

# Análise Geoestatística de Dados Composicionais

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

Orientador: Prof. PhD. Paulo Justiniano Ribeiro Junior - LEG/UFPR  
Programa de Pós-Graduação em Métodos Numéricos em Engenharia

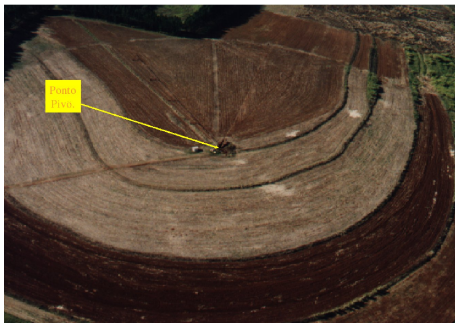
PPGMNE - UFPR

09 de abril de 2010

# TESE

1. Introdução
2. Revisão da Literatura:
  - Modelo Geoestatístico Gaussiano;
    - Predição Linear Espacial - Clássica e Bayesiana;
  - Modelo Multivariado;
  - Dados Composicionais;
3. Métodos:
  - Modelo Geoestatístico Composicional - Clássica e Bayesiana;
  - Análise de Dados Composicionais Simulados;
4. Resultados: Análise de Frações Granulométricas de um Solo;
5. Conclusão e Sugestões de Trabalhos Futuros.

# 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

# Triângulo de Classificação Textural

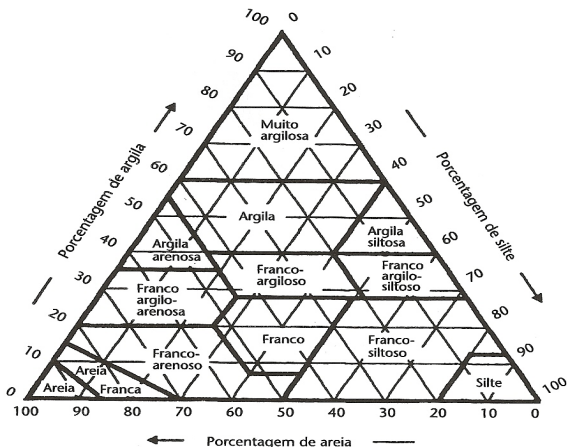


Figura: Diagrama de classificação textural do solo.

Fonte: Lemos e Santos (1996) *apud* Reichardt e Timm (2004).

# Dados Composicionais - Aitchison (1986)

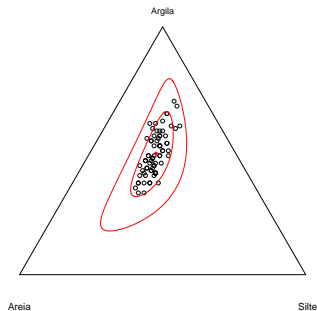


Figura: Diagrama Ternário das composições.

• **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:**

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

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

# Geometria: Diagrama Ternário com altura unitária

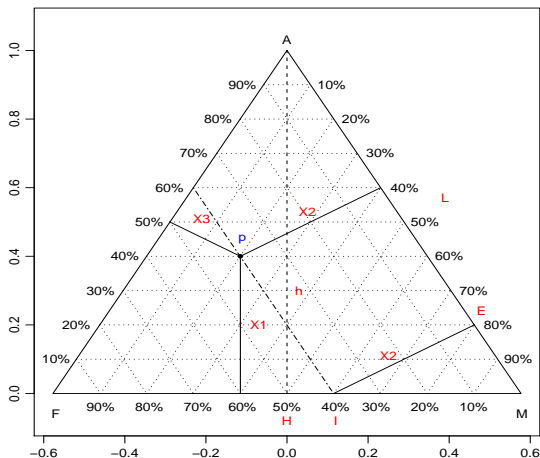


Figura: O ponto  $p$ ,  $(40, 40, 20)\%$  em  $\mathbb{S}^3$  e  $(-0, 115; 0, 40)$  em  $\mathbb{R}^2$ , representado no triângulo de altura unitária e triângulo de classificação textural.

# Geometria: Diagrama Ternário com lado unitário

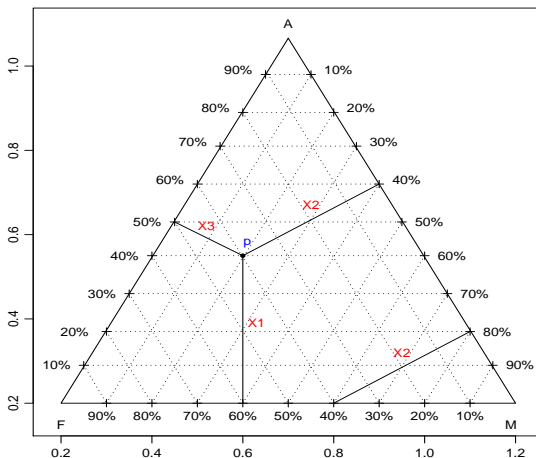


Figura: O ponto  $p$ ,  $(40, 40, 20)\%$  em  $\mathbb{S}^3$  e  $(0, 60; 0, 55)$  em  $\mathbb{R}^2$ , representado no triângulo com lado unitário e triângulo de classificação textural.

# Característica do Diagrama Ternário

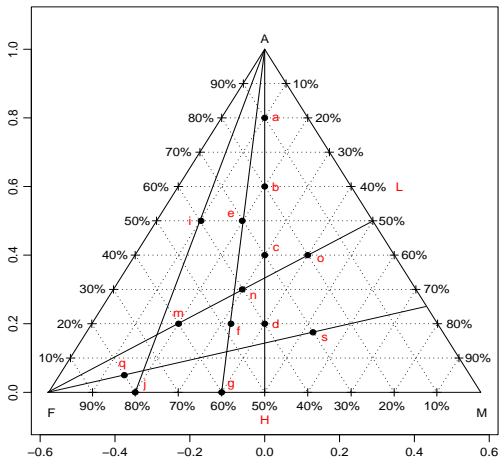


Figura: Diagrama ternário com pontos na mesma linha apresentando mesma razão.



# Análise de Dados Composicionais

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

**Transformação log-razão aditiva (ALR):**

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

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

**Transformação logística generalizada aditiva (AGL):**

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

$$\underline{Y} \longrightarrow \underline{X} = \mathcal{C} \left( \left( \exp \left\{ \ln \left( \frac{X_1}{X_B} \right) \right\}, \dots, \exp\{0\} \right) \right)'$$

Pawlowsky e Olea (2004).



- **Dados composicionais:**
  - Aitchison (1986).
- **Modelos espaciais uni e multivariados:**
  - 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);
    - Lark e Bishop (2007);
  - Tjelmeland e Lund (2003).

# Objetivo Geral

Propor e implementar um modelo geoestatístico para dados espaciais composicionais utilizando estruturas multivariadas como componentes do modelo especificados por função de correlação, e desenvolvendo métodos de inferência baseadas em verossimilhança e sob o enfoque bayesiano.

# Objetivos Específicos

- Fazer modelagem alternativa à de Pawlowsky-Ghlan e Olea (2004);
- Construir um modelo em que a dependência espacial e entre variáveis seja considerada na obtenção de uma função de covariância válida;
- Derivar métodos de inferência baseados em verossimilhança;
- Aplicar métodos bayesianos para a inferência dos parâmetros;
- Desenvolver rotinas computacionais para análise de dados composicionais espacializados;
- Aplicar a metodologia em dados de solo elaborando mapas temáticos de modelos composicionais em estudo de caso.

# Fundamentos Para Análise de Dados Composicionais

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

**Transformação log-razão aditiva (ALR):**

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

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

**Transformação logística generalizada aditiva (AGL):**

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

$$\underline{Y}(\underline{x}) \longrightarrow \underline{X}(\underline{x}) = C \left( \left( \exp \left\{ \ln \left( \frac{X_1(\underline{x})}{X_B(\underline{x})} \right) \right\}, \dots, \exp\{0\} \right)' \right)$$

Pawlowsky e Olea (2004).



# 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;
- $\underline{Y}_{n \times 1}(\underline{x}) = (Y_1(\underline{x}_1), Y_2(\underline{x}_1), \dots, Y_1(\underline{x}_{n_1}), Y_2(\underline{x}_{n_1}))'$ ;
- $U \sim N(\underline{0}; \Sigma_U)$ ,  $\Sigma_U$ : variâncias unitárias e covariâncias em função de  $\rho_U(\phi)$ , exponencial;
- $Z \sim N(\underline{0}; \Sigma_Z)$ ,  $\Sigma_Z$ : variâncias unitárias e covariâncias em função de  $\rho$  induzida pela estrutura composicional.

# Verossimilhança I

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

$$\Sigma = \sigma_1^2 \mathbf{V}$$

- **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:**

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

## Verossimilhança II

- Estimativa obtida por otimização numérica:  $\hat{\underline{\theta}}^* = (\hat{\eta}, \hat{\nu}_1, \hat{\nu}_2, \hat{\phi}, \hat{\rho})'$ ;

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

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

- Hessiano numérico;

- **Método Delta:**

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

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



# Comentários sobre precisão

- Qualidade limitada
  - Aproximações quadráticas - parâmetros de covariância/correlação
- Alternativas: verossimilhança  $\left\{ \begin{array}{l} \textit{perfilhada} \\ \textit{marginal} \\ \textit{condicional} \\ \textit{modificada} \end{array} \right.$
- Tratamento bayesiano.

## Predição em $\mathbb{R}^{B-1}$

- **Cokrigagem** de  $Y_0$  em  $x_0 = (x_{10}, x_{20}, \dots, x_{n_20})$ :

- $\underline{\mu}_{Y_0|Y} = \underline{\mu}_{Y_0} + \Sigma_{Y_0Y} \Sigma_{YY}^{-1} (Y - \underline{\mu}_Y)$

- $\Sigma_{Y_0|Y} = \Sigma_{Y_0Y_0} - \Sigma_{Y_0Y} \Sigma_{YY}^{-1} \Sigma_{YY_0}$

# Predição em $\mathbb{S}^B$ - Aproximação Numérica

## 1. Integração de Gauss-Hermite multivariada de ordem k

- $$\int_{\mathbb{R}^{B-1}} g(\underline{Z}) f(-\underline{Z}'\underline{Z}) d\underline{Z} \approx$$
$$\sum_{i_1=1}^k \sum_{i_2=1}^k \cdots \sum_{i_{B-1}=1}^k \omega_{i_1} \omega_{i_2} \cdots \omega_{i_{B-1}} g(Z_{i_1}, Z_{i_2}, \dots, Z_{i_{B-1}})$$

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)

- Outras referências quadratura de Gauss-Hermite:**
  - Gamerman (2006),
  - Paulino, Turkman e Murteira (2003)

## Predição em $\mathbb{S}^B$ - Aproximação Numérica (cont.)

### Solução numérica - 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  $\Sigma_Y$
- $g_1(\underline{Z}) = \pi^{-\frac{B-1}{2}} \text{agl}(\underline{\mu}_Y + \sqrt{2}R'\underline{Z})$
- $g_2(\underline{Z}) = \pi^{-\frac{B-1}{2}} \left( \text{agl}(\underline{\mu}_Y + \sqrt{2}R'\underline{Z}) - \underline{\mu}_X \right) \left( \text{agl}(\underline{\mu}_Y + \sqrt{2}R'\underline{Z}) - \underline{\mu}_X \right)'$

# Predição em $\mathbb{S}^B$ - Simulação

## 2. Simulação:

- a. gerar dados bivariados de uma distribuição gaussiana multivariada com parâmetros obtidos por cokrigagem;
- b. aplicar a transformação AGL nos dados obtidos no item (a);
- c. construir os mapas de médias de predição para cada componente.

## Funcionais:

- i. máximos e mínimos;
- ii. mediana, 5 e 95% de probabilidade;
- iii. probabilidades a *posteriori* de classificação nas diferentes classes de solo.

# Análise de Frações Granulométricas de Um Solo

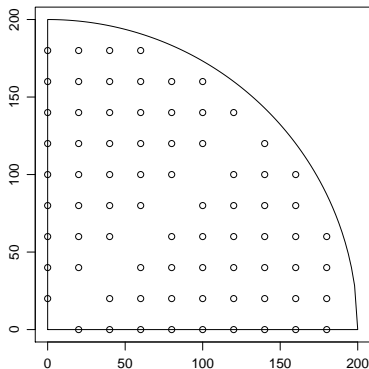


Figura: Distribuição das localizações na área de estudo.

- 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

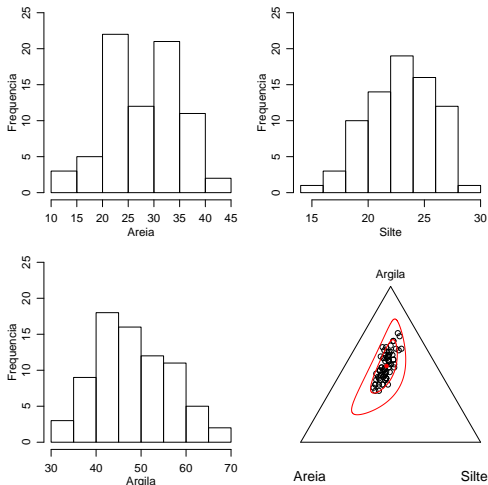


Figura: Distribuição de areia, silte e argila e diagrama ternário das composições.

# Análise de Frações Granulométricas de Um Solo

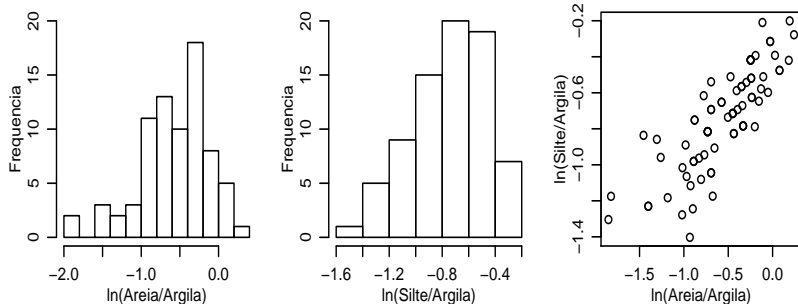


Figura: Distribuição das log-razão e correspondente diagrama de dispersão.



# Análise de Frações Granulométricas de Um Solo

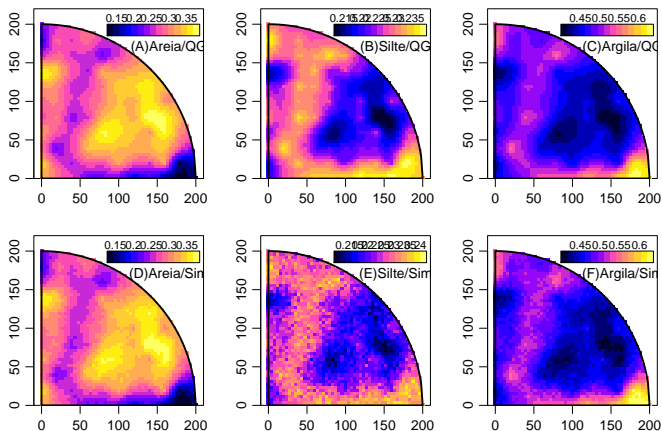
**Tabela:** Estimativas, erros padrão e intervalos de confiança pelo método Delta via método de otimização “L-BFGS-B”.

Parâmetros	Estimativas	Erro Padrão	LI. Delta	LS. Delta
$\mu_1$	-0,759	0,440	-1,200	-0,3195
$\mu_2$	-0,794	0,229	-1,023	-0,5655
$\sigma_1$	0,450	0,121	0,330	0,5712
$\sigma_2$	0,115	0,052	0,064	0,1668
$\tau_1$	0,287	0,045	0,242	0,3321
$\tau_2$	0,267	0,031	0,236	0,2984
$\phi$	66,908	42,554	24,355	109,4623
$\rho$	0,954	0,068	0,886	1,0224

# Análise de Frações Granulométricas de Um Solo

- **Mapas de predição em  $\mathbb{S}^3$** 
  - **Quadratura Gaussiana:** grade de predição de 2500 localizações, ordem de quadratura  $k = 7$ ;
  - **Simulação:** 1000 simulações, grade de predição de 1156 localizações.

# Análise de Frações Granulométricas de Um Solo



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

# Análise de Frações Granulométricas de Um Solo

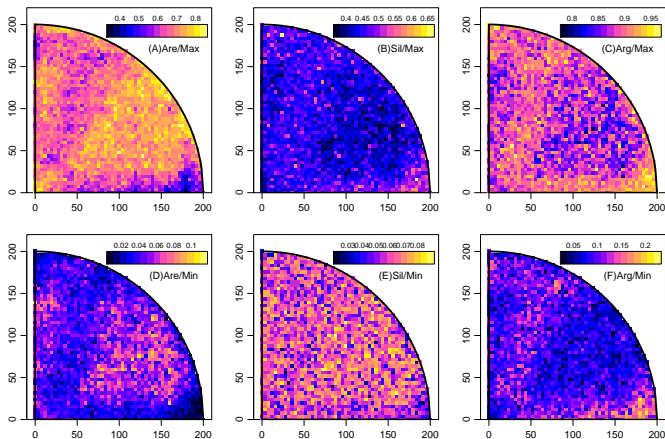


Figura: Mapas das frações máximas e mínimas por componente.

# Análise de Frações Granulométricas de Um Solo

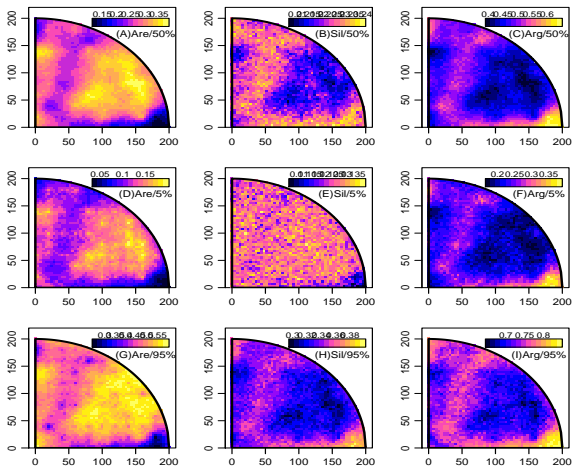


Figura: Mapas das frações medianas, inferiores a 5% e 95% de probabilidade, por componente.



# Análise de Frações Granulométricas de Um Solo

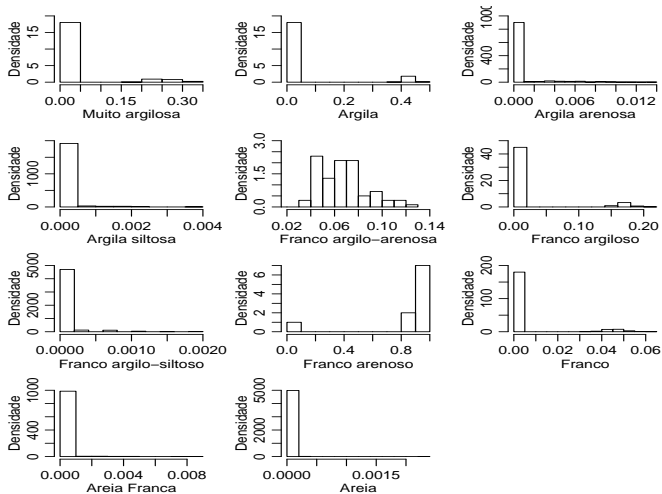


Figura: Probabilidades a *posteriori* de classificação nas diferentes classes de solo.

# Inferência Bayesiana

- $\underline{\theta} = (\mu_1, \mu_2, \sigma_1, \sigma_2, \tau_1, \tau_2, \phi, \rho)'$  → variáveis aleatórias
- **Dist. conjunta** = veros.  $\times$  *priori*:  $P(\underline{Y}, \underline{\theta}) = P(\underline{Y}|\underline{\theta})P(\underline{\theta})$ .
- **Teorema de Bayes**  $\Rightarrow$  distribuição à *posteriori* de  $\underline{\theta}$ :

$$P(\underline{\theta}|\underline{Y}) = \frac{P(\underline{\theta}, \underline{Y})}{P(\underline{Y})} = \frac{P(\underline{Y}|\underline{\theta})P(\underline{\theta})}{P(\underline{Y})} = \frac{P(\underline{Y}|\underline{\theta})P(\underline{\theta})}{\int P(\underline{\theta}, \underline{Y})d\underline{\theta}} \propto P(\underline{Y}|\underline{\theta})P(\underline{\theta}),$$

- **Resumos estatísticos** como esperanças à *posteriori* de funções de  $\underline{\theta}$ :

$$E(f(\underline{\theta})|\underline{Y}) = \frac{\int f(\underline{\theta})P(\underline{\theta})P(\underline{Y}|\underline{\theta})d\underline{\theta}}{\int P(\underline{\theta})P(\underline{Y}|\underline{\theta})d\underline{\theta}} \propto \int f(\underline{\theta})P(\underline{\theta})P(\underline{Y}|\underline{\theta})d\underline{\theta}.$$

# Inferência Bayesiana (cont.)

- **Resolução:**
  - **integração Monte Carlo - cadeias de Markov de Monte Carlo (MCMC):**

Gamerman (2006),  
Lee (2004),  
Gelman, Carlin, Stern e Rubin (2003),  
Gill (2002),  
Gilks, Richardson e Spiegelhalter (1996).



## Inferência Bayesiana (cont.)

- $\underline{\theta} = (\mu_1, \mu_2, \sigma_1, \sigma_2, \tau_1, \tau_2, \phi, \rho)'$
- $\underline{\theta}^* = (\eta, \nu_1, \nu_2, \phi, \rho)$  e  $(\underline{\mu}, \sigma_1^2)$  independentes
- **Distribuição à posteriori** de  $\underline{\theta}$ :

$$P(\underline{\mu}, \sigma_1^2, \underline{\theta}^* | Y) \propto P(Y | \underline{\mu}, \sigma_1^2, \underline{\theta}^*) P(\underline{\mu}, \sigma_1^2) P(\underline{\theta}^*)$$

# Inferência Bayesiana (cont.)

## Prioris:

- **Conjugadas:** levam a uma posteriori da mesma família de distribuições.
- **Casos extremos:**
  - parâmetros conhecidos  $\Rightarrow$  distribuições degeneradas ( $V \equiv 0$ );
  - prioris vagas:
    - prioris **não informativas** (de Jeffrey's) -  $P(\theta) \propto \sqrt{|I(\theta)|}$  ou  $P(\underline{\theta}) \propto |I(\underline{\theta})|^{1/2}$ ;
    - **flat** -  $P(\theta) \propto 1$ ;
    - **imprópia** -  $P(\theta) \propto 1/\sigma^2$ .

Ribeiro Jr e Diggle (1999)

## Prioris:

- $P(\underline{\mu}, \sigma_1^2)$ ;
- $P(\underline{\mu} | \sigma_1^2)$ ;
- $P(\eta)$ ,  $P(\nu_1)$  e  $P(\nu_2)$ ;
- $P(\phi)$ ;
- $P(\rho)$ .

## IB (cont.) - Distribuições a *Posteriori*

- $\int_{\underline{\theta}^*} P(\underline{\mu}, \sigma_1^2, \underline{\theta}^* | \underline{Y}) d\underline{\theta}^* \propto \int_{\underline{\theta}^*} P(\underline{Y} | \underline{\mu}, \sigma_1^2, \underline{\theta}^*) P(\underline{\mu}, \sigma_1^2) P(\underline{\theta}^*) d\underline{\theta}^*$
- $$P(\underline{\mu}, \sigma_1^2 | \underline{Y}) \propto P(\underline{Y} | \underline{\mu}, \sigma_1^2, \underline{\theta}^*) \frac{1}{\sigma_1^2} \int_{\underline{\theta}^*} P(\underline{\theta}^*) d\underline{\theta}^*.$$
- $$\propto \frac{1}{\sigma_1^2} P(\underline{Y} | \underline{\mu}, \sigma_1^2, \underline{\theta}^*)$$
- $[\underline{\mu} | \underline{Y}, \sigma_1^2, \underline{\theta}^*] \sim N(\hat{\underline{\mu}}; \sigma_1^2 (\mathbf{D}'\mathbf{V}^{-1}\mathbf{D})^{-1}) \quad \hat{\underline{\mu}} = (\mathbf{D}'\mathbf{V}^{-1}\mathbf{D})^{-1}(\mathbf{D}'\mathbf{V}^{-1}\underline{Y})$
- $[\sigma_1^2 | \underline{Y}, \underline{\theta}^*] \sim \chi_{\text{Sin}v}^2(n - n_{\hat{\underline{\mu}}}; S^2) \quad S^2 = \frac{(\underline{Y} - \mathbf{D}\hat{\underline{\mu}})' \mathbf{V}^{-1}(\underline{Y} - \mathbf{D}\hat{\underline{\mu}})}{n - n_{\hat{\underline{\mu}}}}.$

## IB (cont.) - Passos Para Inferência Bayesiana

1. Algoritmo Metropolis-Hastings  $\Rightarrow$  matriz de covariância do modelo;

- $\theta^{(t)} \Rightarrow \theta^{(t+1)}$ :  $\theta'$  de  $q(\cdot|\theta^{(t)})$  com probabilidade

$$\alpha(\theta; \theta') = \min \left\{ 1; \frac{P(\underline{Y}|\theta') P(\theta') q(\theta|\theta')}{P(\underline{Y}|\theta) P(\theta) q(\theta'|\theta)} \right\}$$

2. Calcular  $\hat{\underline{\mu}}$  e  $S^2$  e amostrar  $\sigma_1^2$  de  $[\sigma_1^2|\underline{Y}] \sim \chi_{Sinv}^2(n - n_{\underline{\mu}}; S^2)$ ;

3. Calcular  $\text{Var}(\underline{\mu})$ :  $[\underline{\mu}|\underline{Y}, \sigma_1^2] \sim N\left(\hat{\underline{\mu}}; \sigma_1^2 \left(\mathbf{D}'\mathbf{V}_{\underline{Y}}^{-1}\mathbf{D}\right)^{-1}\right)$ ;

4. Amostrar um valor de  $\underline{\mu}$  de  $[\underline{\mu}|\underline{Y}, \sigma_1^2]$ ;

5. Repetir tantas vezes quanto o número de simulações;

## IB (cont.) - Distribuições a *Posteriori*

- $\int_{\underline{\theta}^*} P(\underline{\mu}, \sigma_1^2, \underline{\theta}^* | \underline{Y}) d\underline{\theta}^* \propto \int_{\underline{\theta}^*} P(\underline{Y} | \underline{\mu}, \sigma_1^2, \underline{\theta}^*) P(\underline{\mu}, \sigma_1^2) P(\underline{\theta}^*) d\underline{\theta}^*$
- $P(\underline{\mu}, \sigma_1^2 | \underline{Y}) \propto P(\underline{Y} | \underline{\mu}, \sigma_1^2, \underline{\theta}^*) \frac{1}{\sigma_1^2} \int_{\underline{\theta}^*} P(\underline{\theta}^*) d\underline{\theta}^*.$
- $\propto \frac{1}{\sigma_1^2} P(\underline{Y} | \underline{\mu}, \sigma_1^2, \underline{\theta}^*)$
- $[\underline{\mu} | \underline{Y}, \sigma_1^2, \underline{\theta}^*] \sim N(\hat{\underline{\mu}}; \sigma_1^2 (\mathbf{D}'\mathbf{V}^{-1}\mathbf{D})^{-1}) \quad \hat{\underline{\mu}} = (\mathbf{D}'\mathbf{V}^{-1}\mathbf{D})^{-1}(\mathbf{D}'\mathbf{V}^{-1}\underline{Y})$
- $[\sigma_1^2 | \underline{Y}, \underline{\theta}^*] \sim \chi_{\text{Sinv}}^2(n - n_{\hat{\underline{\mu}}}; S^2) \quad S^2 = \frac{(\underline{Y} - \mathbf{D}\hat{\underline{\mu}})' \mathbf{V}^{-1} (\underline{Y} - \mathbf{D}\hat{\underline{\mu}})}{n - n_{\hat{\underline{\mu}}}}.$

# IB (cont.) - Passos Para Inferência Bayesiana (cont.)

## 6. Predição espacial:

6.1 **Cokrigagem** de  $\underline{Y}_0$  em  $\underline{x}_0 = (\underline{x}_{10}, \underline{x}_{20}, \dots, \underline{x}_{n_20})$  com cada conjunto de parâmetros simulados:

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

6.2 Gerar uma amostra de  $NM \sim (\underline{\mu}_{\underline{Y}_0|\underline{Y}}; \Sigma_{\underline{Y}_0|\underline{Y}})$  para cada conjunto de parâmetros simulados.

6.3 Aplicar transformação AGL.

6.4 Calcular a média das simulações para cada componente.

# IB (cont.) - Diagnóstico de Convergência

## Métodos

- Autocorrelações dos parâmetros a *posteriori*, Gill (2002);
- Gráfico da trajetória da cadeia;
- Gráfico da densidade estimada do parâmetro a *posteriori*;
- Teste de Geweke: Gill (2002), Gamerman e Lopes (2006), teste de *diferença de médias* usando uma aproximação assintótica p/ o erro padrão da diferença. **Preocupação:**  $z_G > |2|$
- Diagnóstico de sequência múltipla de Gelman e Rubin:  
Gill (2002), Gamerman e Lopes (2006),

**Indicador de convergência** - *redução de escala estimado*:

$$\sqrt{F_{RE}} = \sqrt{\frac{\hat{V}ar(\theta)}{V_D}} < 1, 2 \Rightarrow \text{convergência.}$$



# Análise Bayesiana de Frações Granulométricas de Um Solo

- **Prioris:**

- $P(\underline{\mu}, \sigma_1^2) \propto \frac{1}{\sigma_1^2};$

- $P(\underline{\mu}|\sigma_1^2) = 1;$

- $P(\eta), P(\nu_1)$  e  $P(\nu_2)$  → lognormais com médias  
 $\ln(0,24), \ln(0,63), \ln(0,59)$   
e desvios padrão iguais a 0,3

- $P(\phi) = \text{Gama}(66, 1);$

- $P(\rho) = 1$

- nº de simulações= 12000, *burn-in*= 1000 e salto= 10.

# Análise de Frações Granulométricas de Um Solo

**Tabela:** Estimativas obtidas por inferência clássica com respectivos intervalos de 95% de confiança via método Delta e esperanças das 1200 simulações da distribuição à *posteriori* de  $\underline{\theta}$  e intervalos de 95% de credibilidade obtidos por inferência bayesiana.

Par.	Est.Class	LI	LS	Esp.Bayes	LI.cred.	LS.cred
$\mu_1$	-0,759	-1,20	-0,32	-0,764	-1,20	-0,33
$\mu_2$	-0,794	-1,02	-0,57	-0,794	-0,94	-0,67
$\sigma_1$	0,450	0,33	0,57	0,442	0,34	0,56
$\sigma_2$	0,115	0,06	0,17	0,112	0,06	0,18
$\tau_1$	0,287	0,24	0,33	0,287	0,22	0,36
$\tau_2$	0,266	0,24	0,30	0,265	0,23	0,31
$\phi$	66,908	24,35	109,46	65,727	51,53	84,46
$\rho$	0,954	0,89	1,02	0,931	0,83	0,99

# Análise de Frações Granulométricas de Um Solo

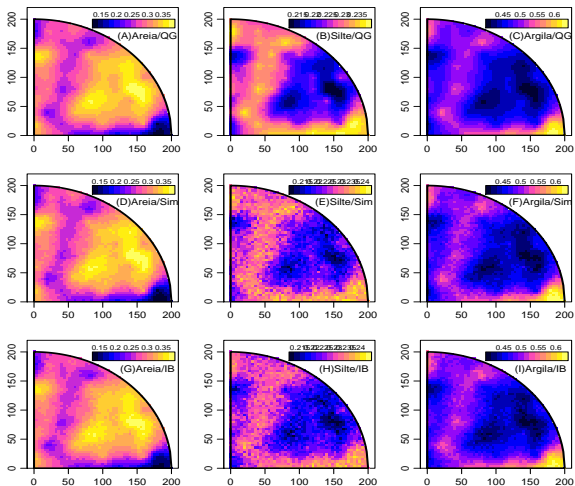


Figura: Mapas das frações de areia, silte e argila obtidos por quadratura gaussiana (GH), por simulação (Sim) e por inferência bayesiana (IB).

# Conclusão

- Modelo captura variações espaciais, induzidas pelas composições e não estruturadas;
- Construção de mapas de areia, silte e argila garantindo que as frações somem 1, nos pontos observados e preditos;
- Metodologia baseada em verossimilhança e declaração explícita do modelo permite fazer inferências sobre parâmetros pelo método clássico e bayesiano considerando nas predições a incerteza associada à estimação dos parâmetros;
- Transformação dos valores preditos para o simplex  $\mathbb{S}^3$  por quadratura de Gauss-Hermite de ordem igual a 7 não diferiram dos obtidos com ordem 20.

## Sugestões de Trabalhos Futuros

1. **Extensões do presente trabalho:** covariáveis, qualidade do modelo, zeros, procedimentos computacionais;
2. **Extensões para outros modelos:** Schmidt e Sansó (2006), pacote *spBayes*;
3. **Aplicações em outras áreas:** proporções de gases que formam efeito estufa, diversidade e riqueza de espécies;
4. **Extensões gerais de estatística espacial:** modelos não estacionários, modelos composicionais espaço-temporais.

# 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.