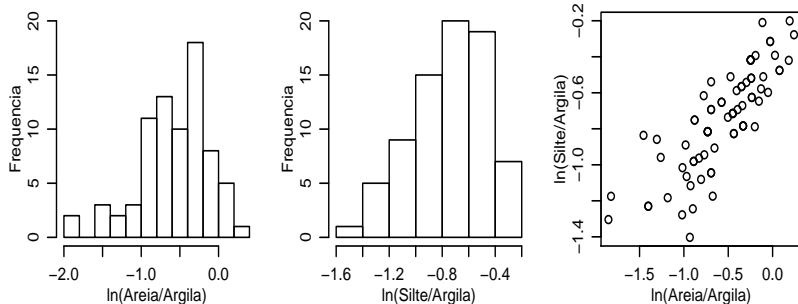


Figura: Distribuição das log-razão e correspondente diagrama de dispersão.

Análise de Frações Granulométricas de Um Solo

Tabela: Estimativas, erros padrão e intervalos de confiança pelo método delta via método de otimização "L-BFGS-B".

Parâmetros	Estimativas	Erro Padrão	LI. Delta	LS. Delta
μ_1	-0,7598	0,44029	-1,20017	-0,3195
μ_2	-0,7941	0,22857	-1,02268	-0,5655
σ_1	0,4504	0,12080	0,32968	0,5712
σ_2	0,1152	0,05153	0,06375	0,1668
τ_1	0,2870	0,04510	0,24191	0,3321
τ_2	0,2669	0,03142	0,23555	0,2984
ϕ	66,9084	42,55386	24,35460	109,4623
ρ	0,9544	0,06803	0,88638	1,0224

Análise de Frações Granulométricas de Um Solo

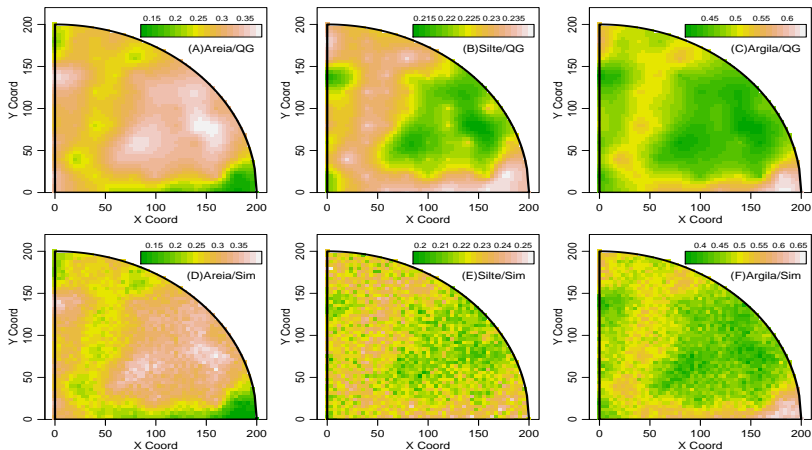


Figura: Mapas das porcentagens de areia, silte e argila obtidos por quadratura de Gauss-Hermite (A-C) e por simulação (D-F).

Conclusão

- Construção de mapas de areia, silte e argila garantindo que as frações somem 1, nos pontos observados e preditos;
- Modelo captura variações espaciais, induzidas pelas composições e não estruturadas;
- Declaração explícita do modelo permite fazer inferências sobre parâmetros;
- Possibilidade de tratamento Bayesiano para considerar nas predições a incerteza associada à estimação dos parâmetros;
- Estimação de outros funcionais;
- Necessidade de investigar alternativas para computação mais eficiente e formas de especificação do modelo para um número maior de componentes.

Bibliografia

- AITCHISON, J., *The statistical analysis of compositional data*, New Jersey: The Blackburn Press, 1986.
- DIGGLE, P.J.; RIBEIRO JR, P.J., *Model-based geostatistics*, USA: Springer Series in Statistics, 2007.
- GONÇALVES, A.C.A., *Variabilidade espacial de propriedades físicas do solo para fins de manejo da irrigação*. 1997. 119p. Tese (Doutorado em Agronomia) - ESALQ-USP, São Paulo, Piracicaba.
- PAWLOWSKY-GLAHN, V.; OLEA, R.A., *Geostatistical analysis of compositional data*, New York: Oxford University Press, Inc., 2004.
- SCHMIDT, A.M.; GELFAND, A.E., A bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research*, v. 108, p. 10-1-18-8, 2003.
- SCHMIDT, A.M.; SANSÓ, B., Modelagem bayesiana da estrutura de covariância de processos espaciais e espaço temporais, In: *17 SINAPE e ABE*, Caxambu: Associação Brasileira de Estatística, 2006, Minicurso.
- TJELMELAND, H.; LUND, K.V., Bayesian modelling of spatial compositional data. *Journal of Applied Statistics*, v. 30, n. 1, p. 87-100, 2003.

OBRIGADA PELA ATENÇÃO!

Agradecimentos

- DES - Departamento de Estatística, UEM.
- PPGMNE - Programa de Pós-Graduação em Métodos Numéricos em Engenharia, UFPR.
- LEG - Laboratório de Estatística e Geoinformação, UFPR.
- CNPQ - Conselho Nacional de Desenvolvimento Científico e Tecnológico.
- FINEP projeto CT-INFRA/UFPR.