

# Aplicação de métodos geoestatísticos na análise do teor de silte do solo 55<sup>a</sup> RBRAS e 15<sup>a</sup> RARG

Ana Patricia Bastos Peixoto <sup>1</sup>, Tiago Viana Flor de Santana <sup>2</sup>, Helga Restum Hissa <sup>3</sup>, Enio Fraga da Silva <sup>4</sup>, José Ronaldo de Macedo <sup>5</sup>, Cláudio Lucas Capeche <sup>6</sup>, Maria Cristina Neves de Oliveira <sup>7</sup>

## 1 Resumo

Sabe-se que variáveis regionalizadas são funções que variam de um lugar a outro no espaço com certa similaridade de continuidade, cujos valores estão relacionados de algum modo com a posição espacial. Em função disso, este trabalho se propõe a estudar a aplicação de métodos geoestatísticos, para tratamento de variáveis georreferenciadas, em dados de que compõe a fração do solo como o silte fornecidos pela Embrapa Solos.

**Palavras-chave:** Dependência espacial, variáveis regionalizadas, fração do solo.

## 2 Introdução

Em um estudo geoestatístico, as variáveis a serem analisadas têm duas características. São aleatórias já que os valores numéricos observados podem variar consideravelmente de um ponto a outro no espaço e são espaciais porque apesar de muito variar dentro do espaço, os valores numéricos observados não são inteiramente independentes (Guerra, 1988). Variáveis com este tipo de estrutura são chamadas de variáveis regionalizadas (VR) ou georreferenciadas, uma definição de Matheron (1963).

O objetivo desse estudo é conhecer melhor os conceitos básicos da geoestatística, assim como identificar se os dados apresentam dependência espacial, realizando simulações para encontrar os parâmetros do semivariograma para os modelos que mais represente os mesmos, com o uso das ferramentas computacionais do software R, através análise espacial dos dados.

## 3 Material e métodos

Os dados experimentais, que fazem parte deste trabalho, foram coletados do levantamento detalhado de solos da Estação Experimental de Campos, Fazenda Angra (Capeche et al., 1997), por pesquisadores da PESAGRO e Embrapa Solos. No estudo pedológico foram avaliadas as características morfológicas, físicas e químicas dos solos, e apresentadas também, informações referentes à distribuição geográfica. A área compreendida pelo levantamento situa-se ao Norte

<sup>1</sup>Professora Assistente do Departamento de Ciências Exatas, UEFS/BA: [apbpeixoto@usp.br](mailto:apbpeixoto@usp.br)

<sup>2</sup>Doutorando em Estatística IMECC - UNICAMP: [tiagodesantana@yahoo.com.br](mailto:tiagodesantana@yahoo.com.br)

<sup>3</sup>Embrapa Solos, RJ, Jd Botânico : [helga@cnps.embrapa.br](mailto:helga@cnps.embrapa.br)

<sup>4</sup>Embrapa Solos, RJ, Jd Botânico : [enio@cnps.embrapa.br](mailto:enio@cnps.embrapa.br)

<sup>5</sup>Embrapa Solos, RJ, Jd Botânico : [jronaldo@cnps.embrapa.br](mailto:jronaldo@cnps.embrapa.br)

<sup>6</sup>Embrapa Solos, RJ, Jd Botânico : [capeche@cnps.embrapa.br](mailto:capeche@cnps.embrapa.br)

<sup>7</sup>Pesquisadora - Embrapa Soja: [mcneves@cnps.embrapa.br](mailto:mcneves@cnps.embrapa.br)

do estado do Rio de Janeiro, à margem esquerda do Rio Paraíba do Sul, no município de Campos dos Goytacazes, entre os paralelos 21°44'47" e 41°18'24" WGr. Para o estudo geoestatístico foi considerada a variável agrônômica teor de silte (%), na camada de 0-20cm em 273 amostras.

Um campo aleatório é um processo estocástico que existe em algum espaço real  $d$ -dimensional, definido por:

$$\{Z(s_i) : s_i \in T \subset \mathbb{R}^d\}$$

em que,  $Z(s_i)$  denota a variável aleatória  $Z$  na localização  $s_i$  do espaço sob uma região de estudo  $T$ . A descrição de um campo aleatório é obtida através das distribuições acumuladas finito-dimensionais  $F$ , para qualquer conjunto de localizações  $(s_1, s_2, \dots, s_n)$  pertencentes à região  $T$  e qualquer inteiro  $n$ :

$$F_{S_1, S_2, \dots, S_n}(z_1, z_2, \dots, z_n) \equiv P(Z(S_1) \leq z_1, Z(S_2) \leq z_2, \dots, Z(S_n) \leq z_n).$$

A quantidade  $\gamma(h)$  é conhecida como semivariância e é a metade da esperança da variância entre pares de pontos separados por uma distância  $h$ . Obedecendo-se às condições da hipótese de estacionariedade, a semivariância pode ser estimada simplesmente através de uma única realização do processo estocástico

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(s_i) - z(s_i + h)]^2,$$

em que  $\gamma(h)$  é a semivariância estimada entre pares de pontos,  $N(h)$  é o número de pares de valores medidos  $z(s_i)$ ,  $z(s_i + h)$ , separados pela distância  $h$ .

Matérn (1986) apresenta uma classe de funções de correlação que é considerada uma das mais completas, por englobar outras funções de correlação, pela simples escolha do parâmetro de diferenciabilidade. Essa família de funções de correlação possui a seguinte forma:

$$\rho(h) = \{2^{k-1} \Gamma(\kappa)\}^{-1} (h/\phi)^\kappa K_\kappa(h/\phi),$$

sendo  $\Gamma(\cdot)$  a função gama,  $\kappa(\cdot)$  é a função Bessel de ordem  $\kappa$ ,  $h$  a distância euclidiana entre duas localizações quaisquer do campo aleatório, os parâmetros dessa função são  $\phi > 0$  e  $\kappa > 0$ .

A função exponencial potência ou família de correlações possui características similares as da família Matérn em  $\phi$  e  $k$  são interpretado de forma análoga. Entretanto, neste caso  $k$  é limitado ao intervalo  $[0, 2]$  e não definindo uma sequência de diferenciabilidade dos processos associados que passam de não diferenciáveis para  $k > 2$ . A forma dessa função de correlação é

$$\rho(\mu) = \exp(\mu|\phi)^k$$

essas duas funções são muito utilizadas devido a capacidade de produzir comportamentos distintos quanto a suavidade do processo, ou seja, se possível modelar processos mais ou menos diferenciáveis. Nesta função, se  $k < 2$ , o processo é contínuo mas não diferenciável e se  $k \geq 2$  pode ser infinitamente diferenciável. Existem dois casos particulares para essa função. No caso  $k = 1$  a função será chamada exponencial, e para  $k = 2$  a função será chamada de gaussiana.

Considerando que em alguma área  $T$  exista um campo aleatório gaussiano  $Z$  latente. Sendo assim, existe um vetor  $Y(s)$  de  $n \times 1$  valores observados em  $s = (s_1, s_2, \dots, s_n)$ , que segundo Diggle & Ribeiro Jr. (2007), sob o pressuposto de gaussianidade, pode ser modelado como:

$$Y(s) = \mu(x) + Z(s) + \varepsilon, \tag{1}$$

em que  $\mu(x) = X\beta$  sendo  $X$  uma matriz  $n \times q$  contendo  $q - 1$  possíveis covariáveis,  $\beta$  um vetor  $q \times 1$  de parâmetros associados a  $X$ ,  $Z(s)$  um campo aleatório gaussiano que possui vetor de médias  $n \times 1$  nulo e matriz de covariâncias  $\Sigma$  de dimensão  $n \times n$ , sendo que cada elemento  $\sum_{i,j}$  é igual a  $Cov(Z(s_i); Z(s_j))$ , para todo  $s_i$  e  $s_j$  pertencentes a  $s$ , e  $\varepsilon$  um vetor  $n \times 1$  de ruídos brancos.

A expressão de uma estimativa empírica para o semivariograma é dada por:

$$\hat{\gamma} = \frac{\sum_{|N(h)|} (Z(s_i) - Z(s_j))^2}{2|N(h)|} \quad (2)$$

em que,  $|N(h)|$  é o número de pontos abrangidos pela distância  $h$ . Por simplicidade matemática, normalmente utiliza-se o logaritmo da função de verossimilhança para fazer a estimação, que associada a (1) tem a seguinte forma:

$$l(\theta; Y(s)) = -0.5(n \ln(2\pi) + \ln(|\Sigma_Y|) + (Y(s) - X\beta)^t \Sigma_Y^{-1} (Y(s) - X\beta)) \quad (3)$$

Quando o interesse é fazer predição para  $Z$  nas localizações  $s^*$ , ou seja, a distribuição de probabilidade de  $(Z(s^*)|Y(s))$  gera as predições com melhor precisão das estimativas. Dessa forma pode-se obter:

$$E(Z(s^*)|Y(s)) = \mu_Z + \Sigma_{Z,Y} \Sigma_Y^{-1} (Y(s) - X\beta) \quad (4)$$

sendo  $\mu_Z$  a média de  $Z(s^*)$  e  $\Sigma_{Z,Y}$  é a matriz de covariâncias cruzadas entre  $Z(s^*)$  e  $Y(s)$ . Além disso a variância preditiva é conhecida:

$$Var(Z(s^*)|Y(s)) = \Sigma_Z - \Sigma_{Z,Y} \Sigma_Y^{-1} \Sigma_{Z,Y}^t \quad (5)$$

sendo  $\Sigma_Z$  a matriz de covariâncias de  $Z(s^*)$ .

## 4 Resultados e discussão

A característica mais evidente observada na Figura 1 é a preponderância de grandes valores de resposta para o extremo leste da região de estudo. Isto sugere que uma superfície de tendência pode ser apropriada. Em algumas aplicações, o contexto específico dos dados poderia sugerir que há algo de espacial na direção norte-leste. Caso contrário, uma superfície de tendência pode ser incluída no modelo em estudo, então ambas as coordenadas espaciais devem contribuir, porque a orientação da região de estudo é essencialmente arbitrária.

Uma maneira de diagnosticar a dependência é através dos variogramas empíricos. Analisando a expressão da função semivariância, pode-se imaginar que quanto mais próximos estiverem os pontos amostrados, maior será a semelhança entre eles e, portanto, menor a semivariância. O semivariograma apresentados na Figura 2 indica estacionaridade de segunda ordem para a variável, porque apresenta patamar claro e bem definido. Se o semivariograma for constante e igual ao patamar para qualquer valor de  $h$ , temos o efeito pepita puro e, neste caso, se a dependência espacial existir, será manifestada à distância ou tempo menor do que o menor espaçamento entre amostras. Também é possível verificar a existência de dependência espacial é através do diagnóstico gráfico utilizando envelopes simulados e variogramas empíricos (Figura 2).

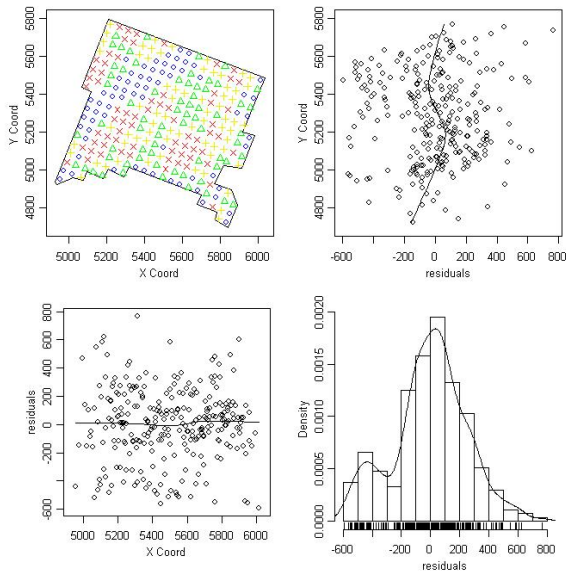


Figura 1: Gráfico das localizações dos pontos do teor de silte (%) no solo.

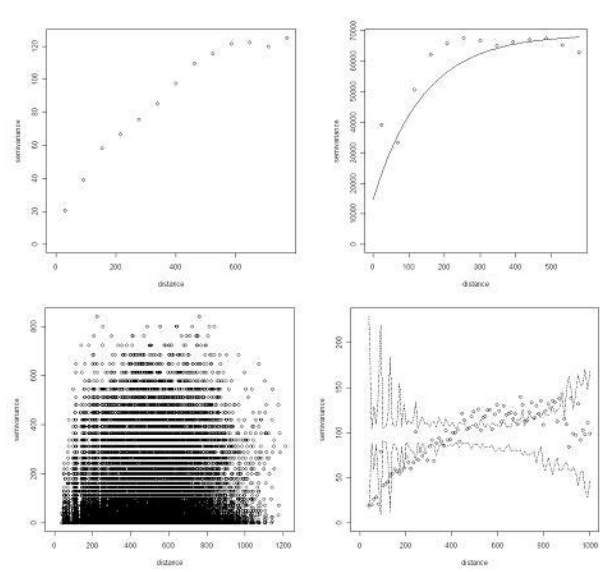


Figura 2: Variogramas e envelope simulando indicador de dependência espacial.

Para que haja dependência, deve haver ao menos um ponto do variograma fora do envelope simulado. Deste modo, se isso ocorrer rejeita-se a hipótese nula, de que não há dependência espacial. Isto sugere que a tendência crescente no variograma empírico é estatisticamente significativa, confirmando a presença de correlação espacial positiva.

Devido a dificuldade de se ajustar um modelo teórico confiável ao semivariograma experimental, optou-se nesse trabalho por se obter uma estimativa pontual para os parâmetros do modelo geostatístico pelo método da máxima verossimilhança (MV). A Tabela 1 mostra os resultados obtidos por MV.

Tabela 1: Estimação dos parâmetros do modelo geostatístico por MV, sendo  $\beta$ : parâmetro do efeito sistemático do modelo,  $\sigma^2$  e  $\phi$ : parâmetros da função de correlação,  $\tau^2$ : parâmetro do erro, AIC: Informação de Akaike e logL: valor de MV.

| Silte(%)  | $\beta$ | $\tau^2$ | $\sigma^2$ | $\phi$ | AIC    | -logL |
|-----------|---------|----------|------------|--------|--------|-------|
| kappa 0.5 | 579,5   | 4589,5   | 171398,7   | 300,0  | 1671,0 | 831,4 |
| kappa 1.5 | 268,1   | 22652,4  | 1573306,8  | 299,9  | 1691,0 | 841,6 |
| kappa 2.5 | -671,8  | 26949,1  | 17858714,6 | 299,8  | 1723   | 859,6 |

É possível observar que as curvas ajustadas da função de correlação de Matérn, foram ajustadas com diferentes valores para o parâmetro de diferenciabilidade. De acordo com os resultados da Tabela 1 foi adotado no modelo, a função de correlação de Matérn com  $k = 1,5$  e média constante, pois apresentou menor AIC.

Após a definição da dependência espacial do o teor de silte (%), realizou-se a krigagem ordinária dos dados (Figura 3) para obtenção dos mapas temáticos dessas seções transversais, estimando valores em locais não medidos e com menor variância da estimada, o que é uma característica desse método (GUIMARÃES, 2000). A krigagem foi estimada partir de modelo ajustado por MV, em 273 pontos de uma malha regular.

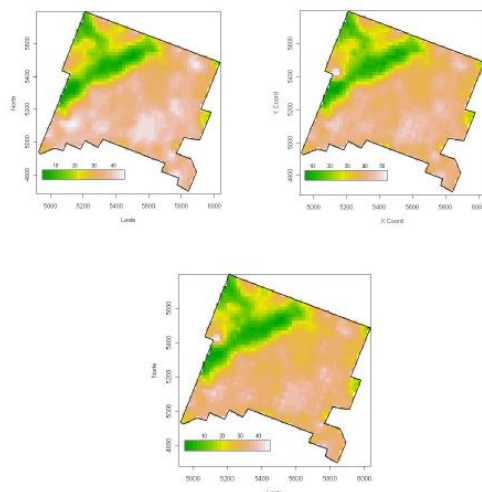


Figura 3: Gráfico das localizações dos pontos (superior esquerdo), os valores observados contra coordenadas (canto superior direito e parte inferior esquerda) e histograma (parte inferior direita) do teor de silte (%) medido para os dados transformados.

## 5 Conclusão

A presença de indícios de teor de silte é mais fortes em torno de 40% nas áreas mais claras dos gráficos que foram observados, indicando que a prevalência da fração total dessa variável tem maior intensidade e interferência na produtividade do solo.

## Referências

- [1] CAPECHE, C.L.; MACEDO, J.R.; MANZATTO, H.R.H.; SILVA, E.F. Caracterização pedológica da fazenda Angra - PESAGRO/RIO - Estação experimental de Campos (RJ). (compact disc). In: CONGRESSO BRASILEIRO DE CIÊNCIA DO SOLO. 26., Informação, globalização, uso do solo; Rio de Janeiro, 1997. trabalhos. Rio de Janeiro: Embrapa/SBCS, 1997.
- [2] DIGGLE, P.; TAWN, J. A.; MOYEED, R. A. Model-based geostatistics. Applied Statistics, V.47, n.3, p.299-350, 1998. ([citeseer.ist.psu.edu/diggle98modelbased.html](http://citeseer.ist.psu.edu/diggle98modelbased.html))
- [3] GUERRA, P.A.G. Geoestatística operacional. Brasília: Ministério das Minas e Energia, 1988. 145p.
- [4] MATHERON, G. Principles of Geostatistics. Economic Geology, v.58, p.1246-1266. 1963.
- [5] MATÉRN, B. Spatial variation. Verlag, Berlin: Springer, 1986. 365 p.
- [6] R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. Vienna, Austria, 2010. (<http://www.R-project.org>)