

# MODELANDO SATURAÇÃO POR BASES E PH DO SOLO: UMA APLICAÇÃO DE MODELOS GEOESTATÍSTICOS MULTIVARIADOS

Bruno Henrique Fernandes Fonseca <sup>1</sup>, Paulo Justiniano Ribeiro Jr. <sup>2</sup>

## 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, que pode existir dependência espacial entre as mensurações de cada atributo, a modelagem geoestatística é um conjunto de ferramentas útil, onde o interesse final é utilizar os valores observados 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. Sendo assim, neste estudo utilizou-se dois modelos bivariados em conjunto com estimação por máxima verossimilhança e krigagem 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 de modelagem bivariada utilizadas se mostraram eficientes para estruturar a correlação espacial dentro e entre as variáveis químicas, o que possibilitou a construção de mapas preditivos para as mesmas em toda a propriedade e a mensuração das precisões dessas predições, as quais mostraram uma qualidade melhor do que as precisões utilizando modelagem geoestatística univariada para cada atributo.

---

<sup>1</sup>ESALQ/USP - PPG em Estatística e Experimentação Agronômica, bhff13@yahoo.com.br

<sup>2</sup>UFPR/Laboratório de Estatística e Geoinformação, paulojus@ufpr.br

## 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.

Esse conceito não é recente. Linley e Bauer, citados por Goering (1993) e por LOWENBERG-DEBOER e SWINTON (1995), propunham, em 1929, o tratamento localizado da correção do pH de uma área de 17 ha com base em dados de 23 amostras de solo ordenadamente coletadas. Através da circular no. 346 da estação Experimental Agrícola da Universidade de Illinois (USA), os autores concluíram que: os requisitos de aplicação variavam bastante para um mesmo talhão em sua necessidade de calcário; e a necessidade de que fossem feitos testes sistemáticos e detalhados no campo, de forma que o calcário pudesse ser aplicado de acordo com a sua necessidade.

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  $\mu$  e  $\Sigma$ , 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. (2006) expõe maiores detalhes.

A matriz de covariâncias  $\Sigma$  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 e função de correlação exponencial potência são muito utilizadas atualmente, por conta da flexibilidade gerada pelos seus parâmetros:  $\phi$  e  $\kappa$ , que estão associados ao alcance das correlações a diferenciabilidade do processo, respectivamente. Ou seja, a maior vantagem em utilizar essas duas funções 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. (2006).

Considerando que existe interesse sob um campo aleatório gaussiano  $Z(s)$  contínuo sob uma área de estudo, porém ele é latente, ou seja, não é possível conhecer os seus verdadeiros valores, sendo assim, pode-se pensar em modelagem geoestatística:

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

sendo  $Y(s)$  um vetor  $n \times 1$  dos valores observados do atributo sob estudo nas  $n$  localizações amostradas  $s = (s_1, \dots, s_n)$ ,  $\mu$  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  $\Sigma$   $n \times n$ , onde cada elemento  $\Sigma_{lk}$  é igual a  $cov(Z(s_l), Z(s_k))$ , para todo  $l$  e  $k = 1, 2, \dots, n$ , e  $\epsilon$  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  $\tau$ .

Utilizando (1) e as suposições feitas sobre  $Z(s)$  é possível encontrar a distribuição de prob-

abilidade da variável resposta,  $Y(s)$  é normal  $n$ -variada, com vetor de médias  $\mu$  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  $l$  e  $k = 1, 2, \dots, n$ ,  $I$  é uma matriz identidade  $n \times n$  e  $\sigma^2$  e  $\tau^2$  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  $\beta^*$  é um vetor de parâmetros associados a  $\mu$  e  $\phi^*$  são os parâmetros associados à função de correlação adotada. A estimação de 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). No entanto, para (1) não é possível encontrar formas analíticas para alguns estimadores em questão, 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. (2006).

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, nome dado em homenagem ao pesquisador sul-africano D.G. Krige que foi um dos pioneiros em estudos de predição espacial. 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. (2006).

Em muitos estudos o interesse não é sobre um único atributo como, por exemplo, avaliação da qualidade do solo, que geralmente, utiliza um processo de amostragem de localizações pertencentes a uma certa região sob estudo para observar diversos atributos relativos ao solo, os quais são importantes para nortear tomadas de decisão quanto ao manejo e trato do solo. Sendo assim, existe mais de um campo aleatório a passar pelo processo de modelagem e krigagem, a intuição inicial é que seja feito esse processo para cada atributo individualmente. Porém, pode haver correlação estatística entre alguns atributos, o que leva a possibilidade de adotar modelos geoestatísticos multivariados, contudo, em muitos casos, 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 agora devem capturar a correlação entre e dentro dos 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  $\Sigma_i$   $n_i \times n_i$ ,  $\mu_i$  é um vetor, que possui os parâmetros de médias associados a  $Y_i$ . Contudo, o interesse é sob o vetor  $Y = (Y_1, Y_2)$ , que 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:

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

onde  $\Sigma_i$  é uma matriz  $n_i \times n_i$  das covariâncias dentro da variável  $Y_i$ ,  $i = 1, 2$ , e  $\Sigma_{1,2}$  uma matriz  $n_1 \times n_2$  com as covariâncias cruzadas entre as respostas.

Na literatura existem algumas abordagens para estruturar  $\Sigma_Y$ , Diggle e Ribeiro Jr. (2006) proporam o modelo gaussiano bivariado com componente de correlação parcialmente comum, BGCCM, que decompõe os termos latentes de (2):

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

sendo  $\sigma^* = (\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2)$  um vetor de parâmetros de dispersão associados a (2) e  $S_{01}$ ,  $S_1$ ,  $S_{02}$  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, as funções de correlação adotadas para  $S_{01}$  e  $S_{02}$  devem ser idênticas, ou seja, existe uma componente de correlação comum as duas respostas, logo, estamos adotando três funções de correlação válidas e utilizando somas das mesmas para gerar uma matriz  $\Sigma_Y$  válida.

Utilizando (3) e definindo  $Y_i(s_l)$  e  $Y_i(s_k)$  como observações feitas do atributo  $Y_i$  em duas localizações quaisquer  $s_l$  e  $s_k$ , que estão separados pela distância espacial  $u$ , para todo  $l, k = 1, 2, \dots, n_i$  e  $i = 1, 2$ , tem-se que o elemento  $\Sigma_{i,(l,k)}$  é a  $cov(u) = \sigma_{0i}^2 \rho_0(u) + \sigma_i^2 \rho_i(u)$ , sendo  $\rho_0$  e  $\rho_i$  as funções de correlação adotadas para  $S_{0i}$  e  $S_i$ , respectivamente. Utilizando propriedades básicas de covariâncias, pode-se encontrar  $\Sigma_{1,2}$ , que é igual a  $\sigma_{01}\sigma_{02}R_0$ , onde  $R_0$  é uma matriz  $n_1 \times n_2$  das correlações cruzadas entre as respostas. Dessa forma  $\Sigma_Y$  fica estruturada completamente, logo,

a distribuição de probabilidade do vetor  $Y = (Y_1, Y_2)$  está estabelecida e depende do vetor de parâmetros  $\theta = (\beta^*, \sigma^*, \phi_0^*, \phi_1^*, \phi_2^*)$ , onde  $\beta^*$  é um vetor de parâmetros associado a  $\mu$  e  $\phi_j^*$  é um vetor de parâmetros associado a escolha da função  $\rho_j$ , para todo  $j = 0, 1, 2$ .

O modelo de co-regionalização, BCRM, proposto inicialmente por Matheron (1982), é uma abordagem concorrente ao BGCCM para propor uma estrutura paramétrica válida para  $\Sigma_Y$ , agora os componentes aleatórios de (2) são decomposto de outra maneira:

$$\begin{aligned} Z_1 &= \sigma_{11}S_{11} \\ Z_2 &= \sigma_{12}S_{12} + \sigma_{22}S_{22} \end{aligned}$$

sendo  $S_{11}$ ,  $S_{12}$  e  $S_{22}$  campos aleatórios gaussianos mutuamente independentes, com vetores de médias nulo, variância unitária e com correlações determinada pela escolha de funções de correlação conhecidamente válidas, sendo que a escolha para  $S_{11}$  e  $S_{12}$  deve ser idêntica, ou seja, agora a variável  $Y_1$  possui apenas um termo de correlação, que é comum as duas variáveis, essa abordagem gera um número menor de parâmetros associados as funções de correlação adotadas, apenas duas, e a variabilidade espacial que agora é  $\sigma^* = (\sigma_{11}, \sigma_{12}, \sigma_{22})$ .

Utilizando o BCRM, o desenvolvimento probabilístico para encontrar a estrutura final de  $\Sigma_Y$  é similar ao exposto para o BGCCM. Onde agora o vetor de parâmetros associados a distribuição de probabilidade de  $Y$  é  $\theta = (\beta^*, \sigma^*, \phi_1^*, \phi_2^*)$ , sendo  $\beta^*$  um vetor de parâmetros associado a  $\mu$  e  $\phi_j^*$  um vetor de parâmetros associado a escolha da função de correlação para  $S_{j2}$ , para todo  $j = 1, 2$ . Maiores detalhes sobre essas technicalidades são expostas por Gelfand et al. (2005).

Independente da abordagem utilizada para estruturar  $\Sigma_Y$ , o próximo passo é estimar o vetor de parâmetros  $\theta$ , onde as technicalidades são similares às adotadas ao modelo univariado, no entanto, para modelos bivariados é necessário um número maior de reparametrizações.

Encontrado um  $\hat{\theta}$  para o modelo (2), novamente o interesse final é fazer a krigagem dos dois campos aleatórios em toda área sob estudo. Sendo que agora utiliza-se a estrutura paramétrica das duas respostas para tal.

Sendo assim esta pesquisa tem o objetivo de expor a utilização de técnicas geoestatísticas multivariadas para gerar boas informações, com embasamento probabilístico, e baratear custos para a agricultura de precisão, para tal estudamos a variabilidade espacial 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, além disso, o custo operacional para observação do pH é menos dispendioso, logo, 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 a informação do pH.

## MATERIAL E MÉTODOS

Os dados utilizados neste estudo 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 em outra. Para fazer observação das variáveis químicas do solo, foram amostradas 67 localizações, com grid regular a cada hectare, georeferenciadas no sistema Universal Transverse Mercatur (UTM) e então utilizando um aparelho GPS foram coletadas as informações 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, logo, a intuição inicial é que seja ajustado individualmente o modelo (1) para  $Y_1$  e  $Y_2$ , notação para os dois vetores  $67 \times 1$  de valores observados do pH e da saturação por base, respectivamente. Sendo assim  $Y_i$ , para todo  $i = 1, 2$ , 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 possíveis  $p - 1$  covariáveis e  $\beta_i^*$  um vetor  $p \times 1$  com os parâmetros associados a  $\mu_i$ .

Utilizando os resultados da análise exploratória, consideramos 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, além disso, foi utilizada a família de funções de correlação de Mátern para estruturar a matriz  $67 \times 67$  de correlações  $R$ , considerando o parâmetro de suavidade  $\kappa$  fixo nos valores 0,5, 1, 1.5, 2, 2.5, logo,  $\phi_i^*$  depende apenas de  $\phi_i$ , proveniente da função de correlação escolhida. Ou seja, ajustamos diversos modelos combinando as diferentes escolhas de  $\kappa$  e de  $X_i$ , sendo que em todos os casos foi utilizada estimação por máxima verossimilhança para os parâmetros, onde reparametrizando  $\nu_i = \tau_i / \sigma_i$ , é possível encontrar formas analíticas para os estimadores de máxima verossimilhança de  $\beta_i^*$  e  $\sigma_i^2$ , que são funções dos parâmetros  $\phi_i$  e  $\nu_i$ , os quais não possuem formas analíticas para os estimadores, então utilizamos o método numérico 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 encontramos as demais estimativas de interesse.

Utilizamos os valores dos máximos estimados da função de máxima verossimilhança con-

centrada e o Critério de Informação de Akaike (AIC) para selecionar o modelo final de cada atributo. Então foram utilizados os parâmetros estimados dos modelos finais para fazer as krigagens ordinárias dos campos aleatórios. Por último, uma análise de resíduos foi realizada para verificação de pressupostos dos ruídos brancos, os quais se comportam conforme o esperado.

Após estudo individual de cada variável química, ajustamos modelos bivariados para o vetor  $Y = (Y_1, Y_2)$ , utilizando as proposições de Diggle e Ribeiro Jr. (2006) e de Matheron (1982). Nas duas abordagens foi utilizado  $\mu = X\beta^*$ , sendo  $X$  uma matriz  $134 \times p$  contendo um intercepto para cada campo aleatório e  $p - 2$  possíveis covariáveis e  $\beta^*$  um vetor  $p \times 1$  com os parâmetros de média associados a  $Y$ , para estruturar  $\Sigma_Y$  somente funções de correlação da família de Matérn foram utilizadas, logo para no BGCCM foram utilizadas três funções de correlação de Matérn e para no BCRM duas, todos os parâmetros  $\kappa$  foram considerados fixos em valores similares aos modelos univariados. Sendo assim, com intuito de selecionar o melhor modelo para cada abordagem, algumas combinações de escolhas de  $\kappa$  e  $X$  foram utilizadas. Então em cada caso o vetor de parâmetros  $\theta$  foi estimado por máxima verossimilhança, onde foram utilizadas as mesmas técnicas dos modelos univariados, sendo que no BGCCM definimos 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  $\nu_2 = \sigma_{22}/\sigma_{11}$ , sendo assim, nos dois casos existe forma analítica para os estimadores de  $\beta^*$  e  $\sigma^2$ , que são função dos demais parâmetros, os quais foram estimados maximizando a função de máxima verossimilhança concentrada por meio do método numérico. E novamente, utilizando a propriedade de invariância dos estimadores foi possível encontrar estimativas para todos os parâmetros dos modelos.

Quanto a seleção de modelos, a krigagem e os testes de pressupostos, as mesmas técnicas dos modelos univariados foram adotadas.

Por último, retiramos do conjunto de dados 20 observações da saturação por bases e conduzimos novamente a modelagem univariada desse atributo e as modelagens bivariadas utilizando a informação completa do pH. Sendo que os parâmetros estimados foram utilizados para fazer as krigagens da saturação por bases nas localizações espaciais omitidas. Com essa análise foi possível verificar que a utilização da informação do pH gera previsões com mais qualidade para a saturação por bases nas localizações retiradas.

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 dos atributos sob estudo, 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 forte simetria 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

A figura (1) acusou um possível padrão espacial nas duas variáveis e um acréscimo nos valores do atributos quando as localizações amostradas possuem coordenada oeste-leste mais elevada, o que coincide com a informação da área de manejo do cultivo de soja, a linha no meio dos mapas separa as mesmas. Além disso, foram conduzidas transformações de Box-Cox nos valores dos atributos para encontrar uma transformação dos dados que valide o pressuposto de gaussianidade dos campo aleatórios, os resultados encontrados indicaram que não é necessário fazer transformações nas variáveis.

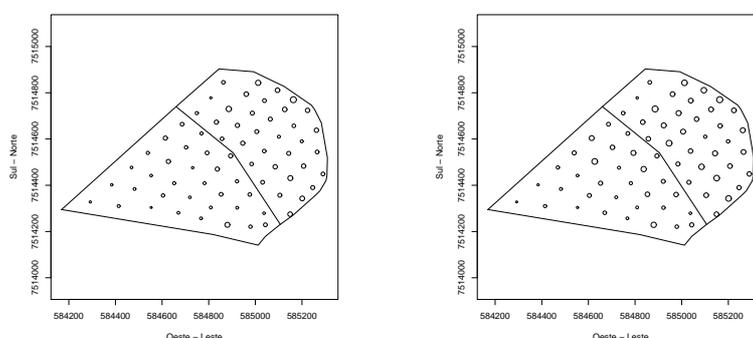


FIGURA 1: Gráficos de círculos - pH a esquerda e saturação por bases a direita

As tabelas 2, 3 e 4 mostram as estimativas de máxima verossimilhança para os parâmetros dos modelos univariados ajustados para a saturação por bases considerando média constante, tendência induzida pela coordenada leste-oeste e tendência induzida pela área de manejo, respectivamente. Utilizando cada abordagem para a média, as diferentes escolhas para  $\kappa$  não produzem

muita diferença nas estimativas dos parâmetros de média e de variabilidade, no entanto o parâmetro  $\phi$  é mais afetado, o que é normal, uma vez que, os parâmetros provenientes da função de correlação Mátern não são ortogonais, fato que gera alcances práticos das correlações parecidos para essas estimativas. Com a inclusão de covariáveis no modelo muda um pouco a relação entre  $\sigma^2$  e  $\tau^2$ , onde os dois valores ficaram bem próximos, além disso, a estimativa  $\phi$  indica que o alcance das correlações é menor.

$\beta$	$\tau^2$	$\sigma^2$	$\phi$	$\kappa$
49,53	47,66	105,33	700,00	0,5
48,53	59,57	120,62	625,58	1
47,98	62,40	124,31	516,78	1,5
47,27	63,29	149,69	489,58	2
47,18	63,51	139,18	400,00	2,5

TABELA 2: EMV - média constante

$\beta_0$	$\beta_1$	$\tau^2$	$\sigma^2$	$\phi$	$\kappa$
-14297,01	0,025	8,98	59,98	54,51	0,5
-14300,66	0,025	27,56	41,46	45,42	1
-14300,92	0,025	33,53	35,53	39,51	1,5
-14300,17	0,025	36,39	32,68	35,33	2
-14298,98	0,025	38,06	31,03	32,19	2,5

TABELA 3: EMV - tendência oeste-leste

$\beta_0$	$\beta_1$	$\tau^2$	$\sigma^2$	$\phi$	$\kappa$
47,91	8,80	24,35	53,68	73,68	0,5
47,92	8,83	38,78	39,28	59,18	1
47,92	8,86	43,35	34,69	50,33	1,5
47,92	8,88	45,53	32,47	44,34	2
47,93	8,89	46,79	31,18	39,98	2,5

TABELA 4: EMV - tendência área de manejo

O modelo final adotado é o que considera a coordenada oeste-leste como covariável e com  $\kappa = 2.5$ , uma vez que, essas estimativas geram o maior máximo estimado do logaritmo da função de verossimilhança concentrada e menor AIC, que foram iguais -234,67 e 479,34, respectivamente. Além disso foi conduzida uma análise de resíduos, utilizando a figura (2) não há evidência contra

os pressupostos do ruído branco, fato corroborado pelo teste de normalidade de Shapiro-Wilk, que gerou um p-valor igual a 0,97, logo, não rejeita-se a hipótese nula de normalidade.

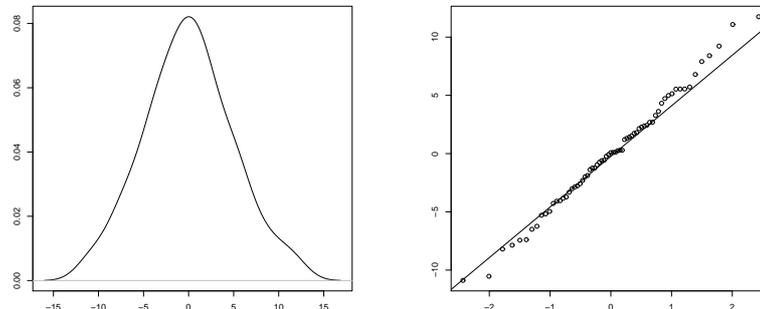


FIGURA 2: qq-plot e densidade dos resíduos

As mesmas técnicas para estimação, seleção de modelo e análise de resíduos foram conduzidas para os valores observados da variável pH, sendo que o modelo final também inclui a coordenada leste-oeste como covariável e possui  $\kappa = 2.5$ . As estimativas foram  $\beta_0 = -608,46$ ,  $\beta_1 = 0,001$ ,  $\tau^2 = 0,015$ ,  $\sigma^2 = 0,097$  e  $\phi = 22,46$ . Com essas estimativas tem-se que para o pH a proporção da variabilidade devido ao ruído branco e o alcance das correlações são menores, além disso, esses valores geram estimativas de -19,35 e 48,69 para o logaritmo da função de verossimilhança concentrada e para o AIC, respectivamente. Com relação aos resíduos, novamente, não houve a evidência suficiente contra as hipóteses feitas.

A figura (3) ilustra os mapas de krigagem, além disso, as variâncias preditivas máximas e mínimas da saturação por bases são iguais a 30,21 e 12,35, respectivamente, já para o pH esses valores são 0,10 e 0,01. Logo, as variâncias preditivas estão num patamar aceitável com relação aos valores preditos.

Com a análise univariada concluída, conduzimos a modelagem bivariada para o problema. Em caráter descritivo foi calculada a correlação de Pearson entre os valores observados dos dois atributos, essa estatística ficou igual a 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, utilizamos as estimativas dos modelos univariados para atribuir valores iniciais ao método numérico de maximização. Sendo que 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

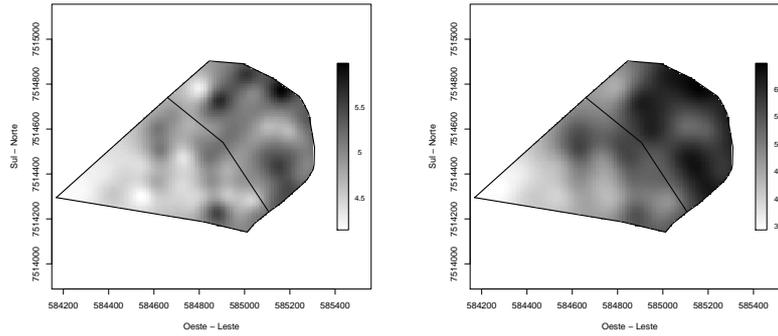


FIGURA 3: Mapas preditivos - pH a esquerda e saturação por bases a direita

correlações geraram diversas estimativas diferentes para os parâmetros dos modelos. Sendo assim, inicialmente, selecionamos as melhores estimativas em cada abordagem para a média. A tabela (5) 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  $\phi_0$ , que para a tendência na área de manejo ficou bem menor que nas outras duas, fato 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.

$\theta$	Constantes	Oeste-leste	Área de manejo
$\beta_{01}$	52,57	$1,54e - 6$	47,83
$\beta_{02}$	4,92	$1,45e - 7$	4,72
$\beta_1$	—	$8,98e - 5$	9,46
$\beta_2$	—	$8,41e - 6$	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
$\sigma_2$	$7e - 5$	$5e - 4$	$1e - 5$
$\phi_0$	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 5: Estimativas de máxima verossimilhança do BGCCM

A tabela (6) mostra os resultados da estimação de máxima verossimilhança utilizando o

BCRM. Para essa modelagem bivariada, com a inclusão da covariável coordenada leste-oeste ocorreu inflacionamento dos parâmetros de variabilidade. Com relação as demais estimativas, os resultados são parecidos com os do BGCCM.

$\theta$	Constantes	Oeste-leste	Área de Manejo
$\beta_{01}$	51,76	$1,51e - 6$	47,93
$\beta_{02}$	4,89	$1,43e - 7$	4,72
$\beta_1$	—	$8,85e - 5$	9,12
$\beta_2$	—	$8,36e - 6$	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

TABELA 6: Estimativas de máxima verossimilhança do BCRM

Nas duas abordagens, o melhor ajuste foi o que considera a área de manejo da propriedade como covariável, os quais geraram AIC igual a 437,35 e 432,96 para o BGCCM e BCRM, respectivamente. Sendo assim, as estimativas finais foram estabelecidas e podemos pensar na krigagem ordinária, para tal, foi utilizado o mesmo grid de pontos utilizados na modelagem univariada.

O processo de krigagem utilizando as estimativas do BGCCM gerou variâncias preditivas relativamente pequenas, sendo que, 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 (4) e (5) ilustram os mapas preditivos utilizando os valores estimados em cada abordagem de modelos bivariados. Não existe muita diferença entre as previsões utilizando modelos univariados ou bivariados, no entanto, como nas abordagens mais complexas foi considerada a covariável área de manejo nos modelos finais, o mapa preditivo tem um salto nos valores, quando passamos de uma região de manejo para outra.

Sendo assim, tem-se que os modelos geoestatísticos bivariados estão ajustando bem os dados observacionais, sendo que o BCRM se mostrou mais eficiente para fazer estimação, onde em todos os casos o AIC é menor com essa abordagem, já do ponto de vista preditivo, o BGCCM apresentou uma menor variabilidade. Finalmente, não existe muita diferença entre os dois modelos tratados nesse estudo, os quais forneceram bons resultados preditivos para o pH e a saturação por bases da fazenda Tupã.

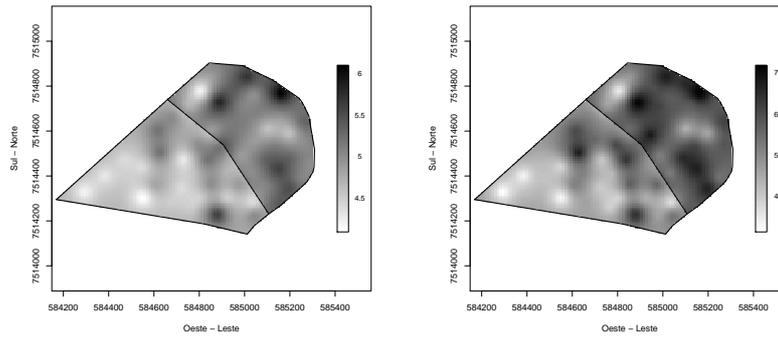


FIGURA 4: Mapas preditivos com BGCCM - pH a esquerda e saturação por bases a direita

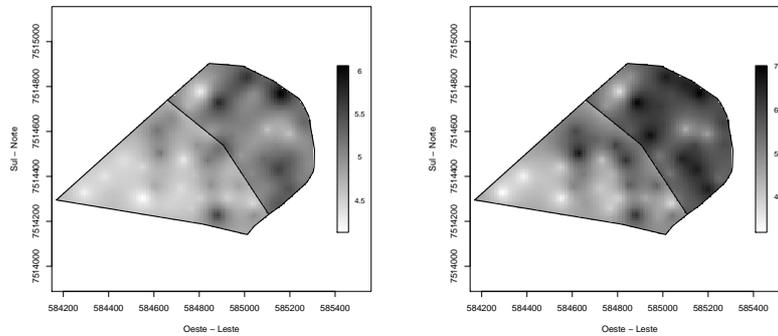


FIGURA 5: Mapas preditivos com BCRM - pH a esquerda e saturação por bases a direita

Omitindo 20 localizações da saturação por bases e estimando novamente, em todas as abordagens, os parâmetros dos modelos finais selecionados, tem-se que não existe muita diferença entre essas estimativas e as com informação completa da saturação por bases. O próximo passo foi fazer a krigagem ordinária da saturação por bases nas 20 localizações retiradas do conjunto de dados, para tal, utilizamos os parâmetros estimados com o modelo univariado, com o BGCCM e com o BCRM. A tabela (7) mostra resultados descritivos das diferenças entre os verdadeiros valores observados e os valores preditos, sendo que, os modelos bivariados geram resultados com melhor qualidade, uma vez que, existe uma menor variabilidade dessas diferenças.

Sendo assim, o objetivo desse estudo foi cumprido, já que foram estruturadas as correlações espaciais cruzadas entre os atributos e além disso, ficou explícito, que a utilização de modelos geoestatísticos bivariados podem gerar mapas preditivos mais precisos para atributos de interesse que sejam correlacionados.

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

TABELA 7: Mediadas descritivas das diferenças entre os observados e as krigagem nos pontos retirados da Saturação por bases

## **AGRADECIMENTOS**

## **LITERATURA CITADA**