

MODELANDO SATURAÇÃO POR BASES E PH DO SOLO: UMA APLICAÇÃO DE MODELOS GEOESTATÍSTICOS BIVARIADOS ¹

Bruno Henrique Fernandes Fonseca ², Paulo Justiniano Ribeiro Jr. ³, Bruno Gherardi ⁴ & Luiz Eduardo Oliveira de Faria ⁵

RESUMO

A agricultura de precisão utiliza diversas formas de pesquisas para tomadas de decisões quanto ao manejo do solo para aumento de produtividade, uma abordagem muito utilizada é a mensuração de atributos do solo da região sob estudo. Neste contexto, onde pode existir padrão espacial de variabilidade nas variáveis químicas, a modelagem geoestatística é um conjunto de ferramentas útil. O interesse final desses modelos é utilizar os valores observados dos atributos em algumas localizações georeferenciadas para propor, com respaldo probabilístico, mapas das variáveis químicas para toda a região sob estudo. Tradicionalmente, um modelo geoestatístico é proposto para cada variável de interesse. No entanto, em alguns casos existe correlação entre as respostas, o que leva à possibilidade de utilizar modelos geoestatísticos multivariados, além da intuição estatística, existem justificativas práticas para tal modelagem, principalmente a redução de custos para mensuração dos atributos do solo em monitoramentos futuros da mesma área sob estudo. Neste estudo utilizou-se modelos univariados e bivariados em conjunto com estimação por máxima verossimilhança e krigagem ordinária para propor mapas preditivos das variáveis pH e a saturação por bases do solo de uma fazenda, localizada em Echaporan/SP e com 51,8ha de área, onde foram amostradas 67 localizações georeferenciadas e mensurados os atributos de interesse. As abordagens univariadas e bivariadas de modelagem se mostraram eficientes para o problema. Por último a amostra da saturação por bases, resposta mais dispendiosa, foi dividida em uma subamostra de modelagem e uma subamostra de controle. A subamostra de modelagem foi utilizada para ajustar novamente os modelos univariados e bivariados, a subamostra de controle foi utilizada para comparar resultados preditivos das abordagens de modelagem. Os modelos bivariados se mostraram mais eficientes para fazer krigagem, o que corroborou a idéia de que em estudos futuros da mesma propriedade agrícola é possível reduzir o custos de amostragem.

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

¹Parte 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.

²Mestrando em Agronomia pela Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo - ESALQ/USP. Av. Pádua Dias 11, CEP 13418-900 Piracicaba (SP). E-mail: bh.fonseca@hotmail.com

³Professor Adjunto da Universidade Federal do Paraná - UFPR. E-mail: paulojus@ufpr.br

⁴Mestrando em Agronomia pela Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo - ESALQ/USP. E-mail: bruno.gherardi@agrofficio.com.br

⁵Bacharel em Agronomia pela Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo - ESALQ/USP. E-mail: luiz.faria@agrofficio.com.br

ABSTRACT

Precision agriculture used some types of researches for decision-making regarding soil management to improve productivity, a approach very used is mensuration of attributes of the soil of region about study. In this context, where there may be spatial variability in the chemical variables, the geostatistical modeling is a useful set of tools. The final interest of these models is use the observed values of the attributes in some referenced locations to propose, with probabilistic support, maps of the variables for all region about study. Traditionally, a geostatistical model is proposed for each variable of interest. However, in some cases there is correlation between responses, then is possible to use multivariate geostatistical models, only the statistical intuition do not justify the increase in complexity of the models, it is necessary that there are practical advantages, mainly the reduction of spending with sampling process. In this study were considered univariate and bivariate models together with maximum-likelihood and ordinary kriging to propose predict maps of the pH and saturation by bases of the soil of a farm localized in Echaporan/SP, where were sampled 67 georeferenced locations to observe the responses. The approaches univariate and bivariate of modeling were efficient to problem. And finality the sample of the saturation by bases, response more expensive, was divided in one subsample of modeling and a one subsample of control. Univariate and bivariate models were fitted with the subsample of modeling. The predict accuracies of the approaches was analyzed with the subsample of control. The bivariate models were more efficient to make kriging, then was confirmed the idea of reduction of spending in the future monitoring of the same farm.

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

INTRODUÇÃO

A agricultura de precisão, de forma bem sucinta, é o conjunto de métodos aplicados ao manejo da variabilidade. SCHUELLER definiu, em 1992, como 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.

De fato, vários autores concluíram que a variabilidade espacial existe, mesmo em áreas consideradas homogêneas. Inúmeros 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, sendo que estes trabalhos mostram que a variabilidade do solo não é puramente aleatória, apresentando correlação ou dependência espacial (VIEIRA et al., 1983; SOUZA et al., 1998; OLIVEIRA et al., 1999; CARVALHO et al., 2003; SILVA et al., 2003).

Diversas abordagens podem ser utilizadas para estruturar a variabilidade espacial das variáveis químicas do solo, em agricultura de precisão, tradicionalmente, utiliza-se apenas medidas descritivas e a intuição pessoal dos pesquisadores para tal, o que pode gerar o problema de não ser possível mensurar

a precisão dos resultados. Uma solução recente para esse tipo de problema, que proporciona resultados com respaldo probabilístico, é a modelagem geoestatística em conjunto com estimação por máxima verossimilhança.

A modelagem geoestatística é conjunto de técnicas que serve para propor uma boa função matemática, com embasamento probabilístico, para um processo estocástico de interesse, que é contínuo sobre uma área de estudo e possui possível padrão espacial de correlação. Logo, $\{Z(s) : s \in G \subset R^d\}$ é a notação para um campo aleatório existente em todas as localizações $s = (s_1, s_2, \dots, s_n)$ de uma área sob estudo G , que está contido no conjunto dos reais d -dimensionais, tradicionalmente $d \leq 3$, por exemplo, se uma pesquisa considera uma estrutura com latitudes e longitudes para as localizações s , logo, $d=2$ e s_i , com $i = (1, 2, \dots, n)$, é um vetor com dois valores que determina a localização georeferenciada de $Z(s_i)$, processo estocástico sob estudo na localização s_i .

Segundo Schmidt e Sansó (2006) e Le e Zidek (2006), a descrição de um campo aleatório é obtida através da sua distribuição de probabilidade acumulada conjunta, para qualquer conjunto de pontos nas localizações s_1, s_2, \dots, s_n pertencentes à região G e qualquer inteiro n . Uma suposição muito utilizada na literatura de geoestatística é que $Z(s)$ possui distribuição de probabilidade gaussiana n -variada. Sendo assim, o processo estocástico é completamente especificado por μ e Σ , vetor de médias $n \times 1$ e matriz de covariâncias $n \times n$, respectivamente. Essa distribuição é muito utilizada devido a maior simplicidade inferencial.

Como na maioria dos estudos não é viável fazer repetições das mensurações de $Z(s)$, existe apenas uma observação do processo estocástico sob estudo, sendo assim, outras suposições, além da gaussianidade, são necessárias para fazer inferências como, por exemplo, estacionariedade e isotropia do campo aleatório. Diggle e Ribeiro Jr. (2007) expõe maiores detalhes.

A matriz de covariâncias Σ deve ser positiva definida e possuir o comportamento empírico utilizado em geoestatística, de que quanto maior a distância entre duas localizações s_k e s_l quaisquer, pertencentes a área sob estudo, menor é a correlação entre $Z(s_k)$ e $Z(s_l)$. Encontrar analiticamente funções de correlação que gerem esse comportamento não é trivial, contudo, na literatura existem diversas funções propostas. A família de funções de correlação de Mátern, proposta por Mátern (1982), é muito utilizada atualmente, por conta da flexibilidade gerada pelos seus parâmetros: ϕ e κ , que estão associados ao alcance das correlações e a diferenciabilidade do processo, respectivamente. Ou seja, a maior vantagem em utilizar essa função é que vários tipos de suavidade dos processos podem ser modeladas. Maiores detalhes sobre essas funções e suas propriedades são expostos por Diggle e Ribeiro Jr. (2007).

Considerando que existe interesse sob um campo aleatório gaussiano $Z(s)$ contínuo sob uma área de estudo, porém latente, pode-se pensar em modelagem geoestatística:

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

sendo $Y(s)$ um vetor $n \times 1$ dos valores observados do atributo sob estudo nas n localizações amostradas $s = (s_1, \dots, s_n)$, μ um vetor $n \times 1$ que contem os parâmetros relativos a média de $Y(s)$, $Z(s)$ um campo aleatório gaussiano que possui vetor de médias nulo $n \times 1$ e matriz de covariâncias Σ $n \times n$, sendo cada elemento $\Sigma_{l,k}$ igual a $cov(Z(s_l), Z(s_k))$, para todo l e $k = 1, 2, \dots, n$, e ϵ um vetor $n \times 1$ de ruídos brancos, que por suposição, são independentes e identicamente distribuídos de uma distribuição de probabilidade normal com média zero e desvio padrão τ . Faraco et al. (2008) mostra a aplicabilidade desse modelo.

Utilizando (1) e as suposições feitas sobre $Z(s)$ é possível encontrar a distribuição de probabilidade da variável resposta, $Y(s)$ é normal n -variada, com vetor de médias μ e matriz de covariâncias $\Sigma_Y = \sigma^2 R + \tau^2 I$, sendo R uma matriz $n \times n$, onde cada $R_{l,k}$ é igual a correlação entre $Z(s_l)$ e $Z(s_k)$, para todo para todo l e $k = 1, 2, \dots, n$, I é uma matriz identidade $n \times n$ e σ^2 e τ^2 são parâmetros de variabilidade do campo aleatório latente e dos ruídos brancos, respectivamente.

Com a distribuição de probabilidade de $Y(s)$ definida é possível fazer estimação por máxima verossimilhança para o vetor de parâmetros $\theta = (\beta^*, \sigma^2, \phi^*, \tau^2)$, onde β^* é um vetor de parâmetros associados a μ e ϕ^* são os parâmetros associados à função de correlação adotada. A estimação por máxima verossimilhança para esse problema consiste em utilizar os valores observados da variável resposta para encontrar um vetor $\hat{\theta}$ que seja o ponto de máximo da função de verossimilhança associada ao vetor de parâmetros de $Y(s)$. Maiores detalhes sobre técnicas e propriedades da estimação por máxima verossimilhança são expostas por Azzalini (1996). Com a função de verossimilhança associada a (1), não é possível encontrar formas analíticas para alguns estimadores, sendo assim, reparametrizações, funções de máxima verossimilhança concentradas, estimadores de mínimos quadrados ponderados e métodos numéricos para encontrar mínimos ou máximos de superfícies podem ser utilizados, maiores detalhes sobre essas técnicas são expostas por Diggle e Ribeiro Jr. (2007).

A maioria das pesquisas de geoestatística tem como interesse final utilizar os valores de $Y(s)$ e $\hat{\theta}$ para propor valores previstos para o campo aleatório em qualquer localização da região sob estudo. Neste contexto, é utilizada a krigagem, processo de predição espacial em localizações não amostradas. As krigagens simples e ordinária são tipos de predição espacial mais utilizada na literatura, maiores detalhes sobre essas ferramentas geoestatísticas são expostas por Diggle e Ribeiro Jr. (2007).

Em muitos estudos o interesse não é sobre um único atributo. A intuição inicial é que seja ajustado (1) para cada atributo individualmente. Porém, pode haver correlação estatística entre algumas respostas, o que leva a possibilidade de adotar modelos geoestatísticos multivariados. Somente com essa justificativa estatística não é vantajoso utilizar tal abordagem, é necessário que existam justificativas e vantagens práticas para esse aumento de complexidade dos modelos geoestatísticos, que neste caso devem modelar a correlação entre e dentre atributos.

Por simplicidade, suponha a existência de dois campos aleatórios gaussianos homogêneos sob

estudo, que podem ser modelados da seguinte forma:

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

Y_i é um vetor $n_i \times 1$ de valores observados campo aleatório gaussiano latente Z_i , que possui vetor de médias nulo $n_i \times 1$ e matriz de covariâncias Σ_i $n_i \times n_i$, μ_i é um vetor, que possui os parâmetros de médias associados a Y_i . 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 Σ_Y , positiva definida que possui o comportamento empírico de correlações utilizado em geoestatística e que pode ser particionada da forma:

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

sendo Σ_i é 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 .

Na literatura existem algumas abordagens para estruturar Σ_Y . Diggle e Ribeiro Jr. (2007) propuseram uma abordagem para problemas geoestatísticos bivariados, a qual é abreviada nesse trabalho por BGCCM e que utiliza as seguintes decomposições dos campos aleatórios latentes de (2):

$$Z_i = \sigma_{0i}S_0 + \sigma_i S_i, \quad i = 1, 2, \quad (3)$$

sendo $\sigma^* = (\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2)$ um vetor de parâmetros de dispersão associados a (2) 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 quaisquer. As quais vão gerar empiricamente as correlações cruzadas entre Y_1 e Y_2 , uma vez que, o campo aleatório S_0 é comum as duas respostas, sendo assim, três funções de correlação válidas devem ser adotadas para gerar uma matriz Σ_Y válida.

Utilizando (3) e 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 , para todo $l, k = 1, 2, \dots, n_i$ 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 ρ_0 e ρ_i as funções de correlação adotadas para S_0 e S_i , respectivamente. 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 Σ_Y fica estruturada completamente, logo, a distribuição de probabilidade do vetor Y está estabelecida e depende do vetor de parâmetros $\theta = (\beta^*, \sigma^*, \phi_0^*, \phi_1^*, \phi_2^*)$, onde β^* é um vetor de parâmetros associado a μ e ϕ_j^* é um vetor de parâmetros associado a escolha da função ρ_j , para todo $j = 0, 1, 2$.

Gelfand et al. (2004) utiliza o modelo de co-regionalização bivariado, abreviado nesse trabalho

por BCRM, para propor uma estrutura paramétrica válida para Σ_Y . Nesta abordagem as componentes latentes de (2) são decomposto da seguinte forma:

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

sendo $\sigma^* = (\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2)$ um vetor de parâmetros de dispersão associados a (2) e S_1 e S_2 campos aleatórios gaussianos independentes, com vetores de médias nulo, variância unitária e com correlações determinadas pela escolha de funções de correlação conhecidamente válidas. Como 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 final de Σ_Y é similar ao exposto para o BGCCM. O vetor de parâmetros associados a distribuição de probabilidade de Y é $\theta = (\beta^*, \sigma^*, \phi_1^*, \phi_2^*)$, sendo β^* um vetor de parâmetros associado a μ e ϕ_j^* um vetor de parâmetros associado a escolha da função de correlação para S_{j2} , para todo $j = 1, 2$.

Sob as duas abordagens de modelos bivariados considerados, o próximo passo é estimar o vetor de parâmetros θ . As technicalidades são similares às adotadas no modelo univariado, no entanto, para modelos bivariados é necessário um número maior de reparametrizações.

Esta pesquisa tem o objetivo de expor a utilização de técnicas geoestatísticas multivariadas para gerar boas informações, com embasamento probabilístico, e baratear custos da agricultura de precisão, para tal foram estudadas as variabilidades espaciais do pH e da saturação por bases do solo de uma propriedade agrícola, essas duas variáveis químicas são muito importantes para nortear decisões sobre o melhor trato do solo. A justificativa para adoção de modelos geoestatísticos bivariados para o problema é baseada na evidência de forte correlação entre os dois atributos e no custo operacional para observação da saturação por bases, que é mais dispendioso. Estabelecida a estrutura de correlação espacial conjunta dos dois processos estocásticos, em monitoramentos futuros da mesma propriedade será possível fazer um número menor de observações da saturação por bases, uma vez que, pode-se inferir sobre essa variável utilizando as informações dela mesmo e do pH.

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 Echaporan/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 Mercatur (UTM). Com a utilização de um aparelho GPS foram medidos os atributos de interesse.

Uma análise exploratória inicial dos dados foi conduzida utilizando gráficos e medidas descritivas,

os quais detectaram padrão espacial das variáveis químicas. Do ponto de vista geoestatístico, o pH e a saturação por bases 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 Σ_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 β_i^* um vetor $p \times 1$ com os parâmetros associados a μ_i .

Utilizando os resultados da análise exploratória, 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 67×67 de correlações, considerando o parâmetro de suavidade κ fixo nos valores 0,5, 1, 1.5, 2, 2.5, logo, ϕ_i^* depende apenas de ϕ_i . Combinando as diferentes escolhas de κ e X_i foram ajustados diversos modelos. As estimativas paramétricas foram calculadas por máxima verossimilhança, onde foi utilizado a reparametrização $\nu_i = \tau_i/\sigma_i$ e o método numérico de Nelder e Mead (1965) para encontrar as estimativas $\hat{\phi}_i$ e $\hat{\nu}_i$ que maximizam a função de máxima verossimilhança concentrada, logo, utilizando a propriedade de invariância dos estimadores em questão foram feitas as demais estimativas de interesse.

Os valores dos máximos estimados da função de máxima verossimilhança concentrada e o Critério de Informação de Akaike (AIC), exposto por AKAIKE (1973), foram utilizados para selecionar o modelo final de cada atributo. Com os parâmetros estimados dos modelos finais 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 as proposições de Diggle e Ribeiro Jr. (2007) e de Gelfand et al. (2004). 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 β^* é um vetor $p \times 1$ com os parâmetros de média associados a Y . Para estruturar Σ_Y somente funções de correlação da família de Matérn foram utilizadas, sendo que o parâmetro κ foi sempre considerado fixo em valores similares aos modelos univariados.

Combinando diferentes de escolhas de κ e X , diversos modelos foram ajustados. Em todos os casos o vetor de parâmetros θ foi estimado por máxima verossimilhança, sendo que foram utilizadas as mesmas técnicas dos modelos univariados. No BGCCM foi lançado mão das 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 das reparametrizações $\sigma = \sigma_1$, $\nu_1 = \sigma_{12}/\sigma_{11}$ e $\nu_2 = \sigma_{22}/\sigma_{11}$. Então, nas duas abordagens existe forma analítica para os estimadores de β^* e σ^2 , que são função dos demais parâmetros, os quais foram estimados pelo método de Nelder e 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 ferramentas dos modelos univariados foram utilizadas.

Por último, para exemplificar as vantagens da abordagem bivariada, 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. Com os parâmetros estimados foram calculadas as predições da saturação por bases nas localizações do grupo de controle, o que possibilitou a comparação entre as metodologias.

Cabe ressaltar que todas as análises estatísticas desse estudo foram conduzidas no ambiente *R* de programação (R Development Core Team, 2005), sendo que o pacote mais utilizado nas análises foi o *geoR* (Ribeiro Jr. e Diggle, 2001).

RESULTADOS E DISCUSSÃO

A Tabela 1 mostra as estatísticas descritivas do pH e da saturação por bases, nos dois casos a mediana é próxima da média e existe uma variabilidade relativamente pequena dos dados. De uma maneira geral, os atributos de solo, estão bem comportados, onde há uma simetria moderada dos dados com relação a média amostral e não existem valores observados muito discrepantes.

Variável	Mediana	Média	D.P.
Saturação	56,00	53,27	10,05
pH	5,00	4,94	0,42

TABELA 1: Estatísticas descritivas do pH e da saturação por bases

A Figura 1 mostra os gráficos de círculos das variáveis sob estudo, onde os círculos representam os valores mensurados dos atributos nas 67 localização amostrada da propriedade agrícola. Quanto maior o círculo mais elevado é o valor observado do atributo. Essa primeira análise gráfica evidência que existe um possível padrão espacial nas variáveis, onde há uma suavidade dos processos ao longo da fazenda. A linha no meio dos gráficos 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. Além disso, os gráficos são bem parecidos para as duas variáveis, o que corrobora a idéia de que os dados são fortemente correlacionados.

Combinando as diferentes abordagens para a matriz de delineamento X e as diferentes escolhas para κ , 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 Tabelas 2 mostra os resultados

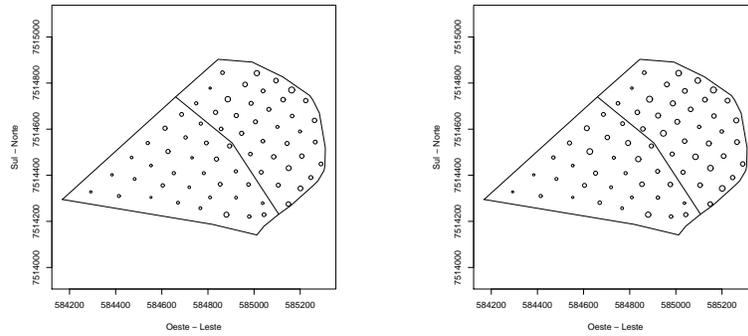


FIGURA 1: Gráficos de círculos - pH a esquerda e saturação por bases a direita

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 ϕ 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 distância qualquer par de localizações possui correlação inferior a 0,05.

Tendência	β_0	β_1	τ^2	σ^2	ϕ	κ	$l(\theta)$	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 2: Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para o pH com média constantes

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, 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.

Para modelar a saturação por bases foi conduzida a mesma metodologia aplicada ao pH. O modelo final também inclui a coordenada leste-oeste 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.

Com os modelos finais estabelecidos, foram calculadas as predições espaciais para as duas variáveis químicas do solo em uma malha de 10000 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.

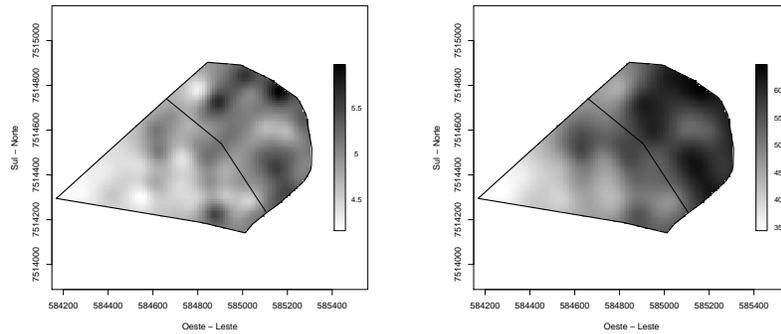


FIGURA 2: Gráficos de predições espaciais, o gráfico da esquerda é relativo ao pH e o gráfico da direita é relativo à saturação por bases

Para validar o pressupostos feitos sob os ruídos brancos dos modelos univariados, foram utilizados testes de normalidade de Shapiro-Wilk, que geraram p-valores de 0,22 e 0,97 para os erros do pH e da saturação por bases, respectivamente.

Com a modelagem univariada concluída, o próximo passo foi ajustar modelos bivariados para as respostas, em caráter descritivo foi calculada a correlação de Pearson entre os atributos, a estatística resultante ficou em 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. As quais combinadas com diversas escolhas para os parâmetros de suavidade das correlações geraram diversas estimativas diferentes para os parâmetros dos modelos. Inicialmente, foram selecionadas as melhores estimativas em cada abordagem para a média. A Tabela 3 mostra os resultados do BGCCM, onde não há muita diferença entre os parâmetros de variabilidade e correlação estimados para cada abordagem de

tendência, a não ser pelo parâmetros ϕ_0 , que para a tendência na área de manejo ficou bem menor do que as demais, fato explicado pela adoção de um κ_0 maior, logo, combinando essas duas informações, tem-se que os alcances de correlações são bem parecidos para as três abordagens de média.

θ	Constantes	Oeste-leste	Área de manejo
β_{01}	52,57	$1,54e - 6$	47,83
β_{02}	4,92	$1,45e - 7$	4,72
β_1	—	$8,98e - 5$	9,46
β_2	—	$8,41e - 6$	0,40
σ_{01}	8,51	8,35	6,85
σ_1	3,47	3,46	3,44
σ_{02}	0,38	0,38	0,31
σ_2	$7e - 5$	$5e - 4$	$1e - 5$
ϕ_0	59,37	57,79	26,73
ϕ_1	46,35	48,15	49,08
ϕ_2	66,82	71,59	86,95
κ_0	1	1	1,5
κ_1	0,5	0,5	0,5
κ_2	0,5	0,5	0,5

TABELA 3: Estimativas de máxima verossimilhança dos parâmetros associados ao BGCCM para a saturação por bases e o pH

A Tabela 4 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 as demais estimativas, os resultados são parecidos com os do BGCCM.

θ	Constantes	Oeste-leste	Área de Manejo
β_{01}	51,76	$1,51e - 6$	47,93
β_{02}	4,89	$1,43e - 7$	4,72
β_1	—	$8,85e - 5$	9,12
β_2	—	$8,36e - 6$	0,40
σ_{11}	8,26	99,53	7,82
σ_{12}	0,31	3,67	0,28
σ_{22}	0,13	1,54	0,14
ϕ_1	100,63	100,21	53,32
ϕ_2	21,43	21,38	19,34
κ_1	0,5	0,5	0,5
κ_2	2,5	2,5	2,5

TABELA 4: Estimativas de máxima verossimilhança dos parâmetros associados ao BCRM para a saturação por bases e o pH

As duas abordagens de modelos bivariados se mostraram 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.

Nas duas abordagens, 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. Com as estimativas finais estabelecidas é possível pensar na krigagem ordinária, para tal, foi utilizado o mesmo grid de pontos da krigagem da modelagem univariada.

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 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.

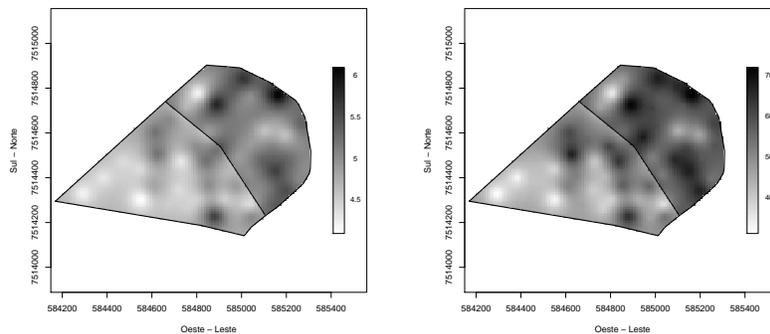


FIGURA 3: Gráficos de previsões espaciais com o BGCCM, o gráfico da esquerda é relativo ao pH e o gráfico da direita é relativo à saturação por bases

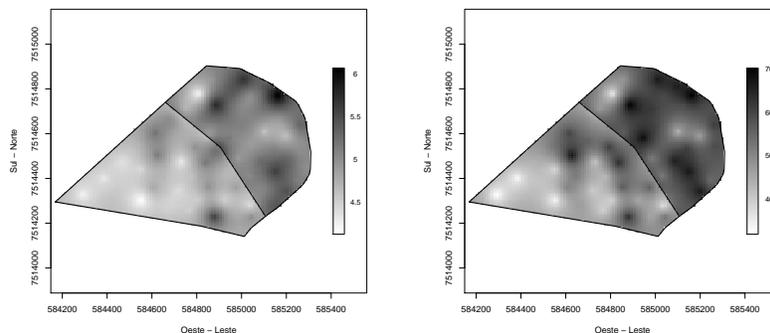


FIGURA 4: Gráficos de previsões espaciais com o BCRM, o gráfico da esquerda é relativo ao pH e o gráfico da direita é relativo à saturação por bases

Para verificar se existem vantagens preditivas com a utilização de modelos mais complexos para o problema, foi conduzido um estudo empírico, que dividiu a amostra da saturação em duas subamostras. Com uma subamostra foram estimados novamente, em todas as abordagens, 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 bases. No entanto, o objetivo dessa análise é avaliar a qualidade preditiva dos modelos univariados e bivariados.

Para verificar o ganho com a utilização de modelos bivariados, foi conduzida a krigagem ordinária da saturação por bases nas 20 localizações retiradas do conjunto de dados sob os três modelos reajustados. A Tabela 5 apresenta os resultados descritivos das diferenças entre os valores observados e os valores preditos 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 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.

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

TABELA 5: Mediadas descritivas das diferenças entre os observados e as krigagem nos pontos retirados da Saturação por bases

CONCLUSÕES

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.

Com o estudo empírico de omissão de algumas localizações, as abordagens bivariadas se mostraram mais eficientes para fazer predição espacial do que a abordagem univariada.

Sendo assim, foi cumprido o objetivo de aplicar as técnicas geoestatísticas para modelar problemas de monitoramentos do solos. Também, foi constado que com a utilização de modelos bivariados é possível diminuir custos com análises de solo mantendo a qualidade das predições espaciais.

AGRADECIMENTOS

À CAPES pelo apoio financeiro, que facilitou a realização dessa pesquisa.

LITERATURA CITADA

AKAIKE, H. Information theory as an extension of the maximum likelihood principle. **International Symposium on Information Theory**, 2. Budapest: Proceedings. Budapest, Akadémia Kiadó, 1973. p.267-281.

AZZALINI, A. **Statistical inference**: based on the likelihood. London: Chapman and Hall, 1996. 341 p.

CARVALHO, M.P. et al. Variabilidade espacial de atributos de um solo sob videira em Vitória Brasil (SP). **Revista Brasileira de Ciência do Solo**, Viçosa, v.27, n.4, p.695-703, 2003.

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. **Revista Brasileira de Ciências do Solo**, Viçosa, v.32, p.463-476, 2008.

GELFAND, A.E.; SCHMIDT, A.M.; BANERJEE S.; SIRMANS, C.F. Nonstationary multivariate process modeling through spatially varying coregionalization. **Sociedad Española de Estadística e Investigación Operativa - Test**, Valencia, v.13, p.263-312, 2004.

LE, D.N.; ZIDEK, J.V. **Statistical analysis of environmental space-time processes**. New York: Springer, 2006. 327p.

MATÉRN, B. **Spatial variation**. Verlag, Berlin: Spinger, 1986. 365 p.

NELDER, J.A.; MEAD R. A simplex method for function minimization. **The Computer Journal**, Oxford, v.7: p.308-313, 1965. Spinger, 1986. 365 p.

OLIVEIRA, J.J. et al. Variabilidade espacial de propriedades químicas em um solo salino-sódico. **Revista Brasileira de Ciência do Solo**, Viçosa, v.23, n.4, p.783-789, 1999.

SCHIMIDT, A.M.; SANSÓ, B. Modelagem bayesiana da estrutura de covariância de processos espaciais e espaço-temporais. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 14, 2006. Caxambú, **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. **Fertilizer Research**, Amsterdam, v.33, p.1-34, 1992.

SILVA, V.R. et al. Variabilidade espacial das características químicas do solo e produtividade de milho em um Argissolo Vermelho-Amarelo distrófico arênico. **Revista Brasileira de Ciência do Solo**, Viçosa, v.27, n.6, p.1013-1020, 2003.

SOUZA, L.S. et al. Variabilidade de fósforo, potássio e matéria orgânica no solo, em relação a sistemas de manejo. **Revista Brasileira de Ciência do Solo**, Viçosa, v.22, n.1, p.77-86, 1998.

VIEIRA, S.R. et al. Geostatistical theory and application to variability of some agronomical properties. **Hilgardia**, Oakland, v.51, n.1, p.1-75, 1983.

R: A Language and Environment for Statistical Computing, R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2006, ISBN 3-900051-07-0, <http://www.R-project.org>.

geoR: A package for geostatistical analysis. RIBEIRO Jr, P.J. and DIGGLE, P.J., R-NEWS, CRAN.R-project.org/doc/Rnews/.v.1, n.2, p.14-18, 2001.