

Seleção e ajuste de modelos espaciais de semivariograma aplicados a dados do pH do solo

Kuang Hongyu¹

1 – Programa de Pós graduação em Estatística e Experimentação Agronômica – Departamento de Ciência Exatas, ESALQ/USP

Resumo:

O uso de métodos geoestatísticos tiveram como ponto de partida a avaliação da dependência espacial por meio do semivariograma. O semivariograma é a parte central da Geoestatística, portanto é importante a ajuste e seleção dos modelos da função de semivariograma. Neste trabalho teve como objetivo avaliar dois métodos de ajuste aplicados a quatro modelos de função de semivariância, a partir das técnicas de validação cruzada e preditiva e do critério de Akaike (Akaike's Information Criterion - AIC). Os dados do pH do solo utilizados para ajuste dos modelos foram coletados em uma área de 71,31 ha, localizada no município de Cascavel, na região Oeste do Paraná, com aproximadamente, 24,95° Sul de latitude e 53,57° Oeste de longitude e altitude média de 650 metros. Foram ajustados quatro modelos da função de semivariância, por meio de dois métodos de ajuste. A seleção do modelo e método de ajuste foi efetuada quanto ao critério de AIC e por meio das técnicas de validações cruzada e preditiva. Segundo as técnicas de seleção avaliadas, o modelo esférico e o método da Máxima Verossimilhança tiveram melhor desempenho. A validação cruzada e a preditiva indicaram o mesmo modelo e o mesmo método de ajuste com melhor desempenho.

1. Introdução

Solos ácidos apresentam grande concentração de íons hidrogênio e alumínio, devido à intensa lixiviação, pela retirada dos nutrientes catiônicos pela cultura sem a devida reposição e pela utilização de fertilizantes de caráter ácido. A acidez do solo pode ser dividida em acidez ativa e acidez potencial. Denomina-se acidez ativa a parte do hidrogênio que está dissociada, ou seja, na solução do solo. O pH ou potencial hidrogeniônico representa a concentração de H⁺, expresso por: $\text{pH} = -\log(\text{H}^+) = \log 1/(\text{H}^+)$. A escala de pH varia de 0 a 14. Em solos podem ser encontrados valores de 3 a 10, mais comuns em solos brasileiros entre 4,0 a 7,5. Segundo Costa e Oliveira (2001) a faixa de pH ideal dos solos para a agricultura esta entre 5,5 e 6,5. Neste intervalo os nutrientes têm maior disponibilidade para a planta (Girardello et al., 2011).

O conhecimento da variabilidade espacial é fundamental para que possa descrever e compreender o comportamento de determinado atributo do solo, o mapeamento de tal atributo em uma região e a compreensão de sua estrutura de variabilidade pode auxiliar na definição de estratégias de manejo deste solo (Ribeiro Jr, 1995). Assim, por meio do estudo geoestatístico obtêm-se a estrutura de dependência espacial dos pontos amostrados e a construção de mapas temáticos de cada atributo da área, possibilitando a aplicação localizada de corretivos considerando a necessidade em cada talhão, o que representa uma das práticas que mais contribui para a eficiência dos fertilizantes e que consequentemente reflete em maior economia para o produtor (Girardello et al., 2011).

O uso de métodos geoestatísticos tiveram como ponto de partida a avaliação da dependência espacial por meio do semivariograma. O semivariograma é a parte central da Geoestatística, capaz de descrever a estrutura de dependência espacial, além de ser o ponto chave na determinação do preditor geoestatístico (krigagem) (McBratney & Webster, 1986). Estes fatos demonstram a importância de se tratar adequadamente a questão da seleção e ajuste de modelos espaciais para semivariogramas experimentais (Mello et al., 2005).

Neste contexto, o objetivo deste trabalho teve como objetivo ajustar e selecionar modelos espaciais de semivariograma para estimativa do pH de solo. Especificamente pretendeu-se, a partir da validação e do critério de AIC (Akaike's Information Criterion), avaliar o desempenho de dois métodos de ajuste, aplicados a quatro modelos de função de semivariância.

2. Material e Métodos

Descrição da área e obtenção dos dados

No ano agrícola 2005/2006, a coleta dos dados foi realizada em uma área de 71,31 ha, localizada no município de Cascavel, na região Oeste do Paraná, com aproximadamente, 24,95° Sul de latitude e 53,57° Oeste de longitude e altitude média de 650 metros. O solo da área é classificado como Latossolo Vermelho Distroférico, com textura argilosa (Embrapa, 2006) e o clima da região é temperado superúmido, tipo climático Cfa (Köppen).

Todas as amostras foram georreferenciadas com auxílio de um receptor de dados GPS (Global Positioning System) GEOEXPLORER 3, da marca Trimble, num sistema espacial de coordenadas (UTM). Foi realizada a amostragem sistemática centrada (lattice plus close pairs), em um grid irregular com distância máxima de 152 metros entre pontos e, em alguns locais pré-determinados, a amostragem foi realizada com distâncias menores, com 50 e 75 metros entre pontos, com um total de 46 observações.

Para compor a amostra composta de solo, que representa cada ponto georreferenciado, foram coletadas quatro sub-amostras num raio de quatro metros do ponto, na profundidade de 0-0,2 m. As sub-amostras foram homogenizadas e colocadas em sacos plásticos, com aproximadamente 500 gramas; compondo a amostra representativa da parcela, os quais foram encaminhados para o laboratório da COODETEC para as análises químicas, obtendo-se, assim, os dados do pH no solo.

Inicialmente, foi feito um estudo exploratório, com uma análise descritiva e gráfica dos dados distribuídos espacialmente. A análise descritiva buscou identificar e avaliar características da distribuição dos dados, tais como a homogeneidade, normalidade, existência de pontos discrepantes e tendências direcionais (Assumpção; 2007).

Semivariograma

Uma das funções mais utilizadas na geoestatística para a determinação da dependência espacial de variáveis é a função de semivariância, que gera o semivariograma. O semivariograma experimental é um gráfico que expressa a variabilidade espacial entre as amostras, sendo uma função que só depende do vetor h , distância entre os pares de casos amostrados. Quando h cresce, o semivariograma aproxima-se da variabilidade total dos dados,

chamado patamar, e havendo estacionariedade de segunda ordem, o semivariograma expressa o grau de dependência entre os pontos amostrais (Opromolla et al., 2006).

Na análise geoestatística foi utilizado o semivariograma, que é um gráfico da função semivariância. O semivariograma permite o cálculo dos parâmetros envolvidos no modelo, a saber a distância da dependência espacial (a), conhecida como range ou alcance da dependência espacial que é o alcance máximo da dependência espacial, indicando que, a partir dessa distância, termina a dependência entre os casos estudados; o efeito pepita (c), conhecido por *nugget effect* que é o valor da semivariância a distância zero, ou seja, representa a variabilidade para escalas menores que a distância de amostragem; e altura (c) ou *sill* que é a distância entre o efeito pepita e o patamar, intervalo esse no qual o semivariograma cresce, havendo dependência espacial. O patamar representa a altura na qual o semivariograma se estabiliza, aproximando-se da variabilidade total dos valores amostrados (Opromolla et al., 2006). O estimador da função semivariância para variáveis regionalizadas com distribuição normal de probabilidade é apresentado na equação 1 (Matheron; 1963):

$$\gamma^*(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [Z(s_i) - Z(s_i + h)]^2 \quad (1)$$

em que, $n(h)$: número de pares em uma mesma distância h ; $Z(s_i)$ e $Z(s_i + h)$ valores da variável na posição s_i e $s_i + h$, com $s \in \mathbf{S} \subset \mathbf{R}^2$.

Os modelos de semivariograma

Os modelos de semivariograma utilizado neste trabalho foram Esférico ou Matheron, Exponencial ou de Formery e Gaussiano ou Parabólico. As expressões destes modelos são apresentadas pelas as equações (2), (3) e (4) a seguir:

$$\rho(\mathbf{h}) = \begin{cases} \tau^2 + \sigma^2 \left[\frac{3}{2} \left(\frac{h}{\phi} \right) - \frac{1}{2} \left(\frac{h}{\phi} \right)^3 \right] & \text{se } 0 \leq \mathbf{h} \leq \phi \\ \tau^2 + \sigma^2 & \text{se } \mathbf{h} > \phi \end{cases} \quad (2)$$

$$\rho(\mathbf{h}) = \begin{cases} \tau^2 + \sigma^2 \left[1 - e^{-\frac{h}{\phi}} \right] & \text{se } 0 \leq \mathbf{h} \leq \phi \\ \tau^2 + \sigma^2 & \text{se } \mathbf{h} > \phi \end{cases} \quad (3)$$

$$\rho(\mathbf{h}) = \begin{cases} \tau^2 + \sigma^2 \left[1 - e^{-\frac{h^2}{\phi^2}} \right] & \text{se } 0 \leq \mathbf{h} \leq \phi \\ \tau^2 + \sigma^2 & \text{se } \mathbf{h} > \phi \end{cases} \quad (4)$$

Métodos de Ajuste

Os métodos de ajustes dos modelos de semivariograma podem dividir em dois grupos: i) Método dos Quadrados Mínimos Ordinários (Ordinary Least Squares - OLS), Método dos Quadrados Mínimos Ponderados (Weight Least Squares - WLS) e método de ajuste denominado de “a sentimento”; ii) Método de ajuste de um modelo direto aos dados - Método da Máxima Verossimilhança (Maximum Likelihood - ML) (Mello et al., 2005).

Método de Quadrados Mínimos Ordinários

O Método de quadrados mínimos ordinários é um dos estimadores mais utilizados na Estatística e baseia-se no princípio dos quadrados mínimos, introduzido por Gauss em 1809 (Bussab & Moretin, 2002). O método consiste em obter os valores dos parâmetros de um modelo que minimizam a soma do quadrado da diferença entre os valores observados e os estimados. A estimativa dos parâmetros do semivariograma pelo Método dos Quadrados Mínimos Ordinário é obtida pela minimização da seguinte expressão (Mello et al., 2005):

$$Q(\theta) = \sum_{i=1}^k [g(h_i) - \gamma(h_i; \theta)]^2 \quad (5)$$

em que θ representa o vetor de parâmetros estimados que definem completamente o semivariograma, sendo cada estimativa denotada por $g(h_i)$ e k refere-se ao número de “lags” do semivariograma experimental. A quantidade $\gamma(h_i; \theta)$ é a semivariância calculada pelo modelo e depende dos parâmetros estimados θ e após minimização, a função $g(h_i)$ se torna a estimativa $\hat{\gamma}(h_i)$ de quadrados mínimos.

Método da Máxima Verossimilhança

O método da Máxima Verossimilhança é uma técnica de estimação usada com frequência porque suas propriedades assintóticas são interessantes, como por exemplo, eficiência e consistência (Wonnacott & Wonnacott, 1972). A idéia do método é obter, a partir de uma amostra, o estimador “mais verossímil” dos parâmetros de um certo modelo probabilístico (Mello et al., 2005). Considera-se um conjunto de observações de volume (Z_1, \dots, Z_n) gerados a partir de um modelo linear gaussiano, volume este que pode ser explicado pela soma de três componentes:

$$Z_i = \mu(x_i) + S(x_i) + \varepsilon_i \quad i = 1, \dots, n \quad (6)$$

Na equação 6, $\mu(x_i)$ é uma componente completamente determinística e é determinado por um modelo linear

$$\mu(x_i) = \sum_{k=1}^p f_k(x_i) \beta_k \quad (7)$$

em que onde, (f_1, \dots, f_p) é um conjunto de p funções que descrevem deterministicamente a variável Z a partir das coordenadas espaciais (x_1, \dots, x_n) . Nesta mesma equação, $S(x_i)$ é uma componente aleatória entendida como um processo estocástico gaussiano autocorrelacionado. A esperança de $S(x_i)$ é igual a zero e a variância constante (σ^2). Esta componente é que confere continuidade espacial aleatória à variável Z , isto é, $S(x)$ é que possui a correlação espacial. A correlação espacial entre dois pontos segue uma função autorizada, para a qual no presente trabalho, foram avaliados os modelos exponencial e Matérn (Mello et al., 2005). O modelo definido na equação 7 para o volume de madeira com uma distribuição Multivariada Normal, que pode se representada da seguinte forma: $Z \sim MVN(F\beta, \sigma^2 \Sigma + \tau^2 I)$. Logo, a função de verossimilhança a ser maximizada tem a seguinte forma:

$$L(\beta, \theta) = -\frac{1}{2} \{ \log |K| + (Z - F\beta)^t [K]^{-1} (Z - F\beta) \} \quad (8)$$

Seleção de Modelos

A avaliação do desempenho de cada modelo na estimativa volumétrica se deu através de dois critérios. No primeiro, avaliou-se o AIC gerado para cada modelo dentro dos respectivos métodos de ajuste. O AIC foi obtido através da função de Máxima Verossimilhança, a partir dos parâmetros ajustados para os modelos conforme os métodos. O AIC para os métodos “a sentimento” e quadrados mínimos, foram obtidos a partir dos respectivos parâmetros obtidos por estes métodos, através da função likGRF do geoR (Mello et al., 2005). O critério de AIC foi desenvolvido a partir da distância ou informação de Kulback e Leibler (1951). Esta distância é uma medida de discrepância entre as linhas do modelo verdadeiro e o modelo aproximado. A distância de Kulback e Leibler com a Máxima Verossimilhança, surgindo o AIC (Akaike, 1983). O AIC é dado pela seguinte expressão:

$$AIC = -2\log L + 2k \quad (9)$$

em que L é a Verossimilhança Maximizada do modelo candidato e K é o número de parâmetros deste modelo.

O segundo critério de comparação do desempenho utilizou as técnicas de validações cruzada e preditiva. Esta técnica consistiu em prever o valor conhecido Z de pH do solo, comparando com o valor observado. Neste trabalho o estimador usado foi o de krigagem, cuja expressão do estimador é:

$$\hat{z}(x_0) = \sum_{i=1}^n \lambda_i z(x_i) \quad (10)$$

em que $\hat{z}(x_0)$ é o valor estimado para o ponto x_0 ; λ_i são os pesos de krigagem definidos conforme os parâmetros do semivariograma e $z(x_i)$ são os valores observados nos pontos amostrados (vizinhança de krigagem).

3. Resultados e discussão

Primeira etapa será feita uma análise exploratória de dados, na tabela 1 foi feita análise descritiva para os dados pH do solo. Pela tabela podemos perceber que os dados de pH apresentam coeficiente de variação inferior a 20%, o que indica que os dados são homogêneos (Pimentel Gomes, 2000). Isso indica que a média é uma boa estatística para representar o conjunto de dados, neste caso, o valor médio de pH para essa área foi de 5,391.

Tabela 1: Estatística descritiva da variável pH.

Estatísticas	Resultados
Média (\bar{x})	5,391
Desvio padrão (S)	0,371
Variância (S^2)	0,137
Coef. Variação (CV)	6,88%
Mínimo	4,800
1º Quartil (Q_1)	5,100
Mediana	5,400
3º Quartil (Q_3)	5,600
Máximo	6,400
N	46

A figura 1 apresenta o gráfico Boxplot para os dados de pH. O gráfico Box-plot é utilizado usualmente para representar dados independentes, porém, neste caso, estamos utilizando-o apenas para ter uma idéia do comportamento dos dados.

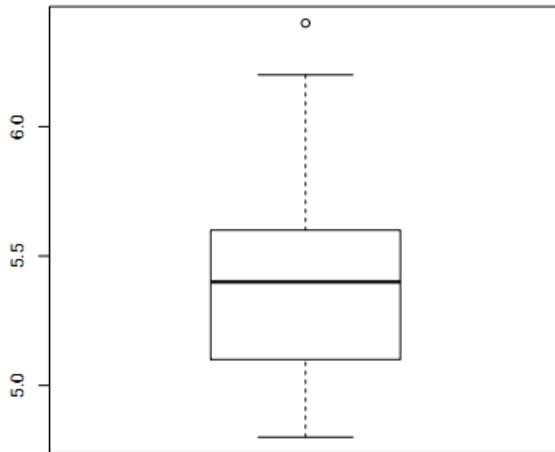


Figura 1 Gráfico Boxplot dos dados de pH

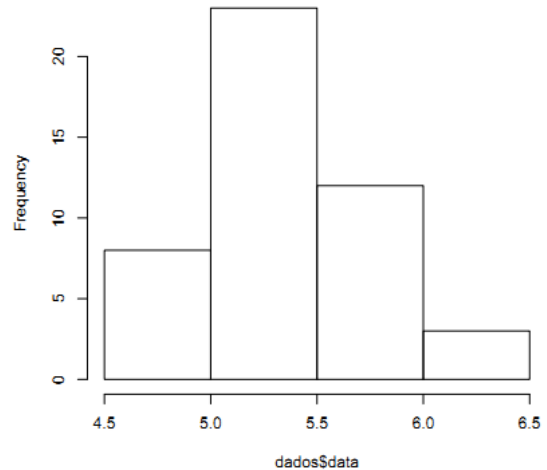


Figura 2 Histograma dos dados de pH

Podemos observar que os dados de pH apresenta um ponto discrepante, porém os dados apresentam um comportamento praticamente simétrico. A figura 2 apresenta o histograma dos dados de pH. Podemos verificar que os dados seguem uma distribuição muito similar à distribuição normal.

Ao aplicar o teste de Shapiro Wilk ao nível de 5% de significância, verificamos este indica que os dados de pH possuem distribuição normal de probabilidade. Assim como o gráfico Boxplot, o teste de Shapiro Wilk é utilizado para amostras independentes, porém, foi utilizada nessa análise, apenas para se ter uma idéia sobre a distribuição dos dados.

Pode-se construir um semivariograma omnidirecional, indicando que os semivariogramas apresentam o mesmo comportamento para todas as direções analisadas. A figura 3 apresenta o semivariograma omnidirecional. O semivariograma é construído considerando o mínimo de 30 pares para o cálculo das semivariâncias (Guerra, 1988). Para o cálculo das semivariâncias foi utilizado o estimador de Matheron (1963).

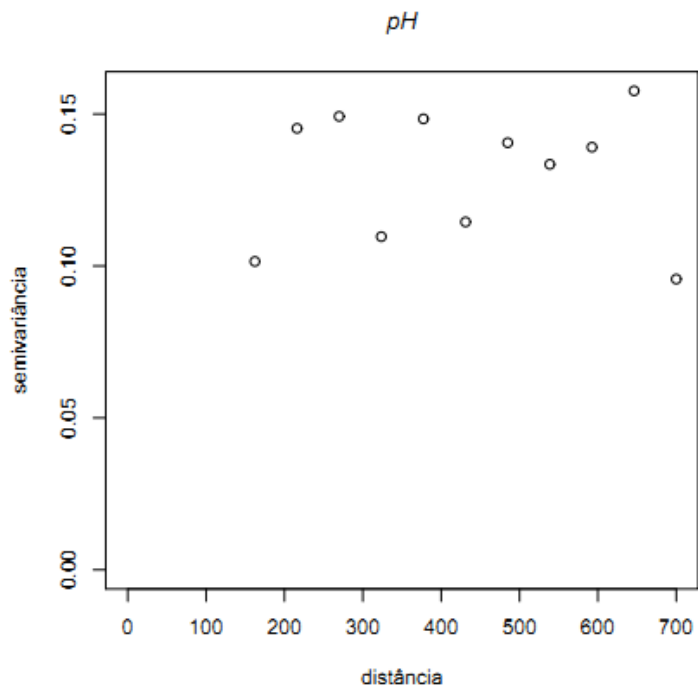


Figura 3 Semivariograma omnidirecional

Conforme a figura 4 apresenta o gráfico de envelopes, este gráfico sugere se há continuidade espacial nos dados. Ou seja, se há pontos fora da linha de confiança, pode-se afirmar que há continuidade espacial nos dados.

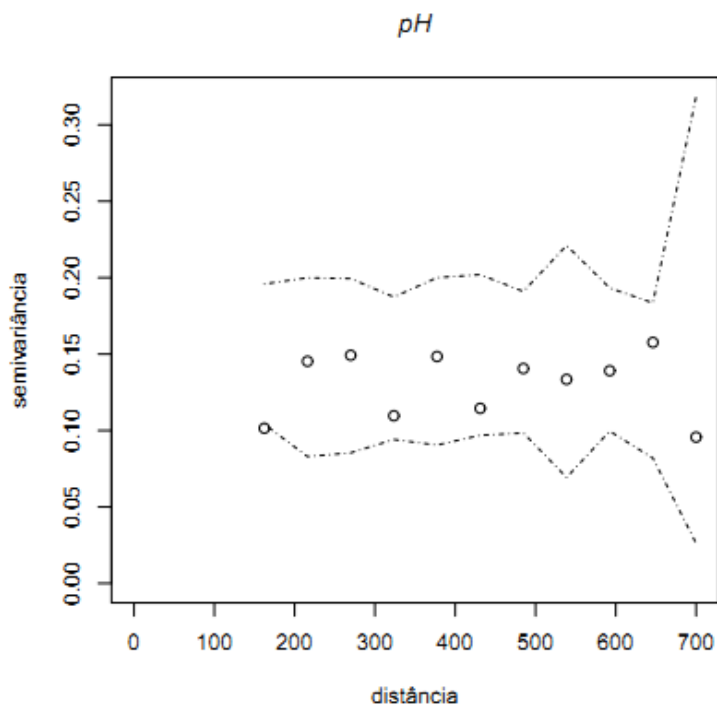


Figura 4 Gráfico de envelopes

Como podemos observar há apenas um ponto fora da linha de confiança (figura 4). Este é apenas um indício de que há continuidade espacial, será realizada a análise para explicitar a estrutura de dependência espacial dos dados de pH.

A seguir serão ajustados os modelos: exponencial, esférico, gaussiano e família Matérn com $k = 0,7$. Para cada um dos métodos de estimação: mínimos quadrados ordinários (OLS) e máxima verossimilhança (ML) para posterior escolha do melhor modelo ajustado.

Mínimos quadrados ordinários - OLS

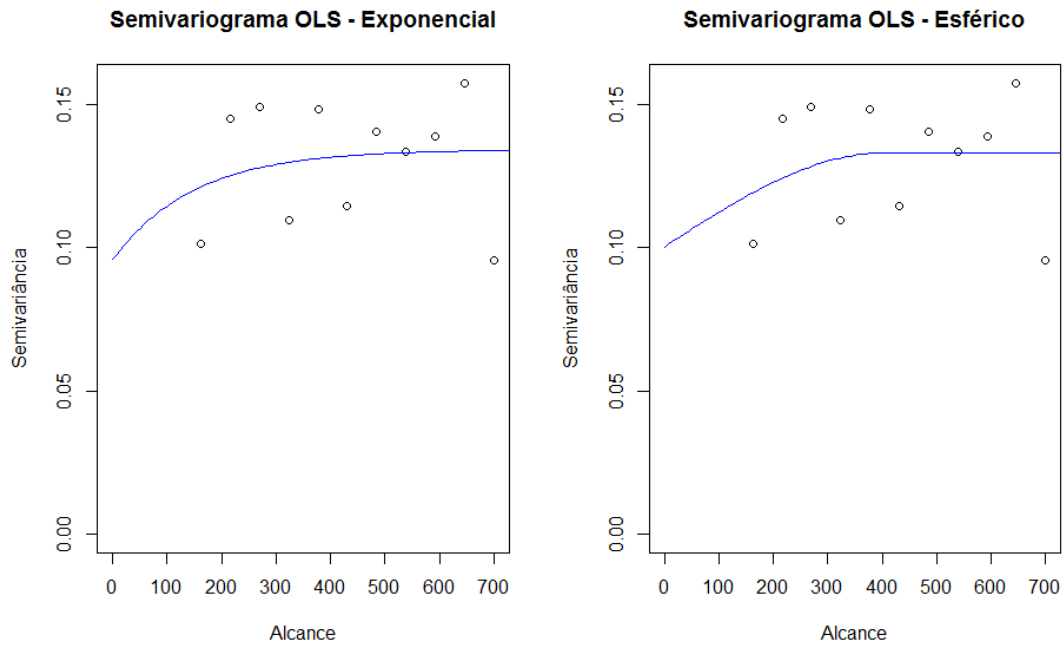


Figura 5 Semivariogramas dos modelos exponencial e Esférico

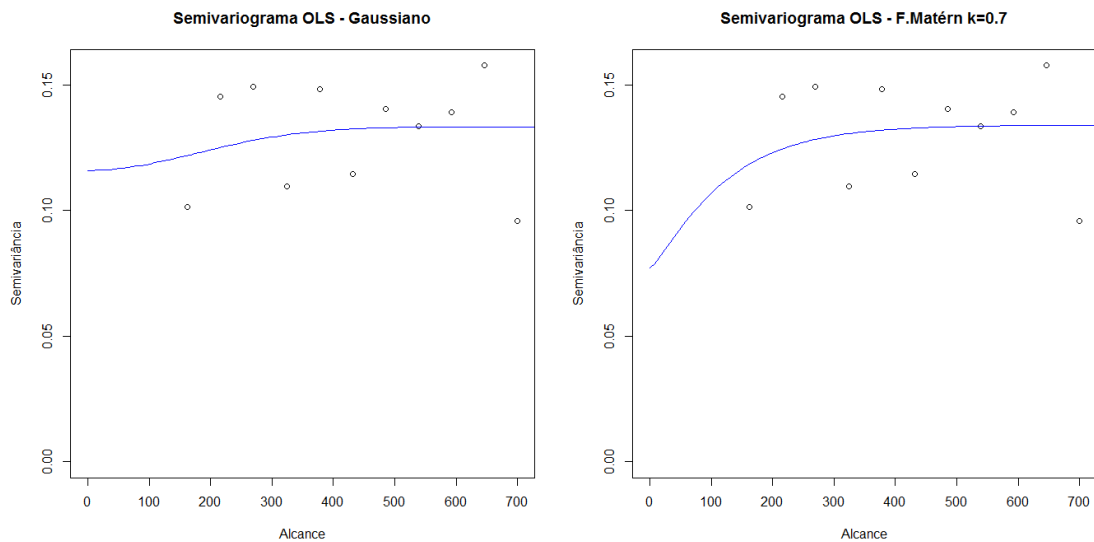


Figura 6 Semivariogramas dos modelos Gaussiano e F.Matérn

Para cada modelo ajustado a partir de um chute inicial, com 99 interações realizadas pelo software R, foram obtidos os valores de efeito pepita (φ_1), contribuição (φ_2) e alcance (a).

Tabela 2: Estimativas de Mínimos quadrados ordinários dos parâmetros associados ao modelo

Método	Modelos	β	φ_1	φ_2	$\varphi_3 = g(a)$	a	EPR
OLS	Exponencial	5,391	0,0959	0,0384	149,99	449,357	0,714
	Esférico	5,391	0,1004	0,0327	399,99	399,999	0,754
	Gaussiano	5,391	0,1160	0,0174	249,99	432,704	0,869
	Matérn (K=0,7)	5,391	0,0569	0,0569	99,99	344,683	0,575

β : média; φ_1 : efeito pepita; φ_2 : contribuição; $\varphi_3 = g(a)$: função do alcance; a : alcance;
 $EPR = \left(\frac{\varphi_1}{\varphi_1 + \varphi_2} \right) \times 100$: efeito pepita relativo.

Com as informações acima, é possível calcular o valor do efeito pepita relativo (EPR) sugerido por Cambardella (1994), que indica o nível grau de dependência para cada modelo (tabela 2).

Tabela 3: Classificação para efeito pepita relativa

$EPR < 0,5$	Forte dependência espacial
$0,5 \leq EPR < 0,75$	Média dependência espacial
$EPR > 0,75$	Fraca dependência espacial

Máxima Verossimilhança – ML

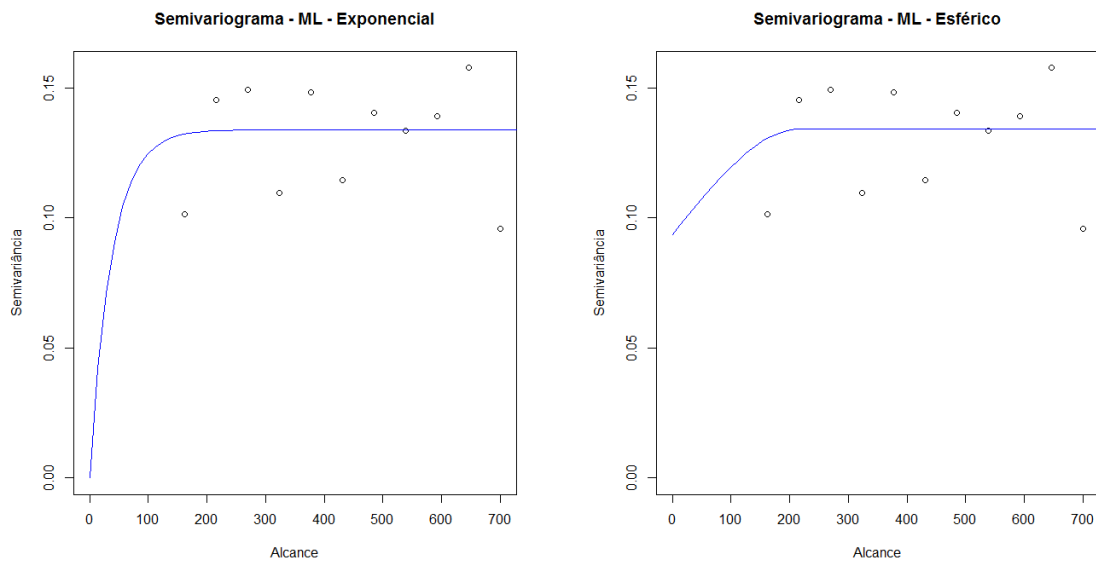


Figura 7 Semivariogramas dos modelos Exponencial e Esférico

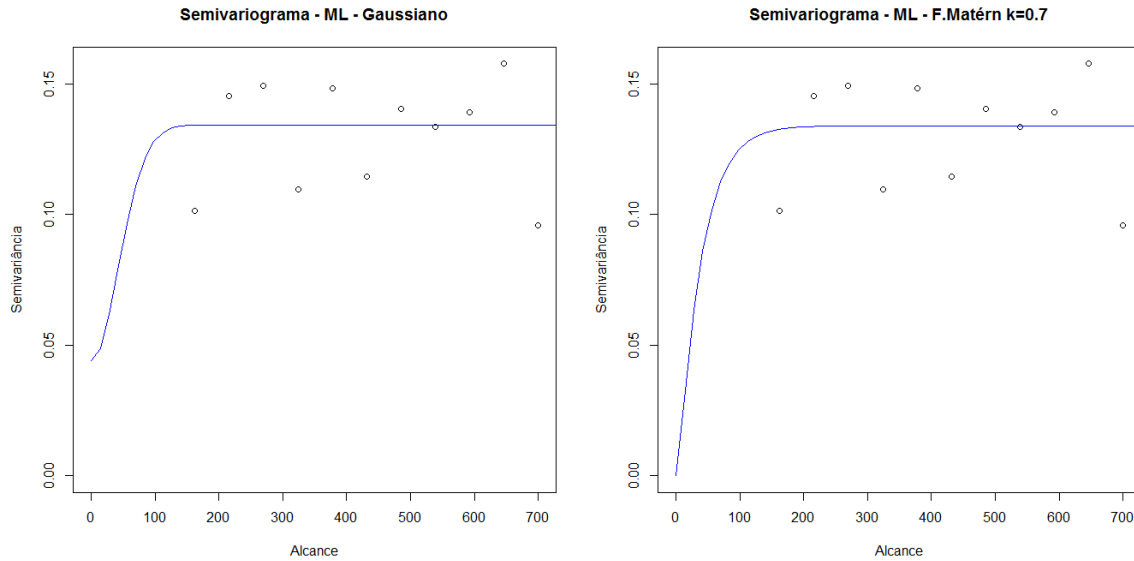


Figura 8 Semivariogramas dos modelos Gaussiano e F.Matérn

Para cada modelo ajustado a partir de um chute inicial, para o método da máxima verossimilhança (ML), foram obtidos os valores da média (β), efeito pepita (φ_1), contribuição (φ_2), alcance (a) e EPR.

Tabela 4: Estimativas de Máxima Verossimilhança dos parâmetros associados ao modelo

Método	Modelos	β	φ_1	φ_2	$\varphi_3 = g(a)$	a	EPR
ML	Exponencial	5,386	0,0000	0,1339	37,14	111,262	-
	Esférico	5,384	0,0936	0,0407	217,58	217,583	0,696
	Gaussiano	5,385	0,0440	0,0902	60,05	103,942	0,327
	Matérn (K=0,7)	5,386	0,0000	0,1340	31,63	109,023	-

β : média; φ_1 : efeito pepita; φ_2 : contribuição; $\varphi_3 = g(a)$: função do alcance; a : alcance;

EPR = $\left(\frac{\varphi_1}{\varphi_1 + \varphi_2}\right) \times 100$: efeito pepita relativo.

Para a escolha do melhor modelo ajustado em cada um dos métodos (OLS e ML) foi utilizado o critério de validação cruzada (Faraco, 2008). O modelo escolhido será aquele que apresentar os menores Erro Médio (EM) e Erro Médio Reduzido (ER), Desvio padrão do erro médio (S_{EM}), menor possível, Desvio padrão do erro reduzido (S_{ER}) mais próximo de 1 e Erro Absoluto (EA) seja o menor possível. Além disso, para o método ML foi calculado os parâmetros de Akaike (AIC) e Akaike corrigido (BIC). A tabela abaixo apresenta os parâmetros calculados pelo método de validação cruzada.

Tabela 5: Parâmetros para escolha dos modelos ajustados

Método	Modelos	EM	ER	S _{EM}	S _{ER}	EA	AIC	BIC
OLS	Exponencial	-0,00027	-0,00038	0,37719	1,05709	14,080	-	-
	Esférico	-0,00028	-0,00040	0,37771	1,06567	14,053	-	-
	Gaussiano	0,00025	0,00035	0,38069	1,05881	14,071	-	-
	Matérn (k=0,7)	-0,00112	-0,00162	0,37494	1,07381	14,064	-	-
ML	Exponencial	-0,00063	-0,00086	0,37090	1,01703	13,853	45,53	52,84
	Esférico	-0,00111	-0,00153	0,36961	1,01630	13,796	45,55	52,86
	Gaussiano	0,00044	0,00059	0,37115	1,01955	13,890	45,44	52,75
	Matérn (k=0,7)	-0,00062	-0,00085	0,37083	1,01773	13,866	45,50	52,81

EM: erro médio; ER: erro reduzido; S_{EM}: desvio padrão do erro médio; S_{ER}: desvio padrão do erro reduzido; EA: erro absoluto; AIC: Método de Akaike; BIC: Método de Akaike modificado.

Assim pela tabela podemos assumir que o modelo que melhor se ajusta ao semivariograma experimental foi o modelo esférico pelo método de máxima verossimilhança (RML).

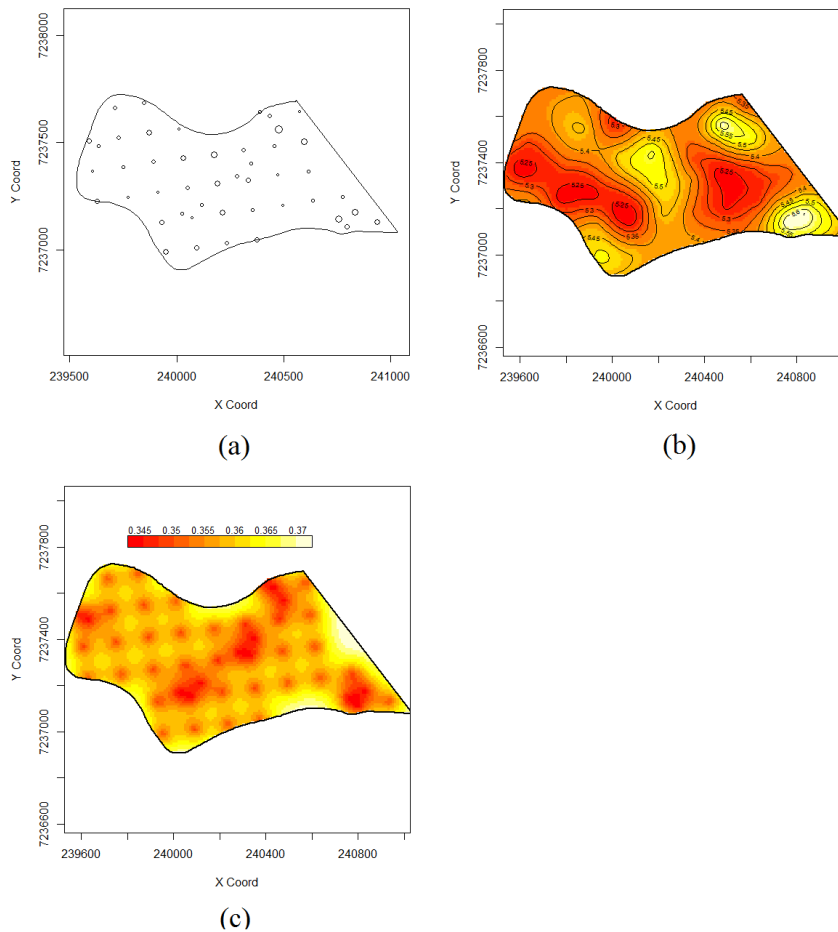


Figura 9. Mapa de *krigagem* (a), mapa dos valores preditos (b) e mapa do erro padrão mostrando a distribuição espacial de pH do solo

O ajuste do semivariograma e a interpolação realizada pela krigagem ordinária possibilitaram a confecção dos mapas (Figura 9). Na Figura 9(b) mostra o mapa dos valores ajustados pelo modelo para os valores pH do solo. Esse mapa indica as regiões em que ocorrem as maiores e as menores estimativas pelo método da krigagem ordinária. Como pode ser observado no mapa, as regiões vermelhas apresentaram baixos índices pH, já no região amarelas tem maior índices de pH do solo. Na Figura9(c) é o mapa do erro padrão, que mostrou pelas cores amarelas nas regiões apresentou altos índices e nas cores vermelhas das regiões apresentaram baixos índices.

Utilizando os algoritmos MCMC da abordagem bayesiana para obter alguns resultados para as análises de geoestatística. Foram executadas 10 mil simulações e o período de aquecimento foi de 100 simulações.

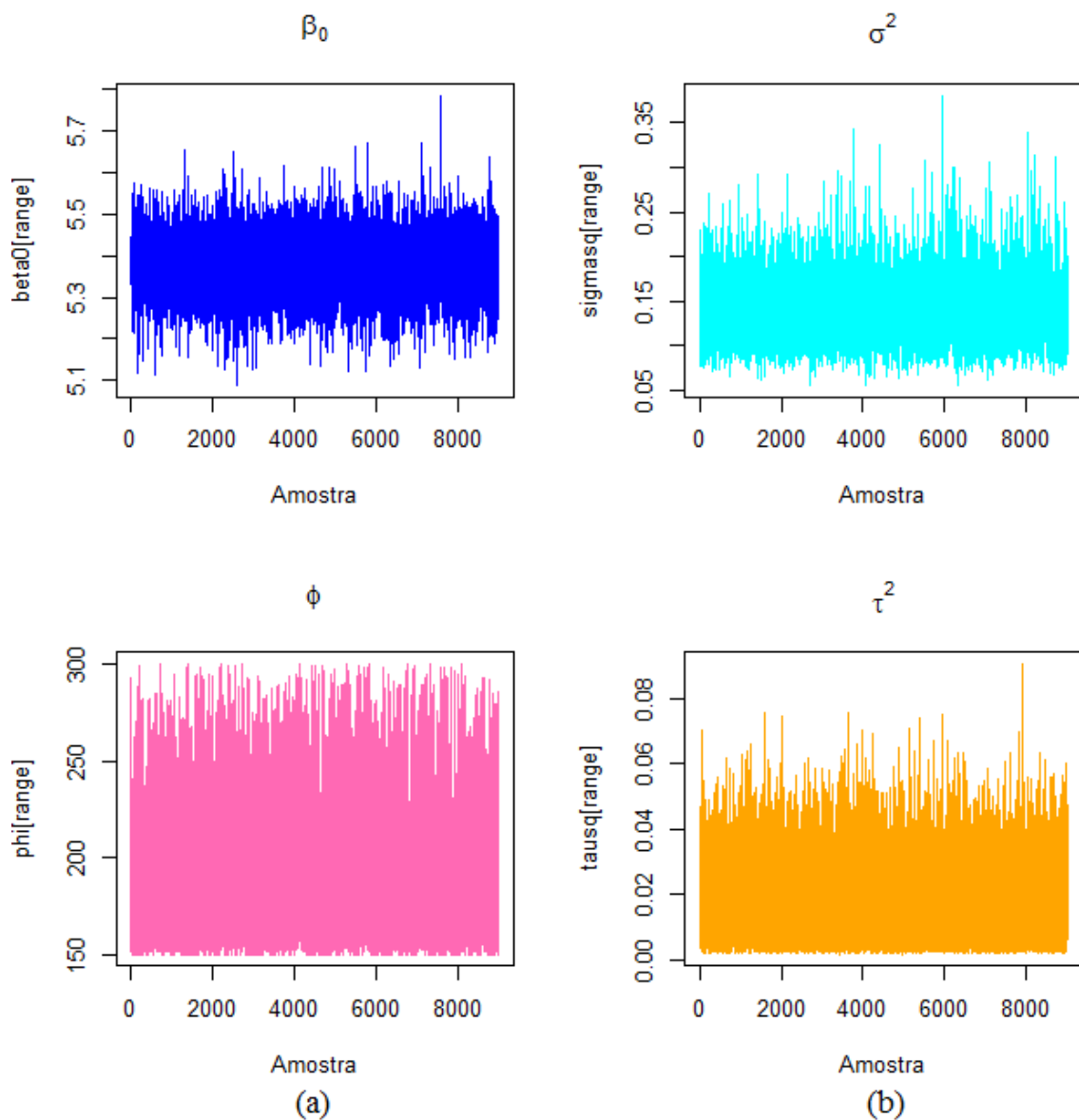


Figura 10 (a) e (b). Traço de uma cadeia a posteriori dos parâmetros

A Figura 10 apresentou os gráficos dos traços de uma cadeia a posteriori para os parâmetros do efeito fixo (a) e dos efeitos aleatórios (b) do modelo ajustado (esférico pelo método de máxima verossimilhança). Podemos observar pela gráfico que a linha foi aleatória e não ficou fixa em nenhuma região ao longo das iterações para β_0 e σ^2 , permitindo verificar que existe fortes evidências para a convergência da cadeia. Já para ϕ e τ^2 não apresentaram o mesmo resultado.

Tabela 6: Resumo da distribuição a posteriori dos parâmetros do modelo ajustado.

Parâmetros	Média	2,5%	Mediana	97,5%
β	5,3752	5,230e+00	5,3757	5,5154
φ_1	0,1374	8,357e-02	0,1319	0,2225
φ_2	0,0194	2,371e-03	0,0171	0,0496
φ_3	188,2722	1,500e+02	170	310

Na Tabela 6 foi apresentado um resumo da distribuição a posteriori de todos os parâmetros do modelo ajustado (esférico pelo método de máxima verossimilhança). Observa-se que as estimativas para os parâmetros pelo método Bayesiano foram próximas das estimativas realizadas na abordagem clássica.

4. Conclusão

Neste trabalho foi feita avaliação de dois métodos de ajuste aplicados a quatro modelos de função de semivariância, a partir das técnicas de validação cruzada e preditiva e do critério de Akaike (Akaike's Information Criterion - AIC), concluímos que o modelo que melhor se ajusta ao semivariograma experimental foi o modelo esférico pelo método de máxima verossimilhança (RML) para os dados de pH do solo.

5. Referências

- AKAIKE, H. Information measures and model selection. **International Statistical Institute, Voorburg**, v.44, p.277-291, 1983.
- ASSUMPÇÃO, R.A.B.; URIBE-OPAZO, M.A.; SOUZA, E.G. JOHANN, J.A. Uso da Uso da krigagem na avaliação da probabilidade de soja segundo os padrões regional, estadual e nacional. **Acta Sci. Agron.** Maringá, v. 29, n. 2, p. 165-171, 2007.
- BUSSAB, W.O.; MORETTIN, P.A. **Estatística básica**. 5.ed. São Paulo: Saraiva, 2002. 526p.
- COSTA, J.M. OLIVEIRA. E.F. **Fertilidade do solo e nutrição de plantas**. 2. ed. Campo Mourão: COAMO/ COODETEC, 2001. 93p.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. Sistema brasileiro de classificação de solos. 2. ed. Rio de Janeiro: Embrapa, 2006. 306 p.

GIRARDELLO, V.C.; AMADO, T.J.C.; NICOLOSO, R.S.; HORBE, T.A.N.; FERREIRA, A.O.; TABALDI, F.M.; LANZANOVA, M.E. Alterações nos atributos físicos de um latossolo vermelho sob plantio direto induzidas por diferentes tipos de escarificadores e rendimento da soja. **R. Bras. Ci. Solo**, 35:2115-2126, 2011

GOMES, F. Curso de estatística experimental. 14. ed. Piracicaba: Livraria Nobel S/A, 477, p. 2000.

GUERRA, P.A.G. **Geoestatística operacional**. Brasília: Ministério das Minas e Energia, 1988.

KULBACK, S.; LEIBLER, R.A. On information and sufficiency. **Annals of Mathematical Statistics**, v.22, p.79-86, 1951.

MATHERON, G. Principles of geostatistics. **Econ. Geol.**, 58:1246-66, 1963.

McBRATNEY, A.G.; WEBSTER, A.G. Choosing functions for semi-variograms and fitting them to sampling estimates. **Journal of Soil Science**, v.37, p.617-39, 1986.

MELLO, J.M.; BATISTA, J.L.F.; RIBEIRO JÚNIOR., P.J.; OLIVEIRA, M.S. Ajuste e seleção de modelos espaciais de semivariograma visando à estimativa volumétrica de *Eucalyptus grandis*. **SCIENTIA FORESTALIS**. n. 69, p.25-37, dez. 2005.

OPROMOLLA, P.A.; DALBEN, I.; CARDIM, M. Análise geoestatística de casos de hanseníase no Estado de São Paulo, 1991-2002. **Rev Saúde Pública**. 2006;40(5):907-13

RIBEIRO JÚNIOR, P.J. **Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo**. 1995. 99p. Dissertação (Mestrado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1995.

WONNACOTT, T.H.; WONNACOTT, R.J. **Introductory statistics**. 2.ed. New York: John Wiley Sons, 1972. 510p.