



Revista Brasileira de Ciência do Solo

ISSN: 0100-0683

revista@sbcs.org.br

Sociedade Brasileira de Ciência do Solo
Brasil

Nunes Nesi, Cristiano; Ribeiro, Andressa; Bonat, Wagner Hugo; Ribeiro Junior, Paulo Justiniano
VEROSSIMILHANÇA NA SELEÇÃO DE MODELOS PARA PREDIÇÃO ESPACIAL

Revista Brasileira de Ciência do Solo, vol. 37, núm. 2, 2013, pp. 352-358

Sociedade Brasileira de Ciência do Solo

Viçosa, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=180226346006>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica

Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal

Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

VEROSSIMILHANÇA NA SELEÇÃO DE MODELOS PARA PREDIÇÃO ESPACIAL⁽¹⁾

Cristiano Nunes Nesi⁽²⁾, Andressa Ribeiro⁽³⁾, Wagner Hugo Bonat⁽⁴⁾ & Paulo Justiniano
Ribeiro Junior⁽⁵⁾

RESUMO

Na modelagem estatística da variabilidade espacial, estimam-se os parâmetros da dependência espacial, que são utilizados na interpolação de valores em locais não amostrados. Para tal, o processo de modelagem deve ser realizado com critérios estatísticos que garantam previsões confiáveis e representem a real variabilidade local. Este trabalho avaliou diferentes formulações do modelo geoestatístico gaussiano para reconstituir a superfície que representa o fósforo (P) na área, a partir de medições dos teores de P em 48 parcelas experimentais localizadas em Xanxerê, SC, destacando o método utilizado nas análises. A combinação da presença de covariáveis no modelo e a necessidade de transformação para normalidade dos dados definiram quatro alternativas para modelagem. Utilizou-se a função de correlação de Matérn, avaliada nos valores 0,5; 1,5; e 2,5 para parâmetro de suavidade. Os modelos foram comparados pelo valor maximizado do logaritmo da função de verossimilhança e também por validação cruzada. O modelo selecionado foi o que incorporou a variável resposta transformada, as coordenadas da área como covariáveis e o valor 0,5 para o parâmetro de suavidade. As medidas de validação cruzada pouco acrescentaram aos resultados de comparação por verossimilhança, que evidenciaram que na modelagem geoestatística, o cuidado com observações globais ou locais atípicas, além da seleção com base em diferentes modelos, deve ser o foco para obter resultados compatíveis com a realidade.

Termos para indexação: modelagem geoestatística, função de correlação, estimação.

⁽¹⁾ Recebido para publicação em 8 de agosto de 2012 e aprovado em 15 de fevereiro de 2013.

⁽²⁾ Engenheiro-Agrônomo, Doutor, Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina - Epagri, Servidão Ferdinando Tusset, s/n, Bairro São Cristóvão. CEP 89801-970 Chapecó (SC), Brasil. E-mail: cristiano@epagri.sc.gov.br

⁽³⁾ Engenheira-Florestal, Doutoranda no Departamento de Ciências Florestais da Universidade Federal de Lavras - UFPA. Caixa Postal 3037. CEP 37200-000 Lavras (MG). E-mail: andressa.florestal@gmail.com

⁽⁴⁾ Estatístico, Mestre. Professor no Setor de Ciências Exatas, Laboratório de Estatística e Geoinformação - LEG, UFPR. Av. Cel. Francisco H. dos Santos, 210, Jardim das Américas. CEP 81531-970 Curitiba (PR). E-mail: wagner@leg.ufpr.br

⁽⁵⁾ Engenheiro-Agrônomo, PhD, Professor no Setor de Ciências Exatas LEG, UFPR. E-mail: paulojus@leg.ufpr.br

SUMMARY: LIKELIHOOD IN THE SELECTION OF MODELS FOR SPATIAL PREDICTION

Statistical modeling of spatial attributes aims to estimate dependence parameters, used to interpolate values at unsampled locations within the study area. Thus, the modeling process is conducted with some statistical criteria that ensure reliable predictions representing the actual local variability. This study evaluates different formulations of the Gaussian geostatistical model to reconstruct the surface of phosphorus over the area from the available measurements on 48 experimental plots located in Xanxerê/SC with emphasis on the methodological framework. Choices of covariates in the model and for data transformation define four modeling options to be assessed. The Matérn correlation function was used, evaluated at values 0.5; 1.5 and 2.5 for smoothness parameter. Models were compared by the maximized logarithm of the likelihood function and also by cross validation. The model with transformed response variable, including coordinates of the area as covariates and the value of 0.5 for the smoothness parameter was selected. The cross validation measures did not add relevant information to the likelihood, and the analysis highlights care must be taken with globally or locally atypical data, as well as the need of objective choice based on different candidate models which ought to be the focus of geostatistical modeling to ensure results compatible with reality.

Index terms: geostatistical modeling, correlation function, estimation.

INTRODUÇÃO

Os solos apresentam variabilidade em seus atributos, quando é analisada a sua estrutura vertical e horizontal. Essa variabilidade não é puramente aleatória, mas geralmente apresenta correlação ou dependência espacial, como pode ser observado nos trabalhos de Souza et al. (2006, 2008), Nanni et al. (2011) e Gontijo et al. (2012). Hoeting et al. (2006) apresentaram os problemas que podem ser encontrados quando a correlação espacial entre as amostras é ignorada e avaliaram o desempenho de alguns critérios para seleção de covariáveis em estudos de ecologia. Considerar amostras georreferenciadas é uma abordagem importante na análise de dados com dependência espacial, uma vez que valores observados em localizações mais próximas são, em geral, mais similares do que aqueles mais distantes e, assim, cada realização da variável contém informação de sua vizinhança. Nesse sentido, os modelos geostatísticos descrevem a distribuição espacial de uma variável, associando o grau de dependência a medidas de distância e de direção entre os pontos amostrados (Diggle & Ribeiro Jr, 2007).

Na modelagem estatística da variabilidade espacial, estimam-se os parâmetros que definem a estrutura da dependência espacial e esses são utilizados na interpolação de valores em locais não amostrados. Para que a interpolação produza previsões confiáveis e represente a real variabilidade local, o processo de modelagem deve ser realizado com critérios estatísticos objetivos como pode ser observado nos trabalhos de Mello et al. (2005), Oda-Souza et al. (2008), Ribeiro Jr et al. (2009), Ávila et al. (2010) e Borssoi et al. (2011). Nesse caso, a inspeção de pontos discrepantes, a verificação da necessidade de transformação da variável resposta, além da correta

seleção do modelo, são procedimentos importantes para a qualidade das inferências.

O estudo da variabilidade espacial das propriedades do solo tem grande importância na escolha da área experimental, na locação das unidades experimentais, na coleta de amostras e interpretação de resultados, no levantamento e na classificação de solos e nos esquemas de uso racional de fertilizantes (Leão et al., 2007). Portanto, o conhecimento da variabilidade espacial do P no solo pode ser determinante para melhor administrar o diagnóstico de áreas com deficiência desse elemento, bem como para a adoção quantitativa da adubação (Carvalho et al., 2003; Bottega et al., 2013). Diante do exposto, o objetivo deste trabalho foi avaliar diferentes formulações do modelo geostatístico Gaussiano para reconstituir a superfície dos teores de P sobre uma área experimental.

MATERIAL E MÉTODOS

As amostras de solo foram coletadas em uma área no município de Xanxerê, SC, cultivada com milho no verão e aveia para cobertura de solo no inverno. O solo do local pertence à classe de Latossolo Vermelho distroférico, o clima é mesotérmico úmido, com verões quentes e invernos frios, e temperatura média anual de 18,7 °C. A área foi dividida em 48 parcelas contíguas, cada uma medindo 2,4 x 5,0 m. Nas parcelas, foram coletadas cinco subamostras simples de solo na profundidade de 0-20 cm, que compuseram uma única amostra, para determinar o teor de P no solo, pelo método de Mehlich-1, conforme descrito por Tedesco et al. (1995). As coordenadas de cada parcela foram estabelecidas de forma arbitrária, análogas ao plano coordenado (X_{1i} e X_{2i} , em que i se refere à parcela).

As bordas da área foram determinadas, considerando-se as coordenadas $X_{10} = 0$ e $X_{20} = 0$ para o primeiro ponto.

O formato básico considerado para os dados foi (x_i, y_i) , $i = 1, 2, \dots, n$, em que $x_i = (x_{1i}, x_{2i})$ identifica a localização espacial; e y_i é o teor de P na posição x_i , que, potencialmente, pode ser medido em qualquer ponto dentro da região de estudo. Os dados foram inicialmente submetidos à análise exploratória espacial, a necessidade de transformação foi verificada pelo método de Box-Cox (Box & Cox, 1964), assim como foi avaliada a tendência espacial em diagramas de dispersão do teor de P (mg dm^{-3}), em razão das coordenadas.

Para o modelo geoestatístico utilizado, assumiu-se que as observações $y_i = y_1, \dots, y_n$ seguem o modelo linear Gaussiano, podendo-se descrever:

$$y_i = D\beta + S(x_i) + e_i \quad (1)$$

em que y_i é uma realização da variável aleatória observada na posição espacial x_i ; D , uma matriz $n \times p$ de covariáveis; e β , o correspondente vetor de parâmetros da regressão. O termo $S(x_i)$ denota um processo estacionário gaussiano latente na posição x_i , com média zero, variância σ^2 e função de correlação $\rho(u; \phi, k)$, com argumento dado pela distância u de separação entre observações e parametrizada por ϕ , que descreve o grau de associação espacial e o parâmetro de suavidade da função de correlação K . Para um conjunto finito de observações, os valores de $S(x_i)$ seguem uma distribuição normal multivariada ($S \sim NMV[0, \sigma^2 R(\phi)]$), em que $R(\phi)$ é a matriz de covariâncias $n \times n$ com (i, j) -ésimo elemento $\rho(u)$, sendo $u = \|x_i - x_j\|$ a distância euclidiana entre x_i e x_j . O termo e_i denota variáveis aleatórias, independentes e normalmente distribuídas com média zero e variância τ^2 . Seguindo a terminologia usual de geoestatística, τ^2 é o efeito pepita (*nugget*); e σ^2 , o patamar (*sill*). A distância a partir da origem até o patamar é chamada de alcance prático (*range*), que é uma função do parâmetro ϕ (Kitanidis, 1997). A relação entre o alcance prático e o parâmetro de escala ϕ depende do $kappa$, com alcances práticos dados por 3ϕ , $4,75\phi$ e $5,92\phi$ para $kappa$ igual a 0,5; 1,5; e 2,5, respectivamente (Diggle & Ribeiro Jr, 2007).

Neste estudo, a presença de covariáveis no modelo (coordenadas da área) e a necessidade de transformação para normalidade dos dados definiram quatro especificações do modelo geoestatístico: com os dados originais; com os dados transformados; com os dados originais e a inclusão de covariáveis (coordenadas); e com os dados transformados e a inclusão de covariáveis (coordenadas).

Os parâmetros dos modelos foram estimados por maximização da função log-verossimilhança, considerando-se o modelo apresentado em (1) com especificação de tendência espacial nas coordenadas ($\mu(x_i)$), feita com a inclusão de uma superfície de tendência linear. Assim, para $\mu(x_i) = D\beta$, $Y \sim N(D\beta,$

$G(\theta)$), com $G(\theta) = \sigma^2 R(\phi) + \tau^2 I$, em que R depende do escalar ou vetor de parâmetros ϕ e do parâmetro de suavidade k (Roa-Ureta & Niklitschek, 2007). Isto posto, a função de log-verossimilhança do modelo geoestatístico é expressa por

$$L(\beta, \tau^2, \sigma^2, \phi) = -0,5\{n \log(2\pi) + \log|G(\theta)|\} + (y - D\beta)^T [G(\theta)]^{-1} (y - D\beta) \quad (2)$$

e para os casos em que não se considera a superfície de tendência, a matriz D é um vetor unitário. A função de correlação utilizada foi a de Matérn (Matérn, 1986), considerando-se possíveis valores do parâmetro de suavidade ($kappa$) iguais a 0,5, 1,5 e 2,5. No caso de $kappa=0,5$, a função de correlação Matérn reduz-se à função de correlação exponencial; aumentando-se o valor de $kappa$, aproxima-se da função de correlação Gaussiana. Os valores de $kappa$ 0,5, 1,5 e 2,5 correspondem, respectivamente, ao processo não diferenciável, uma e duas vezes diferenciáveis. Processos com maior grau de diferenciabilidade apresentam variação espacial mais suave (Diggle & Ribeiro Jr, 2007).

Para comparar a capacidade preditiva dos modelos, é prática comum adotar o procedimento de validação cruzada, que consiste em retirar uma observação, reajustar os modelos e prever a observação retirada, repetindo-se o processo para cada uma das observações que compõem a amostra (Isaaks & Srivastava, 1989). Para cada ponto existirá o valor verdadeiro e o valor estimado e, portanto, o erro de estimação. Com o erro de estimação, foram calculadas medidas de adequabilidade de predição, como o desvio-padrão dos erros médios (DP_{EM}), erro médio reduzido (ER) e desvio-padrão dos erros reduzidos (S_{ER}), conforme proposto por McBratney & Webster (1986):

$$DP_{EM} = \frac{1}{n} \sum_{i=1}^n (z_{x_i} - \hat{z}_{x_i}) \quad (3)$$

$$ER = \frac{1}{n} \sum_{i=1}^n \frac{(z_{x_i} - \hat{z}_{x_i})}{\sigma(\hat{z}_{x_i})} \quad (4)$$

$$S_{ER} = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{|z_{x_i} - \hat{z}_{x_i}|}{\sigma(\hat{z}_{x_i})}} \quad (5)$$

em que n é o número de observações; \hat{z}_{x_i} , o valor observado no ponto x_i ; \hat{z}_{x_i} e $\sigma(\hat{z}_{x_i})$ são, respectivamente, o valor predito por krigagem e o desvio-padrão da krigagem, ambos no ponto x_i sem considerar a observação \hat{z}_{x_i} . Considera-se o melhor ajuste aquele que apresenta menor DP_{EM} , ER mais próximo de zero e S_{ER} mais próximo de um. Para melhor comparação dos modelos, foi calculado o nível de cobertura dos intervalos de predição, definido como o percentual de intervalos construídos que contém o verdadeiro valor da observação retirada (Bonat et al., 2011), assumindo-se o nível nominal de 95 %.

Após o ajuste dos modelos, realizou-se a predição espacial (krigagem) com uma malha de predição de 10.000 pontos, utilizando os modelos ajustados e os

teores de fósforo observados nos pontos distribuídos na área. Com os resultados das predições, elaboraram-se mapas para identificar zonas diferenciadas do teor de fósforo no solo.

Todas as análises foram realizadas com o ambiente R, versão 2.13.2 (R Development Core Team, 2011), utilizando-se o pacote geoR (Ribeiro Jr & Diggle, 2001).

RESULTADOS E DISCUSSÃO

A análise exploratória dos dados evidenciou três valores discrepantes com teores de P acima de 12 mg dm⁻³. Tais observações foram detectadas como discrepantes, mesmo após a transformação dos dados. A presença de valores discrepantes influencia principalmente a parte inicial do variograma, impactando em conclusões errôneas sobre a variância do ruído e, ou, a existência de correlação espacial (Diggle & Ribeiro Jr, 2007). Borssoi et al. (2011) apresentaram que a presença de valores atípicos entre dados amostrados pode exercer forte influência nos mapas, alterando a dependência espacial e, consequentemente, as interpretações. Valores discrepantes quando presentes e isolados são anomalias locais que se não retirados podem influenciar fortemente o modelo global, assumido como válido para toda área. Em virtude disso, as três amostras discrepantes foram excluídas para adequado ajuste dos modelos e interpretação dos mapas. A verificação da normalidade da variável resposta, mesmo sem os valores atípicos, evidenciou a necessidade de transformação dos dados, utilizando como parâmetro de transformação Box-Cox o valor de $\lambda=0,5$ na modelagem espacial. Kitanidis (1997) relatou que a suposição de normalidade é inapropriada em alguns casos, mas para atributos contínuos pode ser contornada com transformações, obtendo-se distribuição aproximadamente normal.

Na figura 1, estão ilustrados os gráficos exploratórios dos dados, após a retirada dos valores discrepantes. No gráfico superior esquerdo, os símbolos representam os quartis de uma distribuição empírica dos valores medidos, em que se observam regiões com teores de P semelhantes (agrupamentos de símbolos). No gráfico superior direito e no inferior esquerdo, tem-se a dispersão dos teores de P amostrados na área, plotados em razão das coordenadas espaciais, com uma linha suave para facilitar a visualização de tendências. Nesses gráficos, observa-se a redução nos teores de P na direção das duas coordenadas, indicando tendência linear em razão das coordenadas. Nesse caso, não há estacionaridade na média e a superfície de tendência deve ser escrita como um modelo de regressão polinomial, usando potências e produtos cruzados das coordenadas cartesianas como variáveis explicativas (Diggle & Ribeiro Jr, 2007). Lark et al. (2006)

destacaram que, na maioria dos estudos reportados na literatura da área de solos, o variograma é estimado com métodos seriamente viesados, se a estrutura de efeitos fixos é mais complexa que apenas um modelo de médias constantes. No estudo de Chai et al. (2008), a inclusão de variáveis topográficas como padrão externo no modelo geoestatístico melhorou a predição espacial do teor de matéria orgânica no solo. O gráfico inferior direito apresentou um histograma dos teores de P medidos na área, com maior densidade entre os valores 4 a 6 mg dm⁻³.

No quadro 1, são apresentadas as estimativas dos parâmetros do modelo geoestatístico e do logaritmo da função de verossimilhança para cada estratégia de modelagem e valor de $kappa$. Os modelos 1 e 2 referem-se ao ajuste do modelo sob pressuposição da média constante, respectivamente sem e com a transformação da variável resposta. Para esses modelos, têm-se as estimativas da média geral do processo sobre a área (β_0); dos parâmetros adicionais que definem a estrutura de covariância dos dados, ou seja, a variância do processo espacial (σ^2) e a variância do ruído ou erro de medição (τ^2); e do parâmetro da função de correlação (ϕ). Incluindo-se uma função linear para descrever a média variando espacialmente (modelos 3 e 4), observa-se que a média do processo estocástico é agora representada por uma relação linear com as coordenadas X_1 e X_2 por meio

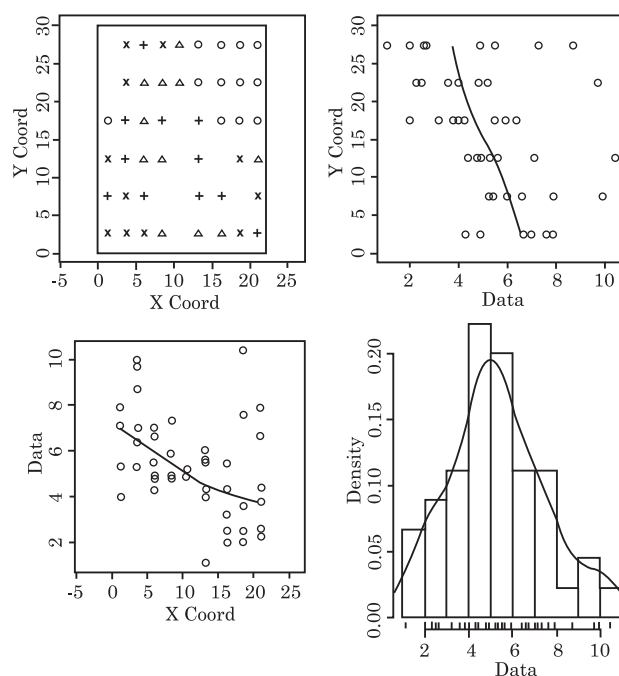


Figura 1. Pontos de locação (superior esquerdo), dispersão dos valores contra as coordenadas da área (superior direito e inferior esquerdo) e histograma (inferior direito) para o teor de fósforo no solo (mg dm⁻³) em 45 pontos amostrais (excluídos os valores discrepantes) na camada 0-20 cm de profundidade do solo.

Quadro 1. Parâmetros do modelo geoestatístico Gaussiano estimados por máxima verossimilhança para o teor de fósforo no solo

Modelo	kappa	Estimativa							AP	L-V
		β_0	β_1	β_2	τ^2	σ^2	$\tau^2+\sigma^2$	ϕ		
Modelo com média constante										
1	0,5	5,557	-	-	1,807	3,010	4,820	8,881	26,64	-92,32
	1,5	5,587	-	-	2,277	2,650	4,930	4,927	23,40	-91,88
	2,5	5,588	-	-	2,324	2,603	4,930	3,596	21,29	-91,70
2	0,5	2,609	-	-	0,297	0,685	0,980	9,749	29,25	-89,92
	1,5	2,623	-	-	0,402	0,601	1,000	5,214	24,77	-89,47
	2,5	2,624	-	-	0,413	0,586	1,000	3,770	22,32	-89,30
Modelo com tendência linear										
3	0,5	8,028	-0,118	-0,085	2,022	1,227	3,250	3,321	9,96	-89,60
	1,5	8,068	-0,121	-0,090	3,210	0,000	3,210	0,000	0	-90,10
	2,5	8,068	-0,121	-0,090	3,210	0,000	3,210	0,000	0	-90,10
4	0,5	3,805	-0,055	-0,042	0,309	0,308	0,620	3,454	10,36	-87,12
	1,5	3,839	-0,057	-0,044	0,605	0,000	0,610	0,000	0	-88,08
	2,5	3,839	-0,057	-0,044	0,605	0,000	0,610	0,000	0	-88,80

β_0 : teor médio de fósforo na área; β_1 e β_2 : parâmetros relacionados às covariáveis (coordenadas); τ^2 , σ^2 e ϕ : parâmetros de covariância; AP: alcance prático, dado em razão de ϕ : 3ϕ , $4,75\phi$ e $5,92\phi$ para kappa 0,5; 1,5; e 2,5, respectivamente; L-V: logaritmo da função de verossimilhança; 1: modelo ajustado aos dados originais; 2: modelo ajustado aos dados transformados; 3: modelo ajustado aos dados originais com covariáveis (coordenadas); e 4: modelo ajustado aos dados transformados com covariáveis (coordenadas).

da equação $\mu(x_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$, em que β_0 , β_1 e β_2 têm suas estimativas apresentadas no quadro 1.

Independentemente do valor do parâmetro de suavidade *kappa*, observou-se redução nas estimativas dos parâmetros de covariância com a transformação. Com a inclusão de covariáveis, há redução nas estimativas de σ^2 e ϕ , além de casos em que foram zeradas, indicativos de ausência de dependência espacial. Quanto maior a proporção da variância proveniente do sinal (σ^2) em relação ao total ($\tau^2 + \sigma^2$), mais forte é a dependência espacial e melhor serão as predições (Bonat et al., 2011). Nessa situação, a variação espacial no teor de P é totalmente explicada pela relação linear com as coordenadas, quando se aumenta o parâmetro de suavidade (1,5 ou 2,5), pois o maior valor de *kappa* substitui a tendência nas coordenadas, o que acarreta em alcances práticos estimados pelos modelos 3 e 4, inferiores aos dos modelos 1 e 2. No modelo de Matérn, a correlação aproxima-se de zero apenas assintoticamente e o alcance prático é definido como a distância para a qual a correlação é 0,05 (Diggle & Ribeiro Jr, 2007). Foram estimados valores de alcance prático de até 29 m para o modelo 2 e *kappa*=0,5 (Quadro 1). Os valores de log-verossimilhança (L-V) aumentaram com a transformação nos dados e com a inclusão das coordenadas no modelo, o que aumenta o número de parâmetros.

No quadro 2, estão listados os resultados de validação cruzada. Pelo critério proposto por McBratney & Webster (1986) e utilizado por Faraco et al. (2008) e Ávila et al. (2010), a escolha do modelo

é difícil, pois não há um modelo que atenda, simultaneamente, ao critério de menor desvio-padrão do erro médio, erro médio reduzido próximo de zero e desvio-padrão do erro reduzido próximo de um.

Observaram-se nos modelos com tendência linear, em geral, maiores desvio-padrões dos erros médios (DP_{EM}), erro médio reduzido (ER) distante de zero e piores coberturas dos intervalos de predição. Esses fatos se devem à maior proporção de τ^2 em relação ao *sill* total e aos maiores valores de *kappa* que, aliados à amostra pequena, tornam mais difícil a estimação e predição, conforme observado por Bonat et al. (2011). Entretanto, com transformação nos dados (modelo 4) e processos menos suaves (*kappa*=0,5), têm-se boas medidas de adequacidade da predição, com menor valor de DP_{EM} e ER próximo a zero, além do nível de cobertura do intervalo de predição próximo a 90 %. Isso corrobora com os resultados, quando o modelo é escolhido pelo maior valor do logaritmo da função de verossimilhança, como observado no quadro 1.

Esses fatos evidenciaram os bons resultados com critérios com base na verossimilhança para a seleção de modelos, principalmente em situações com poucas observações, já que a verossimilhança é uma medida de compatibilidade do modelo com os dados observados. Com os recursos computacionais atuais, o método da verossimilhança deve ser a técnica de estimação padrão, pois esse tem propriedades estatísticas ótimas, ou seja, quanto maior o tamanho da amostra, mais próximas as estimativas estarão dos verdadeiros valores (estimador consistente); sua esperança

Quadro 2. Desvio-padrão do erro médio (DP_{EM}) e do erro médio reduzido (ER) e desvio-padrão do erro reduzido (S_{ER}) e do nível de cobertura dos intervalos de predição obtidos por validação cruzada, para diferentes estratégias de modelagem do teor de fósforo

Modelo ⁽¹⁾	kappa	DP_{EM}	ER	S_{ER}	Cobertura (%)
Modelo com média constante					
1	0,5	0,0210	0,0171	1,0702	93,33
	1,5	0,0183	0,0145	1,0473	93,33
	2,5	0,0166	0,0130	1,0395	93,33
2	0,5	0,0153	0,0116	1,0322	91,11
	1,5	0,0122	0,0091	1,0140	91,11
	2,5	0,0105	0,0079	1,0088	91,11
Modelo com tendência linear					
3	0,5	0,0513	0,0441	1,1279	24,44
	1,5	0,0258	0,0212	1,0701	13,33
	2,5	0,0310	0,0261	1,1093	8,89
4	0,5	0,0078	-0,0057	1,0252	88,89
	1,5	0,1558	-0,1133	1,0029	17,78
	2,5	0,1479	-0,1127	1,0369	2,22

(1) 1: modelo ajustado aos dados originais; 2: modelo ajustado aos dados transformados; 3: modelo ajustado aos dados originais com covariáveis (coordenadas); e 4: modelo ajustado aos dados transformados com covariáveis (coordenadas).

matemática é igual ao valor estimado (não viesado) e é eficiente, uma vez que quanto maior a amostra, maior é a precisão das estimativas (Azzalini, 1996; Oda-Souza et al., 2008). Além disso, esse método possibilita a comparação efetiva de modelos e a construção de testes de hipóteses, o que torna a escolha mais objetiva.

Com base nos valores de log-verossimilhança e da validação cruzada, a escolha da estratégia de modelagem recai no modelo com transformação na variável resposta e inclusão de superfície de tendência (modelo 4), com parâmetro de suavidade $kappa=0,5$. Esse modelo evidencia a dependência espacial “ruidosa” do teor de P e da sua relação com as coordenadas da área. Nesta situação, o patamar no semivariograma seria composto por metade sinal (σ^2) e metade ruído (τ^2).

Em todos os mapas de predição obtidos por krigagem (Figura 2), independentemente do modelo, observaram-se zonas diferenciadas do teor de P no solo, com maiores diferenças entre modelos sem (1 e 2) e com (3 e 4) tendência linear nas coordenadas. Para modelos com médias não constantes (3 e 4) e valores de $kappa$ iguais a 1,5 e 2,5, a variabilidade espacial no teor de P é função apenas das coordenadas. Diante disso, o mapa que proporciona o gerenciamento correto de sítios específicos do teor de P no solo é aquele obtido a partir do ajuste do modelo com tendência nas coordenadas, dados transformados por raiz quadrada e função de correlação exponencial ($kappa=0,5$).

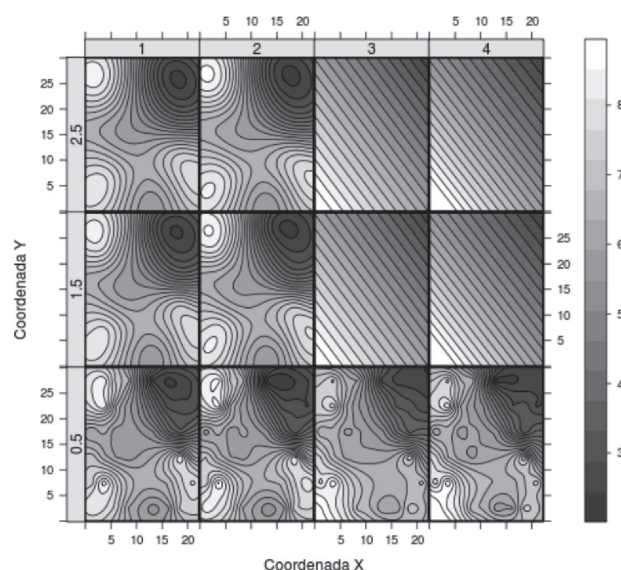


Figura 2. Mapas de teores de fósforo estimados por krigagem em uma malha regular de 10.000 pontos com base em 45 pontos amostrais para diferentes valores do parâmetro de suavidade ($kappa$ 0,5, 1,5 e 2,5). 1: modelo ajustado aos dados originais; 2: modelo ajustado aos dados transformados; 3: modelo ajustado aos dados originais com covariáveis (coordenadas); e 4: modelo ajustado aos dados transformados com covariáveis (coordenadas).

CONCLUSÕES

1. A transformação raiz quadrada na variável resposta, a inclusão das coordenadas da área como covariáveis no modelo de médias não constantes e a inclusão do parâmetro de suavidade adicional para a função de correlação ($kappa=0,5$) definiram o modelo com melhores características de inferência e permitiram aumentar a acurácia na reconstituição da superfície contínua do teor de fósforo.

2. As medidas de validação cruzada pouco acrescentaram aos resultados de comparação por verossimilhança, evidenciando o fato de que na modelagem geoestatística o cuidado com observações global e localmente influentes, além da seleção com base em diferentes modelos, deve ser o foco para obter resultados compatíveis com os dados.

LITERATURA CITADA

- ÁVILA, L.F.; MELLO, C.R. & SILVA, A.M. Continuidade e distribuição espacial da umidade do solo em bacia hidrográfica da Serra da Mantiqueira. R. Bras. Eng. Agríc. Amb., 14:1257-1266, 2010.
- AZZALINI, A. Statistical inference: Based on the likelihood. London, Chapman & Hall, 1996. 352p.

- BONAT, W.H.; RIBEIRO JR, P.J. & ZEVIANI, W.M. Comparando predições por modelos geoestatísticos e aditivo generalizado para reconstituição de superfícies contínuas gaussianas. *R. Eng. Agríc.*, 26:119-128, 2011.
- BORSSOI, J.A.; URIBE-OPAZO, M.A. & GALEA, M. Técnicas de diagnóstico de influência local na análise espacial da produtividade da soja. *Eng. Agríc.*, 31:376-387, 2011.
- BOTTEGA, E.L.; QUEIROZ, D.M.; PINTO, F.A.C. & SOUZA, C.M.A. Variabilidade espacial de atributos do solo em sistema de semeadura direta com rotação de culturas no cerrado brasileiro. *R. Ci. Agron.*, 44:1-9, 2013.
- BOX, G.E.P. & COX, D.R. An analysis of transformations. *J. Royal Stat. Soc.*, 26:211-252, 1964.
- CARVALHO, M.P.; TAKEDA, E.Y. & FREDDI, O.S. Variabilidade espacial de atributos de um solo sob videira em Vitória Brasil (SP). *R. Bras. Ci. Solo*, 27:695-703, 2003.
- CHAI, X.; SHEN, C.; YUAN, X. & HUANG, Y. Spatial prediction of soil organic matter in the presence of different external trends with REML-EBLUP. *Geoderma*, 148:159-166, 2008.
- DIGGLE, P.J. & RIBEIRO JR, P.J. Model-based geostatistics. New York, Springer, 2007. 230p.
- FARACO, M.A.; URIBE-OPAZO, M.A.; SILVA, E.A.A.; JOHANN, J.A. & BORSSOI, J.A. Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja. *R. Bras. Ci. Solo*, 32:463-476, 2008.
- GONTIJO, I.; NICOLE, L.R.; PARTELLI, F.L.; BONOMO, R. & SANTOS, E.O.J. Variabilidade e correlação espacial de micronutrientes e matéria orgânica do solo com a produtividade da pimenta-do-reino. *R. Bras. Ci. Solo*, 36:1093-1102, 2012.
- HOETING, J.A.; DAVIS, R.A.; MERTON, A.A. & THOMPSON, S.E. Model selection for geoestatistical models. *Ecol. Applic.*, 16:87-98, 2006.
- ISAACS, E.H. & SRIVASTAVA, R.M. An introduction to applied geostatistics. New York, Oxford University Press, 1989. 561p.
- KITANIDIS, J.K. Introduction to geostatistics - Applications in hydrogeology. New York, Cambridge University Press, 1997. 249p.
- LARK, R.M.; CULLIS, B.R. & WELHAM, S.J. On spatial prediction of soil properties in the presence of a spatial trend: The empirical best linear unbiased predictor (EBLUP) with REML. *Eur. J. Soil Sci.*, 57:787-799, 2006.
- LEÃO, A.B.; ANDRADE, A.R.S.; CHAVES, L.H.G.; CHAVES, I.B. & GUERRA, H.O.C. Variabilidade espacial de fósforo em solo do perímetro irrigado Engenheiro Arcoverde, PB. *R. Ci. Agron.*, 38:1-6, 2007.
- MATÈRN, B. Spatial Variation. Berlin, Springer, 1986.
- McBRATNEY, A. & WEBSTER, R. Choosing functions for semi-variograms of soil properties and fitting them to sample estimates. *Soil Sci.*, 37:617-639, 1986.
- MELLO, J.M.; BATISTA, J.L.F.; RIBEIRO JR, P.J. & OLIVEIRA, M.S. Ajuste e seleção de modelos espaciais de semivariograma visando à estimativa volumétrica de *Eucalyptus grandis*. *Sci. Flor.*, 69:25-37, 2005.
- NANNI, M.R.; POVH, F.P.; DEMATTÊ, J.A.M.; OLIVEIRA, R.B.; CHICATI, M.L. & CEZAR, E. Optimum size in grid soil sampling for variable rate application in site-specific management. *Sci. Agric.*, 68:386-392, 2011.
- ODA-SOUZA, M.; BARBIN, D.; RIBEIRO JR, P.J. & STAPE, J.L. Aplicação de métodos geoestatísticos para identificação de dependência espacial na análise de dados de um ensaio de espaçamento florestal em delineamento sistemático tipo leque. *R. Árvore*, 32:499-509, 2008.
- R DEVELOPMENT CORE TEAM R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, 2011. Disponível em: <<http://www.R-project.org/>>.
- RIBEIRO JR, P.J. & DIGGLE, P.J. GeoR: A package for geostatistical analysis. *R. News*, 1:14-18, 2001.
- RIBEIRO JR, P.J.; VIOLA, D.N.; DEMÉTRIO, C.G.B.; MANLY, B.F. & FERNANDES, O.A. Spatial pattern detection modeling of thrips (*Thrips tabaci*) on onion fields. *Sci. Agric.*, 66:90-99, 2009.
- ROA-URETA, R. & NIKLITSCHKE, E. Biomass estimation from surveys with likelihood-based geostatistics. *J. Marine Sci.*, 64:1723-1734, 2007.
- SOUZA, Z.M.; MARQUES JÚNIOR, J.; PEREIRA, G.T. & BARBIERI, D.M. Small relief shape variations influence spatial variability of soil chemical attributes. *Sci. Agric.*, 63:161-168, 2006.
- SOUZA, G.S.; LIMA J.S.S.; SILVA, S.A. & OLIVEIRA, R.B. Variabilidade espacial de atributos químicos em um Argissolo sob pastagem. *Acta Sci. Agron.*, 30:589-596, 2008.
- TEDESCO, M.J.; GIANELLO, C. & BISSANI, C.A. Análise de solo, plantas e outros materiais. 2.ed. Porto Alegre, Universidade Federal do Rio Grande do Sul, 1995. 174p. (Boletim Técnico de Solos, 5)