

# APLICAÇÃO DE MODELO GEOESTATÍSTICO BIVARIADO PARA DADOS COMPOSICIONAIS, NA DESCRIÇÃO DA DISTRIBUIÇÃO ESPACIAL DE FRAÇÕES GRANULOMÉTRICAS DE UM ARGILOSO.

Ana Beatriz Tozzo Martins - PPGMNE/LEG, UFPR e DES, UEM <sup>1 2</sup>

Paulo Justiniano Ribeiro Junior - LEG, UFPR

Antônio Carlos Andrade Gonçalves - DAG, UEM

Wagner Hugo Bonat - LEG, UFPR

## 1 Introdução

Os diversos processos naturais relacionados ao desenvolvimento e à produção vegetal dependem, em grande extensão das características do solo relacionadas com a capacidade de armazenar água, além de permitir a infiltração desta; com a capacidade de reter e disponibilizar nutrientes; com a resistência mecânica e com a capacidade de permitir trocas gasosas no seu espaço poroso, o que depende da amplitude e da geometria deste espaço. Além destas características fundamentais do solo, diversas outras podem também exercer influência nestes processos e todas estão intimamente relacionadas com a proporção segundo a qual os seus elementos constituintes o compõem. Esta proporção caracteriza a textura do solo, sendo expressa por meio da composição relativa das suas frações granulométricas, tradicionalmente caracterizadas como areia, silte e argila. Por ser uma característica física de grande importância e pouco susceptível a mudanças ao longo do tempo, a análise granulométrica do solo é uma das primeiras e principais ações a serem conduzidas, para se inferir a respeito das futuras estratégias de manejo do mesmo. Assim, o conhecimento da proporção segundo a qual as frações granulométricas se apresentam no solo contribui, inclusive, para a sua classificação, tornando-se informação de grande importância para o sistema agrícola.

Com o aperfeiçoamento das técnicas de produção, com o avanço do conhecimento científico em torno do agrossistema e com o aprimoramento das tecnologias envolvidas, notadamente em relação ao maquinário empregado, tem-se buscado, frequentemente, a compreensão de que o sistema agrícola não é homogêneo no espaço, mas apresenta uma distribuição espacial das suas características que não pode ser ignorada, quando se busca a otimização do processo de produção, visando atingir metas fixadas em decorrência de conceitos ambientais e econômicos. Assim, descrever a distribuição espacial dos valores das frações granulométricas do solo consiste em uma base de importância primeira para a descrição espacial das suas características relevantes para o processo de produção. Estudos como o de Gonçalves (1997), dentre inúmeros outros, buscam a descrição da distribuição espacial dos valores de areia, silte e argila do solo, utilizando técnicas de abordagem estatística destas distribuições, notadamente com o emprego de ferramentas geoestatísticas. Isto tem possibilitado a descrição independente da distribuição dos valores de cada fração granulométrica, de forma independente das demais.

---

<sup>1</sup>Agradecimento à CAPES pelo apoio financeiro. Esse trabalho foi parcialmente financiado pela FINEP projeto CT-INFRA/UFPR.

<sup>2</sup>Contato: [abtmartins@uem.br](mailto:abtmartins@uem.br)

Entretanto, como se trata de uma proporção relativa entre as frações granulométricas, a análise textural do solo sempre revela um total de 100%, para o somatório destes componentes. Assim, estes componentes podem ser tratados, à luz das ferramentas estatísticas espaciais, como dados composicionais, ou seja, para os quais o somatório é sempre uma constante, o que implica em variações proporcionais dos mesmos. A partir desta abordagem, a descrição da distribuição espacial das três frações granulométricas pode ser feita levando-se em conta a relação entre elas. Isto tem grande potencial de incrementar a confiabilidade desta descrição. Assim, o objetivo do presente trabalho foi o desenvolvimento de metodologia estatística para a descrição da distribuição espacial dos valores das frações granulométricas de um solo, considerando-os como dados composicionais. Isto inclui estender o uso do modelo geoestatístico bivariado para estruturas de dados composicionais, derivando e implementando estimação baseada na verossimilhança e obtendo preditores espaciais que permitam a construção de mapas de predição das frações do solo, ou funcionais destas, na área de estudo.

A Estatística espacial tem se apresentado como uma área de grande importância, para as mais diversas aplicações, inclusive em ciências agrárias, dentro do contexto da denominada Agricultura de precisão. Trabalhos de grande relevância têm sido desenvolvidos, como os de Matheron(1963), Cressie(1993) e muitos outros. Recentemente, em particular na área de geoestatística, surgiram trabalhos como os de Diggle et. al. (1998), Schabenberger e Pierce (2001) e Diggle e Ribeiro Jr (2007) com uma perspectiva baseada em modelos. Com isto métodos clássicos de inferência baseados em verossimilhança foram aplicados para produzir estimativas mais eficientes dos parâmetros desconhecidos e avaliar a incerteza em predições espaciais.

Trabalhos realizados por Aitchison (1986) em análise de dados composicionais apresentam uma metodologia adequada para analisar dados caracterizados por se apresentarem em forma de proporções e somarem um. Recentemente, este tipo de dado vem sendo analisado considerando-se a espacialização das variáveis (PAWLOWSKY-GLAHN e OLEA, 2004).

Este estudo é motivado pelo interesse em modelar e descrever o padrão espacial de dados composicionais de solo, combinando a teoria de dados composicionais com métodos geoestatísticos. A teoria de dados composicionais desenvolvida por Aitchison (1986) para amostras independentes, e estendida por Pawlowsky-Glahn e Olea (2004) considera a dependência espacial segundo métodos geoestatísticos, em uma abordagem que evita declarar completa e explicitamente o modelo multivariado espacial associado às composições. Por outro lado, modelagens multivariadas espaciais descritas em Diggle e Ribeiro Jr (2007), Schmidt e Gelfand (2003), Banerjee, Carlin e Gelfand (2004), Schmidt e Sansó (2006) garantem, por construção, matrizes de covariância definidas positiva. Considera-se aqui o modelo Gaussiano bivariado de componentes comum proposto por Diggle e Ribeiro Jr (2007) e adotado por Bognola et. al. (2008), este no contexto de inventário florestal utilizando uma variável física como informação secundária.

## 2 Material e Métodos

### 2.1 Abordagem estatística

Dados composicionais consistem de vetores, denominados composições, cujos componentes  $X_1, \dots, X_B$  representam frações de algum “todo”, e satisfazem a restrição de que a soma dos componentes é igual a 1 (AITCHISON, 1986), ou seja,

$$X_1 \geq 0, X_2 \geq 0, \dots, X_B \geq 0, \quad \text{e} \quad X_1 + X_2 + \dots + X_B = 1.$$

O espaço amostral para este tipo de dados é o simplex unitário de dimensão  $B$ , uma parte restrita do espaço real.

Um vetor  $\mathbf{W}$  cujos componentes são positivos e medidos na mesma escala denomina-se base e pode se tornar uma composição através do operador fechamento,  $\mathcal{C}$ , dividindo-se cada componente pela soma de todos os componentes. Esta operação garante que a restrição de soma igual

a um seja satisfeita:

Neste espaço amostral, o simplex, as operações matemáticas de soma e multiplicação definidas no espaço real equivalem às operações perturbação

$$\mathbf{X}_1 \oplus \mathbf{X}_2 = (X_{11}, X_{12}, \dots, X_{1B}) \oplus (X_{21}, X_{22}, \dots, X_{2B}) = \mathcal{C}(X_{11}X_{21}, X_{12}X_{22}, \dots, X_{1B}X_{2B}),$$

e potência

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

respectivamente, e a média passa a ser a média geométrica  $g(\mathbf{X}_1) = \sqrt[B]{\prod_{j=1}^B X_{1j}}$ .

Uma característica desse tipo de dados é que estes apresentam um efeito de correlação espúria. A restrição de que a soma dos componentes deve ser igual a 1, implica em correlação negativa entre os componentes fazendo com que as correlações não sejam diretamente interpretáveis (GRAF, 2006), ou seja, as covariâncias estão sujeitas a controles não estocásticos o que implica, segundo Pawlowsky-Glahn e Olea (2004), em singularidade da matriz de covariância de uma composição. Com isto, a aplicação de técnicas estatísticas padrão podem levar a resultados inconsistentes. Para contornar este problema, Aitchison (1986) propôs, dentre outras, a transformação razão log-aditiva (alr) que generaliza a transformação logística para um vetor composicional de duas partes e é dada por:

$$\begin{aligned} \text{alr} : \mathbb{S}^B &\longrightarrow \mathbb{R}^{B-1} \\ \mathbf{X} &\longrightarrow \text{alr}(\mathbf{X}) = \left( \ln \left( \frac{X_1}{X_B} \right), \dots, \ln \left( \frac{X_{B-1}}{X_B} \right) \right)'. \end{aligned}$$

Por outro lado, a transformação inversa denominada transformação logística generalizada aditiva (agl) é dada por

$$\begin{aligned} \text{agl} : \mathbb{R}^{B-1} &\longrightarrow \mathbb{S}^B \\ \text{alr}(\mathbf{X}) &\longrightarrow \text{agl}(\text{alr}(\mathbf{X})) = \mathbf{X} = \left( \exp \left( \ln \left( \frac{X_1}{X_B} \right) \right), \dots, \exp(0) \right)'. \end{aligned}$$

A representação gráfica de uma amostra de composições pode ser feita através do diagrama ternário, por exemplo no caso em que  $B = 3$ , um triângulo equilátero cujos vértices representam os três componentes da composição (BUTLER, 2008). De acordo com Graf (2006), considerando que assintoticamente  $\mathbf{Y} = (\ln(X_1/X_B), \ln(X_2/X_B), \dots, \ln(X_{B-1}/X_B))'$  tem distribuição Gaussiana  $N_{B-1}(\mu_L; \Sigma_L)$  em que  $\mu_L$  é o vetor de médias das log-razões e  $\Sigma_L$  a matriz de covariâncias de ordem  $(B-1) \times (B-1)$ , o domínio de confiança para  $\mathbf{Y}$  é limitado pelo elipsóide

$$B'_{1-\alpha}(\mathbf{Y}) = \{ \mathbf{Y} \in \mathbb{R}^{B-1} \mid (\ln(\mathbf{Y}) - \mu_L)' \Sigma_L^{-1} (\ln(\mathbf{Y}) - \mu_L) \leq \chi_{B-1; 1-\alpha}^2 \}$$

em que  $\chi_{B-1; 1-\alpha}^2$  é o quantil  $(1-\alpha)$  da distribuição qui-quadrado com  $B-1$  graus de liberdade e o domínio correspondente para  $\mathbf{X} = (X_1, X_2, \dots, X_B)$  é um subconjunto do simplex  $\mathbb{S}^B$

$$B'_{1-\alpha}(\mathbf{X}) = \left\{ \mathbf{X} \in \mathbb{S}^B \mid \left( \ln \left( \frac{\mathbf{X}_{-B}}{X_B} \right) - \mu_L \right)' \Sigma^{-1} \left( \ln \left( \frac{\mathbf{X}_{-B}}{X_B} \right) - \mu_L \right) \leq \chi_{B-1; 1-\alpha}^2 \right\}.$$

Para  $\mathbf{X} = (X_1, \dots, X_B)'$  uma composição com  $B$  componentes e  $\mathbf{Y} = \left( \ln \left( \frac{X_1}{X_B} \right), \dots, \ln \left( \frac{X_{B-1}}{X_B} \right) \right)'$  um vetor com  $B-1$  elementos, o modelo geoestatístico com componente comum pode ser obtido seguindo a formulação dada em Diggle e Ribeiro Jr (2007). Neste trabalho, considera-se composições de apenas 3 componentes,  $X_1 = \text{Areia}$ ,  $X_2 = \text{Silte}$  e  $X_3 = \text{Argila}$ , que resultam em vetores bivariados.

A partir do modelo geoestatístico apresentado em Diggle e Ribeiro Jr (2007), propõe-se uma

adaptação e o modelo pode ser escrito como:

$$\begin{cases} Y_1(\underline{x}_i) &= \mu_1(\underline{x}_i) + \sigma_1 U(\underline{x}_i; \phi) + Z_1(\underline{x}_i) \\ Y_2(\underline{x}_{i'}) &= \mu_2(\underline{x}_{i'}) + \sigma_2 U(\underline{x}_{i'}; \phi) + Z_2(\underline{x}_{i'}). \end{cases}$$

em que  $\underline{x}_i, \underline{x}_{i'} \in \mathbb{R}^2$ , são as localizações amostrais  $i, i' = 1, \dots, n_1$ , onde  $n_1$  é o tamanho da amostra;  $Y_1 = \ln(\text{Areia}/\text{Argila})$ ,  $Y_2 = \ln(\text{Silte}/\text{Argila})$  são as variáveis resposta do modelo de modo que  $\underline{Y}_{n \times 1} = (Y_1(\underline{x}_1), Y_2(\underline{x}_1), \dots, Y_1(\underline{x}_{n_1}), Y_2(\underline{x}_{n_1}))'$ . Neste modelo, assume-se que  $U$  é um efeito aleatório de média zero e variância unitária com distribuição Gaussiana multivariada com correlações espaciais dadas pela função de correlação exponencial ( $\rho_U$ ), caracterizada por um parâmetro  $\phi$  que controla o decaimento da correlação como função da separação espacial entre duas localizações. No modelo bivariado geral as unidades de medida são preservadas nas constantes padronizadoras  $\sigma_1$  e  $\sigma_2$ , enquanto que no contexto considerado aqui são adimensionais. Os efeitos aleatórios  $Z_j \sim N(0; \tau_j)$ ,  $j = 1, 2$  capturam a variabilidade não espacial incluindo a correlação ( $\rho$ ) induzida pela estrutura composicional.

Sendo assim,  $\underline{Y} \sim N_2(\underline{\mu}; \underline{\Sigma})$ , com  $\underline{\mu} = (\mu_1, \mu_2)'$  e matriz de covariâncias  $\underline{\Sigma}$  dada por

$$\underline{\Sigma} = \begin{bmatrix} \sigma_1^2 + \tau_1^2 & \sigma_1\sigma_2 + \tau_1\tau_2\rho & \sigma_1^2\rho_U(\underline{x}_1, \underline{x}_2) & \sigma_1\sigma_2\rho_U(\underline{x}_1, \underline{x}_2) \cdots \sigma_1^2\rho_U(\underline{x}_1, \underline{x}_n) & \sigma_1\sigma_2\rho_U(\underline{x}_1, \underline{x}_n) \\ \sigma_1\sigma_2 + \tau_1\tau_2\rho & \sigma_2^2 + \tau_2^2 & \sigma_1\sigma_2\rho_U(\underline{x}_1, \underline{x}_2) & \sigma_2^2\rho_U(\underline{x}_1, \underline{x}_2) \cdots \sigma_1\sigma_2\rho_U(\underline{x}_1, \underline{x}_n) & \sigma_2^2\rho_U(\underline{x}_1, \underline{x}_n) \\ \sigma_1^2\rho_U(\underline{x}_2, \underline{x}_1) & \sigma_1\sigma_2\rho_U(\underline{x}_2, \underline{x}_1) & \sigma_1^2 + \tau_1^2 & \sigma_1\sigma_2 + \tau_1\tau_2\rho \cdots \sigma_1^2\rho_U(\underline{x}_2, \underline{x}_n) & \sigma_1\sigma_2\rho_U(\underline{x}_2, \underline{x}_n) \\ \sigma_1\sigma_2\rho_U(\underline{x}_2, \underline{x}_1) & \sigma_2^2\rho_U(\underline{x}_2, \underline{x}_1) & \sigma_1\sigma_2 + \tau_1\tau_2\rho & \sigma_2^2 + \tau_2^2 \cdots \sigma_1\sigma_2\rho_U(\underline{x}_2, \underline{x}_n) & \sigma_2^2\rho_U(\underline{x}_2, \underline{x}_n) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_1^2\rho_U(\underline{x}_n, \underline{x}_1) & \sigma_1\sigma_2\rho_U(\underline{x}_n, \underline{x}_1) & \sigma_1^2\rho_U(\underline{x}_n, \underline{x}_2) & \sigma_1\sigma_2\rho_U(\underline{x}_n, \underline{x}_2) \cdots \sigma_1^2 + \tau_1^2 & \sigma_1\sigma_2 + \tau_1\tau_2\rho \\ \sigma_1\sigma_2\rho_U(\underline{x}_n, \underline{x}_1) & \sigma_2^2\rho_U(\underline{x}_n, \underline{x}_1) & \sigma_1\sigma_2\rho_U(\underline{x}_n, \underline{x}_2) & \sigma_2^2\rho_U(\underline{x}_n, \underline{x}_2) \cdots \sigma_1\sigma_2 + \tau_1\tau_2\rho & \sigma_2^2 + \tau_2^2 \end{bmatrix}$$

Desta forma, a inferência sobre o vetor de parâmetros  $\underline{\theta} = (\mu_1, \mu_2, \sigma_1, \sigma_2, \tau_1, \tau_2, \phi, \rho)'$  é feita usando a teoria da verossimilhança cuja função de verossimilhança é

$$L(\underline{\theta}, \underline{Y}) = (2\pi)^{-n/2} |\underline{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\underline{Y} - \underline{\mu}_{\underline{Y}})' \underline{\Sigma}^{-1} (\underline{Y} - \underline{\mu}_{\underline{Y}}) \right\}.$$

O processo de maximização desta função, bem como a obtenção dos intervalos de confiança para as estimativas de máxima verossimilhança são descritas em Martins et. al. (2009).

O próximo passo é a realização da predição espacial de  $\underline{Y}_0$  em localizações não amostradas  $\underline{x}_0 = (\underline{x}_{10}, \underline{x}_{20}, \dots, \underline{x}_{n_20})$ . A cokrigagem se baseia nos seguintes resultados da distribuição da distribuição Gaussiana multivariada (DIGGLE e RIBEIRO JR, 2007):

$$\underline{\mu}_{\underline{Y}_0|\underline{Y}} = \underline{\mu}_{\underline{Y}_0} + \underline{\Sigma}_{\underline{Y}_0\underline{Y}} \underline{\Sigma}_{\underline{Y}\underline{Y}}^{-1} (\underline{Y} - \underline{\mu}_{\underline{Y}}) \quad \text{e} \quad \underline{\Sigma}_{\underline{Y}_0|\underline{Y}} = \underline{\Sigma}_{\underline{Y}_0\underline{Y}_0} - \underline{\Sigma}_{\underline{Y}_0\underline{Y}} \underline{\Sigma}_{\underline{Y}\underline{Y}}^{-1} \underline{\Sigma}_{\underline{Y}\underline{Y}_0}. \quad (1)$$

Como os valores de  $\underline{\mu}_{\underline{Y}_0}$  bem como os elementos de  $\underline{\Sigma}$  são desconhecidos, estes são substituídos pelo vetor de médias e matriz de covariância estimadas obtidas no processo de otimização.

Uma vez que a transformação alr foi aplicada aos dados originais e o procedimento de estimação e cokrigagem foi realizada com os dados transformados em  $\mathbb{R}^2$ , deve-se fazer a transformação de volta do vetor de médias e da matriz de covariância para o espaço amostral original, o simplex  $\mathbb{S}^3$ . Uma forma de transformação de volta é descrita em Pawlowsky e Olea (2004) utilizando uma aproximação por quadratura de Gauss-Hermite. O objetivo é calcular para cada localização uma estimativa de

$$\underline{\mu}_{\underline{X}} = E(\underline{X}) = \int_{\mathbb{S}^B} \underline{X} f(\underline{X}) d\underline{X}$$

e

$$\underline{\Sigma}_{\underline{X}} = Cov(\underline{X}, \underline{X}) = \int_{\mathbb{S}^B} (\underline{X} - \underline{\mu}_{\underline{X}})(\underline{X} - \underline{\mu}_{\underline{X}})' f(\underline{X}) d\underline{X}.$$

que no espaço amostral transformado tornam-se

$$\mu_X = \int_{\mathbb{R}^{B-1}} g_1(\mathbf{Z}) f(-\mathbf{Z}'\mathbf{Z}) d\mathbf{Z}$$

e

$$\Sigma_X = \int_{\mathbb{R}^{B-1}} g_2(\mathbf{Z}) f(-\mathbf{Z}'\mathbf{Z}) d\mathbf{Z}$$

Estas integrais são resolvidas por aproximações numéricas, em particular, aproximadas pela integração de Gauss-Hermite multivariada de ordem  $k$ :

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

Nesta aproximação, os pesos  $\omega_{i_1} \omega_{i_2} \cdots \omega_{i_{B-1}}$  e as abscissas  $Z_{i_1}, Z_{i_2}, \dots, Z_{i_{B-1}}$  são conhecidas e seus valores podem ser encontrados, por exemplo, em Abramowitz e Stegun (1972, p. 924). Segundo Gammerman (1997) ordens de quadratura de 6 a 8 são suficientes. Maiores detalhes do procedimento de transformação de volta dos valores preditos para o simplex são encontrados em Martins et al. (2009).

Uma alternativa possível seria fazer a transformação de volta dos valores preditos por cokrigagem através de simulação conforme Diggle e Ribeiro Jr (2007). Primeiramente, para cada localização de predição geraria-se uma amostra aleatória  $\tilde{\mathbf{Z}} = (\mathbf{Z}_1, \mathbf{Z}_2)$  de uma distribuição Gaussiana com média zero e variância um. Em seguida, para cada localização teria-se a matriz de covariância ( $2 \times 2$ ) obtida da matriz de covariância de cokrigagem que seria decomposta utilizando-se a fatoração Cholesky,  $\Sigma = \mathbf{A}\mathbf{A}'$  ou decomposição em valores singulares, por exemplo (BURDEN, 2003). Com isto aplicaria-se a transformação linear  $\mathbf{Y}_0 = \mathbf{A}\tilde{\mathbf{Z}}$ . Este procedimento seria repetido  $n_s$  vezes para cada localização obtendo-se uma distribuição Gaussiana bivariada para cada localização. Por último, com a transformação agl voltaria-se para a escala original dos dados podendo-se então calcular estatísticas bem como probabilidades de interesse.

Finalmente, a predição linear espacial de dados composicionais pode ser resumida nos seguintes passos:

1. Escrever o modelo geoestatístico bivariado para dados composicionais;
2. Escrever a função de verossimilhança para o modelo;
3. Maximizar a função de verossimilhança com a utilização do algoritmo de otimização “L-BFGS-B”, método de Byrd et al. (1995), obtendo-se as estimativas dos parâmetros;
4. Fazer a cokrigagem obtendo os valores preditos no  $\mathbb{R}^2$ ;
5. Utilizar a integração de Gauss-Hermite para obter os valores preditos na escala original,  $\mathbb{S}^3$ ;
4. Construir mapas dos valores esperados preditos de areia, silte e argila.

## 2.2 Conjunto de dados empregados

Os dados empregados para a elaboração do presente trabalho foram obtidos a partir do trabalho de Gonçalves (1997). Em uma área irrigada por pivô-central na Fazenda Areão, Figura 1, pertencente ao campus da Escola superior de agricultura – Luiz de Queiroz, foi demarcado um quadrante na porção mais elevada da área, no qual 76 amostras de solo na profundidade entre 0 e 0,20m foram obtidas em uma malha regular de amostragem, quadrada, de lado igual a 20 metros. Nestas amostras foram medidos os valores das frações granulométricas, areia, silte e argila, dentre os valores de outras variáveis físico-químicas não consideradas aqui.



Figura 1: Foto aérea do campo experimental de irrigação da ESALQ-USP com área de estudo correspondente ao quadrante irrigado por um sistema pivô-central.

Os gráficos da Figura 2 mostram as localizações amostradas com círculos, triângulos, diagonais verticais e cruzadas correspondendo aos quartis das distribuições dos valores das frações granulométricas. Para areia os valores dos quartis foram 24, 29,5, 33 e 42%. Para silte, os percentuais observados foram 21, 23, 25 e 30%, e para argila 42, 48, 55 e 70%.

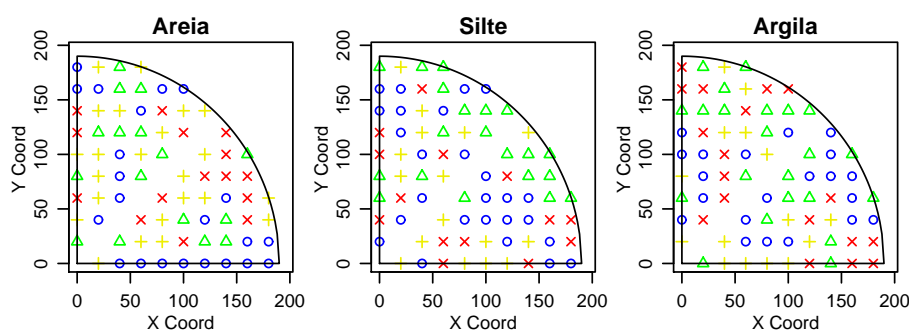


Figura 2: Primeiro (círculos), segundo (triângulos), terceiro (diagonais verticais) e quarto (diagonais cruzadas) quartil da porcentagem de areia (à esquerda), silte (meio) e argila (à direita) nas localizações amostrais.

### 3 Resultados e Discussão

As distribuições dos valores de areia, silte e argila na escala original podem ser vistos nos histogramas da Figura 3. Do diagrama ternário observa-se que nas composições amostradas, o silte foi o componente que apresentou o menor valor devido o afastamento dos pontos deste vértice. É possível notar também que a variabilidade da razão argila/areia foi maior que a razão silte/areia pela amplitude de variação dos pontos projetados sobre os lados que ligam os vértices areia-argila e areia-silte. O mesmo raciocínio quanto a variabilidade das razões pode ser aplicado aos outros componentes. As regiões de 2 e 4-desvios-padrão de confiança são delimitadas pelas

isolinhas que circundam os pontos.

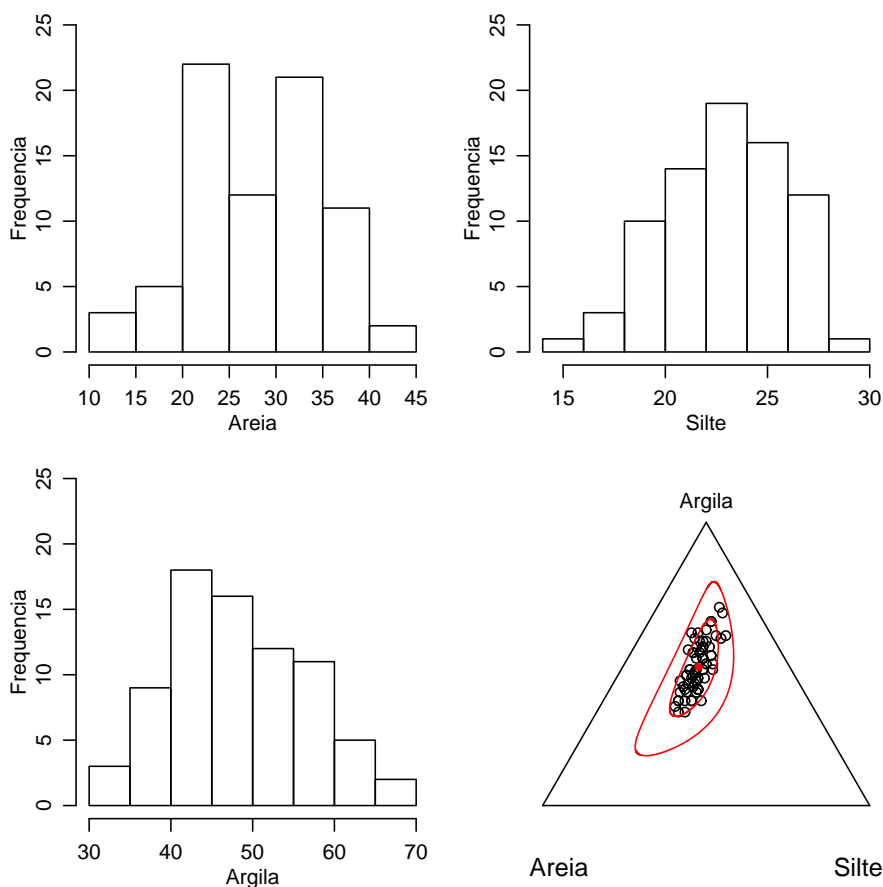


Figura 3: Distribuição de areia, silte e argila e diagrama ternário das composições.

Com a transformação  $\ln$  o espaço amostral passou a ser o  $\mathbb{R}^2$  e as novas variáveis  $Y_1 = \ln(\text{Areia}/\text{Argila})$  e  $Y_2 = \ln(\text{Silte}/\text{Argila})$  apresentaram distribuição aproximadamente Gaussiana (Figura 4). Nesta figura, o diagrama de dispersão evidenciou uma correlação linear positiva entre  $Y_1$  e  $Y_2$ .

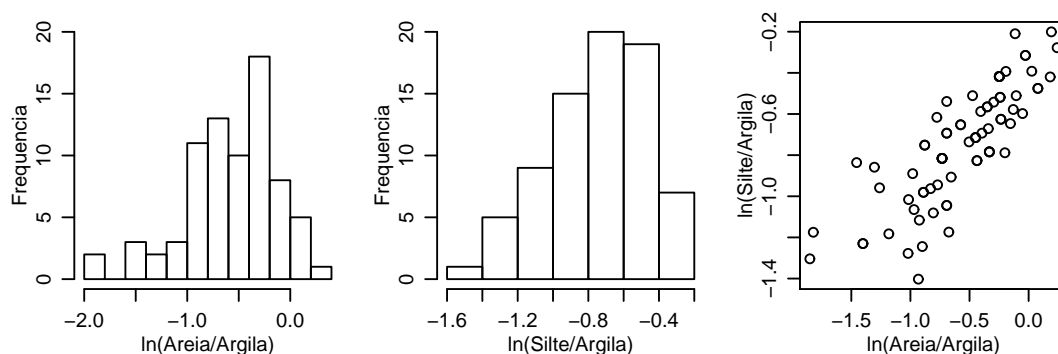


Figura 4: Distribuição das log-razão e correspondente diagrama de dispersão.

Usando a proposta de Martins et. al. (2009), as estimativas obtidas para os parâmetros e os respectivos intervalos de confiança do modelo proposto são apresentados na Tabela 1. Pode-se observar que para cada localização existe uma dependência espacial até uma distância de aproximadamente 70 metros e uma correlação forte entre  $Y_1$  e  $Y_2$ . Para estes parâmetros, os

valores altos implicaram no afastamento dos pontos do vértice argila e o não espalhamento dos pontos no diagrama, respectivamente. As estimativas das médias negativas e quase de mesma magnitude fizeram com que a nuvem de pontos se localizasse no centro do diagrama e aproximação dos pontos para o vértice areia e não para o silte ocorreu devido a  $\sigma_2 < \sigma_1$ . Por último, os baixos valores para  $\tau_1$  e  $\tau_2$  fizeram com que os dados não se deslocassem para os lados do triângulo.

Tabela 1: Estimativas, erros padrão e intervalos de confiança do modelo.

parâmetros	Estimativas	Erro padrão	LI. Delta	LS. Delta
$\mu_1$	-0,7598748	0,44029631	-1,20017113	-0,3195785
$\mu_2$	-0,7941109	0,22857184	-1,02268277	-0,5655391
$\sigma_1$	0,4504836	0,12080170	0,32968188	0,5712853
$\sigma_2$	0,1152904	0,05153622	0,06375416	0,1668266
$\tau_1$	0,2870196	0,04510491	0,24191470	0,3321245
$\tau_2$	0,2669836	0,03142454	0,23555903	0,2984081
$\phi$	66,9084688	42,55386352	24,35460526	109,4623323
$\rho$	0,9544251	0,06803953	0,88638554	1,0224646

No processo de otimização os valores iniciais para o vetor de médias foram considerados como as médias dos valores observados para  $Y_1$  e  $Y_2$ . Metade da variância calculada para  $Y_1$  foi atribuída para o efeito espacial e a outra metade para o efeito composicional. Da mesma forma, procedeu-se com  $Y_2$ . O valor inicial para  $\rho$  foi calculado como o coeficiente de correlação de Pearson e  $\phi = \min + 0,2(\max - \min)$ , onde “min” e “max” foram, respectivamente, a menor e maior distância entre duas localizações. Estes valores resultaram numa log-verossimilhança de  $-42,5541$ . Após o processo de otimização, a log-verossimilhança maximizada foi  $0,8292$ .

As estimativas de máxima verossimilhança foram substituídas em (1) obtendo-se para cada localização, como resultado da cokrigagem, o vetor de médias e a matriz de covariância no espaço  $\mathbb{R}^2$ . A volta dos valores preditos para o simplex  $S^3$  foi feita em uma grade constituída de 2500 pontos usando aproximação de Gauss-Hermite com ordem de quadratura  $k = 7$  (Figura 5).

Pode-se observar na extremidade direita dos mapas de argila que houve um crescimento do teor de argila devido ao fato de que nesta região inicia-se um processo de transição de um argissolo para um nitossolo. Esta ocorrência pode ser justificada pelo fato já observado no diagrama ternário de que as porcentagens para este componente são as menores em relação a areia e argila. O tempo computacional gasto para a construção dos mapas usando quadratura Gaussiana foi de um dia, em média.

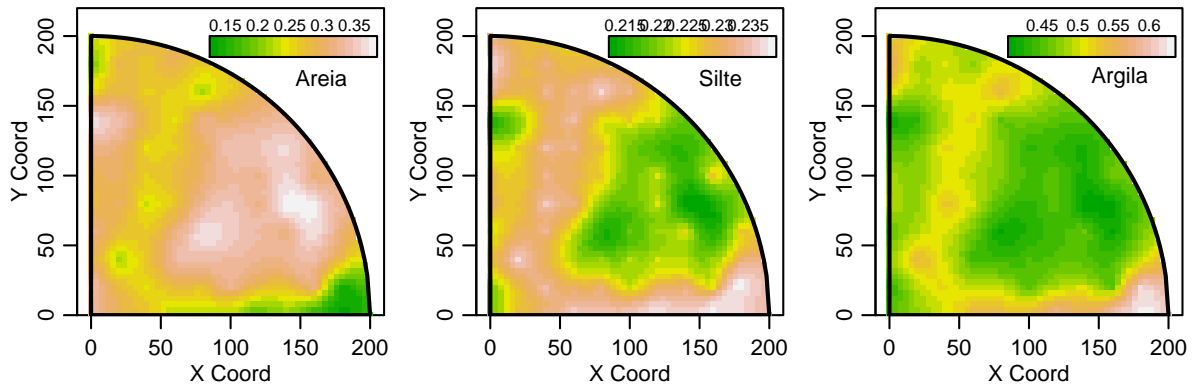


Figura 5: Mapas das porcentagens de areia, silte e argila obtidos por quadratura de Gauss-Hermite.

Todo o trabalho foi realizado utilizando recursos de *software* livre em ambiente operacional



GNU/Linux; no ambiente estatístico R (R Development Core Team, 2008), utilizando o pacote *geoR* (RIBEIRO JR; DIGGLE, 2001), *compositions* (BOOGAART; TOLOSANA; BREN, 2008), *statmod* (SMYTH; HU; DUNN, 2009) e rotinas desenvolvidas especificamente para o desenvolvimento deste trabalho.

## 4 Conclusão

Os procedimentos adotados permitiram descrever, com qualidade, no domínio de um argis-solo a distribuição dos teores de areia, silte e argila. Isto foi feito através da construção de mapas de areia, silte e argila por uma metodologia que implicitamente garante a restrição de que as frações somem um, não só nos pontos observados como nos pontos preditos.

O modelo proposto captura variações espaciais, induzidas pelas composições e não estruturadas.

A declaração explícita do modelo permite que sejam feitas inferências sobre parâmetros de forma usual.

Há a necessidade de se continuar investigando a interação solo-água bem como alternativas para computação mais eficiente e considerar outras formas de especificação do modelo multivariado para o caso de maiores números de componentes.

## Referências

- [1] ABRAMOWITZ, M.; STEGUN, I.A., *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, USA: Milton Abramowitz and Irene A. Stegun, 1972.
- [2] AITCHISON, J., *The statistical analysis of compositional data*, New Jersey: The Blackburn Press, 1986.
- [3] AZZALINI, A., *Statistical inference based in the likelihood*, London: Chapman & Hall, 1996.
- [4] BANERJEE, S.; CARLIN, B.P.; GELFAND, G.E., *Hierarchical modelling and analysis for spatial data*, Boca Raton: Chapman and Hall, 2004.
- [5] BOGNOLA, I.A.; RIBEIRO Jr, P.J.; SILVA, E.A.A; LINGNAU, C., HIGA, A.R. Modelagem uni e bivariada da variabilidade espacial de rendimento de pinus taeda l., *Floresta*, v. 38, n. 2, p. 373-385, 2008.
- [6] BOOGAART, G.v.d.; TOLOSANA, R.;BREN, M., *compositions: compositional data analysis*. Disponível em: <<http://www.stat.boogaart.de/compositions>>. Acesso em 25 maio 2009.
- [7] BUTLER, A.; GLASBEY, C., A latent Gaussian model for compositional data with zeros, *Journal of the Royal Statistical Society, Series C*, v. 57, n. 5, p. 505-520, 2008.
- [8] BURDEN, R.L.; FAIRES, J.D., *Análise numérica*, são Paulo: Pioneira Thomson Learning, 2003, 736p.
- [9] BYRD, R.H.; LU, P.; NOCEDAL, J.; ZHU, C., A limited memory algorithm for bound constrained optimization. *SIAM J. Scientific Computing*, v. 16, p. 1190-1208, 1995.
- [10] COX, D.R.; HINKLEY, D.V., *Theoretical Statistics*, London: Chapman & Hall, 1974.
- [11] DEGROOT, M.H.; SCHERVISH, M.J., *Probability and Statistics*, USA: Addison-Wesley, Third Edition, 2002.

- [12] DIGGLE, P.J.; RIBEIRO JR, P.J., *Model-based geostatistics*, USA: Springer Series in Statistics, 2007.
- [13] GAMMERMAN, D., *Markov chain monte carlo*, London: Chapman and Hall, 1997.
- [14] 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, 1997.
- [15] GRAF, M., Precision of compositional data in a stratified two-stage cluster sample: comparison of the swiss earnings structure survey 2002 and 2000. *Survey Research Methods Section, ASA*, Session 415: Sample Survey Quality V, p. 3066–3072, 2006.
- [16] MARTINS, A.B.T., RIBEIRO JR, P.J.; BONAT, W.A., Um modelo geoestatístico bivariado para dados composicionais. *Revista Brasileira de Biometria*, Session 415: Sample Survey Quality V, p. 3066–3072, 2009.
- [17] PAWITAN, Y., *In all likelihood: statistical modelling and inference using likelihood*, New York: Oxford University Press, 2001.
- [18] PAWLOWSKY-GLAHN, V.; OLEA, R.A., *Geostatistical analysis of compositional data*, New York: Oxford University Press, Inc., 2004.
- [19] R Development Core Team., *R: a language and environment for statistical computing*, Vienna, Austria, 2009. Disponível em:<<http://www.R-project.org>>. Acesso em: 25 maio 2009.
- [20] RIBEIRO JR, P.J.; DIGGLE, P.J., geoR: a package for geostatistical analysis, *R-NEWS*, v. 1, n. 2, p. 14-18, 2001. Disponível em:<<http://CRAN.R-project.org/doc/Rnews/>>. Acesso em 25 maio 2009.
- [21] 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.
- [22] SCHMIDT, A.M.; SANSÓ, B., Modelagem bayesiana da estrutura de covariância de processos espaciais e espaço temporais, In: *17 SINAPE e ABE-Associação Brasileira de estatística*, Caxambu: Associação Brasileira de estatística, 2006. Minicurso.
- [23] SMYTH, G.; HU, Y.; DUNN, P., *statmod: statistical modeling*. Disponível em:<<http://www.statsci.org/r>>. Acesso em 25 maio 2009.

## MAPAS DE FUNCIONAIS

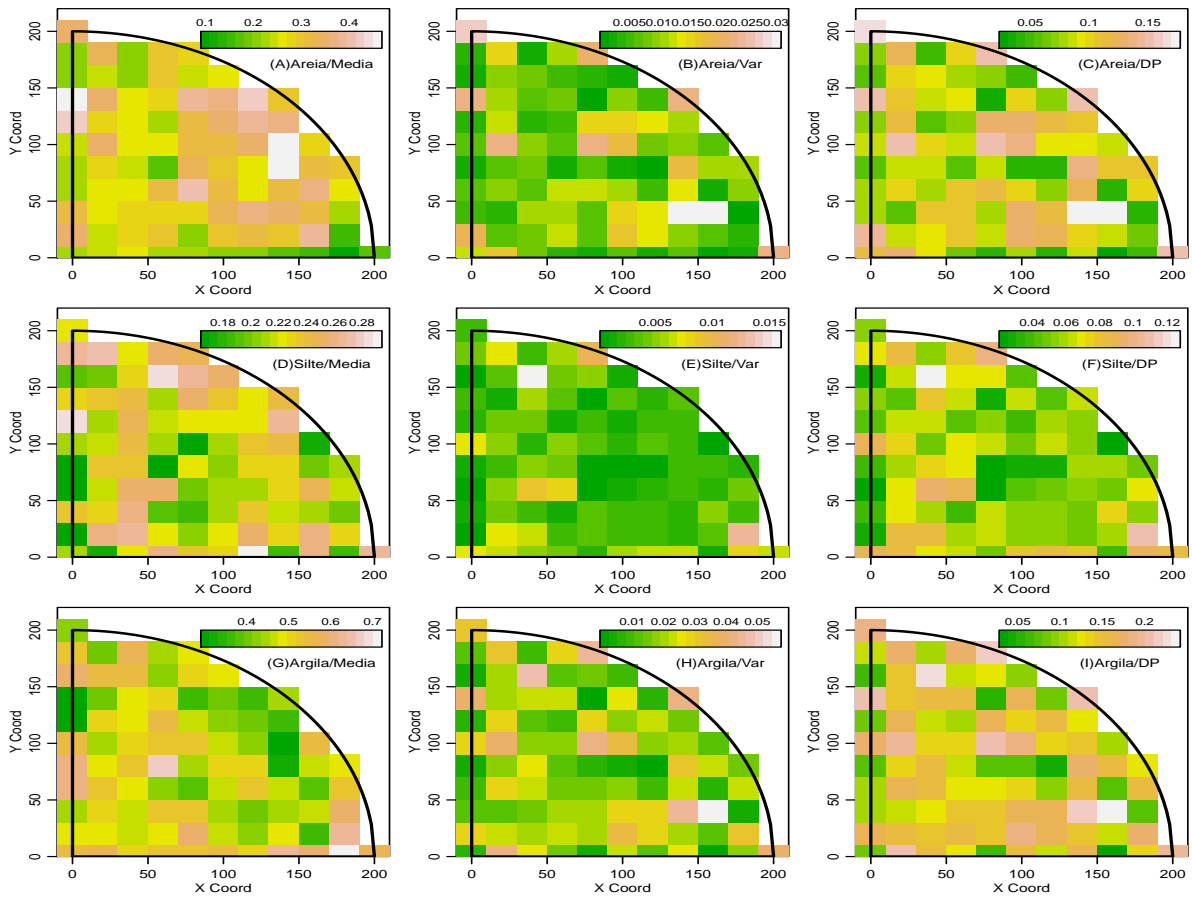


Figura 6: Mapas de médias, variâncias e desvios-padrão das porcentagens de areia (A-C), silte (D-F) e argila (G-I) obtidos por simulação.

O histograma da Figura 9 mostra as porcentagens da área em que a argila é maior que 50% em 200 simulações. Pode-se dizer que em 70 simulações, de 45 a 50% da área apresenta argila maior que 50%.

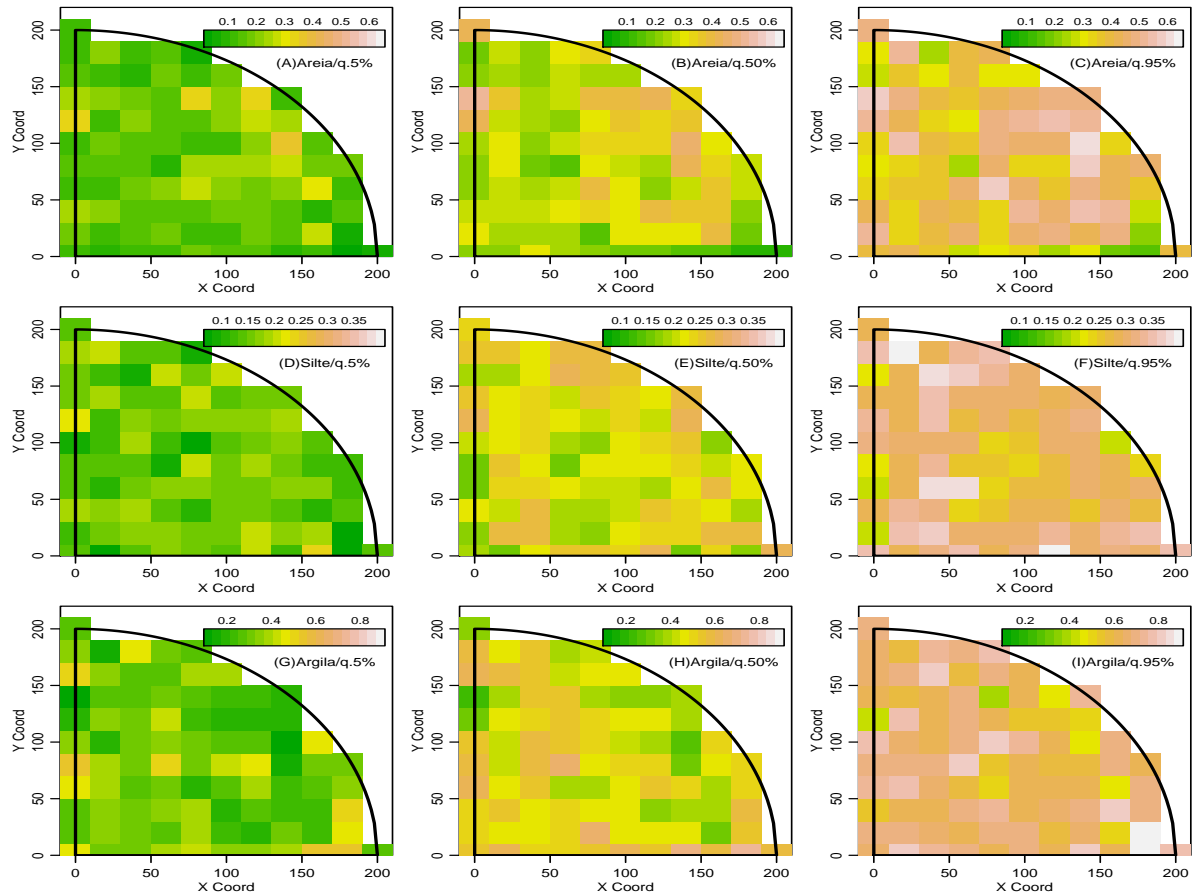


Figura 7: Mapas dos quartis 5, 50 e 95% das porcentagens de areia (A-C), silte (D-F) e argila (G-I) obtidos por simulação.

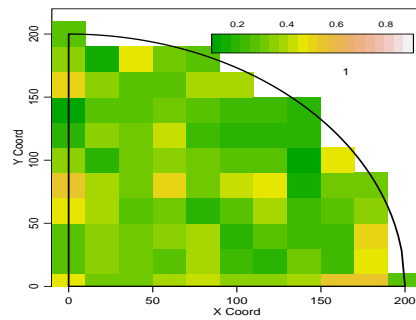


Figura 8: Mapas da probabilidade de areia menor que 50%,  $20\% < \text{silte} < 50\%$  e  $50\% < \text{argila} < 75\%$ .

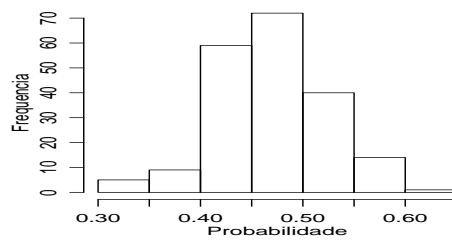


Figura 9: distribuição das probabilidades de argila  $> 50\%$  em 200 simulações.