

# ESTIMAÇÃO DOS PARÂMETROS DE UMA FUNÇÃO DE COVARIÂNCIA NO ESTUDO DA VARIABILIDADE ESPACIAL DO TEOR DE CÁLCIO NO SOLO

Maria Cristina Neves de OLIVEIRA<sup>1</sup>  
Décio BARBIN<sup>2</sup>  
Enio Fraga da SILVA<sup>3</sup>  
José Ronaldo de MACEDO<sup>3</sup>  
Cláudio Lucas CAPECHE<sup>3</sup>  
Helga Restum HISSA<sup>3</sup>

- RESUMO: Até a década 1970, o problema da variabilidade espacial, de muitas variáveis agronômicas em experimentos de campo, foi solucionado com o uso dos delineamentos em blocos ao acaso associados aos princípios da aleatorização e da repetição. Embora este método seja eficiente para avaliar a variabilidade experimental não considera a posição espacial entre as observações. Quando estas variáveis apresentam aspectos aleatórios e espaciais, é necessário avaliar a dependência espacial, pelas ferramentas geoestatísticas usando estimadores não viciados. A proposta deste trabalho foi avaliar a dependência espacial do teor de cálcio no solo utilizando uma estrutura de covariância exponencial e as estimativas dos parâmetros obtidas por máxima verossimilhança e máxima verossimilhança restrita.
- PALAVRAS-CHAVE: Geoestatística; dependência espacial; máxima verossimilhança; máxima verossimilhança restrita; estrutura de covariância exponencial.

## 1 Introdução

O estudo da variabilidade espacial de muitas variáveis agronômicas em experimentação de campo é antigo e a solução para este problema foi o uso dos delineamentos em blocos ao acaso associados aos princípios da aleatorização e da repetição (Fisher, 1935). Este método foi criticado quando ignorada a posição espacial entre as observações experimentais (Vieira, Neilsen e Biggar, 1981; Reichardt, Vieira e Libardi, 1986), mas é tão eficiente quanto as ferramentas geoestatísticas, se considerada a estrutura de correlação espacial, obtendo-se inferências mais precisas para os efeitos de tratamentos, no caso de experimento de campo (Grondona e Cressie, 1991). O estudo geoestatístico fundamenta-se na teoria das variáveis regionalizadas, proposta por (Matheron, 1963) e a ferramenta utilizada para identificar a

<sup>1</sup>EMBRAPA - Pesquisa & Desenvolvimento - Embrapa Soja, CEP 86001-970, Londrina, PR, Brasil, E-mail: [mcneves@cnpso.embrapa.br](mailto:mcneves@cnpso.embrapa.br)

<sup>2</sup>Departamento de Ciências Exatas, Universidade de São Paulo, campus de Piracicaba – ESALQ/USP, CEP 13418-900, Piracicaba, SP, Brasil, E-mail: [debarbin@esalq.usp.br](mailto:debarbin@esalq.usp.br)

<sup>3</sup>EMBRAPA – Pesquisa & Desenvolvimento - Embrapa Solos, CEP 22460-000, Rio de Janeiro - RJ, Brasil, E-mail: [enio@cnpso.embrapa.br](mailto:enio@cnpso.embrapa.br)

dependência é o variograma. Dois aspectos são apresentados quando se avalia a continuidade espacial, o primeiro, considera a modelagem da dependência espacial sob suposições Gaussianas obtendo-se estimativas paramétricas por meio dos estimadores não-viesados como os de máxima verossimilhança e inferência Bayesiana (Diggle e Ribeiro Junior, 2000, 2002; Ploner e Dutter, 2000). O segundo é o não-paramétrico, que independe de métodos rígidos, e que se baseia na análise visual para obter os parâmetros de um modelo. Neste método, a inferência somente será possível se forem assumidas as hipóteses de estacionariedade ou a hipótese intrínseca (Sad e Quade, 1986). O método empírico utilizado no estudo da dependência espacial é limitado devido a sua subjetividade e por ser pouco rigoroso nas estimativas dos parâmetros (Ribeiro Junior, 1995). O variograma, utilizado para identificar a dependência espacial é uma função que descreve o valor da variância entre duas variáveis regionalizadas em função da distância  $\mathbf{h}$  e seu estimador, proposto por Matheron (1963), é dado por:

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{i=1}^{N(\mathbf{h})} (z(x_i) - z(x_j))^2, \quad \mathbf{h} \in \mathfrak{R}^d,$$

em que,  $\mathbf{h}$  é um vetor,  $|N(\mathbf{h})|$  é o número de pares de amostras  $[Z(x_i), Z(x_j)]$  separados pelo vetor  $\mathbf{h}$  elementos distintos de  $N(\mathbf{h})$

$$e N(\mathbf{h}) \equiv [(x_i, x_j) : x_i - x_j = \mathbf{h} \quad i, j = 1, \dots, n].$$

Pode-se também utilizar um estimador robusto se existem valores discrepantes ou “outliers”, (maiores detalhes são encontrados em Cressie, 1993). Uma vez identificada a dependência espacial pelo variograma o objetivo da análise espacial é modelar esta dependência a fim de fazer predições sobre os dados. Sendo assim, foi utilizado o interpolador de *krigagem* ordinária, que possibilitou obter a partir de observações pontuais, informações para toda região. Este viabiliza a predição e os erros-padrões da predição para estes pontos e não requer o prévio conhecimento da média. Os objetivos gerais deste trabalho foram: a) determinar os valores iniciais dos parâmetros dos variogramas teóricos pelo método não-paramétrico; b) estudar o modelo exponencial na identificação da dependência espacial do cálcio, permitindo assim, melhor interpretação dos fatores que interferem na produção das culturas; c) obter as estimativas dos parâmetros do modelo exponencial, por máxima verossimilhança e máxima verossimilhança restrita e, d) obter os mapas das predições e das variâncias das predições que melhor representem a real distribuição espacial do cálcio no solo.

## 2 Material e métodos

Os dados experimentais, que fazem parte deste trabalho, foram coletados do levantamento detalhado de solos da Estação Experimental de Campos (Capeche et al., 1997), por pesquisadores da Pesagro e Embrapa Solos. Neste estudo foi considerado o teor de cálcio ( $\text{mmol}_c\text{dm}^{-3}$ ), na camada de 0-20cm. O objetivo principal do levantamento foi o planejamento adequado do uso dos solos possibilitando um manejo racional e sustentável, evitando o processo de erosão. As amostras de solo foram coletadas com um trado tipo holandês, em uma malha regular, sob as coordenadas leste e norte, distando nestas direções 900 metros, espaçadas de 50 metros, totalizando 178 amostras, na camada de 0-20cm. A área

experimental foi dividida em três regiões em função dos diferentes períodos de adubação realizados conforme indicados na Figura 1. Definiu-se como região 1 aquela em que a adubação do solo foi realizada no primeiro período e regiões 2 e 3 para o segundo e terceiro períodos.

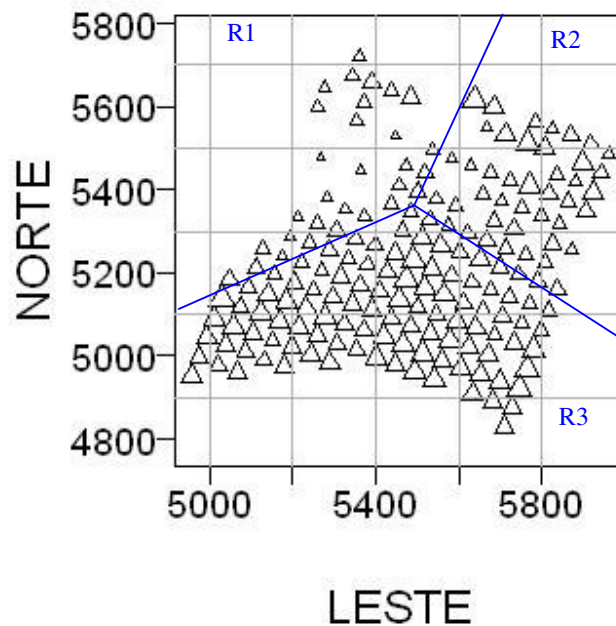


Figura 1 - Mapa de amostragem em malha regular confiável para mensurações do cálcio ( $\text{mmolcdm}^{-3}$ ), em três regiões, na camada de 0-20 cm do solo.

Nos solos brasileiros é grande a variabilidade nos teores de micronutrientes devido as manchas de fertilidade e, nos experimentos de campo este fato não permite detectar diferenças estatísticas entre os tratamentos. A solução encontrada para detectar tal variabilidade e a existência de dependência espacial foi avaliar estes nutrientes considerando a posição espacial de cada amostra. Considerou-se a estrutura de covariância exponencial para melhor identificar e interpretar a dependência espacial do cálcio no solo, incluindo também a covariável região e tendência linear nas coordenadas.

## 2.1 Descrição dos modelos

Considerando-se que as observações medidas para o teor de cálcio nos diferentes pontos amostrais apresentam dependência espacial e não existem repetições em cada ponto para prever o que acontece na região toda, pode-se pensar num processo de pares de distâncias, obtendo-se com isso as repetições necessárias para a predição em pontos não amostrados. Assim, tem-se um vetor multidimensional de observações correlacionadas em duas dimensões espaciais, as quais foram analisadas por um modelo linear misto, com a estrutura de covariância exponencial, o qual pode ser representado na forma matricial

como  $\mathbf{Y} = \mathbf{XB} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$  em que:  $\mathbf{Y} = (Y_1, \dots, Y_k)$  é o vetor dos valores da variável resposta nas posições  $(x_1, \dots, x_k)$ ;  $\boldsymbol{\beta}$  é o vetor de parâmetros desconhecidos associados aos efeitos fixos;  $\boldsymbol{\gamma}$  é o vetor de parâmetros desconhecidos associados aos efeitos aleatórios e,  $\boldsymbol{\varepsilon}$  é o vetor dos erros independentes e normalmente distribuídos.

### 2.1.1 Modelo com a covariável região

Neste modelo foi incluído o efeito de regiões como covariável e, desta forma, a expressão para o modelo misto é apresentado por:

$$Y_i = \beta_0 + \beta_1 \text{região2}(x_i) + \beta_2 \text{região3}(x_i) + S(x_i) + \varepsilon_i, i = 1, \dots, n \quad (1)$$

em que:  $Y_i$  é valor da variável resposta na posição  $x_i$ ;  $\beta_0$  é uma constante;  $\beta_1$  e  $\beta_2$  são os parâmetros de regressão desconhecidos associados à covariável região sendo:  $(\beta_1 + \beta_2)$  o efeito da região 1,  $\beta_1$  o efeito da região 2 e  $\beta_2$  o efeito da região 3;  $S(x_i)$  é um processo estacionário Gaussiano com variância  $\sigma^2$  e função de correlação  $\rho(\mathbf{h}; \phi)$  em que,  $\phi =$  alcance  $\mathbf{h} = \|x_i - x_j\|$  e,  $\varepsilon_i$  é o ruído ou erro, mutuamente independentes,  $\varepsilon_i \sim N(0, \tau^2)$ .

### 2.1.2 Modelo com a covariável região e tendência linear nas coordenadas

Neste modelo além da covariável região foi incluída a tendência linear nas coordenadas Leste e Norte, uma vez que na área experimental estudada foi observada uma dependência espacial mais acentuada na direção Norte. Sendo assim, o modelo foi expresso como segue:

$$Y_i = \beta_0 + \beta_1 \text{região2}(x_i) + \beta_2 \text{região3}(x_i) + \beta_3 \text{leste}(x_i) + \beta_4 \text{norte}(x_i) + S(x_i) + \varepsilon_i, i = 1, \dots, n \quad (2)$$

em que, além dos efeitos considerados na eq. (1) tem-se:  $\beta_3$  é o parâmetro de regressão desconhecido associado a coordenada Leste e  $\beta_4$  é o parâmetro de regressão desconhecido associado a coordenada norte. Esses modelos podem ser escritos na forma matricial e expressos por:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{S}(\mathbf{x}) + \boldsymbol{\varepsilon} \quad (3)$$

em que:  $\mathbf{Y} = (Y_1, \dots, Y_k)$  é o vetor dos valores da variável resposta nas posições  $(x_1, \dots, x_k)$ ;  $\boldsymbol{\beta}$  é o vetor de parâmetros desconhecidos associados aos efeitos fixos;  $S(x)$  é um processo estocástico estacionário e,  $\boldsymbol{\varepsilon}$  é o vetor dos erros independentes e normalmente distribuídos. Para  $\mathbf{Y}$  é assumida distribuição normal multivariada e pode ser indicada por:  $MVN \sim (\mathbf{XB}, \sigma^2 \mathbf{R}_\phi + \tau^2 \mathbf{I})$  em que,  $\mathbf{I}$  é uma matriz identidade de dimensão  $n \times n$ ,  $\mathbf{R}_\phi$  é a matriz de variâncias e covariâncias de dimensão  $n \times n$  com  $(i, j)$ -ésimo elemento  $\rho(h_{ij})$  com  $h_{ij} = \|x_i - x_j\|$  a distância Euclidiana entre  $x_i$  e  $x_j$ .

Para a estrutura de covariância exponencial (Diggle e Ribeiro Junior, 2000), a função de correlação  $\rho(\mathbf{h})$  é expressa por:

$$\rho(\mathbf{h}, \phi) = \exp\left(-\frac{\mathbf{h}}{\phi}\right) \quad (4)$$

## 2.2 Métodos de estimação dos parâmetros

A estimativa dos parâmetros para os modelos com diferentes covariáveis foi realizada pelos estimadores de máxima verossimilhança e máxima verossimilhança restrita, utilizando os valores iniciais obtidos pelo método de ajuste informal subjetivo. Estas estimativas foram obtidas sob a suposição Gaussiana para o vetor  $\mathbf{Y}$  conforme apresentado no item 2.1.

### 2.2.1 Máxima verossimilhança

Considerando os modelos (1 e 2) e a estrutura de covariância exponencial e assumindo que  $\mathbf{Y} = (Y_1, \dots, Y_k)$  segue uma distribuição multivariada Gaussiana a função de densidade de probabilidades de  $\mathbf{Y}$ , a função de verossimilhança e o logaritmo da função de verossimilhança  $L(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y})$  são expressos por:

$$f(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\theta}) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right) \quad (5)$$

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y}) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right) \quad (6)$$

$$l(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y}) = \ln L(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y}) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (7)$$

em que:  $\mathbf{y}$  é um vetor de observações experimentais;  $\boldsymbol{\beta}$  é o vetor de parâmetros desconhecidos associados aos efeitos fixos;  $\boldsymbol{\theta}$  são os parâmetros da matriz de covariância com  $\boldsymbol{\theta} = (\tau^2, \sigma^2, \phi)$  se a estrutura de covariância é a exponencial; T é a matriz transposta e,  $|\boldsymbol{\Sigma}|$  é o determinante da matriz de variâncias e covariâncias com dimensão  $n \times n$ , das  $n$  observações experimentais (Hartley e Rao, 1967; Patterson e Thompson, 1971; Kitanidis e Shen, 1996; Xavier, 2000; Diggle e Ribeiro Junior, 2000). As estimativas dos parâmetros de  $\boldsymbol{\theta}$  são obtidas maximizando-se  $l(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}^2; \mathbf{y})$  e o vetor de estimativas de máxima verossimilhança pelo procedimento de maximização numérica pela subrotina geoR (Pardo-Igúzquiza, 1997; Ribeiro Junior e Diggle, 2001).

### 2.2.2 Máxima verossimilhança restrita

O estimador de máxima verossimilhança restrita ou residual tem sido recomendado nos estudos geoestatísticos devido a estimação dos componentes de variância serem realizadas para os efeitos fixos e aleatórios, independentemente (Patterson e Thompson, 1971; Martins et al., 1993; Xavier, 2000 e Diggle e Ribeiro Junior, 2000). Além disso, utiliza-se as transformações lineares  $\mathbf{Y}^* = \mathbf{P}\mathbf{Y}$ , um projetor ortogonal dado por:  $\mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  de forma que este estimador para  $\boldsymbol{\theta}$  pode ser calculado por maximizar a seguinte expressão:

$$L^*(\boldsymbol{\theta}) = -\frac{n-p}{2} \ln(2\pi) + \frac{1}{2} \ln(\mathbf{X}^T \mathbf{X}) - \frac{1}{2} \left( \ln|G(\boldsymbol{\theta})| + \ln \left| \mathbf{X}^T [G(\boldsymbol{\theta})]^{-1} \mathbf{X} \right| + \left( \mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}} \right)^T [G(\boldsymbol{\theta})]^{-1} \left( \mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}} \right) \right) \quad (8)$$

em que,  $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$  e  $G(\boldsymbol{\theta}) = \mathbf{I}\tau^2 + \sigma^2\mathbf{R}(\phi)$ .

### 2.2.3 Perfil de verossimilhança

Na investigação da variabilidade dos parâmetros dos modelos deste estudo foi usado o perfil de verossimilhança somente sob um conjunto deles, pois a dimensão da superfície de verossimilhança não permite inspeção direta. A expressão do perfil de verossimilhança utilizada foi  $L_p(\beta) = L[\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\beta}})] = \max_{\boldsymbol{\theta}} [L(\boldsymbol{\beta}, \boldsymbol{\theta})]$  (Diggle e Ribeiro Junior, 2000).

### 2.2.4 Critério de informação de Akaike e teste da razão de verossimilhanças na comparação de modelos com dependência espacial

Foi utilizado, para a seleção dos modelos, o Critério de Informação de Akaike (AIC) conforme apresentado por Xavier (2000), o qual é baseado na teoria de decisão e definido por:

$$AIC = -2l + 2p \quad (9)$$

em que:  $l$  é o ln da função de verossimilhança e  $p$  é o número de parâmetros do modelo considerado. Por este critério, o melhor modelo é aquele que apresenta o menor valor de AIC. Da mesma forma que o AIC, o teste da razão de verossimilhanças permite comparar modelos com parâmetros estimados pelo mesmo estimador (Pinheiro e Bates, 2000; Xavier, 2000). Neste último, considera-se a diferença entre os números de parâmetros dos modelos a serem comparados pelo valor crítico do  $\chi^2_{(2,0,05)}$  e a estatística do teste da razão de verossimilhanças é sempre positiva como a seguir:

$$2 \log(L_2/L_1) = 2[\log(L_2) - \log(L_1)] \quad (10)$$

O modelo adequado será aquele em que o resultado do teste for menor que o valor do que  $\chi^2_{(2,0,05)}$  tabelado.

### 2.2.5 Método da validação cruzada

Avaliou-se pelo método subjetivo da validação cruzada se o fenômeno foi modelado adequadamente, de maneira que sejam obtidas informações fidedignas da amostra. Este método consiste em suprimir do conjunto de dados, alguns valores e estimá-los a partir dos remanescentes. Calcula-se o erro pela diferença entre o valor predito e o medido ( $\hat{Z}_i - Z_i$ ) comparando com o desvio ( $\hat{Z}_i - Z_i$ ) /  $\sigma_i$ . Calcula-se pelo método da *krigagem*, o valor estimado do ponto  $x_i$ , empregando somente os dados restantes, seja  $\hat{Z}_i$ . Na validação cruzada verifica-se se os erros obtidos pela krigagem têm média nula e variância unitária (Samper e Ramirez, 1990; Ribeiro Junior, 1995 e Gonçalves, 1997).

### 2.2.6 Método de análise estatística

Os métodos de análise estatística usados estão apresentados nos itens 2.1 e 2.2 e os programas que geraram as estatísticas e gráficos foram: GeoEAS (Englund e Sparks, 1988), SAS versão 8.0, a subrotina geoR versão 3.7 (Diggle e Ribeiro Junior, 2000; Ribeiro Junior e Diggle, 2001), e o programa Statistica).

## 3 Resultados e discussão

### 3.1 Análise descritiva

Os resultados da Tabela 1 e das Figuras 2a e 2b indicam leve assimetria, os valores para média e mediana são próximos e o coeficiente de variação foi 21,86%. Os valores para os coeficientes de assimetria e curtose estão próximos de zero.

Tabela 1 - Resultados das estatísticas descritivas para a variável teor de cálcio ( $\text{mmol}_c\text{dm}^{-3}$ ), na camada de 0-20 cm do solo

Estatísticas	Cálcio 0-20cm
Número de observações	178
Média	50,67
Mediana	50,50
Mínimo	21,00
Máximo	78,00
Amplitude total	57,00
Amplitude interquartílica	15,00
Variância	122,77
Coefficiente de assimetria	-0,098
Coefficiente de curtose	-0,33
Desvio padrão	11,08
C.V.(%)	21,86

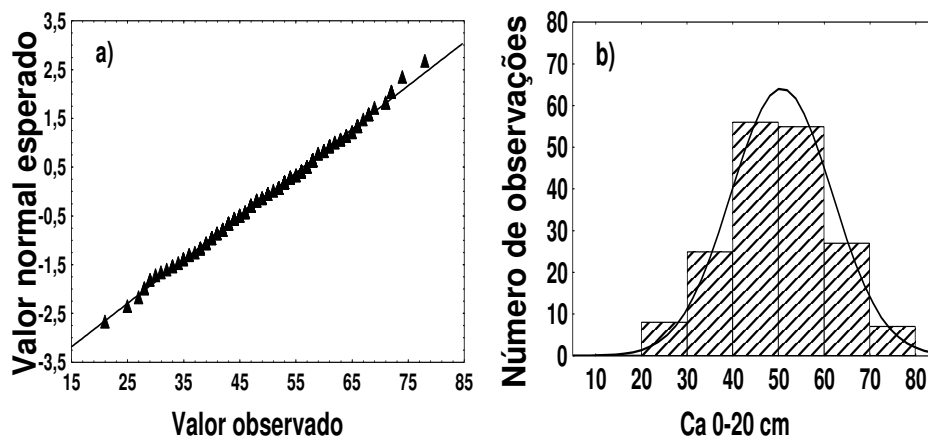


Figura 2 - Gráficos da probabilidade normal (a), distribuição de freqüências sob a curva normal (b), para o teor de cálcio ( $\text{mmol}_c\text{dm}^{-3}$ ) na camada de 0-20 cm do solo.

Os resultados gráficos utilizados para a verificação da normalidade foram coincidentes com aqueles obtidos pelo teste de Shapiro-Wilk, conforme os valores da estatística  $W=0,9946$  e  $p\text{-valor} < 0,70$ . Sendo assim, não foi rejeitada a hipótese de nulidade e os teores de cálcio podem ser estudados pela distribuição normal. Este teste de normalidade é considerado mais rigoroso devido à sensibilidade na presença de dados repetidos e pode ser utilizado para testar qualquer distribuição de probabilidade (Ortiz, 2002).

### 3.2 Análise exploratória espacial

Com os resultados da análise descritiva (Tabela 1) obteve-se uma idéia geral do comportamento da dispersão dos dados de cálcio, mas estes resultados são penalizados pois não estão levando em conta a posição geográfica. A média e a variância não foram suficientes para medir a continuidade espacial ou avaliar as hipóteses de estacionaridade, mas sabe-se, pela literatura, que na presença de tendência não existe estacionaridade (Câmara et al., 2001).

Observou-se que existe tendência linear da dependência espacial do cálcio na direção norte na camada 0-20 cm e, na direção Leste, os dados apresentaram-se mais dispersos (Figura 3a, 3b). Os valores para os coeficientes de correlação entre o teor de cálcio e coordenadas Norte e Leste foram respectivamente  $r = -0,519$  e  $r = -0,488$  e significativos ( $p \leq 0,05$ ). Tais informações permitiram afirmar que houve tendência para a aceitação da estacionaridade, assim sendo esta tendência deve ser inserida no modelo a ser estudado, para a variável resposta cálcio.



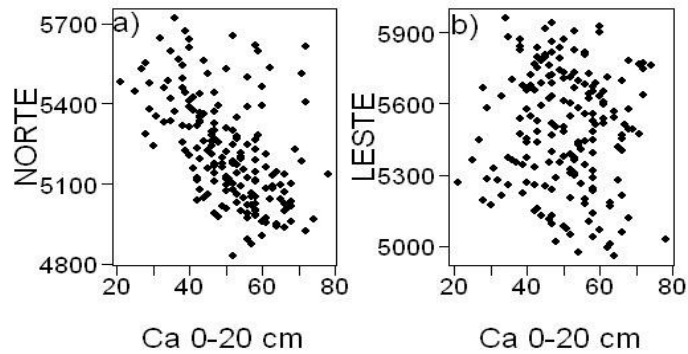


Figura 3 - Perfil do teor de cálcio na direção Norte (a), e na direção Leste (b), para a camada 0-20 cm do solo.

Este diagnóstico estatístico descritivo e espacial associado aos conhecimentos agrônômicos da região deu subsídios para selecionar quais variáveis explanatórias deveriam ser implementadas no modelo a ser estudado, visando melhor interpretação da fertilidade do solo e deficiência de alguma propriedade química.

### 3.3 Análise geostatística para modelos com covariável e estrutura de covariância exponencial

Se existe correlação e tendência no teor de cálcio é importante caracterizar a dependência espacial, e assim, fundamentando-se nos resultados das análises exploratórias foram estudados modelos que representaram a dependência espacial para esta variável resposta. Foram considerados os modelos com a covariável região e com tendência linear nas coordenadas Leste e Norte na estrutura de covariância exponencial.

Verificou-se também se algum possível valor discrepante influenciou a resposta do cálcio e, métodos exploratórios como variogramas empíricos de nuvens e pontuais foram realizados. O primeiro, pelo método clássico de Matheron (1963) e, o segundo, o variograma robusto, proposto por Cressie (1993) (Figura 4 a, b, c e d).

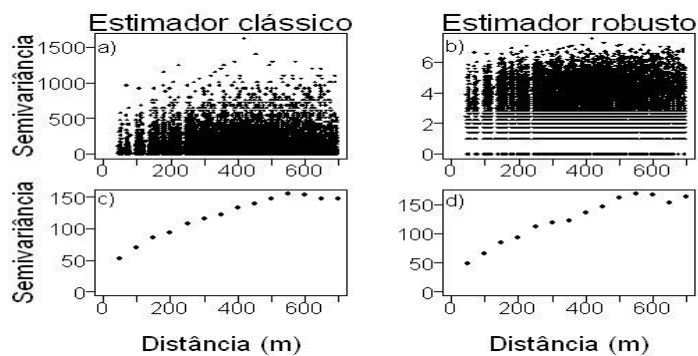


Figura 4 - Variogramas empíricos de nuvens (a, b) e pontuais (c, d), para o teor de cálcio na camada 0-20 cm do solo.

Embora existam valores que distanciam da massa de dados, não devem ser considerados *outliers*, pois são valores repetidos e muito pequenos, oriundos da região em que a adubação não foi recente. Mesmo com variogramas empíricos erráticos em função dos diferentes números de pares que geram cada ponto do variograma, verificou-se a existência de dependência espacial para o cálcio, mas foi necessário obter variogramas baseados em modelos e seus parâmetros obtidos por máxima verossimilhança. Os valores iniciais dos parâmetros do modelo exponencial para o efeito pepita, patamar e alcance foram respectivamente  $\tau^2=15$ ,  $\sigma^2=160$  e  $\phi=550$ .

Os valores das estimativas do efeito pepita e do alcance, para os dois estimadores, foram sempre menores quando utilizado o modelo considerando os efeitos de regiões e coordenadas (modelo 2) e, no caso das estimativas da variância espacial  $\sigma^2$ , os valores foram similares (Tabela 2).

Tabela 2 - Estimativas dos parâmetros, AIC e logaritmo da função verossimilhança para o teor de cálcio, com estrutura de covariância exponencial utilizando-se os estimadores de máxima verossimilhança (ML) e máxima verossimilhança restrita (REML)

Estimadores	ML		REML	
	Região	Região + Coordenadas	Região	Região + Coordenadas
Efeito pepita ( $\tau^2$ )	3,26	0,00	13,07	7,35
Patamar ( $\sigma^2$ )	98,70	94,77	104,52	108,70
Alcance ( $\phi$ )	72,61	63,02	104,19	92,57
AIC	1270,65	1271,56	1246,92	1233,06
Verossimilhança	-629,33	-627,78	-617,46	-608,53
Número de Parâmetros	6	8	6	8

Os resultados para o teste da razão de verossimilhanças quando compararam-se os modelos com os efeitos de regiões e coordenadas e com menor número de parâmetros foi 3,1 inferior ao valor crítico de  $\chi^2_{(2,0,05)} = 5,99$ . Sendo assim, ao nível de 5%, conclui-se que não houve diferenças estatísticas significativas entre os modelos. Esses resultados foram similares aos obtidos pelo AIC com os dois modelos, quando os parâmetros foram estimados por máxima verossimilhança. Indicou ainda, que não existe forte preferência entre os dois modelos, mas existe evidência agrônômica de que o teor de cálcio foi influenciado pela adubação em períodos distintos e pelas coordenadas geográficas, desse modo, o modelo mais adequado foi o 2, que inclui a covariável região e tendência linear das coordenadas.

Com o estimador de máxima verossimilhança restrita obtiveram-se estimativas da variância do efeito pepita ( $\tau^2$ ) superiores às encontradas por máxima verossimilhança e, no modelo 2, esta estimativa reduziu-se a um pouco mais que a metade. Os valores das estimativas do patamar e alcance foram próximos para os dois modelos com os diferentes

estimadores (Tabela 2). O valor do teste da razão de verossimilhanças de 17,86 foi superior ao valor crítico de  $\chi^2_{(2,0,05)} = 5,99$ .

O valor do AIC foi menor utilizando-se o estimador de máxima verossimilhança restrita. Com estes resultados conclui-se, ao nível de 5%, que o modelo (2) foi o selecionado para estudar a dependência espacial do cálcio.

Os variogramas ajustados e experimentais para os modelos com covariáveis e estimados por máxima verossimilhança e máxima verossimilhança restrita encontram-se na Figura 5.

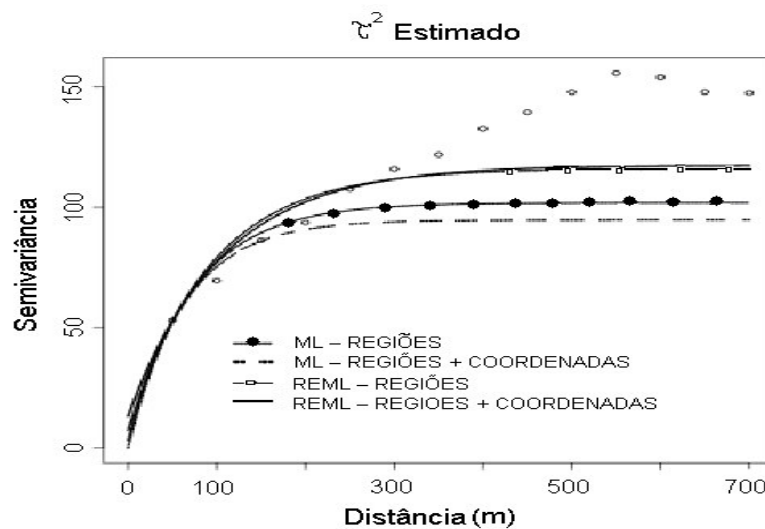


Figura 5 - Variogramas ajustados para os modelos com a covariável região e região mais a tendência linear nas coordenadas para o teor de cálcio ( $\text{mmolc.dm}^3$ ) na camada de 0-20 cm do solo.

Com os resultados obtidos (Tabela 2) é possível em amostragens futuras assegurar maior precisão das pesquisas tomando como base os alcances obtidos neste estudo, bem como o modelo que melhor represente a distribuição espacial do cálcio, como sugerido por Souza et al. (2001).

Foi inspecionada a incerteza das estimativas dos parâmetros da estrutura de covariância pelos perfis de verossimilhança para os modelos conforme as equações (1) e (2), estimados por máxima verossimilhança e máxima verossimilhança restrita, para o cálcio conforme apresentadas nas Figura 6a e b. O perfil para  $\sigma^2$  (patamar) foi levemente assimétrico, o intervalo de confiança para o alcance ( $\phi$ ) foi amplo, apresentando maior variabilidade quando utilizado o estimador de máxima verossimilhança ao contrário dos resultados obtidos pelo segundo estimador. Ainda neste gráfico as duas linhas horizontais indicaram os níveis de significância a 10% e 5% para a distribuição de  $\chi^2$  (Figura 6).

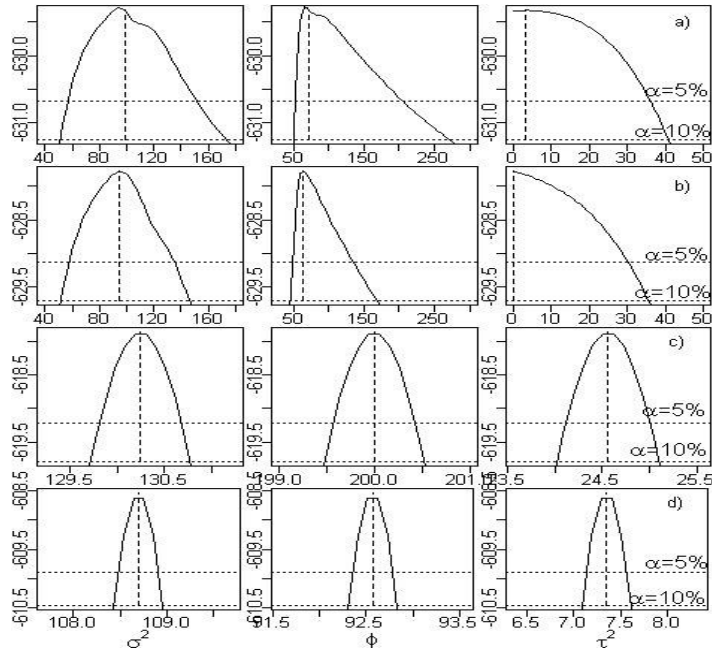


Figura 6 - Perfil de verossimilhança para os parâmetros do modelo exponencial com o teor de cálcio (0-20 cm). Parâmetros estimados por máxima verossimilhança com os modelos com covariável região (a), e, com a covariável região mais a tendência linear nas coordenadas (b). Parâmetros estimados por máxima verossimilhança restrita com covariável região (c) e região mais tendência linear nas coordenadas (d).

As predições e variâncias das predições para o teor de cálcio estão apresentadas nas Figura 7a e b, respectivamente para os modelos com a covariável região e com tendência linear nas coordenadas com as estimativas dos parâmetros obtidos por máxima verossimilhança. Quando incluídas aos modelos as covariáveis regiões e coordenadas Leste e Norte pode-se observar que o padrão da predição mudou e, apresentou pequenas diferenças variando um pouco mais as cores de tons mais fortes para os maiores valores de cálcio e os menores para os tons mais fracos (Figura 7a e c). A variância da predição apresentou boa confiabilidade, entretanto, nas bordas, por ausência de informações esta estatística é maior, quando associada com as cores mais fortes (Figura 7b e d). Uma das justificativas é que as amostras foram realizadas numa malha regular de forma transversal de acordo com as variações topográficas do terreno no sentido Norte-Nordeste e não no sentido Norte-Sul das coordenadas geográficas (Figura 7b e d). Observou-se que as variâncias estão muito próximas e que nos pontos onde as cores são mais claras elas são menores.

A predição obtida utilizando-se os parâmetros ajustados por máxima verossimilhança restrita, com os modelos com covariáveis apresentou regiões mais bem definidas nos locais onde encontram-se os maiores valores de cálcio (Figura 8a e b). As variâncias das predições foram menores quando comparadas aos resultados obtidos por máxima verossimilhança e permitirão que sejam identificadas corretamente as regiões onde existe maior deficiência do cálcio e reduzindo os custos na instalação de experimentos de campo.

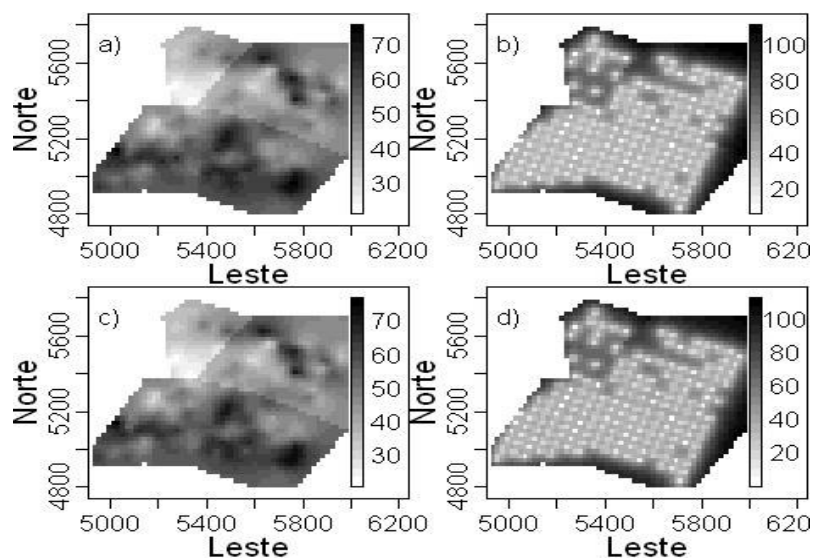


Figura 7 - Resultados da predição (a e c) e variância da predição (b e d), para o teor de cálcio (0-20cm). Modelo com a covariável região (a e b) e região mais tendência linear nas coordenadas (c e d). Parâmetros estimados por máxima verossimilhança e com covariância exponencial.

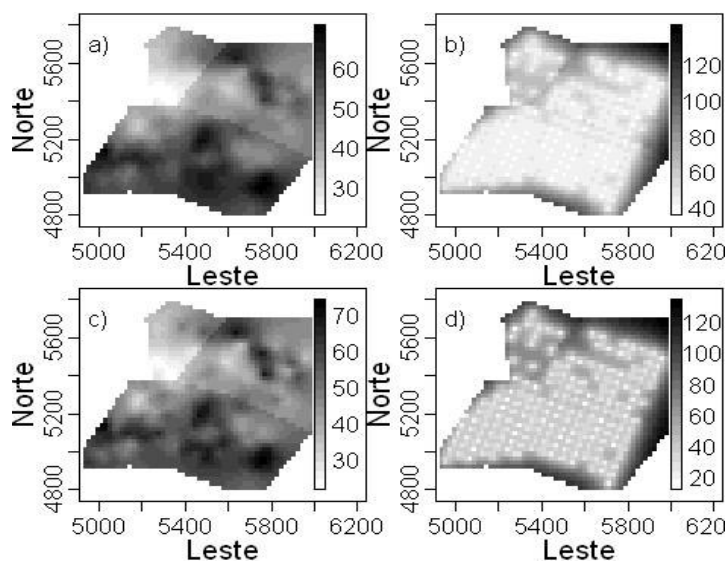


Figura 8 - Resultados da predição (a) e (c) e variância da predição (b) e (d), para o teor de cálcio (0-20 cm). Modelo com a covariável região (a) e (b) e região mais tendência linear nas coordenadas (c) e (d). Parâmetros estimados por máxima verossimilhança restrita e com covariância exponencial.

## Conclusões

Pelas metodologias empregadas neste trabalho e os resultados obtidos, pode-se concluir que: a) foi possível identificar a dependência espacial da variável cálcio no solo; b) as análises exploratórias descritivas e espaciais foram importantes na identificação e escolha do modelo que melhor interpretou a distribuição espacial do cálcio; c) o melhor modelo ajustado foi o modelo (2); d) com o estimador de máxima verossimilhança restrita obtiveram-se as menores variâncias de predições na região estudada; e) o estimador de máxima verossimilhança não foi eficiente para detectar diferenças entre os modelos 1 e 2 e, f) pelos métodos aplicados considerando a posição geográfica foi possível indicar com precisão os locais da distribuição do calcário, dando subsídios para instalação de futuros experimentos com custos reduzidos.

OLIVEIRA, M. C. N. de; BARBIN, D.; SILVA, E. F. da; MACEDO, J. R.; CAPECHE, C. L.; HISSA, H. R. Parameter estimation of a covariance function in the study of spatial variability of calcium content in the soil. *Rev. Mat. Est.*, São Paulo, v.24, n.4, p.117-132, 2006.

- **ABSTRACT:** *Until the 1970s, the problem of spatial variability of many agronomic variables in field experiments was solved by using randomized block designs associated with the randomization and repetition principles. Although this method is efficient to evaluate experimental variability, it does not consider the spatial position among observations. When random and spatial aspects are presented for these variables, an evaluation of spatial dependence is required and geostatistical tools considering unbiased estimators can do it. The purpose of this paper was to evaluate the spatial dependence of calcium content using the exponential covariance function. The parameter estimates were obtained by maximum likelihood and restricted maximum likelihood.*
- **KEYWORDS:** *Geostatistics; spatial dependence; variogram; maximum likelihood; restricted maximum likelihood; exponential covariance function; calcium in the soil.*

## Referências

- CÂMARA, G. et al. *Análise espacial de dados geográficos*. 2. ed. São José dos Campos: INPE, 2001.182 p.
- CAPECHE, C. L. et al. Caracterização pedológica da fazenda Angra - PESAGRO/RIO - Estação experimental de Campos (RJ). In: CONGRESSO BRASILEIRO DE CIÊNCIA DO SOLO. 26., 1997, Rio de Janeiro. *Resumos...* Rio de Janeiro: Sociedade Brasileira de Ciência do Solo, 1997. p.317. CD-ROM
- CRESSIE, N. A. C. *Statistics for spatial data*. New York: John Wiley, 1993. 900p.
- DIGGLE, P. J.; RIBEIRO JUNIOR, P. J. Model based geostatistics. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 14., 2000. Caxambu. *Minicurso...* Caxambu: SINAPE, 2000. 129p.
- DIGGLE, P. J.; RIBEIRO JUNIOR, P. J. Bayesian inference in Gaussian model-based geostatistics. *Geogr. Environ. Modell.*, Abingdon, v.6, n. 2, p.131-148, 2002.
- ENGLUND, E.; SPARKS, A. *Geo-EAS: Geostatistical environmental assessment software*. Las Vegas: U.S. Environmental Protection Agency, 1988. Não paginado. (EPA/600/4-88/033a).

- FISHER, R. A. *The design of experiments*. Edinburgh: Oliver and Boyd, 1935. 252p.
- GONÇALVES, A. C. A. *Variabilidade espacial de propriedades físicas do solo para fins de manejo da irrigação*. 1997. 118f. Tese (Doutorado em Agronomia) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1997.
- GRONDONA, M. O.; CRESSIE, N. Using spatial considerations in the analysis of experiments. *Technometrics*, Washington, v.33, n.4, p.381-391, 1991.
- HARTLEY, H. O.; RAO, J. N. K. Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika*, London, v.54, p.93-108, 1967.
- KITANIDIS, P. K.; SHEN, K.F. Geostatistical interpolation of chemical concentration. *Adv. Water Res.*, Southampton, v.19, n.6, p.369-78, 1996.
- MARTINS, E. N. et al. *Modelo linear misto*. Viçosa: Universidade Federal de Viçosa, Departamento de Zootecnia, 1993. 46p.
- MATHERON, G. Principles of geostatistics. *Econ. Geol.*, Lancaster, v.58, p.1246-1266, 1963.
- ORTIZ, G. C. *Aplicação de métodos geoestatísticos para identificar a magnitude e a estrutura da variabilidade espacial de variáveis físicas do solo*. 2002. 75f. Dissertação (Mestrado em Agronomia) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2002.
- PARDO-IGÚZQUIZA, E.; MLREML: A computer program for the inference of spatial covariance parameters by maximum likelihood and restricted maximum likelihood. *Comp. Geosci.*, New York, v.23, n.2, p.153-162, 1997.
- PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. *Biometrika*, London, v.58, p.545-554, 1971.
- PINHEIRO, J. C.; BATES, D. M. *Statistics and computing: mixed-effects models in S and S-PLUS*. New York: Springer, 2000. 528p.
- PLONER, A.; DUTTER, R. New directions in geostatistics. *J. Stat. Plan. Infer.*, Amsterdam, v.91, p.499-509, 2000.
- REICHARDT, K; VIEIRA, S. R.; LIBARDI, P. L. Variabilidade espacial de solos e experimentação de campo. *Rev. Bras. Ciênc. Solo*, Campinas, v.10, n.1, p1-6, 1986.
- RIBEIRO JUNIOR, P. J. *Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo*. 1995. 99f. Dissertação (Mestrado em Agronomia) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1995.
- RIBEIRO JUNIOR, P. J.; DIGGLE, P. J. GEOR: A package for geostatistical analysis. *News: Rev.*, London, v.11, p.15-18, 2001.
- SAD, J. H. G.; QUADE, H. *Fundamentos sobre a variabilidade dos depósitos minerais*. Rio de Janeiro: DNPM/CPRM, 1986. 141p.
- SAMPER, F. J. S.; RAMIREZ, J. C. *Geoestadística: aplicaciones a la hidrogeología subterránea*. Barcelona: Gráficas Torres, 1990. 484p.
- SOUZA, Z. M. et al. Variabilidade espacial de atributos físicos de um latossolo vermelho distrófico sob semeadura em Selvíria (MS). *Rev. Bras. Ciênc. Solo*, Viçosa, v.25, p.699-707, 2001.

VIEIRA, S. R.; NIELSEN, D. R.; BIGGAR, J. W. Spatial variability of field-measured infiltration rate. *Soil Sci. Am. J.*, Madison, v.45, n.3, p.1040-8, 1981.

XAVIER, L. H. *Modelos univariado e multivariado para análise de medidas repetidas e verificação da acurácia do modelo univariado por meio de simulação*. 2000. 90f. Dissertação (Mestrado em Agronomia) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2000.

Recebido em 18.08.2006.

Aprovado após revisão em 23.02.2007.