

# MODELANDO SATURAÇÃO POR BASES E PH DO SOLO: UMA APLICAÇÃO DE MODELOS GEOESTATÍSTICOS BIVARIADOS<sup>(1)</sup>

Bruno Henrique Fernandes Fonseca<sup>(2)</sup>, Paulo Justiniano Ribeiro Jr.<sup>(3)</sup>, Bruno Gherardi<sup>(4)</sup> e Luiz Eduardo Oliveira de Faria<sup>(5)</sup>

**Resumo** – Segundo SCHUELLER (1992) a agricultura de precisão é um método de administração cuidadosa e detalhada do solo e da cultura para adequar as diferentes condições encontradas em cada pedaço de lavoura, tendo em vista a desuniformidade intrínseca dos solos. Trabalhos de campo têm mostrado a importância do estudo das variações das condições do solo como aspecto fundamental para se implementar uma agricultura mais eficiente e rentável. Os modelos geoestatísticos bivariados possibilitam explorar a correlação estatística entre duas variáveis que possuem padrão espacial de variabilidade, fato que pode diminuir os custos para analisar atributos do solo. Esses modelos definem funções aleatórias para dois processos estocásticos com localizações espaciais conhecidas. Pode-se adotar a suposição da existência de um campo aleatório gaussiano latente para as duas variáveis aleatória conjuntamente. A suposição de gaussianidade do processo latente é conveniente para inferências sobre parâmetros do modelo e para obtenção de previsões espaciais. A matriz de covariância dessa distribuição deve estruturar a variabilidade espacial entre e dentro os atributos. Gelfand et al. (2004) e Diggle & Ribeiro Jr. (2007) propuseram estratégias para estruturar essa matriz, porém não existem muitos relatos sobre o uso e avaliações comparativas entre essas abordagens. Neste estudo foram utilizados dados provenientes da análise de solo de uma propriedade agrícola com 51,8ha de área, onde foram amostradas 67 localizações georeferenciadas. Foram utilizados os valores mensurados do pH e da saturação por bases do solo, que foram submetidas à análise descritiva espacial, modelagens geoestatísticas univariadas, bivariadas e previsões espaciais. Além disso, foram utilizadas medidas de erro de previsão para verificar se existem vantagens em adotar modelos bivariados ao invés de modelos univariados, o que pode gerar vantagens financeiras para análises dos atributos de solo sob estudo.

*Termos de Indexação: agricultura de precisão, estimadores de máxima verossimilhança, modelos geoestatísticos bivariados*

**Abstract** - According SCHUELLER (1992) precision agriculture is a method of careful and detailed management of the soil and crop to suit the different conditions found in each piece of

---

<sup>(1)</sup>Parte da Tese de Mestrado em Agronomia com Área de Concentração em Estatística e Experimentação Agronômica da Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo - ESALQ/USP. Pesquisa financiada pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, <sup>(2)</sup>Mestrando em Agronomia pela ESALQ/USP. Av. Pádua Dias 11, CEP 13418-900 Piracicaba (SP). E-mail: [bh.fonseca@hotmail.com](mailto:bh.fonseca@hotmail.com), <sup>(3)</sup>Professor Adjunto da Universidade Federal do Paraná - UFPR. E-mail: [paulojus@ufpr.br](mailto:paulojus@ufpr.br), <sup>(4)</sup>Mestrando em Agronomia pela ESALQ/USP. E-mail: [bruno.gherardi@agrofficio.com.br](mailto:bruno.gherardi@agrofficio.com.br), <sup>(5)</sup>Bacharel em Agronomia pela ESALQ/USP. E-mail: [luiz.faria@agrofficio.com.br](mailto:luiz.faria@agrofficio.com.br)

farming, in view of the inherent unevenness of the soil. Field work has shown the importance of studying the variations of soil conditions as a fundamental aspect to make agriculture more efficient and profitable. Bivariate geostatistical models enable to explore the statistical correlation between two variables that have spatial pattern of variability, which may reduce the price to analyses attributes soil. These models define random functions for two stochastic processes with known spatial locations. Existence of a Gaussian random field latent can be assumed for two random variables jointly. The covariance matrix of this distribution should be positive-definite and to have the spatial variability structure between and among the attributes. Gelfand et al. (2004) and Diggle & Ribeiro Jr. (2007) suggested strategies for structuring this matrix, however there are few reports on comparing approaches. This work examined sample soil data of pH and base saturation of 67 spatial locations of a farm with 51.8 hectares. The sample data were submitted to spatial descriptive analysis, univariate and bivariate modeling and spatial prediction. In addition, error measures were used to verify the advantages in adopting bivariate model rather than univariate models, which can generate economic benefits for analysis of soil attributes under study.

*Index Terms: precision agriculture, maximum likelihood estimators, bivariate geostatistical models*

## INTRODUÇÃO

Trabalhos de agricultura de precisão mostram que mesmo em áreas consideradas homogêneas a variabilidade do solo não é puramente aleatória, apresentando correlação ou dependência espacial (OLIVEIRA et al., 1999; CARVALHO et al., 2003; SILVA et al., 2003). Uma possível abordagem para estudar esse tipo de problema de variabilidade, é a modelagem geoestatística de atributos de solo e em particular, em conjunto com métodos de análise baseados em verossimilhança.

A modelagem geoestatística é um conjunto de técnicas que tenta encontrar uma boa função aleatória para um ou mais atributos que possuem localizações espaciais conhecidas. Sendo assim, essas ferramentas são úteis para capturar a correlação entre as observações das variáveis aleatórias de interesse, onde existe uma forte suspeita de que pontos espaciais mais próximos possuem valores dos atributos mais parecidos.

Um campo aleatório é um processo estocástico que existe em algum espaço real  $d$ -dimensional, geralmente bi ou tri-dimensional sua definição é dada por  $\{Z(s_i) : s_i \in G \subset R^d\}$ , sendo que  $Z(s_i)$  denota a variável aleatória  $Z$  na localização  $s_i$  do espaço sob estudo  $G$ .

Segundo Schmidt & Sansó (2006) e Le & Zidek (2006), 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  $G$  e qualquer inteiro  $n$ :

63           Devido a simplicidade inferencial, a distribuição de probabilidade gaussiana é uma das mais utilizadas na literatura. Sendo assim, um campo aleatório é dito ser gaussiano se  $Z$  segue uma distribuição normal em todas as localizações do espaço sob estudo  $G$ . Para qualquer conjunto finito de localizações  $s = (s_1, s_2, \dots, s_n)$  pertencente a  $G$ ,  $Z(s)$  segue uma distribuição normal  $n$ -variada e é completamente especificado pelo vetor de média  $n \times 1$ , notado por  $\mu$ , e pela matriz de covariâncias  $n \times n$  notada por  $\Sigma$ , que no contexto de geoestatística, em geral possui o comportamento de que quanto maior a distância euclidiana entre duas localizações  $s_l$  e  $s_k$  quaisquer, menor a correlação entre  $Z(s_l)$  e  $Z(s_k)$ , a especificação de um modelo com tal comportamento deve assegurar essa matriz seja positiva definida. Além disso, em geral não é possível ter mais de uma realização do processo, sendo assim, outras suposições devem ser impostas sobre o campo aleatório gaussiano para a realização de inferências. A restrição mais utilizada é a de que o processo estocástico é homogêneo. Um campo aleatório gaussiano é dito ser homogêneo se ele for estacionário e seu padrão de correlações não depende das direções, ou seja, isotrópico. Utilizando essa suposição, o processo estocástico, embora restritivo, consegue modelar diversos problemas naturais.

70           Encontrar uma função de correlação para estruturar a matriz  $\Sigma$  válida no contexto de geoestatística pode ser um procedimento complexo. Na prática, o usual é adotar famílias de funções ditas permissíveis. Proposta por Matérn (1986), a família de funções de correlação de Matérn é muito utilizada devido a capacidade de produzir comportamentos distintos quanto a suavidade do processo. Além disso, sob estacionariedade fraca, essas funções possuem propriedade conhecidas e desejáveis. Schabenberger & Gotway (2005) apresentam e discutem detalhes sobre tais propriedades. Silva (2006), em um trabalho discutindo a construção de funções de correlação válidas para modelos geoestatísticos espaço temporais apresenta diversas outras funções de correlação conhecidamente válidas e bem como uma extensa lista de referências sobre o tema.

84           Considerando que em alguma área  $G$  exista um campo aleatório gaussiano  $Z$  latente, ou seja, o processo existe, mas não é observável, é necessário fazer uma amostragem de  $n$  localizações espaciais dentro da área  $G$  e observar valores de atributo de interesse nas localizações amostradas. 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:

$$Y(s) = \mu + Z(s) + \epsilon \quad (1)$$

91           sendo  $\mu = X\beta$ , onde  $X$  é 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  $n \times n$ ; e  $\epsilon$  um vetor  $n \times 1$  de ruídos brancos, que por suposição são independentes e identicamente distribuídos com distribuição de probabilidade normal com média zero e desvio padrão  $\tau$ . Faraco et al. (2008) mostra a aplicabilidade desse modelo.

98 Utilizando (1) é possível encontrar a distribuição de probabilidade de  $Y(s)$ , que é gaussiana  $n$ -variada, com vetor de médias  $X\beta$  e matriz de covariâncias  $\Sigma_Y = \Sigma + \tau^2 I$ , sendo  $I$  uma matriz identidade  $n \times n$ . Logo, existe um vetor de parâmetros  $\theta = (\beta, \sigma^2, \phi^*, \tau^2)$  a ser estimado, onde  $\sigma^2$  é a variância de  $Z(s)$  e  $\phi^*$  é um vetor de parâmetros associados à função de correlação adotada.

Especificado um modelo adequado para os dados em questão, o próximo passo é fazer a estimação dos parâmetros. Assumindo que o campo aleatório possui estacionariedade forte, pode-se optar por estimadores baseados na função de verossimilhança ou verossimilhança restrita. Esta  
105 abordagem consiste em utilizar os valores  $Y(s)$  para encontrar um vetor  $\hat{\theta}$  que seja o ponto de máximo da função de verossimilhança associada a  $\theta$ . Por simplicidade matemática, normalmente utiliza-se o logaritmo da função de verossimilhança para fazer a estimação, que associada a (1) tem à seguinte forma:

$$l(\theta; Y(s)) = -0,5 (n \ln(2\pi) + \ln(|\Sigma_Y|) + (Y(s) - X\beta)^t \Sigma_Y^{-1} (Y(s) - X\beta)). \quad (2)$$

No contexto de geoestatística, Diggle & Ribeiro Jr. (2007) adotam a utilização da reparametrização  $v = \tau/\sigma$ , a qual facilita a estimação de  $\theta$ . O vetor de parâmetros passa a ser  
112  $\theta^* = (\beta, \sigma^2, \phi^*, v^2)$  e  $\Sigma_Y$  pode ser escrita como  $\sigma^2 V$ , onde  $V$  é uma matriz  $n \times n$  que depende apenas de  $v$  e  $\phi^*$ . Existem formas analíticas apenas para os estimadores de  $\beta$  e  $\sigma^2$ , que são funções dos demais parâmetros. Para  $\phi^*$  e  $v^2$  não existem estimadores fechados. Utilizando  $\hat{\beta}$  e  $\hat{\sigma}^2$  em (2), tem-se o logaritmo da função de verossimilhança concentrada, que depende apenas de  $\theta_c = (\phi^*, v^2)$ .

Para encontrar  $\hat{\theta}_c$  é possível utilizar métodos numéricos de maximização de funções, como por exemplo, o método de Nelder & Mead (1965), o qual calcula numericamente  $\hat{\phi}^*$  e  $\hat{v}^2$ . Com os  
119 parâmetros da função de máxima verossimilhança concentrada estimados, por invariância é possível encontrar as estimativas de  $\beta$ ,  $\sigma^2$  e  $\tau^2$ .

No entanto, geralmente o objetivo final dos estudos com dados georeferenciados é calcular as predições espaciais em localizações não amostradas. Esse processo de predição é denominado de krigagem e é baseado nas propriedades da distribuição normal multivariada (DIGGLE & RIBEIRO Jr., 2007).

Quando existem dois campos aleatórios gaussianos de interesse, a intuição inicial é que seja  
 126 ajustado um modelo geoestatístico univariado para cada vetor de valores observados. Porém, em  
 alguns casos existe correlação estatística entre as variáveis aleatórias, o que leva a possibilidade de  
 adotar modelos geoestatísticos bivariados.

Supondo a existência de dois campos aleatórios gaussianos de interesse, pode-se utilizar  
 seguinte o modelo:

$$Y_i = \mu_i + Z_i, \quad i = 1,2 \quad (3)$$

sendo  $Y_i$  um vetor  $n_i \times 1$  de valores observados do campo aleatório gaussiano latente  $Z_i$ , que possui  
 133 vetor de médias nulo  $n_i \times 1$  e matriz de covariâncias  $\Sigma_i$ ,  $\mu_i$  é um vetor, que possui os parâmetros de  
 médias associados a  $Y_i$ . Observe que em (3) não existem termos de ruídos brancos.

A distribuição de  $Y = (Y_1, Y_2)$  é de interesse final e possui distribuição gaussiana  $n$ -variada,  
 sendo  $n = n_1 + n_2$ , com vetor de médias  $\mu = (\mu_1, \mu_2)$  e matriz de covariâncias  $\Sigma_Y$ , positiva  
 definida que possui o comportamento empírico de correlações utilizado em geoestatística e que  
 pode ser particionada da seguinte forma:

$$\Sigma_Y = \begin{bmatrix} \Sigma_1 & \Sigma_{1,2} \\ \Sigma_{1,2}^t & \Sigma_2 \end{bmatrix}$$

140 sendo  $\Sigma_i$  uma matriz  $n_i \times n_i$  das covariâncias da variável  $Y_i$ ,  $i=1,2$ , e  $\Sigma_{1,2}$  uma matriz  $n_1 \times n_2$  com  
 as covariâncias cruzadas entre  $Y_1$  e  $Y_2$ .

Nesse contexto a especificação do modelo exige uma matriz  $\Sigma_Y$  válida. A abordagem por  
 modelos separáveis é a mais simples para resolver esse problema e utiliza decomposições dos  
 termos latentes de (3).

Diggle & Ribeiro Jr. (2007) propuseram uma abordagem para problemas geoestatísticos  
 bivariados, a qual é abreviada nesse trabalho por *BGCCM* (*bivariate Gaussian common component*  
 147 *model*) e que utiliza as seguintes decomposições dos campos aleatórios latentes de (3):

$$Z_i = \sigma_{0i} S_0 + \sigma_i S_i, \quad i = 1,2$$

sendo  $\sigma^* = (\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2)$  um vetor de parâmetros de dispersão associados a (3) e  $S_0$ ,  $S_1$  e  $S_2$   
 campos aleatórios gaussianos mutuamente independentes, com vetores de médias nulos, variâncias  
 unitárias e correlações determinadas por funções de correlação válidas. As correlações destes  
 termos, quando combinadas desta forma, vão gerar correlações cruzadas válidas entre as variáveis  
 observáveis  $Y_1$  e  $Y_2$ , uma vez que, o campo aleatório  $S_0$  é comum as duas respostas.

154 Definindo  $Y_i(s_l)$  e  $Y_i(s_k)$  como observações de  $Y_i$  feitas em duas localizações quaisquer  $s_l$  e  
 $s_k$ , que estão separados por uma distância euclidiana  $h = h_{l,k}$ ,  $l, k = 1,2, \dots, n_i$  e  $i = 1,2$ , tem-se

que o elemento  $\Sigma_{i,(l,k)}$  é dado por  $Cov(h) = \sigma_{0i}^2 \rho_0(h) + \sigma_i^2 \rho_i(h)$ , sendo  $\rho_0$  a função de correlação adotada para  $S_0$  e  $\rho_i$  as funções de correlação adotadas para  $S_i$ ,  $i = 1,2$ . Utilizando propriedades básicas de covariâncias, pode-se encontrar  $\Sigma_{1,2}$ , que é igual a  $\sigma_{01} \sigma_{02} R_0$ , sendo  $R_0$  uma matriz  $n_1 \times n_2$  das correlações cruzadas entre as respostas e depende da função de correlação adotada para  $S_0$ . Dessa forma  $\Sigma_Y$  fica completamente estruturada e a função de verossimilhança associada a (3) depende de  $\theta = (\beta^*, \sigma^*, \phi_0^*, \phi_1^*, \phi_2^*)$ , onde  $\beta^*$  é um vetor de parâmetros associado a  $\mu = (\mu_1, \mu_2)$  e  $\phi_i^*$  os parâmetros associados à escolha da função de correlação para  $S_i$ ,  $i = 0,1,2$ .

Gelfand et al. (2004) propôs outra abordagem para encontrar uma estrutura paramétrica válida para  $\Sigma_Y$ , que é abreviada aqui por *BCRM* (*bivariate coregionalisation model*). Nesta abordagem os termos latentes de (3) são decompostos da seguinte forma:

$$\begin{cases} Z_1 = \sigma_{11} S_1 \\ Z_2 = \sigma_{12} S_1 + \sigma_{22} S_2 \end{cases}$$

sendo  $S_1$  e  $S_2$  campos aleatórios gaussianos mutuamente independentes, com vetores de médias nulo, variância unitária e com correlações determinadas pela escolha de funções de correlação conhecidamente válidas. Como o campo aleatório  $S_1$  é comum as duas variáveis, é induzida a correlação empírica entre as respostas.

Utilizando o *BCRM*, o desenvolvimento probabilístico para encontrar a estrutura de  $\Sigma_Y$  é similar ao exposto para o *BGCCM*. O vetor de parâmetros associados a distribuição de probabilidade de  $Y$  é  $\theta = (\beta^*, \sigma_{11}, \sigma_{12}, \sigma_{22}, \phi_1^*, \phi_2^*)$ , onde  $\beta^*$  é um vetor de parâmetros associado a  $\mu = (\mu_1, \mu_2)$  e  $\phi_i^*$  os parâmetros associados à escolha da função de correlação para  $S_i$ ,  $i = 1,2$ .

Com  $\Sigma_Y$  especificada sob o *BGCCM* ou o *BCRM*, os próximos passos são estimar o vetor de parâmetros  $\theta$  e realizar a krigagem. As técnicas são similares às adotadas no modelo univariado, no entanto, nesse caso é necessário um número maior de reparametrizações para fazer as estimações.

Esse trabalho utiliza metodologias geoestatísticas para estudar a variabilidade espacial de duas variáveis químicas do solo de uma propriedade agrícola. A utilização de modelos geoestatísticos bivariados é possível devido a natureza dos atributos, que são fortemente correlacionados e, além disso, uma das respostas é considerada ser mais dispendiosa para ser observada. Os resultados fornecem mapas de interesse prático descrevendo o comportamento dos atributos na região, com mensurações de incertezas associadas. Além disto, com a estrutura conjunta de correlação espacial estabelecida, em monitoramentos futuros da mesma propriedade agrícola, será possível avaliar possíveis diminuições nos gastos com a coleta de informações.

## MATERIAL E MÉTODOS

Os dados analisados com as técnicas apresentadas foram obtidos por meio de uma pesquisa realizada em julho 2006 na fazenda Tupã, localizada no município de Echaporã/SP, que possui 51,8ha de área, solo argissolo de textura média e dois históricos de manejo distintos, soja numa região e pastagem na outra. Para analisar as variáveis químicas do solo, foram amostradas 67 localizações, com grid regular a cada hectare, georeferenciadas no sistema Universal Transverse Mercator (UTM).

Dois parâmetros químicos do solo levantados pelos pesquisadores foram modelados nesse trabalho: saturação por bases e pH do solo, sendo que, existe uma forte correlação entre essas duas respostas, o que justifica a tentativa de modelagem bivariada.

Do ponto de vista geoestatístico, foi assumido que o pH e a saturação são dois campos aleatórios gaussianos e latentes. A intuição inicial é que seja ajustado individualmente o modelo (1) para cada vetor de valores observados, que foram notados por  $Y_1$  e  $Y_2$  para a saturação por base e o pH, respectivamente. Então  $Y_i$  segue uma distribuição gaussiana 67-variada, com matriz de covariâncias  $\Sigma_i$  e vetor de médias  $\mu_i = X_i\beta_i$ , onde  $X_i$  é uma matriz  $67 \times p$  com o intercepto e  $p - 1$  possíveis covariáveis e  $\beta_i$  um vetor  $p \times 1$  com os parâmetros associados a  $\mu_i$ .

Foram consideradas três formas diferentes para  $X_i$ , média constante, média com tendência induzida pela área de manejo e média com tendência induzida pela coordenada oeste-leste das localizações amostradas. Foi utilizada a família de funções de correlação de Matérn para estruturar a matriz  $67 \times 67$  de correlações, considerando o parâmetro de suavidade  $\kappa$  fixo nos valores 0,5, 1, 1.5, 2, 2.5, logo,  $\phi_i^* = \phi_i$ . Combinando as diferentes escolhas de  $\kappa$  e  $X_i$  foram ajustados diversos modelos. As estimativas paramétricas foram calculadas por máxima verossimilhança, onde foi utilizado a reparametrização  $v_i = \tau_i/\sigma_i$  e o método numérico de Nelder & Mead (1965) para encontrar as estimativas  $\hat{\phi}_i$  e  $\hat{v}_i$  que maximizam a função de máxima verossimilhança concentrada a partir das quais são encontradas as demais estimativas de parâmetros de interesse.

Os valores dos máximos estimados da função de verossimilhança concentrada e o Critério de Informação de Akaike (AIC) foram utilizados para selecionar o modelo univariado final para cada atributo. Com os parâmetros estimados desses modelos foram conduzidas as krigagens ordinárias para as variáveis químicas do solo. Por último, uma análise de resíduos foi realizada para verificação de pressupostos dos ruídos brancos, os quais se comportaram conforme o esperado.

Após estudo individual de cada variável química, foram ajustados modelos bivariados para o vetor  $Y = (Y_1, Y_2)$ , utilizando *BGCCM* e o *BCRM*. Nas duas abordagens foi utilizado  $\mu = X\beta^*$ , onde

$X$  é uma matriz  $134 \times p$  contendo um intercepto para cada resposta e  $p - 2$  possíveis covariáveis e  $\beta^*$  é um vetor  $p \times 1$  com os parâmetros de média associados a  $Y$ . Para estruturar  $\Sigma_Y$  somente funções de correlação da família de Matèrn foram utilizadas, sendo que o parâmetro  $\kappa$  foi sempre considerado fixo em valores similares aos modelos univariados.

224 Diversos modelos foram ajustados para combinações de diferentes escolhas de  $\kappa$  e  $X$ . Em todos os casos o vetor de parâmetros  $\theta$  foi estimado por máxima verossimilhança. No *BGCCM* utilizou-se as reparametrizações  $\sigma = \sigma_{01}$ ,  $\eta = \sigma_{02}/\sigma_{01}$ ,  $\nu_1 = \sigma_1/\sigma_{01}$  e  $\nu_2 = \sigma_2/\sigma_{01}$  e no *BCRM* as reparametrizações  $\sigma = \sigma_{11}$ ,  $\nu_1 = \sigma_{12}/\sigma_{11}$  e  $\nu_2 = \sigma_{22}/\sigma_{11}$ . Nas duas abordagens existe forma analítica para os estimadores de  $\beta^*$  e  $\sigma^2$ , que são função dos demais parâmetros, os quais foram estimados pelo método de Nelder & Mead (1965). Utilizando a propriedade de invariância dos estimadores foi possível encontrar estimativas para todos os parâmetros dos modelos. Quanto a seleção de modelos e a krigagem, as mesmas técnicas dos modelos univariados foram utilizadas.

231 Para exemplificar o possível ganho com a adoção de modelos bivariados, a amostra da saturação por bases foi dividida em uma subamostra de modelagem e uma subamostra de controle, que contêm 47 e 20 observações, respectivamente. O grupo de modelagem foi utilizado para modelar novamente a saturação por bases sob as abordagens uni e bivariadas. Cabe ressaltar que nos modelos bivariados foi utilizada a informação completa do pH. Com os parâmetros estimados foram calculadas as previsões da saturação por bases nas localizações do grupo de controle. Para comparar as metodologias foram calculados os erros de krigagem produzidos em cada caso.

238 Todas as análises e resultados foram obtidos através do ambiente R de programação (*R* Development Core Team, 2006), utilizando o pacote *geoR* (Ribeiro Jr. & Diggle, 2001). Assim como são disponibilizados em [www.leg.ufpr.br/papercompanions](http://www.leg.ufpr.br/papercompanions) os códigos, resultados e dados utilizados aqui, que corresponde ao complemento online deste artigo.

## RESULTADOS E DISCUSSÃO

245 Inicialmente foi conduzida uma análise gráfica evidencia dos valores observados do pH nas 67 localizações amostradas da propriedade agrícola. A Figura 1 evidencia que existe um possível padrão espacial na variável, onde há uma suavidade do processo ao longo da fazenda. A linha no meio do gráfico separa as áreas com históricos de manejo distintos: cultivo de soja e pastagem, para as regiões a direita e a esquerda, respectivamente. A primeira área possui valores mais elevados, o que leva a possibilidade de ser considerada essa informação como covariável no processo de modelagem.

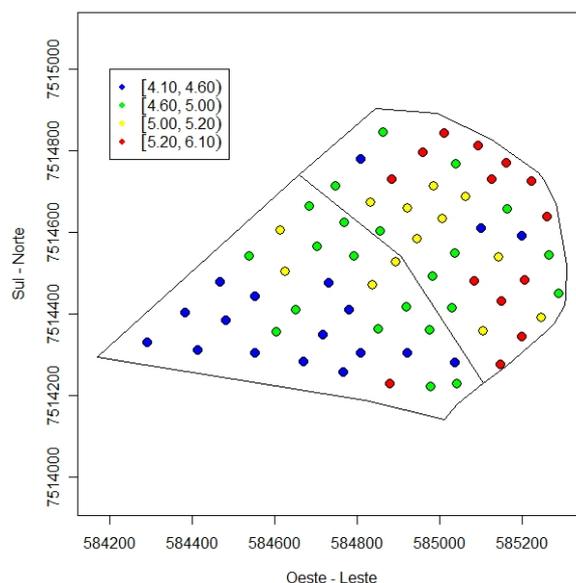


Figura 1: Gráfico das localizações espaciais amostradas e dos quartis da variável pH

252

Para a saturação por bases as características são muito similares as do pH, fato corroborado pela correlação marginal de 0,92 entre essas variáveis. As médias dos valores observados são 4,94 e 53,27 para o pH e a saturação, respectivamente. De uma maneira geral os atributos de solo não mostram nítidos desvios de Gaussianidade havendo uma simetria dos dados com relação a média amostral e não existem valores observados muito discrepantes.

259

Considerando as diferentes abordagens para a matriz  $X$  e as diferentes escolhas para  $\kappa$ , inicialmente foram ajustados diversos modelos geoestatísticos univariados para o pH. Foram utilizados os valores estimados dos máximos do logaritmo da função de verossimilhança para selecionar o melhor modelo dentro de cada metodologia para a tendência na média. A Tabela 1 mostra os resultados do melhor modelo univariado ajustado para o pH em cada uma das formas da matriz  $X$ . Os parâmetros de média e de variabilidade não são muito afetados com as diferentes modelagens, já o parâmetro  $\phi$  apresentou resultados com maior diferença entre as abordagens, fato explicado, em parte, pela não ortogonalidade entre os parâmetros da função de correlação de Matérn. Para selecionar o modelo final foram utilizados os valores estimados do AIC, o modelo que considera tendência induzida pela coordenada oeste-leste foi o escolhido, pois foi o menos penalizado. O modelo final gera um alcance prático estimado de 132,93, ou seja, a partir dessa

266

Tendência	$\beta_0$	$\beta_1$	$\sigma^2$	$\tau^2$	$\phi$	$\kappa$	$l(\theta)$	AIC
Constante	4,903	-	0,106	0,297	491,1	2,5	-25,36	58,73
Área de manejo	4,723	0,403	0,014	0,111	28,43	1,5	-23,33	56,66
Oeste-Leste	-608,5	0,001	0,015	0,097	22,46	2,5	-19,35	48,70

Tabela 1: Estimativas de máxima verossimilhança dos parâmetros da modelagem do pH

Além de selecionar o modelo final dentre os que consideram o padrão espacial, foi conduzida uma regressão linear simples considerando a coordenada oeste-leste como covariável, ou seja, considerando que o padrão dos dados é completamente aleatório. Sobre esse enfoque o AIC ficou igual a 49,8, o que afirma a idéia de que o pH possui variabilidade espacial significativa.

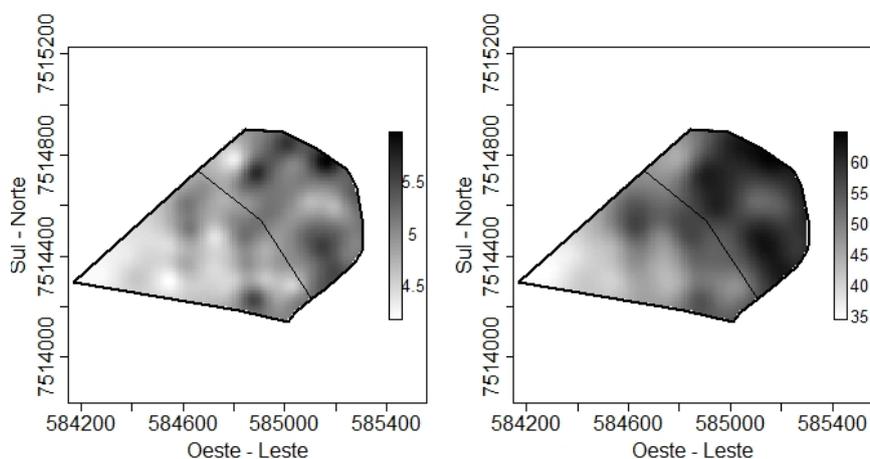
273

Para modelar a saturação por bases foi conduzida a mesma metodologia aplicada ao pH. O modelo final também inclui a coordenada oeste-leste como covariável e possui  $\kappa = 2,5$ . As estimativas foram  $\beta_0 = -14298,98$ ,  $\beta_1 = 0,025$ ,  $\tau^2 = 38,06$ ,  $\sigma^2 = 31,03$  e  $\phi = 32,19$ . Para essa variável o alcance prático estimado ficou em 190,52. Além disso, o valor do AIC com o modelo linear simples para a saturação foi maior do que o modelo geoestatístico.

Com os modelos univariados estabelecidos, foram calculadas as predições espaciais para as duas variáveis químicas do solo em uma malha de localizações espaciais, sendo que, as estimativas paramétricas foram substituídas nas fórmulas de krigagem e variância preditiva. A Figura 2 mostra os mapas preditivos dos campos aleatórios. As imagens ficaram condizentes com os dados observados, ou seja, localizações com menores valores para a coordenada oeste-leste possuem, em sua maioria, valores preditos menores. Além disso, as variâncias preditivas da saturação por bases estão todas no intervalo [12,35; 30,21], já para o pH os valores máximo e mínimo das variâncias de krigagens ficaram iguais a 0,10 e 0,01, respectivamente. Esses valores das variâncias preditivas são relativamente pequenos com relação a grandeza de escala das variáveis, sendo assim, a precisão das krigagens ficaram em um patamar aceitável.

287

Para validar os pressupostos feitos sob os ruídos brancos dos modelos univariados, foi utilizado o teste de normalidade de Shapiro-Wilk, que gerou p-valores de 0,22 e 0,97 para os erros do pH e da saturação por bases, respectivamente. Logo, esse pressuposto está garantido.



294 *Figura 2: Gráficos de predições espaciais, o gráfico da esquerda é relativo ao pH e o gráfico da direita é relativo à saturação por bases*

Concluída a modelagem univariada, foram ajustados modelos bivariados para as respostas. A correlação de Pearson entre os atributos é 0,92, o que indica uma correlação positiva e quase perfeita entre as duas variáveis químicas em questão. Para ajustar o *BGCCM* e o *BCRM*, foram utilizadas as estimativas dos modelos univariados para atribuir valores iniciais ao método numérico de maximização. Três abordagens para modelar as médias dos campos aleatórios foram consideradas: médias constantes, médias com tendências induzidas pelas coordenadas oeste-leste e médias com tendências induzidas pela área de manejo do solo. Tais possibilidades combinadas com diversas escolhas para os parâmetros de suavidade das funções de correlação geraram diversas estimativas diferentes para os parâmetros dos modelos. Inicialmente, foram selecionadas as melhores estimativas em cada abordagem para a média. A Tabela 2 mostra os resultados do *BGCCM*, onde não há muita diferença entre os parâmetros de variabilidade e correlação estimados para cada abordagem de tendência, exceto para o parâmetro  $\phi_0$ , que para a tendência na área de manejo ficou bem menor do que as demais. Esse fato pode ser explicado pela adoção de um  $\kappa_0$  maior, logo, combinando essas duas informações, tem-se que os alcances de correlações são bem parecidos para as três abordagens de média.

$\theta$	Constante	Oeste-leste	Área de manejo
$\beta_{01}$	52,57	1,54e-6	47,83
$\beta_{02}$	4,92	1,45e-7	4,72
$\beta_{11}$	-	8,98e-5	9,46
$\beta_{12}$	-	8,41e-6	0,40
$\sigma_{01}$	8,51	8,35	6,85
$\sigma_1$	3,47	3,46	3,44
$\sigma_{02}$	0,38	0,38	0,31
$\sigma_2$	7.00e-05	5.00e-04	1.00e-05
$\phi_0$	59,37	57,79	26,73
$\phi_1$	46,35	48,15	49,08
$\phi_2$	66,82	71,59	86,95
$\kappa_0$	1	1	1,5
$\kappa_1$	0,5	0,5	0,5
$\kappa_2$	0,5	0,5	0,5

Tabela 2: Estimativas de máxima verossimilhança dos parâmetros associados ao *BGCCM* para a saturação por bases e o pH

A Tabela 3 mostra os resultados da estimação por máxima verossimilhança utilizando o *BCRM*. Com esse modelo, a inclusão da covariável coordenada leste-oeste gerou inflacionamento dos parâmetros de variabilidade. Com relação às demais estimativas, os resultados são parecidos com os do *BGCCM*.

As duas abordagens de modelos bivariados se mostraram computacionalmente estáveis para fazer estimação dos parâmetros, uma vez que, as estimativas dos parâmetros de variabilidade e de correlação apresentaram resultados que geram características similares às encontradas nos modelos univariados. O melhor ajuste foi o que considera a área de manejo da propriedade como covariável, os quais geraram AIC iguais a 437,35 e 432,96 para o *BGCCM* e *BCRM*, respectivamente. O *BCRM* foi menos penalizado pelo AIC, esse modelo se comportou melhor para estimação paramétrica.

$\theta$	Constante	Oeste-leste	Área de manejo
$\beta_{01}$	51,76	1,51e-6	47,93
$\beta_{02}$	4,89	1,43e-7	4,72
$\beta_{11}$	-	8,85e-5	9,12
$\beta_{12}$	-	8,36e-6	0,40
$\sigma_{11}$	8,26	99,53	7,82
$\sigma_{12}$	0,31	3,67	0,28
$\sigma_{22}$	0,13	1,54	0,14
$\phi_1$	100,63	100,21	53,32
$\phi_2$	21,43	21,38	19,34
$\kappa_1$	0,5	0,5	0,5
$\kappa_2$	2,5	2,5	2,5

322 Tabela 3: Estimativas de máxima verossimilhança dos parâmetros associados ao *BCRM* para a saturação por bases e o pH

O processo de krigagem utilizando as estimativas do *BGCCM* gerou variâncias preditivas relativamente pequenas, para a saturação por bases os valores mínimos e máximos são 0,47 e 58,40, respectivamente. Já para o pH esses valores ficaram iguais a 0,0001 e 0,096. Utilizando o *BCRM* as variâncias preditivas da saturação por bases e do pH ficaram dentro dos intervalos (2;60,35) e (0,002;0,098), respectivamente. As Figuras 3 e 4 ilustram os mapas preditivos utilizando os valores  
329 estimados em cada abordagem de modelos bivariados. Não existe muita diferença entre as previsões espaciais utilizando modelos univariados ou bivariados, mas o *BGCCM* gerou menores variâncias de krigagem.

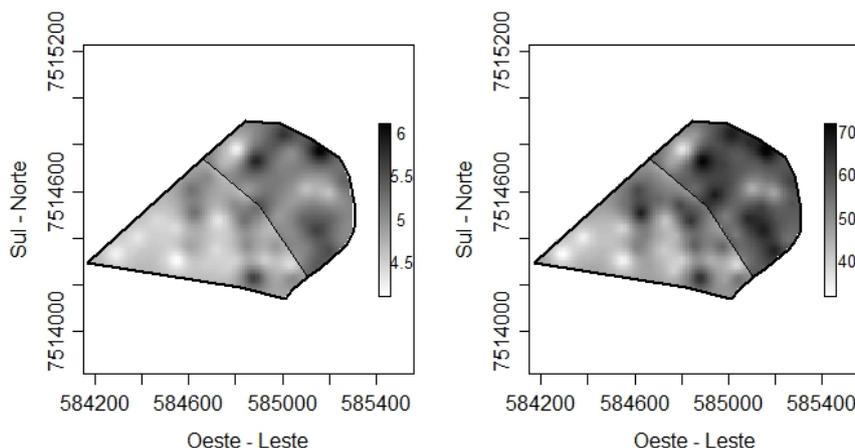
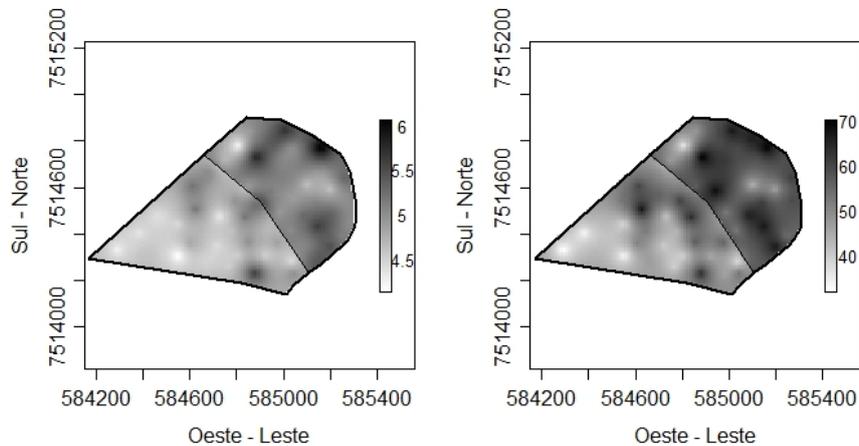


Figura 3: Gráficos de predições espaciais com o BGCCM, o gráfico da esquerda é relativo ao pH e o gráfico da direita é relativo à saturação por bases



336 Figura 4: Gráficos de predições espaciais com o BCRM, o gráfico da esquerda é relativo ao pH e o gráfico da direita é relativo à saturação por bases

Além da fundamentação estatística para considerar modelos bivariados para o problema dos dados do solo, existe a justificativa prática para adoção de tal abordagem. A coleta de informação da saturação por bases é mais dispendiosa, e como a propriedade agrícola continuará sendo monitorada futuramente, é possível que nas próximas avaliações do solo seja amostrado um número menor de localizações espaciais para observação da saturação por bases, pois é possível utilizar a  
343 informação contida no pH para fazer inferências sobre a saturação por bases. Ou seja, talvez ocorra diminuição de gastos em monitoramentos de solos se forem utilizados os modelos geoestatísticos bivariados para problemas com variáveis químicas correlacionadas.

Para ilustrar esse fato, foram omitidas do conjunto de dados 20 localizações espaciais da saturação por bases, e novamente, em todas as abordagens, estimou-se os parâmetros dos modelos finais selecionados. Ocorreu uma diferença moderada entre os resultados das estimativas calculadas com a omissão de localizações e os das calculadas com a informação completa da saturação por  
350 bases. No entanto, o objetivo é avaliar a qualidade preditiva dos modelos univariados e bivariados.

Com intuito de verificar as vantagens utilizando os modelos bivariados, foi conduzida a krigagem ordinária da saturação por bases nas 20 localizações da subamostra de controle sob os três modelos reajustados. As Tabelas 4 e 5 apresentam os valores observados e preditos sob todas as abordagens consideradas e os resultados descritivos dos erros de predição nas três abordagens de modelagem. Os modelos bivariados se mostraram mais eficientes para fazer a krigagem, uma vez que, os desvios padrões dos erros de krigagem são menores do que o valor gerado pela abordagem  
357 univariada, ou seja, há evidências de que é melhor utilizar a informação do pH para fazer as inferências sobre a saturação por bases com menos observações.

Observado	Univariado	BGCCM	BCRM
31	37,97514	36,70374	36,41034
52	50,31664	48,47015	48,75127
41	38,7341	43,28838	42,6973
51	42,32669	50,60234	50,66902
41	56,75724	39,95526	40,90431
43	53,03821	46,20176	46,51603
39	44,36711	43,17711	42,64546
51	56,26729	48,47547	48,79987
64	46,88141	68,33242	68,03238
53	47,06461	53,68137	52,71941
47	46,36761	47,39061	46,98018
67	58,88574	70,07791	69,85622
57	54,87683	59,06358	58,93524
56	57,6176	52,66449	52,93289
59	60,95767	54,67692	55,20092
57	56,82567	56,92622	57,16953
52	57,78833	52,64306	52,94522
61	56,3847	61,31782	61,35151
59	56,07022	54,4587	54,30496

Tabela 4: Valores observados e preditos da saturação por bases nas 20 localizações da subamostra de controle

Estatística	Univariado	BGCCM	BCRM
Média	-0,34	-0,50	-0,50
Desvio Padrão	7,52	3,02	2,84

Tabela 5: Médias e desvios padrões dos erros de krigagem da saturação por bases nas 20 localizações da subamostra de controle

## CONCLUSÕES

364           Esse trabalho teve o objetivo de avaliar o comportamento de modelos geoestatísticos bivariados, em conjunto com estimação por máxima verossimilhança e krigagem ordinária. Foram feitas implementações computacionais e comparações entre dois modelos bivariados, bem como procurou verificar a aplicabilidade desses modelos com dados observacionais e suas vantagens com relação aos modelos univariados.

371           Os modelos geoestatísticos univariados e bivariados em conjunto com estimadores por máxima verossimilhança se mostraram eficientes para gerar mapas preditivos do pH e da saturação por bases da fazenda Tupã.

          A análise dos dados mostrou que os modelos geoestatísticos bivariados são uma ferramenta útil para a agricultura de precisão, pois sua utilização pode baratear os custos dos monitoramentos de solos e ao mesmo tempo gerar mapas preditivos com alta qualidade.

## LITERATURA CITADA

- CARVALHO, M.P. et al. Variabilidade espacial de atributos de um solo sob videira em Vitória Brasil (SP). R. Bras. Ci. Solo, Viçosa, 27:695-703, 2003.
- 378 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. Nonstationary multivariate process modeling through spatially varying coregionalization. R. Bras. Ci Solo, Viçosa, 32:463-476, 2008.
- GELFAND, A.E.; SCHMIDT, A.M.; BANERJEE S. & SIRMANS, C.F. Nonstationary Multivariate process modeling through spatially varying coregionalization. Test, Valencia, 13:263-312, 2004.
- 385 LE, D.N. & ZIDEK, J.V. Statistical analysis of environmental space-time processes. New York, Springer, 2006. 327p.
- MATÈRN, B. Spatial variation. Berlin, Springer, 1986. 365 p.
- NELDER, J.A. & MEAD R. A simplex method for function minimization. Comput. J., Oxford, 7:308-313, 1965.
- OLIVEIRA, J.J. et al. Variabilidade espacial de propriedades químicas em um solo salino-sódico. R. Bras. Ci. Solo, Viçosa, 23:783-789, 1999.
- 392 R: A Language and Environmental for Statistical Computing, R Development Core Team, ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>. Acesso em 28 mai 2007.
- RIBEIRO Jr, P.J. & DIGGLE, P.J. geoR: A package for geostatistical analysis. R-NEWS, Vienna, 1:14-18, 2001. ISSN 1609-3631.
- SCHABENBERGER, O. & GOTWAY, C.A. Statistical methods for spatial data analysis. Boca Raton, Chapman and Hall / CRC, 2005. 488p.
- SCHIMIDT, A.M. & SANSÓ, B. Modelagem bayesiana da estrutura de covariância de processos  
399 espaciais e espaço-temporais. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 14, Caxambú, 2006. Minicurso... São Paulo: Associação Brasileira de Estatística, 2006. 151 p.
- SCHULLER, J.K. A review and integrating analysis of spatially-variable crop control of crop production. Fert. Res., Amsterdam, 33:1-34, 1992.
- SILVA, A.S. Modelos gaussianos geoestatísticos espaço-temporais e aplicações, 2006. 70 p. Dissertação (Mestrado em Estatística e Experimentação Agrônômica) – ESALQ/USP, 2006.
- 406 SILVA, V.R. et al. Variabilidade espacial das características químicas do solo e produtividade de milho em um Argissolo Vermelho-Amarelo distrófico arênico. R. Bras. Ci. Solo, Viçosa, 27:1013-1020, 2003.