

VEROSSIMILHANÇA NA SELEÇÃO DE MODELOS PARA PREDIÇÃO ESPACIAL⁽¹⁾

Cristiano Nunes Nesi⁽²⁾, Andressa Ribeiro⁽³⁾, Wagner Hugo Bonat⁽⁴⁾ & Paulo Justiniano Ribeiro Júnior⁽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 Matèrn, 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.

1⁰ Recebido para publicação em 8 de agosto de 2012 e aprovado em 15 de fevereiro de 2013.

2⁰ Engenheiro-Agrônomo, Doutor, Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina - Epagri, Servidão Ferdinando Tusset, s/nº, Bairro São Cristóvão, CEP 89801-970 Chapecó (SC), Brasil. E-mail: cristiano@epagri.sc.gov.br

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

4⁰ 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

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

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

28

29 **SUMMARY:** LIKELIHOOD IN THE SELECTION OF MODELS FOR SPATIAL PREDICTION

30

31 *Statistical modeling of spatial attributes aims to estimate dependence parameters, used to*
32 *interpolate values at unsampled locations within the study area. Thus, the modeling process is*
33 *conducted with some statistical criteria that ensure reliable predictions representing the actual*
34 *local variability. This study evaluates different formulations of the Gaussian geostatistical model to*
35 *reconstruct the surface of phosphorus over the area from the available measurements on 48*
36 *experimental plots located in Xanxerê/SC with emphasis on the methodological framework.*
37 *Choices of covariates in the model and for data transformation define four modeling options to be*
38 *assessed. The Matèrn correlation function was used, evaluated at values 0,5; 1.5 and 2.5 for*
39 *smoothness parameter. Models were compared by the maximized logarithm of the likelihood*
40 *function and also by cross validation. The model with transformed response variable, including*
41 *coordinates of the area as covariates and the value of 0.5 for the smoothness parameter was*
42 *selected based. The cross validation measures did not add relevant information to the likelihood,*
43 *and the analysis highlights care must be taken with globally or locally atypical data, as well as the*
44 *need of objective choice based on different candidate models which ought to be the focus of*
45 *geostatistical modeling to ensure results compatible with reality.*

46

47 **Index terms:** geoestatistical modeling, correlation function, estimation.

48

49

INTRODUÇÃO

50

51 Os solos apresentam variabilidade em seus atributos, quando é analisada a sua estrutura
52 vertical e horizontal. Essa variabilidade não é puramente aleatória, mas geralmente apresenta
53 correlação ou dependência espacial, como pode ser observado nos trabalhos de Souza et al. (2006,
54 2008), Nanni et al. (2011) e Gontijo et al. (2012). Hoeting et al. (2006) apresentaram os problemas
55 que podem ser encontrados quando a correlação espacial entre as amostras é ignorada e avaliaram o
56 desempenho de alguns critérios para seleção de covariáveis em estudos de ecologia. Considerar
57 amostras georreferenciadas é uma abordagem importante na análise de dados com dependência
58 espacial, uma vez que valores observados em localizações mais próximas são, em geral, mais

12

59 similares do que aqueles mais distantes e, assim, cada realização da variável contém informação de
60 sua vizinhança. Nesse sentido, os modelos geoestatísticos descrevem a distribuição espacial de uma
61 variável, associando o grau de dependência a medidas de distância e de direção entre os pontos
62 amostrados (Diggle & Ribeiro Júnior, 2007).

63 Na modelagem estatística da variabilidade espacial, estimam-se os parâmetros que definem a
64 estrutura da dependência espacial e esses são utilizados na interpolação de valores em locais não
65 amostrados. Para que a interpolação produza previsões confiáveis e represente a real variabilidade
66 local, o processo de modelagem deve ser realizado com critérios estatísticos objetivos como pode
67 ser observado nos trabalhos de Mello et al. (2005), Oda-Souza et al. (2008), Ribeiro Junior et al.
68 (2009), Ávila et al. (2010) e Borssoi et al. (2011). Nesse caso, a inspeção de pontos discrepantes, a
69 verificação da necessidade de transformação da variável resposta, além da correta seleção do
70 modelo, são procedimentos importantes para a qualidade das inferências.

71 O estudo da variabilidade espacial das propriedades do solo tem grande importância na
72 escolha da área experimental, na locação das unidades experimentais, na coleta de amostras e
73 interpretação de resultados, no levantamento e na classificação de solos e nos esquemas de uso
74 racional de fertilizantes (Leão et al., 2007). Portanto, o conhecimento da variabilidade espacial do P
75 no solo pode ser determinante para melhor administrar o diagnóstico de áreas com deficiência desse
76 elemento, bem como para a adoção quantitativa da adubação (Carvalho et al., 2003; Bottega et al.,
77 2013). Diante do exposto, o objetivo deste trabalho foi avaliar diferentes formulações do modelo
78 geoestatístico Gaussiano para reconstituir a superfície dos teores de P sobre uma área experimental.

79

80

MATERIAL E MÉTODOS

81

82 As amostras de solo foram coletadas em uma área no município de Xanxerê, SC, cultivada
83 com milho no verão e aveia para cobertura de solo no inverno. O solo do local pertence à classe de
84 Latossolo Vermelho distroférrico, o clima é mesotérmico úmido, com verões quentes e invernos
85 frios, e temperatura média anual de 18,7 °C. A área foi dividida em 48 parcelas contíguas, cada uma
86 medindo 2,4 x 5,0 m. Nas parcelas, foram coletadas cinco subamostras simples de solo na
87 profundidade de 0-20 cm, que compuseram uma única amostra, para determinar o teor de P no solo,
88 pelo método de Mehlich-1, conforme descrito por Tedesco et al. (1995). As coordenadas de cada
89 parcela foram estabelecidas de forma arbitrária, análogas ao plano coordenado (X_{1i} e X_{2i} , em que i
90 se refere à parcela). As bordas da área foram determinadas, considerando-se as coordenadas $X_{10} = 0$

13

3

91 e $X_{20} = 0$ para o primeiro ponto.

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

98 Para o modelo geoestatístico utilizado, assumiu-se que as observações $y_i = y_{1,\dots,n}$ seguem o
99 modelo linear Gaussiano, podendo-se descrever:

$$100 \quad y_i = D\beta + S(x_i) + e_i \quad (1)$$

101 em que y_i é uma realização da variável aleatória observada na posição espacial x_i ; D , uma matriz $n \times$
102 p de covariáveis; e β , o correspondente vetor de parâmetros da regressão. O termo $S(x_i)$ denota um
103 processo estacionário gaussiano latente na posição x_i , com média zero, variância σ^2 e função de
104 correlação $\rho(u; \phi, \kappa)$, com argumento dado pela distância u de separação entre observações e
105 parametrizada por ϕ , que descreve o grau de associação espacial e κ parâmetro de suavidade da
106 função de correlação κ . Para um conjunto finito de observações, os valores de $S(x_i)$ seguem uma
107 distribuição normal multivariada ($S \sim NMV [0, \sigma^2 R(\phi)]$), em que $R(\phi)$ é a matriz de covariâncias n
108 $\times n$ 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
109 denota variáveis aleatórias, independentes e normalmente distribuídas com média zero e variância
110 τ^2 . Seguindo a terminologia usual de geoestatística, τ^2 é o efeito pepita (*nugget*); e σ^2 , o patamar
111 (*sill*). A distância a partir da origem até o patamar é chamada de alcance prático (*range*), que é uma
112 função do parâmetro ϕ (Kitanidis, 1997). A relação entre o alcance prático e o parâmetro de escala
113 ϕ depende do *kappa*, com alcances práticos dados por 3ϕ , $4,75\phi$ e $5,92\phi$ para *kappa* igual a 0,5;
114 1,5; e 2,5, respectivamente (Diggle & Ribeiro Junior, 2007).

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

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

123 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
 124 de parâmetros ϕ e do parâmetro de suavidade κ (Roa-Ureta & Niklitschek, 2007). Isto posto, a
 125 função de log-verossimilhança do modelo geoestatístico é expressa por

$$126 \quad L(\beta, \tau^2, \sigma^2, \phi) = -0,5 \left\{ n \log(2\pi) + \log |G(\theta)| + (y - D\beta)^T [G(\theta)]^{-1} (y - D\beta) \right\} \quad (2)$$

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

135 Para comparar a capacidade preditiva dos modelos, é prática comum adotar o procedimento
 136 de validação cruzada, que consiste em retirar uma observação, reajustar os modelos e prever a
 137 observação retirada, repetindo-se o processo para cada uma das observações que compõem a
 138 amostra (Isaaks & Srivastava, 1989). Para cada ponto existirá o valor verdadeiro e o valor estimado
 139 e, portanto, o erro de estimação. Com o erro de estimação, foram calculadas medidas de
 140 adequacidade de predição, como o desvio-padrão dos erros médios (DP_{EM}), erro médio reduzido
 141 (ER) e desvio-padrão dos erros reduzidos (S_{ER}), conforme proposto por McBratney & Webster
 142 (1986):

$$143 \quad DP_{EM} = \frac{1}{n} \sum_{i=1}^n (z_{x_i} - \hat{z}_{x_i}); \quad (3)$$

$$144 \quad ER = \frac{1}{n} \sum_{i=1}^n \frac{(z_{x_i} - \hat{z}_{x_i})}{\sigma(\hat{z}_{x_i})}; \quad (4)$$

$$145 \quad S_{ER} = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{|z_{x_i} - \hat{z}_{x_i}|}{\sigma(\hat{z}_{x_i})}}, \quad (5)$$

146 em que n é o número de observações; z_{x_i} , o valor observado no ponto x_i ; e \hat{z}_{x_i} e $\sigma(\hat{z}_{x_i})$ são,
 147 respectivamente, o valor predito por krigagem e o desvio-padrão da krigagem, ambos no ponto x_i
 148 sem considerar a observação z_{x_i} . Considera-se o melhor ajuste aquele que apresenta menor DP_{EM} ,
 149 ER mais próximo de zero e S_{ER} mais próximo de um. Para melhor comparação dos modelos, foi
 150 calculado o nível de cobertura dos intervalos de predição, definido como o percentual de intervalos

151 construídos que contém o verdadeiro valor da observação retirada (Bonat et al., 2011), assumindo-
152 se o nível nominal de 95 %.

153 Após o ajuste dos modelos, realizou-se a predição espacial (krigagem) com uma malha de
154 predição de 10.000 pontos, utilizando os modelos ajustados e os teores de fósforo observados nos
155 pontos distribuídos na área. Com os resultados das predições, elaboraram-se mapas para identificar
156 zonas diferenciadas do teor de fósforo no solo.

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

159

160

RESULTADOS E DISCUSSÃO

161

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

177 Na figura 1, estão ilustrados os gráficos exploratórios dos dados, após a retirada dos valores
178 discrepantes. No gráfico superior esquerdo, os símbolos representam os quartis de uma distribuição
179 empírica dos valores medidos, em que se observam regiões com teores de P semelhantes
180 (agrupamentos de símbolos). No gráfico superior direito e no inferior esquerdo, tem-se a dispersão
181 dos teores de P amostrados na área, plotados em razão das coordenadas espaciais, com uma linha
182 suave para facilitar a visualização de tendências. Nesses gráficos, observa-se a redução nos teores

183 de P na direção das duas coordenadas, indicando tendência linear em razão das coordenadas. Nesse
184 caso, não há estacionaridade na média e a superfície de tendência deve ser escrita como um modelo
185 de regressão polinomial, usando potências e produtos cruzados das coordenadas cartesianas como
186 variáveis explicativas (Diggle & Ribeiro Junior, 2007). Lark et al. (2006) destacaram que, na
187 maioria dos estudos reportados na literatura da área de solos, o variograma é estimado com métodos
188 seriamente viesados, se a estrutura de efeitos fixos é mais complexa que apenas um modelo de
189 médias constantes. No estudo de Chai et al. (2008), a inclusão de variáveis topográficas como
190 padrão externo no modelo geoestatístico melhorou a predição espacial do teor de matéria orgânica
191 no solo. O gráfico inferior direito apresentou um histograma dos teores de P medidos na área, com
192 maior densidade entre os valores 4 a 6 mg dm⁻³.

193 No quadro 1, são apresentadas as estimativas dos parâmetros do modelo geoestatístico e do
194 logaritmo da função de verossimilhança para cada estratégia de modelagem e valor de *kappa*. Os
195 modelos 1 e 2 referem-se ao ajuste do modelo sob pressuposição da média constante,
196 respectivamente sem e com a transformação da variável resposta. Para esses modelos, têm-se as
197 estimativas da média geral do processo sobre a área (β_0); dos parâmetros adicionais que definem a
198 estrutura de covariância dos dados, ou seja, a variância do processo espacial (σ^2) e a variância do
199 ruído ou erro de medição (τ^2); e do parâmetro da função de correlação (ϕ). Incluindo-se uma função
200 linear para descrever a média variando espacialmente (modelos 3 e 4), observa-se que a média do
201 processo estocástico é agora representada por uma relação linear com as coordenadas X_1 e X_2 por
202 meio da equação $\mu(x_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$, em que β_0 , β_1 e β_2 têm suas estimativas apresentadas no
203 quadro 1.

204 Independentemente do valor do parâmetro de suavidade *kappa*, observou-se redução nas
205 estimativas dos parâmetros de covariância com a transformação. Com a inclusão de covariáveis, há
206 redução nas estimativas de σ^2 e ϕ , além de casos em que foram zeradas, indicativos de ausência de
207 dependência espacial. Quanto maior a proporção da variância proveniente do sinal (σ^2) em relação
208 ao total ($\tau^2 + \sigma^2$), mais forte é a dependência espacial e melhor serão as predições (Bonat et al.,
209 2011). Nessa situação, a variação espacial no teor de P é totalmente explicada pela relação linear
210 com as coordenadas, quando se aumenta o parâmetro de suavidade (1,5 ou 2,5), pois o maior valor
211 de *kappa* substitui a tendência nas coordenadas, o que acarreta em alcances práticos estimados pelos
212 modelos 3 e 4, inferiores aos dos modelos 1 e 2. No modelo de Matérn, a correlação aproxima-se de
213 zero apenas assintoticamente e o alcance prático é definido como a distância para a qual a
214 correlação é 0,05 (Diggle & Ribeiro Junior, 2007). Foram estimados valores de alcance prático de

215 até 29 m para o modelo 2 e $kappa=0,5$ (Quadro 1). Os valores de log-verossimilhança (L-V)
216 aumentaram com a transformação nos dados e com a inclusão das coordenadas no modelo, o que
217 aumenta o número de parâmetros.

218 No quadro 2, estão listados os resultados de validação cruzada. Pelo critério proposto por
219 McBratney & Webster (1986) e utilizado por Faraco et al. (2008) e Ávila et al. (2010), a escolha do
220 modelo é difícil, pois não há um modelo que atenda, simultaneamente, ao critério de menor desvio-
221 padrão do erro médio, erro médio reduzido próximo de zero e desvio-padrão do erro reduzido
222 próximo de um.

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

232 Esses fatos evidenciaram os bons resultados com critérios com base na verossimilhança para
233 a seleção de modelos, principalmente em situações com poucas observações, já que a
234 verossimilhança é uma medida de compatibilidade do modelo com os dados observados. Com os
235 recursos computacionais atuais, o método da verossimilhança deve ser a técnica de estimação
236 padrão, pois esse tem propriedades estatísticas ótimas, ou seja, quanto maior o tamanho da amostra,
237 mais próximas as estimativas estarão dos verdadeiros valores (estimador consistente); sua esperança
238 matemática é igual ao valor estimado (não viesado) e é eficiente, uma vez que quanto maior a
239 amostra, maior é a precisão das estimativas (Azzalini, 1996; Oda-Souza et al., 2008). Além disso,
240 esse método possibilita a comparação efetiva de modelos e a construção de testes de hipóteses, o
241 que torna a escolha mais objetiva.

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

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

255

256

CONCLUSÕES

257

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

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

267

268

LITERATURA CITADA

269

270 ÁVILA, L.F.; MELLO, C.R. & SILVA, A.M. Continuidade e distribuição espacial da umidade do
271 solo em bacia hidrográfica da Serra da Mantiqueira. R. Bras. Eng. Agríc. Amb., 14:1257-1266,
272 2010.

273 AZZALINI, A. Statistical inference: Based on the likelihood. London, Chapman & Hall, 1996.
274 352p.

275 BONAT, W.H.; RIBEIRO JÚNIOR, P.J. & ZEVIANI, W.M. Comparando predições por modelos
276 geoestatísticos e aditivo generalizado para reconstituição de superfícies contínuas gaussianas.
277 R. Eng. Agric., 26:119-128, 2011.

278 BORSSOI, J.A.; URIBE-OPAZO, M.A. & GALEA, M. Técnicas de diagnóstico de influência local

279 na análise espacial da produtividade da soja. *Eng. Agríc.*, 31:376-387, 2011.

280 BOTTEGA, E.L.; QUEIROZ, D.M.; PINTO, F.A.C. & SOUZA, C.M.A. Variabilidade espacial de
281 atributos do solo em sistema de semeadura direta com rotação de culturas no cerrado
282 brasileiro. *R. Ci. Agron.*, 44:1-9, 2013.

283 BOX, G.E.P. & COX, D.R. An analysis of transformations. *J. Royal Stat. Soc.*, 26:211-252, 1964.

284 CARVALHO, M.P.; TAKEDA, E.Y. & FREDDI, O.S. Variabilidade espacial de atributos de um
285 solo sob videira em Vitória Brasil (SP). *R. Bras. Ci. Solo*, 27:695-703, 2003.

286 CHAI, X.; SHEN, C.; YUAN, X. & HUANG, Y. Spatial prediction of soil organic matter in the
287 presence of different external trends with REML-EBLUP. *Geoderma*, 148:159-166, 2008.

288 DIGGLE, P.J. & RIBEIRO JÚNIOR, P.J. Model-based geostatistics. New York, Springer, 2007.
289 230p.

290 FARACO, M.A.; URIBE-OPAZO, M.A.; SILVA, E.A.A.; JOHANN, J.A. & BORSSOI, J.A.
291 Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos
292 físicos do solo e produtividade da soja. *R. Bras. Ci. Solo*, 32:463-476, 2008.

293 GONTIJO, I.; NICOLE, L.R.; PARTELLI, F.L.; BONOMO, R. & SANTOS, E.O.J. Variabilidade
294 e correlação espacial de micronutrientes e matéria orgânica do solo com a produtividade da
295 pimenta-do-reino. *R. Bras. Ci. Solo*, 36:1093-1102, 2012.

296 HOETING, J.A.; DAVIS, R.A.; MERTON, A.A. & THOMPSON, S.E. Model selection for
297 geo-statistical models. *Ecol. Applic.*, 16:87-98, 2006.

298 ISAAKS, E.H. & SRIVASTAVA, R.M. An introduction to applied geostatistics. New York,
299 Oxford University Press, 1989. 561p.

300 KITANIDIS, J.K. Introduction to geostatistics - Applications in hydrogeology. New York,
301 Cambridge University Press, 1997. 249p.

302 LARK, R.M.; CULLIS, B.R. & WELHAM, S.J. On spatial prediction of soil properties in the
303 presence of a spatial trend: The empirical best linear unbiased predictor (E-BLUP) with
304 REML. *Eur. J. Soil Sci.*, 57:787-799, 2006.

305 LEÃO, A.B.; ANDRADE, A.R.S.; CHAVES, L.H.G.; CHAVES, I.B. & GUERRA, H.O.C.
306 Variabilidade espacial de fósforo em solo do perímetro irrigado Engenheiro Arcoverde, PB.
307 *R. Ci. Agron.*, 38:1-6, 2007.

308 MATÈRN, B. *Spatial Variation*. Berlin, Springer, 1986.

309 McBRATNEY, A. & WEBSTER, R. Choosing functions for semi-variograms os soil properties
310 and fitting them to sample estimates. *Soil Sci.*, 37:617-639, 1986.

311 MELLO, J.M.; BATISTA, J.L.F.; RIBEIRO JÚNIOR, P.J. & OLIVEIRA, M.S. Ajuste e seleção de
312 modelos espaciais de semivariograma visando à estimativa volumétrica de *Eucalyptus*
313 *grandis*. Sci. Flor., 69:25-37, 2005.

314 NANNI, M.R.; POVH, F.P.; DEMATTÊ, J.A.M.; OLIVEIRA, R.B.; CHICATI, M.L. & CEZAR,
315 E. Optimum size in grid soil sampling for variable rate application in site-specific
316 management. Sci. Agric., 68:386-392, 2011.

317 ODA-SOUZA, M.; BARBIN, D.; RIBEIRO JÚNIOR, P.J. & STAPE, J.L. Aplicação de métodos
318 geoestatísticos para identificação de dependência espacial na análise de dados de um ensaio
319 de espaçamento florestal em delineamento sistemático tipo leque. R. Árvore, 32:499-509,
320 2008.

321 R DEVELOPMENT CORE TEAM R: A language and environment for statistical computing. R
322 Foundation for Statistical Computing, Vienna, 2011. Disponível em: <[http://www.R-](http://www.R-project.org/)
323 [project.org/](http://www.R-project.org/)>.

324 RIBEIRO JUNIOR, P.J. & DIGGLE, P.J. GeoR: A package for geostatistical analysis. R. News,
325 1:14-18, 2001.

326 RIBEIRO JUNIOR, P.J.; VIOLA, D.N.; DEMÉTRIO, C.G.B.; MANLY, B.F. & FERNANDES,
327 O.A. Spatial pattern detection modeling of thrips (*Thrips tabaci*) on onion fields. Sci. Agric.,
328 66:90-99, 2009.

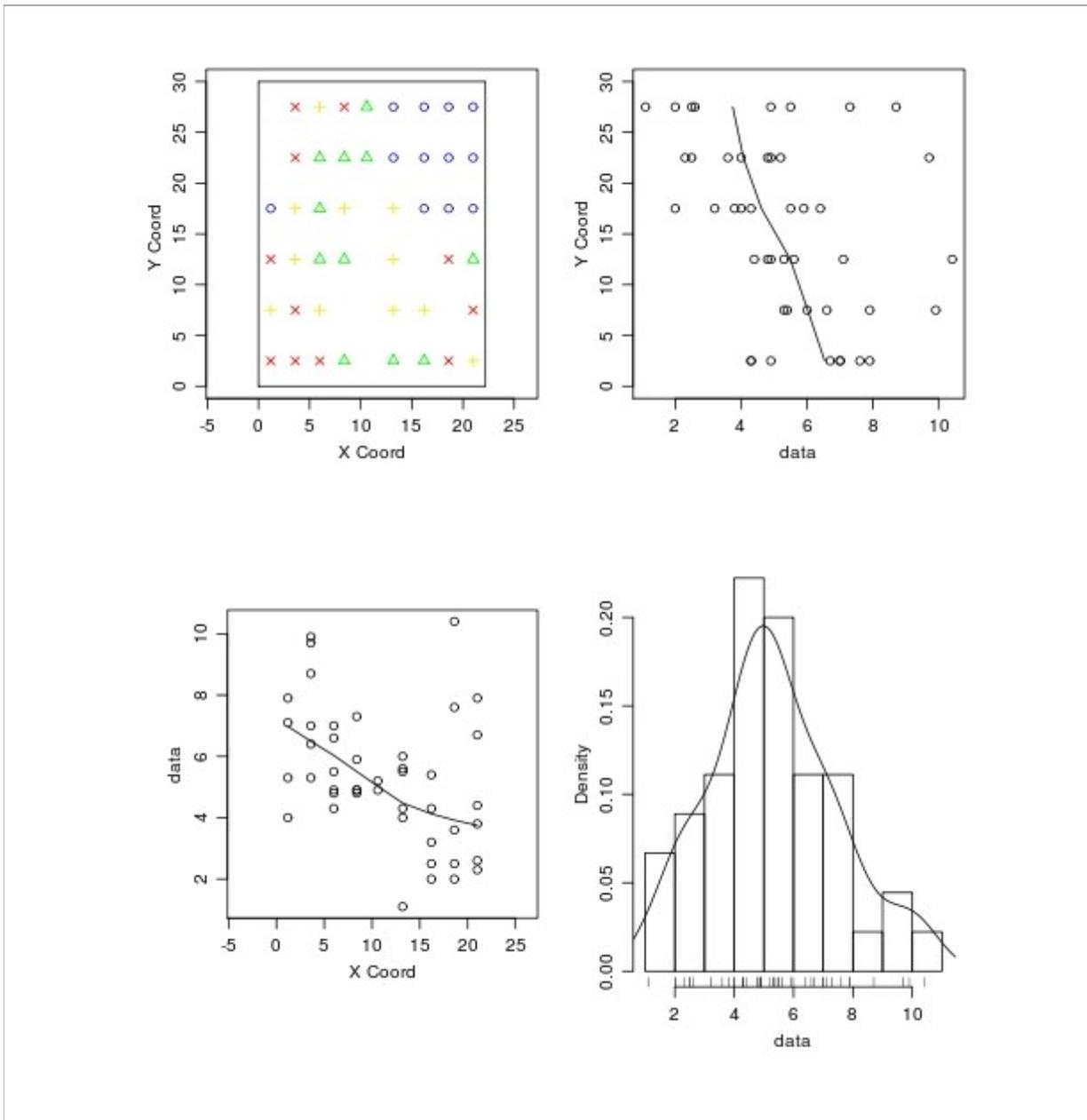
329 ROA-URETA, R. & NIKLITSCHK, E. Biomass estimation from surveys with likelihood-based
330 geostatistics. J. Marine Sci., 64:1723-1734, 2007.

331 SOUZA, Z.M.; MARQUES JÚNIOR, J.; PEREIRA, G.T. & BARBIERI, D.M. Small relief shape
332 variations influence spatial variability of soil chemical attributes. Sci. Agric., 63:161-168,
333 2006.

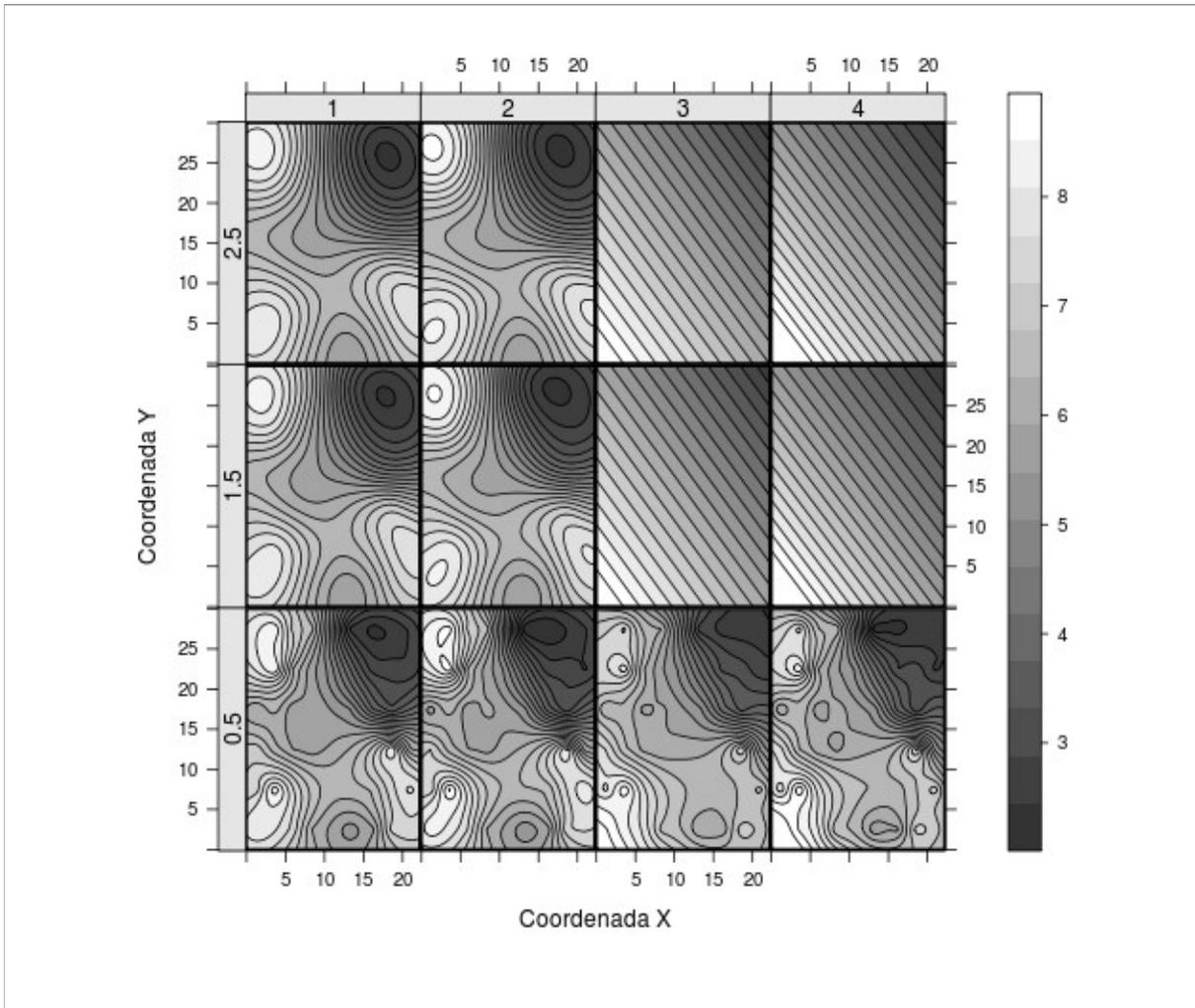
334 SOUZA, G.S.; LIMA J.S.S.; SILVA, S.A. & OLIVEIRA, R.B. Variabilidade espacial de atributos
335 químicos em um Argissolo sob pastagem. Acta Sci. Agron., 30:589-596, 2008.

336 TEDESCO, M.J.; GIANELLO, C. & BISSANI, C.A. Análise de solo, plantas e outros materiais.
337 2.ed. Porto Alegre, Universidade Federal do Rio Grande do Sul, 1995. 174p. (Boletim
338 Técnico de Solos, 5)

339



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



345 **Figura 2. Mapas de teores de fósforo estimados por krigagem em uma malha regular de**
 346 **10.000 pontos com base em 45 pontos amostrais para diferentes valores do parâmetro de**
 347 **suavidade (κ 0,5, 1,5 e 2,5). 1: modelo ajustado aos dados originais; 2: modelo**
 348 **ajustado aos dados transformados; 3: modelo ajustado aos dados originais com**
 349 **covariáveis (coordenadas); e 4: modelo ajustado aos dados transformados com**
 350 **covariáveis (coordenadas).**

351

352 **Quadro 1. Parâmetros do modelo geostatístico Gaussiano estimados por máxima**
 353 **verossimilhança para o teor de fósforo no solo**

354

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

355 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
 356 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,
 357 respectivamente; L-V: logaritmo da função de verossimilhança; 1: modelo ajustado aos dados originais; 2: modelo
 358 ajustado aos dados transformados; 3: modelo ajustado aos dados originais com covariáveis (coordenadas); e 4: modelo
 359 ajustado aos dados transformados com covariáveis (coordenadas).

360

361

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

365

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

366 ⁽¹⁾ 1: modelo ajustado aos dados originais; 2: modelo ajustado aos dados transformados;
 367 3: modelo ajustado aos dados originais com covariáveis (coordenadas); e 4: modelo
 368 ajustado aos dados transformados com (coordenadas).

369

370