

**MÉTODOS DE ESTIMAÇÃO DE PARÂMETROS EM MODELOS
GEOESTATÍSTICOS COM DIFERENTES ESTRUTURAS DE
COVARIÂNCIAS: UMA APLICAÇÃO AO TEOR DE CÁLCIO NO SOLO**

MARIA CRISTINA NEVES DE OLIVEIRA

Tese apresentada à Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, para obtenção do título de Doutor em Agronomia, Área de Concentração: Estatística e Experimentação Agronômica.

P I R A C I C A B A
Estado de São Paulo - Brasil
Janeiro - 2003

**MÉTODOS DE ESTIMAÇÃO DE PARÂMETROS EM MODELOS
GEOESTATÍSTICOS COM DIFERENTES ESTRUTURAS DE
COVARIÂNCIAS: UMA APLICAÇÃO AO TEOR DE CÁLCIO NO SOLO**

MARIA CRISTINA NEVES DE OLIVEIRA

Licenciado em Matemática

Orientador: Prof. Dr. **DÉCIO BARBIN**

Tese apresentada à Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, para obtenção do título de Doutor em Agronomia, Área de Concentração: Estatística e Experimentação Agronômica.

P I R A C I C A B A
Estado de São Paulo - Brasil
Janeiro - 2003

Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP

Oliveira, Maria Cristina Neves de

Métodos de estimação de parâmetros em modelos geoestatísticos com diferentes estruturas de covariâncias: uma aplicação ao teor de cálcio no solo / Maria Cristina Neves de Oliveira. -- Piracicaba, 2003.

140 p. : il.

Tese (doutorado) - Escola Superior de Agricultura Luiz de Queiroz, 2003.
Bibliografia.

1. Cálcio 2. Distribuição espacial 3. Distribuição (teoria da probabilidade) 4. Função exponencial 5. Geoestatística 6. Método da máxima (verossimilhança) 7. Solo 8. Variabilidade espacial I. Título

CDD 519.535

“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”

A Deus e seus arcanjos
que me iluminaram.

Aos meus queridos pais,
o amor, a dedicação o carinho.

Ao meu filho Rodrigo,
a paciência em suportar minha ausência.

À minha família,
que a todo momento me incentivou.

DEDICO

AGRADECIMENTOS

Ao professor Dr. Décio Barbin a orientação, os ensinamentos, as sugestões, a amizade e o crédito em meu trabalho.

À Empresa Brasileira de Pesquisa Agropecuária a oportunidade dada para a realização deste curso de pós-graduação.

À professora Dr^a Clarice G. B. Demétrio do Departamento de Ciências Exatas da ESALQ/USP o estímulo, a amizade, a compreensão e as sugestões ao meu trabalho.

Ao Professor Dr. Paulo Justiniano Ribeiro Junior do Departamento de Estatística da Universidade Federal do Paraná a paciência, os ensinamentos, a amizade, e as incansáveis discussões e sugestões ao meu trabalho.

Aos pesquisadores da Embrapa Solos por terem disponibilizado os dados utilizados nesse trabalho, em especial ao Dr. Enio Fraga da Silva e ao Dr. José Ronaldo de Macedo pelas incansáveis discussões.

Aos meus amigos do curso de pós-graduação, André Jales Monteiro, Cláudia Cristina Paro de Paz, Glaucy da Conceição Ortiz, Heyder Diniz, Jeanete Moreira, Denise Santana, Jomar Antonio Camarinha Filho, Silvano Cesar da Costa e Suely Ruiz Giolo a união, o incentivo, a amizade, o carinho, a compreensão e a convivência.

Ao grande amigo e professor Dr. Norberto Antonio Lavorenti do Centro de Ciências Agrárias - UFSCar - Araras o apoio, o carinho, a paciência, a amizade e a dedicação sem restrição.

Às minhas grandes amigas Maria Luisa Meneguetti Calçada, Maria Imaculada Lima Montebelo e respectivas famílias a amizade, o carinho e o apoio permitindo que esta estada em Piracicaba fosse a melhor possível.

À professora Dra. Sonia Maria De Stefano Piedade a atenção, a amizade e o carinho nas diferentes etapas do curso de pós-graduação.

À Rosa Maria Alves, Luciane Brajão e Solange Sabadin a amizade e aos demais colegas, professores e funcionários do Departamento de Ciências Exatas da ESALQ/USP a atenção e o respeito.

À bibliotecária Eliana Maria Garcia Sabino a revisão das normas.

A todos que de forma direta ou indireta contribuíram para a realização deste novo desafio em minha vida agradeço.

SUMÁRIO

	Página
RESUMO.....	ix
SUMMARY.....	xi
1 INTRODUÇÃO.....	1
2 REVISÃO DE LITERATURA.....	3
2.1 Teoria das variáveis regionalizadas.....	4
2.1.1 Aspectos gerais.....	4
2.1.2 Localização e continuidade.....	5
2.1.3 Função aleatória.....	6
2.1.4 Características das variáveis regionalizadas.....	6
2.1.5 Função de distribuição e momentos de uma função aleatória.....	8
2.1.6 Hipóteses que possibilitam a inferência de uma função aleatória.....	9
2.1.6.1 Hipóteses restritivas.....	10
2.1.6.2 Relações entre variâncias, covariância e semivariância.....	12
2.2 Exemplos de dados agronômicos com dependência espacial.....	15
2.3 Análise exploratória de dados.....	19
2.4 Análise estrutural.....	20
2.4.1 Objetivos.....	20
2.4.2 Variogramas e covariâncias.....	21
2.4.3 Estimador robusto do variograma.....	22
2.4.4 Tipos de variograma.....	22
2.4.5 Características dos parâmetros que descrevem um semivariograma.....	23
2.4.6 Modelos anisotrópico e isotrópico.....	25
2.4.7 Modelos teóricos de semivariogramas.....	26
2.4.7.1 Modelos com patamar.....	26

2.4.7.2 Modelos sem patamar.....	28
2.5 Modelo Matérn.....	28
2.6 Método de interpolação pela krigagem.....	29
2.7 Outros métodos de estimação dos parâmetros do semivariograma	32
2.7.1 Método de estimação empírico, mínimos quadrados ordinários, mínimos quadrados generalizados e mínimos quadrados ponderados.....	32
2.7.2 Método de estimativas dos parâmetros de modelos com dependência espacial.....	34
2.7.2.1 Máxima verossimilhança.....	34
2.7.2.2 Máxima verossimilhança restrita.....	39
2.8 Perfil de verossimilhança.....	41
2.9 Critério de informação de Akaike na seleção de modelos com dependência espacial.....	43
2.10 Teste da razão de verossimilhanças.....	43
2.11 Método da validação cruzada.....	44
3 METODOLOGIA.....	46
3.1 Material.....	46
3.2 Local experimental e método de amostragem.....	46
3.3 Métodos.....	48
3.3.1 Descrição dos modelos.....	48
3.3.2 Métodos de estimação dos parâmetros.....	50
3.3.2.1 Máxima verossimilhança.....	50
3.3.2.2 Máxima verossimilhança restrita.....	53
3.4 Perfil de Verossimilhança.....	55
3.5 Critério de Informação de Akaike na seleção de modelos com dependência espacial	55
3.6 Teste da Razão de Verossimilhanças.....	55
3.7 Método da Validação Cruzada.....	56
3.8 Método de análise estatística.....	56
3.9 Programas estatísticos utilizados.....	56
4 RESULTADOS E DISCUSSÃO.....	57
4.1 Generalidades.....	57
4.2 Análise estatística.....	57
4.3 Análise geoestatística.....	62

4.4 Análise geoestatística para modelos com covariável	65
4.4.1 Estimadores de máxima verossimilhança e máxima verossimilhança restrita com a estrutura de covariância exponencial	65
4.4.2 Estimadores de máxima verossimilhança e máxima verossimilhança restrita com a estrutura de covariância pela família Matérn para $k = 1$ e $k = 2$	83
5 CONCLUSÕES.....	99
ANEXOS.....	101
REFERÊNCIAS BIBLIOGRÁFICAS.....	132
APÊNDICE.....	139

MÉTODOS DE ESTIMAÇÃO DE PARÂMETROS EM MODELOS GEOESTATÍSTICOS COM DIFERENTES ESTRUTURAS DE COVARIÂNCIAS: UMA APLICAÇÃO AO TEOR DE CÁLCIO NO SOLO

Autora: MARIA CRISTINA NEVES DE OLIVEIRA

Orientador: Prof. Dr. DÉCIO BARBIN

RESUMO

A compreensão da dependência espacial das propriedades do solo vem sendo cada vez mais requerida por pesquisadores que objetivam melhorar a interpretação dos resultados de experimentos de campo fornecendo, assim, subsídios para novas pesquisas a custos reduzidos. Em geral, variáveis como, por exemplo, o teor de cálcio no solo, estudado neste trabalho, apresentam grande variabilidade impossibilitando, na maioria das vezes, a detecção de reais diferenças estatísticas entre os efeitos de tratamentos. A consideração de amostras georreferenciadas é uma abordagem importante na análise de dados desta natureza, uma vez que amostras mais próximas são mais similares do que as mais distantes e, assim, cada realização desta variável contém informação de sua vizinhança. Neste trabalho, métodos geoestatísticos que baseiam-se na modelagem da dependência espacial, nas pressuposições Gaussianas e nos estimadores de máxima verossimilhança são utilizados para analisar e interpretar a variabilidade do teor de cálcio no solo, resultado de um experimento realizado na Fazenda Angra localizada no Estado do Rio de Janeiro. A área experimental foi dividida em

três regiões em função dos diferentes períodos de adubação realizadas. Neste estudo foram utilizados dados do teor de cálcio obtidos das camadas 0-20cm e 20-40cm do solo, de acordo com as coordenadas norte e leste. Modelos lineares mistos, apropriados para estudar dados com esta característica, e que permitem a utilização de diferentes estruturas de covariâncias e a incorporação da região e tendência linear das coordenadas foram usados. As estruturas de covariâncias utilizadas foram: a exponencial e a Matérn. Para estimar e avaliar a variabilidade dos parâmetros utilizaram-se os métodos de máxima verossimilhança, máxima verossimilhança restrita e o perfil de verossimilhança. A identificação da dependência e a predição foram realizadas por meio de variogramas e mapas de *krigagem*. Além disso, a seleção do modelo adequado foi feita pelo critério de informação de Akaike e o teste da razão de verossimilhanças. Observou-se, quando utilizado o método de máxima verossimilhança, o melhor modelo foi aquele com a covariável região e, com o método de máxima verossimilhança restrita, o modelo com a covariável região e tendência linear nas coordenadas (modelo 2). Com o teor de cálcio, na camada 0-20cm e considerando-se a estrutura de covariância exponencial foram obtidas as menores variâncias *nugget* e a maior variância espacial (*sill – nugget*). Com o método de máxima verossimilhança e com o modelo 2 foram observadas variâncias de predição mais precisas. Por meio do perfil de verossimilhança pode-se observar menor variabilidade dos parâmetros dos variogramas ajustados com o modelo 2. Utilizando-se vários modelos e estruturas de covariâncias, deve-se ser criterioso, pois a precisão das estimativas, depende do tamanho da amostra e da especificação do modelo para a média. Os resultados obtidos foram analisados, com a subrotina geoR desenvolvida por Ribeiro Junior & Diggle (2000), e por meio dela pode-se obter estimativas confiáveis para os parâmetros dos diferentes modelos estimados.

**PARAMETER ESTIMATION METHODS IN GEOSTATISTIC MODELS WITH
DIFFERENT COVARIANCE STRUCTURES: AN APPLICATION TO THE CALCIUM
CONTENT IN THE SOIL**

Author: MARIA CRISTINA NEVES DE OLIVEIRA

Adviser: Prof. Dr. DÉCIO BARBIN

SUMMARY

The understanding of the spatial dependence of the properties of the soil becomes more and more required by researchers that attempt to improve the interpretation of the results of field experiments supplying subsidies for new researches at reduced costs. In general, variables as, for example, the calcium content in the soil, studied in this work, present great variability disabling, most of the time, the detection of real statistical differences among the treatment effects. The consideration of georeferenced samples is an important approach in the analysis of data of this nature, because closer samples are more similar than the most distant ones and, thus, each realization of this variable contains information of its neighborhood. In this work, geostatistics methods that are based on the modeling of the spatial dependence, under the Gaussian assumptions and the maximum likelihood estimators, are used to analyze and to interpret the variability of calcium content in the soil, obtained from an experiment carried on at Fazenda Angra, located in Rio de Janeiro, Brazil. The experimental area was divided in three areas depending on the different periods of fertilization. In this study, data of the calcium soil content from the layers 0-20cm and 20-40cm, were used, according to the north and east

coordinates. Mixed linear models, ideal to study data with this characteristic, and that allow the use of different covariance structures, and the incorporation of the region and linear tendency of the coordinates, were used. The covariance structures were: the exponential and the Matérn. Maximum likelihood, maximum restricted likelihood and the profile of likelihood methods were used to estimate and to evaluate the variability of the parameters. The identification of the dependence and the prediction were realized using variograms and kriging maps. Besides, the selection of the appropriate model was made through the Akaike information criterion and the likelihood ratio test. It was observed that when maximum likelihood method was used the most appropriate model was that with the region covariate and, with the maximum restricted likelihood method, the best model was the one with the region covariate and linear tendency in the coordinates (model 2). With the calcium content, in the layer 0-20cm and considering the exponential covariance structure, the smallest *nugget* variances and the largest spatial variance (*sill - nugget*) were obtained. With the maximum likelihood method and with the model 2 more precise prediction variances were observed. Through the profile of likelihood method, smaller variability of the adjusted variogram parameters can be observed with the model 2. With several models and covariance structures being used, one should be very critical, because the precision of the estimates depends on the size of the sample and on the specification of the model for the average. The obtained results were analyzed, with the subroutine geoR developed by Ribeiro Junior & Diggle (2000), and through this subroutine, reliable estimates for the parameters of the different estimated models can be obtained.

1 INTRODUÇÃO

A preocupação com a variabilidade espacial de muitas variáveis agronômicas é antiga e a forma encontrada para solucionar este problema, até a década de setenta, em experimentos de campo, foi o uso dos delineamentos em blocos ao acaso associado aos princípios da aleatorização e da repetição (Fisher, 1935). Este método foi alvo de várias críticas enquanto ignorada a posição espacial entre as observações experimentais (Vieira, 1981; Reichardt et al., 1986). Embora criticado, o método de Fisher, não deve ser descartado, pois é tão eficiente quanto as ferramentas geoestatísticas, se considerada a estrutura de correlação espacial nos métodos de análise de experimentos dos delineamentos convencionais obtendo-se inferências mais precisas para os contrastes entre tratamentos (Grondona & Cressie, 1991).

Nos casos em que as variáveis agronômicas não são obtidas através dos delineamentos formais e, apresentam aspectos aleatórios e espaciais, surgiu a necessidade de avaliar a dependência espacial, por um método alternativo, associado aos conceitos básicos estatísticos denominado de geoestatística. Esse último é fundamentado na teoria das variáveis regionalizadas (VR), em que as amostras coletadas da variável em estudo são georreferenciadas (Matheron, 1963). A variável regionalizada é uma função que varia de um lugar a outro no espaço com certa aparência de continuidade e, cujos valores estão relacionados com a posição espacial que ocupam. A dificuldade maior deste método é quanto à inferência estatística com uma única realização para cada variável coletada em uma malha regular ou irregular (Guerra, 1988).

No estudo geoestatístico dois aspectos são apresentados quando se avalia a continuidade espacial, o primeiro, fundamenta-se na modelagem da dependência espacial considerando as pressuposições da distribuição Gaussiana e determinando estimativas paramétricas através dos estimadores de: máxima verossimilhança, máxima verossimilhança restrita, verossimilhança perfilada e inferência Bayesiana (Diggle & Ribeiro Junior, 2000; Ploner & Dutter, 2000). O segundo é o não-paramétrico, que independe de métodos rígidos, e que

baseiam-se em tratamentos de dados ordenados, de tal forma que cada valor é representado por um número que caracteriza a sua posição na seqüência. Neste método (empírico ou a *sentimento*), a inferência somente será possível se forem assumidas as hipóteses de estacionaridade ou a hipótese intrínseca (Sad, 1986; Diggle & Ribeiro Junior, 2000). Em geral, o método empírico utilizado no estudo da dependência espacial em geologia, hidrologia e agronomia é considerado limitado devido a sua subjetividade e com pouco rigor (Ribeiro Junior, 1995). Estas duas linhas de estudos utilizam-se do semivariograma para identificar a dependência espacial, e do interpolador de *krigagem*, que possibilita a partir de observações pontuais, informações para grandes extensões de terra, baseadas nas observações da variável a ser estimada em pontos não amostrados.

Portanto, uma proposta deste trabalho foi utilizar os estimadores paramétricos de máxima verossimilhança e máxima verossimilhança restrita, tomando-se como valores iniciais para os parâmetros do semivariograma, aqueles obtidos pelo método não-paramétrico (empírico) para avaliar a estrutura da dependência espacial em variáveis de fertilidade do solo. O método de interpolação utilizado foi o da *krigagem* ordinária, pois o mesmo viabiliza a predição e os erros-padrões da predição para os pontos não amostrados, e não requer o prévio conhecimento da média .

Os objetivos gerais deste trabalho são:

- a) determinar os valores iniciais dos parâmetros dos semivariogramas teóricos pelo método empírico;
- b) estudar os modelos exponencial e Matérn na identificação da dependência espacial do cálcio no solo, permitindo, assim, melhor interpretação dos fatores que interferem na produção das culturas e,
- c) obter as estimativas dos parâmetros dos modelos com estrutura de covariância exponencial e Matérn, pelos estimadores de máxima verossimilhança e máxima verossimilhança restrita na obtenção dos semivariogramas.

2 REVISÃO DE LITERATURA

As contribuições das teorias e métodos de Fisher (1935) para a estatística, que fundamentam-se na independência, são importantes na aplicação prática da teoria de inferência estatística na pesquisa científica, se for considerado todas as exigências necessárias para o uso destes métodos tais como: os princípios da aleatorização, da repetição, que permite a redução dos erros e aumento da precisão do experimento, e do controle local.

A variabilidade espacial de muitas propriedades do solo, encontrada em experimentos agronômicos está associada a heterogeneidade natural dos solos. Sendo assim, Fisher (1935) discutiu os aspectos do planejamento, a definição dos delineamentos e a análise destes experimentos como solução do problema em questão. Neste estudo, o autor ainda comentou que ao se escolher uma área para a experimentação não se tinha conhecimento do histórico da área, somente da extensão de manchas de solo e, que nas parcelas vizinhas, as produções das culturas eram mais semelhantes do que naquelas mais afastadas.

As contribuições de outros autores também foram importantes para adequação e melhoria dos métodos estatísticos e algumas críticas surgiram relacionados aos métodos de Fisher. Uma das questões discutida por Yates (1938) foi que, embora a aleatorização controlasse o viés indesejado e neutralizasse o efeito da correlação espacial, ela amenizaria a variabilidade mas, não poderia removê-la e, as produções de culturas em parcelas de campo, mesmo sendo distribuídas ao acaso, não eram independentes apresentando correlação positiva entre parcelas vizinhas.

Após algumas décadas a variabilidade foi também questionada nas pesquisas realizadas com mineração e, o estudo da variabilidade espacial surgiu com as pesquisas de

prospeção de jazidas de ouro, na África do Sul, por Krige¹, citado por David (1977), Guerra (1988), Cressie (1993) e Ribeiro Junior (1995). O autor era da opinião, quando analisava dados de concentração de ouro, que no cálculo das variâncias deveria ser considerada a distância entre as amostras coletadas. Além disso, concluiu que:

- a) as variâncias encontradas nos estudos com dependência espacial utilizando malha regular são sistematicamente menores do que as obtidas de uma malha irregular;
- b) a variância total é a soma de uma variância aleatória e uma espacial ;
- c) para efeito de parametrização de reservas deve ser utilizada a média geométrica devido à influência de valores extremos e,
- d) devido à existência de dois tipos de variâncias (aleatória e espacial), o número de amostras necessárias no estudo geoestatístico é sempre menor do que a solicitada pela estatística clássica.

Com essa experiência da África do Sul houve prosseguimento dessas pesquisas pelo matemático probabilístico Matheron (1963), em que o autor em uma linguagem rigorosa, formalizou as observações experimentais anteriores, intitulando de teoria das variáveis regionalizadas (VR), a qual também recebeu o nome de geoestatística devido a sua aplicação ter sido direcionada, inicialmente, a problemas de Geologia e Mineração.

2.1 Teoria das variáveis regionalizadas

2.1.1 Aspectos gerais

A teoria da variável regionalizada (VR) desenvolvida por Matheron (1971) vem sendo aplicada em diferentes áreas de pesquisas, como: nematologia, entomologia, ecologia, hidrologia, fertilidade dos solos e agrometeorologia, quando as variáveis estudadas apresentam dependência espacial.

Diz-se que toda variável distribuída no espaço e que apresenta uma estrutura espacial correlacionada, é regionalizada. Assim, se $z(x)$ é o valor de uma variável aleatória Z em um ponto x , $Z(x)$ é uma variável regionalizada (VR). Definido desta forma, o conceito de VR é puramente descritivo e não tem nenhuma conotação probabilística. Do ponto de vista

¹ KRIGE, D. A statistical approach to some basic mine valuation problems on the Witwatersrand. **Journal of the Chemical, Metallurgical and Mining Society of South Africa**, v. 52, p.119-139, 1951.

matemático, uma VR é simplesmente uma função $Z(x)$ que assume um valor para cada x no espaço n-dimensional \Re^n (Trangmar, 1985; Braga, 1990; Samper & Ramirez, 1990; Ribeiro Junior, 1995; Pannatier, 1996; Vieira, 1996).

Alguns aspectos das variáveis regionalizadas como a localização e a continuidade são também importantes e devem ser também considerados no estudo da dependência espacial.

2.1.2 Localização e continuidade

Uma das características qualitativas da VR diretamente ligada à estrutura do fenômeno espacial, conforme Guerra (1988), é a localização ou suporte geométrico. Esta variável assume seus valores importantes no campo geométrico em que a mesma foi definida, ou seja, no local onde a variável é suscetível de tomar valores definidos e no interior do qual será avaliada sua variação. A escolha do espaço geométrico não é arbitrário e deve ser definido de forma que atenda às condições de homogeneidade física, que é de difícil formulação na prática.

Outra característica importante é a continuidade espacial, que para a VR se manifesta sobretudo pela tendência de apresentarem valores mais próximos em dois pontos amostrados, quanto menos afastados eles estiverem. Estas variáveis que se apresentam com continuidade mais ou menos estável em sua variação espacial, podem ser expressas por meio de uma flutuação entre os valores de amostras vizinhas, e tal flutuação reflete o grau de dependência entre um valor e outro (Guerra, 1988). Segundo Soares (2000) a continuidade espacial, é a quantificação da forma como se dispersam espacialmente as características de um atributo.

Mesmo considerando a complexidade das flutuações para as diversas variáveis regionalizadas existe, em geral, uma continuidade em média e esta foi expressa por Valente (1989) por:

$$\lim_{x \rightarrow x_0} E[Y(x) - Y(x_0)] = 0$$

Em muitas situações esta continuidade não se verifica surgindo assim, o efeito pepita ($t^2 \neq 0$), que é definido como:

$$\lim_{x \rightarrow x_0} E[Y(x) - Y(x_0)]^2 = t^2 \neq 0$$

2.1.3 Função aleatória

Seja uma região D onde são extraídas amostras e realizadas mensurações de atributos de interesse. Destas amostras resulta um conjunto de dados com distribuição espacial e isto implica em medidas deste atributo seguidas de suas coordenadas, as quais possibilitam o cálculo das distâncias euclidianas entre os pontos observados.

Desta forma, o processo estocástico é expresso por:

$$\{Z(x_i) : x_i \in D \subset \mathbb{R}^d\} \quad (1)$$

em que, Z é a variável aleatória que varia continuamente em D , x é a localização da variável, considerada fixa, D é a região em estudo e \mathbb{R}^d é o espaço d-dimensional ($d = 1, 2, 3$ ou 4).

Para os casos em que $d = 1$ os dados estão na transeção da malha, para $d = 2$, em um plano, para $d = 3$, em um volume e para $d = 4$ inclui-se o tempo.

Assim, o conjunto de valores $z(x_i)$ para todos os pontos x_i na região do domínio D , ou seja, a variável regionalizada $Z(x_i)$, é considerada uma particular realização do conjunto de variáveis aleatórias $\{Z(x_i) : x_i \in D \subset \mathbb{R}^d\}$ (Soares, 2000). O conjunto de variáveis aleatórias é chamado de função aleatória e pode ser expresso como $Z(x)$ (Ribeiro Junior, 1995).

A função aleatória expressa aspectos aleatórios e estruturados de uma variável regionalizada de tal forma que: a) localmente, para um ponto x_i , $Z(x_i)$ é uma variável aleatória e, b) $Z(x)$ é também uma função aleatória no sentido de que para cada par de pontos (x_i) e $(x_i + \mathbf{h})$, as correspondentes variáveis regionalizadas $Z(x_i)$ e $Z(x_i + \mathbf{h})$ são, em geral, correlacionadas, expressando a estrutura espacial da variável regionalizada $Z(x_i)$.

2.1.4 Características das variáveis regionalizadas

A teoria das variáveis regionalizadas formalizada por Matheron (1963), foi também chamada de geoestatística devido a sua aplicação na Geologia e Mineralogia.

A geoestatística modela os valores de um atributo, dentro de uma região, como uma função aleatória como apresentada por (1). Para cada posição $x \in D$ o valor do atributo de um

dado espacial é modelado como uma variável aleatória $Z(x)$. Isto indica, que nesta posição, esta variável aleatória pode assumir distintos valores para o atributo, e cada uma dessas variáveis está associada uma distribuição (Samper & Ramirez, 1990; Felgueiras, 1999).

Sua função de distribuição acumulada multivariada, composta por variáveis aleatórias, é definida por:

$$F_{Z(x_1), Z(x_2), \dots, Z(x_k)}(Z_1, Z_2, \dots, Z_k) = \text{Prob}[Z(x_1) \leq Z_1, Z(x_2) \leq Z_2, \dots, Z(x_k) \leq Z_k]. \quad (2)$$

A interpretação probabilística de uma variável regionalizada como realização de uma função aleatória $Z(x)$ tem sentido somente se for possível inferir, pelo menos em parte, qual é a função de distribuição ou lei de probabilidade de $Z(x)$. Em geral, uma das limitações dos dados com dependência espacial está relacionada com a impossibilidade de repetir indefinidamente um experimento e realizar inferência a partir de uma única realização, e da mesma maneira que não é possível reconstituir a função de distribuição de uma variável aleatória, por exemplo, retirar resultados de uma observação. Um exemplo é a precipitação de chuva diária ou mesmo realizar novas amostras de um bloco de minério já processado (Guerra, 1988; Braga, 1990; Cressie, 1993).

Tais hipóteses da VR estão relacionadas com a homogeneidade espacial da função aleatória e supor que a função aleatória é estacionária é o mesmo que pensar que a função aleatória repete-se no espaço de tal forma que esta proporcione a informação equivalente a muitas realizações da mesma função aleatória, possibilitando assim a inferência estatística (Samper & Ramirez, 1990).

Dois métodos foram propostos por Matheron (1971), a fim de possibilitar inferências para as variáveis regionalizadas. O primeiro método, chamado de transitivo tem como premissa as considerações determinísticas e o segundo baseia-se na teoria das funções aleatórias. Neste último é introduzida uma interpretação probabilística da variável regionalizada, exigindo uma série de hipóteses a respeito da função aleatória (estacionaridade etc) sendo que ambos os métodos conduzem a resultados equivalentes, o que é importante do ponto de vista metodológico.

A preferência de Matheron pelas funções aleatórias tem como justificativas:

- a) o enfoque probabilístico permite levar em conta as flutuações erráticas da variável regionalizada e,

- b) o método transitivo é geral, requerendo certas hipóteses no momento de calcular as variâncias de estimação. O significado da certeza do conhecimento científico destas hipóteses é idêntico à necessidade de considerar a esperança matemática. Portanto, dado que não é realmente possível evitar a interpretação probabilística, parece mais adequado considerá-la explicitamente.

A adoção do enfoque probabilístico sugere duas questões fundamentais:

- tem sentido considerar um fenômeno natural com uma única observação de uma função aleatória, ou seja, uma entre as infinitas realizações possíveis?,
- é possível usar a inferência estatística e determinar os parâmetros estatísticos que definem a lei de probabilidade do fenômeno a partir de uma única observação de uma função aleatória? (Samper & Ramirez, 1990).

2.1.5 Função de distribuição e momentos de uma função aleatória

Na geoestatística, em que as estimativas de pontos não amostrados são obtidas por meio de uma combinação linear, são suficientes os dois primeiros momentos da distribuição de $Z(x)$, sendo que, na maioria das aplicações práticas, a informação disponível não permite deduzir momentos de ordem maior.

O momento de primeira ordem de uma função aleatória $Z(x)$ é a esperança matemática, definida como:

$$E[Z(x)] = m(x). \quad (3)$$

Além deste, existem três momentos de segunda ordem considerados em geoestatística, definidos por Samper & Ramirez (1990) e Braga (1990) os quais são apresentados como:

- a variância de $Z(x)$ quando existe, pode ser definida como o momento de segunda ordem em relação à média $m(x)$.

$$\mathbf{s}^2 = \text{Var}[Z(x)] = E\{[Z(x) - m(x)]^2\}; \quad (4)$$

b) covariância - sejam duas variáveis aleatórias $Z(x_i)$ e $Z(x_j)$, nos pontos x_i e x_j , respectivamente, então elas têm covariância como função das localizações x_i e x_j e, consequentemente têm variância finita. Assim a expressão da covariância é dada por:

$$C(x_i, x_j) = E \left[[Z(x_i) - m(x_i)][Z(x_j) - m(x_j)] \right], \quad (5)$$

ou ainda, de forma equivalente,

$$C(x_i, x_j) = E \left[[Z(x_i) - E(Z(x_i))][Z(x_j) - E(Z(x_j))] \right]. \quad (6)$$

A covariância, dada por (5), deve ser definida positiva para assegurar a existência de todas as distribuições acumuladas e é, também, chamada de função de auto-covariância;

c) semivariograma $\mathbf{g}(x_i, x_j)$

$$\mathbf{g}(x_i, x_j) = \frac{1}{2} E \left[[Z(x_i) - Z(x_j)]^2 \right] \quad (7)$$

O variograma é, portanto, apresentado por $2\mathbf{g}(x_i, x_j)$, mas existem autores que usam indistintamente ambas as formas para referir-se à função $\mathbf{g}(x_i, x_j)$.

Tanto a variância quanto o semivariograma são sempre positivos enquanto que a covariância (5), ou (6), pode apresentar valores negativos (Samper & Ramirez, 1990).

2.1.6 Hipóteses que possibilitam a inferência de uma função aleatória

Quando se deseja estudar uma variável aleatória, necessita-se então, de várias realizações desta variável (a função aleatória) e da distribuição que a rege, pois tal função é o próprio processo estocástico e é reapresentado por (1).

Um processo estocástico apresenta distribuições de várias ordens, sendo a distribuição de primeira ordem caracterizada por $F_z(z, s)$ e a de segunda ordem por $F_{z_1, z_2}(z_1, z_2; s_1, s_2)$ e, assim, sucessivamente para as demais ordens.

Em dados com distribuição espacial, o interesse está em distribuições de segunda ordem, em que são incluídas as funções de média, de covariância, de correlação e de semivariância.

Na impossibilidade de estimar as funções de primeira e segunda ordens devem ser estimadas as propriedades do processo.

No caso em questão, como não se pode fazer inferência somente com uma realização, são necessárias algumas hipóteses restritivas que permitam conhecer um fenômeno da natureza que tem continuidade espacial. Dentre elas existem as hipóteses de estacionaridade, que se divide em estacionaridade de segunda ordem e hipótese intrínseca.

Além de se assumirem as hipóteses mencionadas, é possível obter várias realizações da variável é por meio da diferença entre os valores em pontos separados por uma distância \mathbf{h} , ou seja, $[Z(x + \mathbf{h}) - Z(x)]$ (David, 1977; Cressie, 1993; Ribeiro Junior, 1995; Gonçalves, 1997; Kitanidis, 1997).

2.1.6.1 Hipóteses restritivas

a) Estacionaridade

Uma função aleatória é estritamente estacionária se sua função de distribuição de probabilidades dada por (2) é invariante em relação a qualquer translação do vetor \mathbf{h} , ou então, se a função de distribuição do vetor aleatório $[Z(x_1), Z(x_2), \dots, Z(x_k)]$ é idêntica à do vetor $[Z(x_1 + \mathbf{h}), Z(x_2 + \mathbf{h}), \dots, Z(x_k + \mathbf{h})]$ para qualquer \mathbf{h} .

Embora, enfatizado que a geoestatística empírica baseia-se nos dois primeiros momentos da função aleatória, é suficiente supor que estes momentos existam e limitem-se nas hipóteses de estacionaridade dos dois primeiros momentos (Valente, 1989; Samper & Ramirez, 1990).

b) Estacionaridade de primeira ordem

Uma função aleatória é estacionária de primeira ordem (3) se admitirmos que todas as variáveis aleatórias $Z(x_i)$, $i = 1, \dots, N$, têm a mesma média para todo os locais em toda a região de estudo. Esta hipótese definida por Trangmar et al. (1985) é expressa por:

$$E[Z(x_1)] = E[Z(x_2)] = \dots = E[Z(x_i)] = E[Z(x)] = m.$$

c) Estacionaridade de segunda ordem

Uma função aleatória $Z(x)$ é estacionária de segunda ordem se, além de atender a estacionaridade de 1ª ordem, apresentar a característica de que, para cada par de valores $Z(x)$ e $Z(x + \mathbf{h})$, a covariância dada por (5) ou (6), existe e depende apenas da distância h (Journel & Huijbregts, 1978; Vieira, 1983; Trangmar et al., 1985; Vieira, 1996; Gonçalves, 1997). As equações relacionadas a esta hipótese como a esperança matemática, covariância e variância (eq. 4) são:

$$E[Z(x)] = m, \quad \text{para todo } x \tag{8}$$

$$C(x + \mathbf{h}, x) = C(\mathbf{h}) = E[Z(x + \mathbf{h})Z(x)] - m^2, \tag{9}$$

ou

$$C(\mathbf{h}) = E[Z(x) - m][Z(x + \mathbf{h}) - m]. \tag{10}$$

A hipótese de estacionaridade de segunda ordem pressupõe a existência de covariância e em consequência, uma variância *a priori* finita. A expressão para variância, neste caso, será:

$$\text{Var}[Z(x)] = E[Z(x) - m]^2 = C(0). \tag{11}$$

Na natureza existem numerosos fenômenos que apresentam dispersão ilimitada. Neste caso, tais fenômenos não podem ser descritos de forma correta se lhes forem atribuída uma variância finita (Samper & Ramirez, 1990).

Esta hipótese de variância finita é muito forte e raramente assumida e, é usualmente substituída por uma hipótese mais fraca chamada de hipótese intrínseca.

d) Hipótese intrínseca

Esta hipótese não é mais definida através das distribuições das variáveis aleatórias pontuais e sim através das diferenças entre elas, também chamada de estacionaridade dos incrementos ou estacionaridade fraca (Ribeiro Junior, 1995).

A hipótese intrínseca é definida pelas seguintes condições:

- a) o incremento entre dois pontos tem esperança nula

$$E[Z(x + \mathbf{h}) - Z(x)] = 0 . \quad (12)$$

- b) para qualquer vetor \mathbf{h} o incremento tem uma variância finita independente da posição x .

$$\text{Var}[Z(x + \mathbf{h}) - Z(x)] = E[Z(x + \mathbf{h}) - Z(x)]^2 = 2\gamma(\mathbf{h}), \quad (13)$$

sendo $2\gamma(\mathbf{h})$ a definição de variograma.

Um processo estacionário de segunda ordem é intrinsecamente estacionário, mas o inverso não é verdadeiro. Assim, nos estudos geoestatísticos é suficiente que uma realização de um processo estocástico seja pelo menos intrinsecamente estacionário.

2.1.6.2 Relações entre variância, covariância e semivariância

Com as medidas de continuidade espacial e definidas as hipóteses de estacionaridade apresentadas nas Eqs. (8) a (13) é possível fazer uma relação entre elas.

Para valores grandes de h , a covariância espacial $C(\mathbf{h})$ dada por (9) e (10) decresce (5) e Figura (1). A estacionaridade de $C(\mathbf{h})$ implica na estacionaridade da variância amostral s^2 sendo a covariância espacial nula em distâncias superiores ao *range* ou alcance (f) (Vieira et al., 1983; Trangmar et al., 1985; Vieira, 1996; Gonçalves, 1997).

A estacionaridade de 2^a ordem não se aplica se não existirem as covariâncias e a variância finita apresentadas, respectivamente, em (9), (10) e (11), como em situações em que ocorrem tendências no fenômeno. Nestes casos, assume-se uma fraca estacionaridade (hipótese intrínseca), a qual permite o uso de resultados de uma variável regionalizada por estimação pelo método dos momentos (Trangmar et al., 1985; Ribeiro Junior & Diggle, 1999).

A hipótese intrínseca apresentada em (12) e (13) requer que para todos os vetores de \mathbf{h} , a variância do acréscimo $[Z(x + \mathbf{h}) - Z(x)]$, dada por (12), seja finita e independente da localização dentro da região.

Dividindo (13) por dois, obtém-se a semivariância $\mathbf{g}(\mathbf{h})$, a qual depende do vetor de distâncias \mathbf{h} e é de se esperar que $\mathbf{g}(\mathbf{h})$ aumente com a distância h até um valor máximo no qual a semivariância estabiliza (Figura 1).

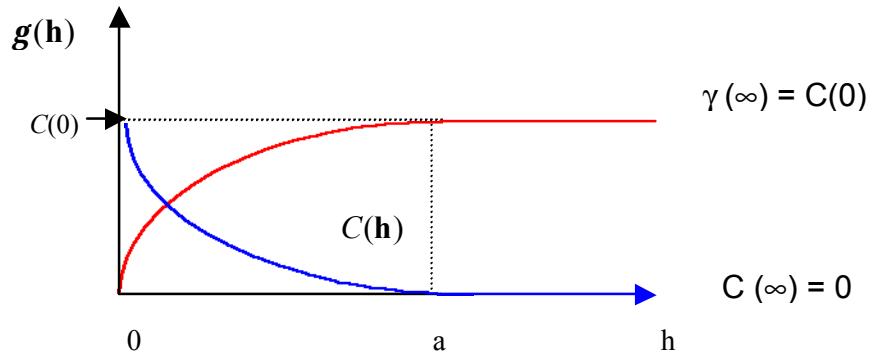


Figura 1- Relação entre a covariância espacial $C(\mathbf{h})$ e a semivariância $\mathbf{g}(\mathbf{h})$.

Fonte: Journel e Huijbregts (1978).

Fundamentando-se nas definições apresentadas em (8), (12) e (13) e definindo-se o segundo momento da variável aleatória $[Z(x) - Z(x + \mathbf{h})]$ podem-se determinar algumas relações entre as medidas de continuidade. Assim, o segundo momento pode ser escrito como:

$$M_2 = E \left[[Z(x) - Z(x + \mathbf{h})] - \mathbf{m}_z \right]^2,$$

ou

$$M_2 = E \left[[Z(x + \mathbf{h}) - Z(x)]^2 \right] \left[[Z(x + \mathbf{h}) - Z(x)] \right]^2 \quad (14)$$

Considerando (12) obtém-se a equação do segundo momento (14) que é escrita da forma que se segue:

$$M_2 = E[Z(x) - Z(x + \mathbf{h})]^2,$$

$$\begin{aligned}
&= E[Z(x)]^2 - 2[E[Z(x)Z(x+\mathbf{h})]] + [E[Z(x+\mathbf{h})]]^2, \\
&= E[Z(x)]^2 - 2E[Z(x)Z(x+\mathbf{h})] + E[Z(x+\mathbf{h})]^2.
\end{aligned}$$

Pela eq. (11) e supondo-se que a variância de $Z(x)$ é dada por:

$$\mathbf{s}_{[Z(x)]}^2 = E[Z(x)]^2 - \{E[Z(x)]\}^2 = C(0), \quad (15)$$

então,

$$E[Z(x)]^2 = C(0) + \{E[Z(x)]\}^2. \quad (16)$$

A covariância entre $Z(x)$ e $Z(x+\mathbf{h})$, assumindo-se a estacionaridade de 2ª ordem, definida por Journel e Huijbregts (1978), pode ser escrita como:

$$C(\mathbf{h}) = \text{Cov}[Z(x), Z(x+\mathbf{h})],$$

$$C(\mathbf{h}) = E[Z(x)Z(x+\mathbf{h})] - [E[Z(x)]] [E[Z(x+\mathbf{h})]], \quad (17)$$

Reescrevendo a eq. (17) pode-se ter:

$$E[Z(x)Z(x+\mathbf{h})] = C(\mathbf{h}) - [E[Z(x)]] [E[Z(x+\mathbf{h})]]$$

Considerando-se as eqs. (11), (15) e (17) e substituindo-se na equação do segundo momento (14) obtém-se:

$$M_2 = C(0) + [E[Z(x)]^2] - 2[C(\mathbf{h}) - [E[Z(x)]] [E[Z(x+\mathbf{h})]]] + E[Z(x+\mathbf{h})]^2.$$

Portanto,

$$M_2 = C(0) + m^2 - 2[C(\mathbf{h}) + m^2] + C(0) + m^2,$$

$$M_2 = 2C(0) + 2m^2 - 2C(\mathbf{h}) - 2m^2.$$

Logo, a função de semivariância representada por $2g(\mathbf{h})$ é escrita como:

$$2g(\mathbf{h}) = 2C(0) - 2C(\mathbf{h}). \quad (18)$$

Para uma função estacionária, a expressão (18) é considerada como a relação entre o variograma e a função de covariância (Kitanidis, 1997).

A função de semivariância é a esperança da variável aleatória $[Z(x) - Z(x + \mathbf{h})]$ (Journel e Huijbregts, 1978), ou seja:

$$g(x, h) = \frac{1}{2} E[Z(x + \mathbf{h}) - Z(x)]^2 \quad (19)$$

A expressão (19) originou-se da definição do momento de segunda ordem, considerando-se a estacionaridade de primeira ordem (3).

Conhecida a ferramenta que permitirá avaliar a dependência espacial, salienta-se que é importante conhecer melhor os dados por meio de estatísticas descritivas e gráficas, ou seja, de uma análise exploratória. Este método de análise, que será descrito no próximo item, permite verificar se as hipóteses básicas para um estudo geoestatístico estão sendo observadas.

2.2 Exemplos de dados agronômicos com dependência espacial

Nas duas últimas décadas, os resultados de pesquisas, que consideram a ocorrência de dependência espacial, têm sido observados nas diferentes áreas da ciência e, atualmente, na ciência do solo, esta continuidade espacial vem se destacando na nutrição e fertilidade do solo, no manejo da cultura, irrigação e drenagem entre outras áreas.

Alguns autores afirmam que na coleta de dados de uma determinada variável é importante considerar a posição espacial da amostra, seja em uma malha regular ou irregular e, no Brasil, um dos primeiros trabalhos utilizando a geoestatística com dados georreferenciados, foi o estudo da variabilidade de medidas de taxas de infiltração de água no solo (Vieira et al., 1981). Os autores realizaram a amostragem regionalizada numa malha de 160 x 55 m e as

estimativas de pontos não amostrados foram obtidas pela *krigagem*. A *krigagem* é um método de interpolação linear, não tendencioso e de variância mínima, em que utiliza-se de uma ponderação proporcional às distâncias euclidianas entre o ponto estimado e os pontos conhecidos.

Estudos com o propósito de identificar a variabilidade espacial no solo foram realizados por Reichardt et al. (1986), em que os autores afirmaram que a variabilidade espacial intrínseca do solo deve ser bem definida e que a teoria das variáveis regionalizadas é importante às metodologias existentes. Enfatizaram ainda, que os métodos geoestatísticos apresentaram respostas a várias questões impossíveis de serem respondidas pelos métodos clássicos.

Em geral, com o surgimento de novas metodologias, é comum considerá-las como a única a solucionar questões da pesquisa experimental, ignorando os métodos clássicos que também podem ser utilizados nas situações em que existe variabilidade espacial. Salienta-se que muitas metodologias apresentam limitações e, as mesmas devem ser associadas, para melhor interpretação dos resultados de uma pesquisa que investiga um fenômeno da natureza, com dependência espacial. Na década de 80, a maioria dos autores, analisaram as variáveis agronômicas com essa característica, pelo método geoestatístico empírico para identificar a continuidade espacial. Para estimar as semivariâncias, a exigência do estimador de Matheron (1963) é que a variável a ser analisada deve ter distribuição normal aproximada, pressuposição apresentada por Fisher (1935). Sendo assim, ambos os métodos podem ser utilizados para melhor interpretação da dependência espacial. Essa coerência pode ser avaliada no trabalho de Grondona et al. (1991), em que os autores incorporaram a estrutura de dependência, na análise de dados agronômicos provenientes de delineamentos experimentais planejados conforme Fisher (1935).

²Cressie et al. citados por Ribeiro Junior (1995) e Zamboti (2001), compararam 10 estimadores para seis conjuntos de dados simulados para uma transeção num processo autoregressivo e, concluíram que o estimador de Matheron, no cálculo das semivariâncias das variáveis, sob a pressuposição de normalidade, apresentava-se mais estável. Este fato vem confirmar que a estatística clássica é imprescindível e seus pressupostos são considerados também nos estudos geoestatísticos.

²CRESSIE, N.; HAWKINS, D.M. Robust estimation of the variogram, I. *Journal of the International Association for Mathematical Geology*, v.12, p.115-125, 1980.

Na mineração, os estudos da dependência espacial foram iniciados por Valente (1989), e não houve maior utilização devido aos escassos artigos teóricos envolvendo funções matemáticas com um certo grau de complexidade (Guerra, 1988; Braga, 1990).

Na condutividade hidráulica e porosidade de aquífero, de acordo com Samper & Ramirez (1990), é comum encontrar a variável regionalizada com aspecto aleatório associado a variações erráticas da variável e um aspecto geral estruturado, que reflete de certa forma as características globais de variação do fenômeno regionalizado. Estas variações da VR impedem seu tratamento ou descrição mediante uma função contínua e, acredita-se que no estudo desta devem ser levados em consideração dois aspectos da VR a possibilidade de uma representação simples da variabilidade espacial, e a consistência dos pontos de vista teórico, operativo e prático.

Em ecologia, o estudo da dinâmica espacial de diferentes espécies no ambiente também foi considerada (Rossi et al., 1992). Os autores enfatizaram, inicialmente, a necessidade da análise exploratória dos dados devido à sua importância, principalmente, quando existem dados discrepantes, pois estes afetam a interpretação da dependência espacial ao usar o variograma. Os autores utilizaram dados botânicos e zoológicos para descrever e ilustrar novos métodos geoestatísticos. Sob o ponto de vista dos autores é importante incorporar aos modelos as medidas de covariâncias, correlações e semivariâncias para a obtenção de resultados de análises geoestatísticas com interpretação correta do fenômeno em estudo .

Uma comparação entre os métodos que consideram a posição espacial de amostras e os experimentos clássicos foi realizada por Cressie (1993). O autor apresentou vários enfoques de dados espaciais, alguns métodos de estimativa por máxima verossimilhança e considerações sobre o método de Fisher (1935) que avalia a variabilidade experimental em ensaios de campo. Afirmou que os delineamentos em blocos completos casualizados utilizados para avaliar a variabilidade espacial são ineficientes, principalmente quando a estrutura de correlação espacial era maior ou igual a 0,5. Segundo esse autor, estes delineamentos são preferidos devido à facilidade de construção e foram muito utilizados em pesquisas com fertilidade do solo, visando controlar a heterogeneidade.

Nas pesquisas com irrigação utilizando o pivô central, por exemplo, em que foram utilizadas as ferramentas geoestatísticas, para avaliar a dependência espacial de atributos como: o silte, a argila e a areia, o método de amostragem considerado foi o sistematizado. No estudo destas propriedades do solo a aleatoriedade não garante uma boa distribuição dos pontos de amostragem na área. Tais mensurações são variáveis contínuas cujo valor em determinada

posição deve ser função da direção e da distância em relação às parcelas vizinhas. A correlação espacial só poderá ser avaliada se os valores das variáveis estiverem associados às coordenadas de uma malha regular (Cressie, 1993; Gonçalves, 1997).

Em mineração, por exemplo, se as posições das amostras são ignoradas, considerando que todas são aleatórias e com mesma probabilidade de serem escolhidas, as zonas mais ricas poderão ser descartadas impedindo assim, a obtenção de maiores produções de um referido minério. Em consequência deste fato, amostras provenientes de pontos adjacentes apresentando valores mais similares, não são levadas em conta. Além disso, existe a necessidade do conhecimento parcial da função densidade de probabilidade que representa a ocorrência da variável regionalizada (Landim, 1998).

Na agrometeorologia, a precipitação pluviométrica foi estudada por Atkinson & Lloyd (1998) que utilizaram a análise exploratória, a *krigagem* ordinária e a *krigagem* indicatriz. A *krigagem* ordinária não requer o prévio conhecimento da média que é calculada internamente e, por meio deste método, estima-se apenas a esperança da variável Z nos pontos desconhecidos. A *krigagem* indicatriz é uma das alternativas que possibilita não só a estimativa da esperança, mas de toda a função de distribuição acumulada da variável em cada ponto, tomando como base a transformação do conjunto de dados em variáveis binárias indicadoras, mantendo-se os procedimentos básicos da *krigagem* ordinária. Como os dados de chuva não apresentaram distribuição normal, os autores usaram uma distribuição lognormal. Os dados pluviométricos apresentaram anisotropia geométrica, ou seja, o comportamento da dependência espacial nas diferentes direções foi distinto e, o método de interpolação selecionado para estimar pontos amostrais não observados foi a *krigagem* ordinária.

Na agricultura, além da carência nutricional do solo existem outros fatores que limitam a obtenção de produções ótimas nas diferentes culturas como as pragas e as doenças que, geralmente, são de difícil controle.

O nematóide, por exemplo, é um parasita que vem danificando muitas culturas, e seu controle, até há alguns anos, era realizado com a aplicação de produtos químicos, os quais poluem o ambiente. No estudo desse parasita, na cultura do algodão, a geoestatística foi fundamental para o conhecimento da distribuição espacial e o monitoramento dos nematóides, permitindo ao pesquisador utilizar a rotação de culturas na redução populacional de nematóides, com precisão, nas áreas mais infestadas (Farias, 1999). O autor sugeriu o método geoestatístico associado aos conhecimentos biológicos e agronômicos para auxiliar no controle dos nematóides do algodoeiro, bem como no programa de agricultura de precisão.

A amostragem de nematóides foi realizada em uma malha regular em duas direções. Os modelos de variogramas teóricos ajustados foram: o exponencial, o esférico e o gaussiano, e o método de interpolação de pontos não amostrados foi a *krigagem*. Além disso, o autor acrescentou que, foi possível analisar as áreas de riscos, interpolar valores em mapas e identificar a distribuição espacial dos nematóides, sendo o modelo esférico o mais adequado nessa pesquisa com um alcance médio de 15 metros (Farias, 1999).

A dependência espacial, foi também avaliada no estudo de alguns micronutrientes do solo sob o pivô central no Mato Grosso do Sul (Couto & Klamt, 1999). Esses autores discutiram a deficiência de nutrientes como: boro, cobre, ferro, zinco e manganês e que os mesmos estão relacionados com a quantidade insuficiente de adubos aplicada. Por meio dos métodos geoestatísticos, como a *krigagem* e o semivariograma, detectou-se que o alcance da dependência espacial, no caso do manganês, aumentou com a profundidade, sendo respectivamente nas camadas de 0-20cm e 20-40cm de 385m e 412m. Afirmaram que as técnicas de manejo de solo ao longo dos anos interferem na estrutura da dependência espacial destas propriedades, na variabilidade destes micronutrientes e os valores dos coeficientes de variação expressam esta magnitude, quando observados coeficientes de variação superiores a 50% para o cobre e manganês e no caso do zinco, na camada de 20-40cm, o coeficiente obtido foi de 128%.

2.3 Análise exploratória de dados

Alguns autores enfatizam, que anterior à pesquisa geoestatística, e para certificar-se de que existe realmente uma estrutura de dependência espacial nos dados de pesquisa deve ser realizada uma análise exploratória elaborando histogramas, regressões e envelopes simulados sem considerar a posição geográfica. Esta análise permite: identificar *outliers* que interferem nos resultados de testes de normalidade e homogeneidade dos dados, indicar a forma da distribuição; orientar na escolha da análise a ser aplicada, auxiliar na decisão do tipo de estacionaridade que pode ser assumida, caracterizar a variabilidade, identificar em que local essa variabilidade é mais acentuada (Hamlett et al., 1986; Rossi et al., 1992; Ribeiro Junior, 1995; Libardi et al., 1996; Oliveira et al., 1998, Diggle & Ribeiro Junior, 2000), realizar mapas de contorno, gráficos tridimensionais e gráficos de médias *versus* variâncias, aplicar estatísticas baseadas em janelas

móveis, avaliar a normalidade, assimetria e curtose, retirar dados atípicos principalmente quando existem erros grosseiros com causas desconhecidas e adotar técnicas robustas de análise.

Além destas técnicas os autores descreveram outras técnicas como gráficos caule-e-folhas, gráficos de caixa, médias resumo, dispersão média *versus* variância e transformação de dados para adaptá-los às exigências do estudo espacial.

Maiores detalhes sobre este assunto podem ser encontrados em Isaaks & Srivastava (1989); Hoaglin et al. (1992); Cressie (1993); Ribeiro Junior (1995).

2.4 Análise estrutural

2.4.1 Objetivos

A análise estrutural comprehende uma fase muito importante para um estudo geoestatístico, pois condiciona os resultados da *krigagem* e simulação.

Os objetivos principais desta análise são os de:

- a) construir um modelo estatístico que descreva a variabilidade espacial do fenômeno em estudo e,
- b) analisar a continuidade espacial a qual se divide em duas partes:
 - b₁) determinar medidas da continuidade espacial a partir de dados disponíveis como: os semivariogramas (7) e correlogramas experimentais e
 - b₂) modelar as medidas experimentais através do ajuste por um modelo analítico, para posterior uso nas etapas de estimação e simulação.

Dentre as medidas de continuidade espacial o uso dos semivariogramas eq. (7) é preferido e a covariância é indicada nos casos em que não se pode obter a ergodicidade de médias, ou seja, quando um processo estocástico não pode ser representado por uma única realização da variável aleatória (Cressie, 1993).

2.4.2 Variograma e Covariância

O variograma é uma função matemática definida para representar o nível de dependência entre duas variáveis aleatórias regionalizadas locais e, para exemplificar, pode ser o grau de continuidade das propriedades da fertilidade do solo. Ele é usado para modelar dois

valores correlacionados no espaço ou no tempo e é pré-requisito para a *krigagem*. A modelagem do variograma é uma modelagem de cada estrutura de correlação espacial.

Para se obter a estimativa dos parâmetros de um modelo baseado na teoria apresentada no item (4), nas condições de (12), (13) e (19) pode-se obter um estimador para o variograma, que depende somente da separação do vetor \mathbf{h} da seguinte forma:

$$2\hat{\alpha}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{i=1}^{N(\mathbf{h})} \left[[z(x_i) - z(x_j)] \right]^2, \quad \mathbf{h} \in \mathbb{R}^d \quad (20)$$

em que, \mathbf{h} é um vetor, $|N(\mathbf{h})|$ é o número de elementos distintos de $N(\mathbf{h})$

$$N(\mathbf{h}) \equiv \{(x_i, x_j) : x_i - x_j = \mathbf{h}; i, j = 1, \dots, n\}. \quad (21)$$

A expressão apresentada em (20) dividida por 2, chamada de estimador de Matheron, vem sendo utilizada por vários autores para medir a estrutura da dependência espacial (Journel & Huijbregts, 1978; Braga, 1990; Cressie, 1993; Ribeiro Junior, 1995).

Outra medida de continuidade espacial importante é o estimador da covariância centrada apresentada pela seguinte expressão:

$$\hat{C}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{i=1}^{N(\mathbf{h})} [z(x_i) - m_z][z(x_j) - m_z] \quad (22)$$

em que, m é a média, \mathbf{h} é um vetor e $|N(\mathbf{h})|$ é o número de elementos distintos de $N(\mathbf{h})$ apresentada em (21).

2.4.3 Estimador Robusto do Variograma

Através do estimador de Matheron, dado pela eq. (20), pode ser obtido o estimador robusto o qual é usado quando existem valores discrepantes ou “outliers” (Cressie, 1993). Quando a variável aleatória segue uma distribuição Gaussiana, o incremento quadrático $[Z(x + \mathbf{h}) - Z(x)]^2$ tem distribuição χ_1^2 com um grau de liberdade e média igual ao variograma.

Um estimador robusto menos sensível aos valores extremos sugerido por alguns autores para solucionar alguns problemas inerentes ao variograma amostral, é dado por:

$$2\bar{g}(\mathbf{h}) = \frac{\left[\frac{1}{|N(\mathbf{h})|} \sum_{i=1}^{N(\mathbf{h})} |Z(\mathbf{x}_i) - Z(\mathbf{x}_i + \mathbf{h})|^{\frac{1}{2}} \right]^4}{0,457 + 0,494} \quad |N(\mathbf{h})| \quad (23)$$

e, que pode ser escrito como:

$$2\tilde{g}(\mathbf{h}) \equiv [med \{ |Z(x_i) - Z(x_i + \mathbf{h})|^{\frac{1}{2}} : (x, x + \mathbf{h}) \in N(\mathbf{h}) \}]^4 * \frac{1}{B(\mathbf{h})}, \quad (24)$$

em que, med {} denota a mediana da seqüência {} e $B(\mathbf{h})$ é a correção do viés (assintoticamente, $B(\mathbf{h}) = 0,457$).

Uma vez conhecidas as hipóteses restritivas (eqs. de 8 a 13) para realizar inferência entre as variáveis aleatórias regionalizadas e o semivariograma amostral, é necessário conhecer novos modelos e seus parâmetros para se estudar a estrutura de dados com continuidade espacial.

2.4.4 Tipos de Variograma

Num estudo em que se deseja averiguar a dependência espacial de amostras georreferenciadas por meio do variograma ou semivariograma têm-se três tipos de situações variográficas, isto é, o variograma observado, o variograma verdadeiro e o variograma teórico (Guerra, 1988).

O variograma observado ou experimental é a primeira informação gráfica que se obtém sobre os dados, o qual contém informação da escala de flutuações da variável e, por meio dele, pode-se inferir a variabilidade da distribuição espacial em relação às escalas espaciais. Este é proveniente do conjunto de dados da pesquisa realizada, originados de um processo de amostragem sobre coordenadas geográficas, antes de qualquer ajuste de modelos.

O variograma verdadeiro é aquele que representa a situação real de um micronutriente no solo do campo experimental, ou do depósito de algum minério e é sempre desconhecido (Kitanidis, 1997). Assim, na análise estrutural o objetivo é estudar qual é o

variograma teórico que melhor se ajusta ao variograma experimental, de forma que a partir do variograma teórico, realizam-se inferências sobre o variograma verdadeiro (Guerra, 1988).

2.4.5 Características dos parâmetros que descrevem um semivariograma

A construção do semivariograma para identificar a dependência espacial origina-se nas diferenças entre dois pontos georreferenciados no espaço, separados por uma distância \mathbf{h} .

Na presença de dependência espacial, em geral, duas medidas próximas são mais similares que duas mais afastadas permitindo afirmar que $g(\mathbf{h})$ aumenta à medida que \mathbf{h} aumenta até estabilizar. Este ponto, onde os dados deixam de apresentar dependência espacial, é chamado de patamar ou *sill* e, no semivariograma, a distância, a partir da origem até o patamar é chamada de alcance ou *range* denotado neste trabalho por \mathbf{f} .

Pela definição $g(\mathbf{h})$ para $\mathbf{h}=0$ deverá ser zero, mas na prática observa-se que há casos em que $\mathbf{h}=0$ e $g(\mathbf{h})$ aproxima-se de um valor positivo denominado efeito pepita, ou efeito *nugget* (Cressie, 1993; Ribeiro Junior, 1995; Kitanidis, 1997; Gonçalves, 1997). O efeito pepita ou efeito *nugget* é simbolizado por t^2 e evidencia a descontinuidade do semivariograma para medidas menores do que a menor medida das amostras e é o valor de γ para $\mathbf{h}=0$ (Figura 2).

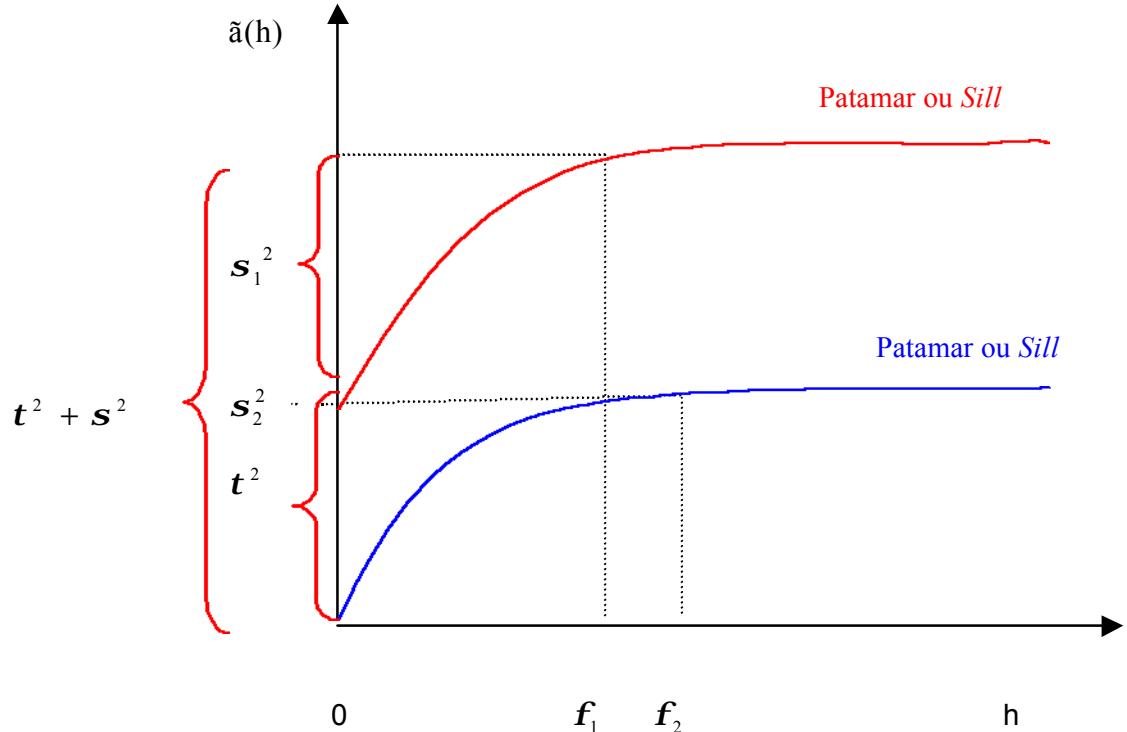


Figura 2 - Semivariograma hipotético apresentando variância *nugget* diferente de zero = $t^2 + s^2$ e com variância *nugget* igual a zero passando pela origem ($\hat{\sigma}^2$).

Na Figura 2 podem-se identificar os parâmetros dos semivariogramas, com e sem efeito pepita, da forma que segue : t^2 é o efeito pepita ou *nugget*, s_1^2 e s_2^2 são as variâncias totais com e sem efeito pepita e patamar ou *sill* f_1 e f_2 são os alcances ou *ranges* para o gráfico com e sem efeito pepita ou *nugget*.

Alguns autores afirmam que o fato de ocorrer efeito pepita, pode estar associado a algum erro de medição e, que resultados com efeito pepita maior do que zero com tangente vertical têm sido encontrados na precipitação pluvial (Vieira,1996). Este comportamento também pode ser encontrado nas pesquisas com minerais de distribuição geralmente errática, como depósitos de ouro (Guerra, 1988). Na Figura 2 pode-se observar, que nos dois gráficos, com e sem efeito pepita, os valores hipotéticos de $g(h)$ foram sempre crescentes.

Em geral, quando o semivariograma apresenta as características da Figura 2, patamar claro e bem definido, subjetivamente, indica que as amostras estudadas atendem à

hipótese de estacionaridade de segunda ordem (eqs. de 8 a 13). Outra característica é a de que se existe dependência entre dois pares de variáveis aleatórias, existe a covariância e, consequentemente, variância finita (C_0), dada por (15), características preconizadas para atender à hipótese mencionada anteriormente.

Existe, também, o caso de variável com efeito pepita puro, ou seja, como se o fenômeno fosse totalmente aleatório permitindo ser utilizada a teoria das probabilidades. O gráfico apresenta uma reta paralela ao eixo das abscissas cortando o eixo da semivariância. Este caso, não se aplica na indústria de mineração e pode estar associado a uma malha de sondagem que não foi suficientemente apertada para detectar a estrutura (Guerra, 1988).

Outra preocupação também a ser considerada num estudo geoestatístico é a de como estão os atributos na região de estudo, ou seja, em que direção as distâncias são medidas. Justifica desta forma avaliar se o fenômeno é isotrópico ou anisotrópico.

2.4.6 Modelos anisotrópico e isotrópico

A análise da dependência espacial, usando-se o semivariograma, tem contribuído para a compreensão de muitos aspectos referentes ao estudo da fertilidade do solo, e um deles é a variação que pode ser anisotrópica ou isotrópica. Antes de iniciar esta pesquisa da anisotropia é necessário que seja verificado se a variável aleatória $Z(x)$ não é estacionária ou nem ao menos intrínseca, pois é a condição necessária para um estudo de dados com dependência espacial.

Sabe-se que o semivariograma depende da direção em que é calculado e é uma ferramenta que permite verificar a anisotropia que pode estar ocorrendo no solo pesquisado. Esta também aparece com freqüência nas pesquisas geológicas, e a sugestão para estudá-la é calcular o semivariograma em várias direções. Se os semivariogramas direcionais são distintos, há que se pensar na presença de anisotropia (Guerra, 1988; Samper & Ramirez, 1990). Nos casos em que os semivariogramas variam de maneira semelhante em todas as direções, é chamado de isotrópico, e estes dependem somente da distância \mathbf{h} (Trangmar et al., 1985).

Existem dois tipos de anisotropia, a anisotropia geométrica e a zonal. A anisotropia é dita geométrica quando variogramas construídos em diferentes direções apresentam a mesma variabilidade global (em particular o mesmo patamar), mas apresentando diferentes alcances (ϕ).

No caso da anisotropia zonal os variogramas construídos sobre distintas direções apresentam características de variabilidade diferentes, isto é, diferentes patamares, mas mantêm o alcance constante (Guerra, 1988; Ribeiro Junior, 1995).

A anisotropia é muito comum também nos fenômenos naturais geológicos, pedológicos, meteorológicos e hidrogeológicos.

2.4.7 Modelos teóricos de semivariograma

Para se obterem estimativas de reservas de minérios, ou da quantidade de nutrientes no solo, ou nível de contaminação de poluentes é necessário ajustar semivariogramas teóricos que representem os semivariogramas experimentais. Os modelos teóricos que existem na literatura não surgiram espontaneamente e nem foram estabelecidos arbitrariamente. Eles correspondem a diferentes regionalizações que se apresentam na natureza (Guerra, 1988).

O método de ajuste de variogramas teóricos a variogramas experimentais é denominado de método das aproximações sucessivas, pois o processo de ajuste finaliza quando as discrepâncias entre os valores experimentais e teóricos forem mínimas.

Existem dois tipos de modelos de variogramas teóricos: os modelos com e sem patamar e, estes podem, ou não, apresentar o efeito pepita (Journel & Huijbregts, 1978; Guerra, 1988). Dentre os que apresentam patamar os mais utilizados são: o modelo esférico (Matheron), o modelo exponencial (Formery) e o modelo gaussiano (Gauss). Entre os modelos sem patamar encontram-se o linear e o logarítmico (Wijs).

2.4.7.1 Modelos com patamar

Os modelos com patamar têm como parâmetros: t^2 que é o efeito pepita, σ^2 que é a variância espacial e f é o alcance. Dentre eles têm-se:

a) Modelo Esférico ou Modelo de Matheron

Este modelo apresenta um crescimento rápido na origem, alcança um patamar a uma distância h finita e é representado por:

$$g(h) = \begin{cases} t^2 + s^2 \left[\frac{3h}{2f} - \frac{1}{2} \left(\frac{h}{f} \right)^3 \right] & \text{para } 0 < h \leq f, \\ t^2 + s^2 & \text{para } h > f \\ 0 & \text{para } h = 0 \end{cases}$$

A desvantagem desta estrutura é a de apresentar superfícies de verossimilhança multimodal (Laslett, 1994).

b) Modelo Exponencial ou de Formery

Este modelo apresenta um comportamento linear na origem, alcança seu patamar somente assintoticamente e é representado por:

$$g(h) = \begin{cases} t^2 + s^2 \left[1 - e^{-\frac{h}{f}} \right] & \text{para } 0 < h \leq f, \\ t^2 + s^2 & \text{para } h > f \\ 0 & \text{para } h = 0 \end{cases}$$

c) Modelo Gaussiano ou Parabólico

O modelo gaussiano apresenta comportamento parabólico na origem e tem como equação o modelo a seguir:

$$g(h) = \begin{cases} t^2 + s^2 \left[1 - e^{-\frac{h^2}{f^2}} \right] & \text{para } 0 < h \leq f, \\ t^2 + s^2 & \text{para } h > f \\ 0 & \text{para } h = 0 \end{cases}$$

2.4.7.2 Modelos sem patamar

a) Modelo linear

O modelo linear é definido para o intervalo $0 < f < 2$ e $s^2 > 0$ e é expresso por:

$$\mathbf{g}(h) = s^2 * h^f$$

sendo que s^2 e f têm o mesmo significado que nos modelos anteriores.

b) Modelo logarítmico ou de Wijs

Foi um dos primeiros modelos a ser utilizado em aplicações geoestatísticas realizadas para depósitos de ouro e urânio, por Krige³, citado por (Guerra, 1988), devido às suas propriedades analíticas. É inadequado para $h=1$ e é representado por:

$$\gamma(h) = \hat{\sigma}^2 + 3\alpha \ln(h)$$

em que : τ^2 é o efeito pepita e α é uma constante conhecida como dispersão absoluta.

2.5 Modelo Matérn

Outras metodologias que estudam a dependência espacial e a obtenção de semivariogramas ajustados foi proposto também por Diggle & Ribeiro Junior (2000), em que os autores sugerem este estudo através da modelagem utilizando a distribuição Gaussiana. Comentam sobre as diferentes estruturas de covariância utilizadas no estudo espacial e sugerem uma nova estrutura da função de correlação denominada de família Matérn. Esta função apresenta um parâmetro k chamado de ordem do modelo Matérn o qual determina a suavização analítica do sinal $S(x)$ num modelo Gaussiano. A função de correlação de Matérn foi definida por:

$$r(h, f, k) = \left\{ 2^{k-1} \Gamma(k) \right\}^{-1} \left(\frac{h}{f} \right)^k K_k \left(\frac{h}{f} \right)$$

em que,

(f, k) são parâmetros e,

K_k é a função de Bessel de terceiro tipo de ordem k

³ KRIGE, D. A statistical approach to some basic mine valuation problems on the Witwatersrand. **Journal of the Chemical, metallurgical and Mining Society of South Africa**, v. 52, p.119-139, 1951.

A família Matérn é válida para f e $k > 0$ e, no caso em que $\kappa = 0,5$, corresponde a função de correlação Exponencial, apresentada por $r(h) = \exp\left(\frac{-h}{f}\right)$ e para $k \rightarrow \infty$ equivale a função de correlação Gaussiana.

Os autores enfatizaram que a vantagem de se utilizar a família Matérn é que o parâmetro k controla a diferenciabilidade do processo latente $S(x)$ de uma maneira direta e, sendo assim, a parte inteira de k dá o número de vezes que $S(x)$ tem quadrado médio diferenciável. Os autores concluíram que a família Matérn é provavelmente a melhor escolha quanto uma função de correlação somente com dois parâmetros para uso generalizado.

2.6 Método de interpolação pela *krigagem*

No estudo geoestatístico não-paramétrico, além de se avaliar a estrutura da variabilidade espacial, por meio do variograma ou semivariograma, existe o interesse na estimativa não viesada e de variância mínima dos parâmetros para a variável regionalizada e nas propriedades estruturais do semivariograma obtido para um determinado conjunto de dados (Trangmar et al., 1985; Ribeiro Junior, 1995; Gonçalves, 1997). Isso está relacionado ao fato de que se deseja o conhecimento de um determinado atributo em estudo fora da região amostrada. A estimativa é obtida através do estimador linear geoestatístico denominado de *krigagem*, sendo uma combinação linear de valores amostrados pelo peso atribuído a cada vizinho. O preditor de *krigagem* pressupõe ausência de erros de medidas e variogramas corretos. É a média móvel ponderada dos valores observados na vizinhança em que os vizinhos mais próximos têm peso maior e, os mais afastados, pesos cada vez menores, nulos ou até negativos, (Samper & Ramirez, 1990; Menk, 1990; Cressie, 1993; Ribeiro Junior, 1992, 1995). O método para determinar a vizinhança na estimativa de *krigagem* é um misto entre os métodos da distância constante e dos vizinhos constantes (Vieira, 1996).

Este preditor foi utilizado pela primeira vez por Matheron (1963), e este nome foi dado em homenagem aos trabalhos pioneiros de dependência espacial na mineração, (Soares, 2000). É obtido por:

$$\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i z(x_i) \quad (25)$$

em que: $\hat{Z}(x_0)$ é o preditor para um ponto x_0 , n é o número de vizinhos utilizados na predição, λ_i é o peso atribuído a cada vizinho e $z(x_i)$ é o valor observado em cada vizinho. Deve-se ter $\sum_{i=1}^n \lambda_i = 1$, para garantir uma estimativa não viesada.

Vários são os interpoladores para estimar valores em pontos não amostrados. Existe uma família que cobre os estimadores não-estacionários (*krigagem simples*, *krigagem ordinária*, *krigagem universal*, *krigagem indicadora* e o estimador de funções de distribuição de probabilidade) e dentre os estimadores não-lineares encontram-se a *krigagem multigaussiana* e a *disjuntiva* (Soares, 2000).

Um outro fator considerado pelo autor é que a *krigagem* além de levar em conta a distância ao ponto a ser estimada, também considera as distâncias entre os vizinhos agrupados, sendo esta característica denominada de desagregação.

Considerando que $\sum_{i=1}^n \lambda_i = 1$, o preditor $\hat{Z}(x_0)$ utiliza alguma medida de associação

entre os pontos que está relacionada com a estacionaridade de segunda ordem ou com a hipótese intrínseca aos dados, e esta medida pode ser tanto a covariância quanto a semivariância como apresentadas em (19) e (20). Estas estatísticas não podem ser estimadas a partir de pontos amostrais e, portanto, é necessário um modelo. O sistema de *krigagem* com solução única será assegurado somente se a matriz de covariâncias for positiva definida, indicando assim que o seu determinante deve ser positivo ou zero. Se, ao invés de utilizar a covariância, for usada a semivariância na matriz, o modelo de semivariograma deve ser negativo semidefinido condicional-NSDC (Gonçalves, 1997).

Como ilustração será aqui apresentada a obtenção do preditor de *krigagem* ordinária pontual conforme apresentado por Ribeiro Junior (1995), em que se assumiu a hipótese intrínseca. A Figura 3 ilustra uma situação em que o valor da variável em um ponto x_0 será obtido pela ponderação dos valores $[z(x_1), z(x_2), z(x_3), \dots, z(x_n)]$ observados em n vizinhos. Os pesos λ_i serão diretamente obtidos ao assegurar as condições de preditor ótimo.

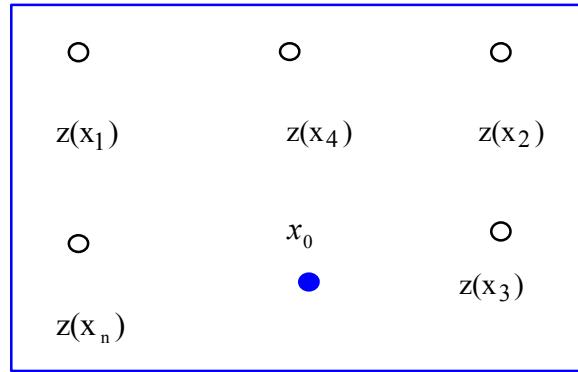


Figura 3- Ponto a ser estimado numa região D e valores observados na vizinhança.

Usando-se o estimador dado por (25), $Z(x_0)$ o verdadeiro e desconhecido valor da variável no ponto x_0 e as hipóteses apresentadas nos itens 2.1.6 e 2.1.6.1 satisfazendo-se a condição de não-tendenciosidade, pode-se obter a variância de estimação da diferença entre o estimador e o valor verdadeiro bem como o sistema de equações de *krigagem* apresentados na seqüência (Samper & Ramirez, 1990; Ribeiro Junior, 1995).

Sabendo-se que:

$$\begin{aligned} E[\hat{Z}(x_0) - Z(x_0)] &= 0, \\ E\left[\sum_{i=1}^n \lambda_i Z(x_i) - Z(x_0)\right] &= 0, \\ \sum_{i=1}^n \lambda_i E[Z(x_i)] - E[Z(x_0)] &= 0. \end{aligned} \quad (26)$$

Aplicando a esperança matemática em (26) e isolando-se o somatório dos pesos obtém-se a prova da unicidade mencionada anteriormente.

$$\sum_i^n I_i \mathbf{m} - \mathbf{m} = 0, \quad \mathbf{m} \sum_i^n I_i - \mathbf{m} = 0, \quad \mathbf{m} \sum_i^n I_i = \mathbf{m}, \quad \sum_i^n I_i = 1,$$

Os coeficientes de ponderação são obtidos, minimizando-se o erro da estimação $[\hat{Z}(x_0) - Z(x_0)]$, sendo a variância da estimação dada por:

$$\text{Var}[\hat{Z}(x_0) - Z(x_0)] = E[\hat{Z}(x_0) - Z(x_0)]^2 - \{E[\hat{Z}(x_0) - Z(x_0)]\}^2. \quad (27)$$

Através das hipóteses apresentadas nos itens 2.1.6 e 2.1.6.1 e das eqs. (25) e (27) tem-se a seguinte expressão para a variância:

$$\text{Var}[\hat{Z}(x_0) - Z(x_0)] = - \sum_i \sum_j \lambda_i \lambda_j \gamma(x_i, x_j) + 2 \sum_i \lambda_i \gamma(x_i, x_0). \quad (28)$$

Incluiu-se em (28) o multiplicador de Lagrange que, permitiu solução para o sistema de equações de *krigagem* (Samper & Ramirez, 1990; Ribeiro Junior, 1995) como na expressão a seguir.

$$\text{Var}[\hat{Z}(x_0) - Z(x_0)] = - \sum_i \sum_j I_i I_j g(x_i, x_j) + 2 \sum_i I_i g(x_i, x_0) - 2m(\sum_i I_i).$$

Desta expressão, calculando-se as derivadas parciais de primeira ordem em relação aos parâmetros e igualando-se a zero obtém-se o sistema de equações de *krigagem* como segue:

$$\left\{ \begin{array}{l} \sum_{i=1}^n \hat{\lambda}_i \gamma(x_i, x_0) + \hat{\mu} = \gamma(x_0, x_0) \\ \sum_{i=1}^n \hat{I}_i = 1. \end{array} \right. \quad (29)$$

2.7 Outros métodos de estimação dos parâmetros do semivariograma

2.7.1 Método de estimação empírico, mínimos quadrados ordinários, mínimos quadrados generalizados e mínimos quadrados ponderados.

Na literatura mencionam-se alguns métodos de ajuste de modelo de semivariograma e o mais comum é o ajuste a *sentimento*. Por meio deste escolhe-se um modelo e seus

parâmetros em que é sobreposto aos pontos estimados, e a partir de uma nuvem de pontos obtém-se a estimativa dos parâmetros do semivariograma por meio de uma avaliação visual (Ribeiro Junior, 1995). O autor sugere que o processo deve ser repetido até que se obtenha um modelo adequado, e recomendou que seja utilizado o ajuste empírico como valor inicial para estimar os parâmetros do modelo de interesse por outro método em que as estimativas dos parâmetros sejam mais precisas. Afirma ainda, que este ajuste subjetivo é questionado e a cada vez menos será utilizado.

Alguns métodos de ajuste que minimizam as somas de quadrados foram também utilizados, como o de mínimos quadrados ordinários e, considerado insensível para o propósito de ajuste, entretanto é um dos que permite a implementação de forma simplificada (Ribeiro Junior, 1995).

McBratney & Webster (1986) e Ribeiro Junior (1995), afirmam ser o método de mínimos quadrados, pouco rigoroso e não fidedigno para estimar os parâmetros de um semivariograma. Os autores estudaram também os métodos dos mínimos quadrados generalizados e os de máxima verossimilhança e, apresentaram para os três métodos de ajuste dos parâmetros do semivariograma a expressão que permite minimizar os resíduos. Para o método de quadrados mínimos tem-se a seguinte expressão:

$$L(\mathbf{t}; \mathbf{g}_k) = \sum_{j=1}^k \{\mathbf{g}(\mathbf{h}_j) - \mathbf{g}(\mathbf{h}_j; \mathbf{t})\}^2. \quad (30)$$

em que, \mathbf{t} representa o vetor de estimativas de parâmetros de dimensões $p \times 1$, \mathbf{g}_k representa o vetor de estimativas de semivariâncias de dimensões $k \times 1$, cada estimativa inicial denotada por $\gamma(\mathbf{h}_j)$ e, $\gamma(\mathbf{h}_j; \mathbf{t})$ é a j -ésima semivariância esperada bidimensional a partir do modelo ajustado e depende das estimativas dos parâmetros de \mathbf{t} .

Neste método de ajuste, McBratney & Webster (1986) assumiram que os resíduos obtidos pela eq. (30) são normalmente distribuídos e independentes, e todas as semivariâncias estimadas com a mesma variância. Segundo eles estas pressuposições são criticadas, e que, além disso, através deste método não se leva em consideração que as semivariâncias entre dois pontos sejam correlacionadas. Samper & Ramirez (1990), concordam que se não for considerada a correlação, os resultados são comprometidos quando se aumenta o número de observações.

Esta polêmica foi contornada quando McBratney & Webster (1986), que utilizaram o método de mínimos quadrados generalizados, obtiveram estimativas eficientes assintoticamente, mas com a exigência de inversão de uma matriz de grandes dimensões. A expressão utilizada para o ajuste dos parâmetros é a que segue:

$$G(\mathbf{t}; \tilde{\mathbf{a}}_k) = \{ \tilde{\mathbf{a}}_k - \tilde{\mathbf{a}}(\mathbf{t}) \}^T \tilde{\mathbf{O}}_k^{-1} \{ \tilde{\mathbf{a}}_k - \tilde{\mathbf{a}}(\mathbf{t}) \}$$

em que, $\tilde{\mathbf{a}}_k$ representa o vetor de estimativas de semivariâncias $k \times 1$, $\tilde{\mathbf{a}}(\mathbf{t})$ é o vetor $k \times 1$ das semivariâncias estimadas que depende do vetor das estimativas dos parâmetros, $(\tilde{\mathbf{a}}_k - \tilde{\mathbf{a}}(\mathbf{t}))$ é o vetor dos resíduos de dimensões $k \times 1$ e, $\tilde{\mathbf{O}}_k$ é a matriz de variâncias e covariâncias de dimensões $k \times k$ entre as estimativas de $\gamma(\mathbf{h}_j; \tau)$.

Os autores, neste mesmo trabalho estudando o método dos mínimos quadrados ponderados, consideraram-no mais real e computacionalmente eficiente e que o modelo é ajustado minimizando a seguinte equação:

$$W(\mathbf{t}; \mathbf{g}_k) = \frac{\sum_{j=1}^k \{ \mathbf{g}(\mathbf{h}_j) - (\mathbf{g}(\mathbf{h}_j; \mathbf{t}) \}^2}{\text{var}[\mathbf{g}(\mathbf{h}_j)]}$$

em que, $\text{var}[\mathbf{g}(\mathbf{h}_j)]$ é a variância da estimativa da semivariância.

2.7.2 Métodos de estimativas dos parâmetros de modelos com dependência espacial.

2.7.2.1 Máxima Verossimilhança

No estudo de modelos mistos, para a predição e avaliação de tendências é necessário ter conhecimento dos componentes de variâncias e da covariância das observações. Vários são os métodos estudados para solucionar estas questões e dois métodos têm sido abordados na literatura, com os avanços da computação científica e o desenvolvimento de algoritmos iterativos mais eficientes: o estimador de máxima verossimilhança (Hartley & Rao, 1967) e de máxima verossimilhança restrita (Patterson & Thompson, 1971).

O método de máxima verossimilhança consiste em maximizar a função de densidade de probabilidades, em relação aos efeitos fixos e aos componentes de variâncias dos efeitos aleatórios do modelo. Esse método é iterativo e fornece sempre estimativas positivas de componentes de variância e, no melhoramento genético elimina o viés atribuído às mudanças nas freqüências gênicas, se for considerado o parentesco entre indivíduos. Existem alguns problemas no uso deste estimador devido ao fato deste não considerar a perda de graus de liberdade resultante da estimação dos efeitos fixos do modelo.

McBratney & Webster (1986), utilizaram este estimador e afirmaram que, geralmente, é mais eficiente. Mardia⁴ & Marshall (1984), citados pelos autores anteriores desenvolveram teoria estatística para a estimação de máxima verossimilhança para o caso em que uma função aleatória $Z(x)$ é Gaussiana e de estacionaridade fraca. De acordo com os autores, as estimativas de máxima verossimilhança de μ , a média, e \mathbf{b} os p parâmetros do covariograma, são maximizados pela equação que se segue:

$$L_N(\mathbf{b}; \mathbf{z}_N) = -0,5 \ln |\mathbf{\hat{O}}| - 0,5 (\mathbf{z}_N - \mathbf{1}\mu)^T \mathbf{\hat{O}}^{-1} (\mathbf{z}_N - \mathbf{1}\mu)$$

em que, \mathbf{b} é um vetor de parâmetros combinados de dimensões $(1+p) \times 1$, isto é, $\hat{\mathbf{a}} = [i, i]$, $\mathbf{Z}_N = [z(x_1), z(x_2), \dots, z(x_N)]^T$ é o vetor de observações de N locais amostrados de dimensão $N \times 1$; $\mathbf{\hat{O}}_N$ é a matriz de variâncias e covariâncias entre todos os locais, e $\mathbf{1}$ é um vetor de uns de dimensão $N \times 1$.

Uma dificuldade encontrada em pesquisa na área de hidrologia é o número pequeno de dados, principalmente dos atributos condutividade elétrica, porosidade e ponto piezométrico (Pardo-Igúzquiza, 1997). Os valores destes atributos, em geral, apresentam variabilidades espaciais e a modelagem estocástica destas variabilidades é de grande interesse prático. De acordo com o autor na estimação por máxima verossimilhança assume-se que os dados seguem uma distribuição multivariada Gaussiana com função de densidade de probabilidades conjunta dada por:

⁴ MARDIA, K.V. & MARSHALL, R.J. Maximum likelihood models for residual covariance in spatial regression. **Biometrika**. v.71, N° 1, p.135-146, 1984.

$$f(\mathbf{y}; \hat{\mathbf{a}}, \hat{\mathbf{e}}) = (2\delta)^{-\frac{n}{2}} |\hat{\mathbf{O}}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\hat{\mathbf{a}})^T \hat{\mathbf{O}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{a}})\right] \quad (31)$$

em que, \mathbf{y} é o vetor de dimensão $n \times 1$ dos dados experimentais, $\hat{\mathbf{e}}$ são os parâmetros da covariância e $|\hat{\mathbf{O}}|$ é a matriz de variância e covariância de dimensões $n \times n$ das n observações.

Neste caso os autores apresentaram a matriz de variância e covariância como:

$$\hat{\mathbf{O}} = E[(\mathbf{y} - \mathbf{X}\hat{\mathbf{a}})(\mathbf{y} - \mathbf{X}\hat{\mathbf{a}})^T] = E[\mathbf{e}\mathbf{e}^T].$$

A matriz $\hat{\mathbf{O}}$ pode ser fatorada como $\hat{\mathbf{O}} = \mathbf{s}^2 \mathbf{Q}$, sendo \mathbf{s}^2 a variância e \mathbf{Q} a matriz de correlação. Considerando a inversa $\hat{\mathbf{O}}^{-1} = \frac{1}{\delta^2} * \mathbf{Q}^{-1}$ e $|\hat{\mathbf{O}}|^{-\frac{1}{2}} = \delta^{-n} |\mathbf{Q}|^{-\frac{1}{2}}$ a função de densidade de probabilidades pode ser escrita como:

$$f(\mathbf{y}; \hat{\mathbf{a}}, \delta^2, \hat{\mathbf{e}}) = (2\delta)^{-\frac{n}{2}} \delta^{-n} |\mathbf{Q}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2\delta^2} [\mathbf{y} - \mathbf{X}\hat{\mathbf{a}}]^T \mathbf{Q}^{-1} [\mathbf{y} - \mathbf{X}\hat{\mathbf{a}}]\right] \quad (32)$$

em que $\hat{\mathbf{e}}$ é o conjunto de parâmetros da matriz de covariâncias excluindo a variância.

Através da eq. (32) pode-se ainda escrever a função de verossimilhança a qual é dada por $L(\hat{\mathbf{a}}, \sigma^2, \hat{\mathbf{e}}; \mathbf{y}) = \prod_{i=1}^n f(y_i; \hat{\mathbf{a}}, \sigma^2, \hat{\mathbf{e}})$.

Considerando-se o logarítmico da função de verossimilhança, $l = \ln(L(\hat{\mathbf{a}}, \sigma^2, \hat{\mathbf{e}}; \mathbf{y}))$, estima-se o conjunto de parâmetros $(\hat{\mathbf{a}}, \delta^2, \hat{\mathbf{e}})$, como os valores que maximizam

$$\begin{aligned} l(\hat{\mathbf{a}}, \sigma^2, \hat{\mathbf{e}}; \mathbf{y}) &= \ln\left[L(\hat{\mathbf{a}}, \sigma^2, \hat{\mathbf{e}}; \mathbf{y})\right] = -\frac{n}{2} \ln(2\pi) - n \ln(\sigma) - \\ &- \frac{1}{2} \ln |\mathbf{Q}| - \frac{1}{2} [\mathbf{y} - \mathbf{X}\hat{\mathbf{a}}]^T \mathbf{Q}^{-1} [\mathbf{y} - \mathbf{X}\hat{\mathbf{a}}], \end{aligned} \quad (33)$$

Obtém-se as estimativas de máxima verossimilhança $\hat{\mathbf{a}}$ e σ^2 a partir das equações de verossimilhança, as quais são obtidas pelas derivadas parciais de $l(\hat{\mathbf{a}}, \sigma^2, \mathbf{e}; \mathbf{y})$ em relação a esses parâmetros $\hat{\mathbf{a}}$ e σ^2 . Estas equações quando igualadas a zero fornecem:

$$\hat{\mathbf{a}} = (\mathbf{X}^T \mathbf{Q}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Q}^{-1} \mathbf{y}, \text{ e} \quad (34)$$

que coincide com a estimativa de mínimos quadrados generalizados de $\hat{\mathbf{a}}$.

$$\hat{\mathbf{s}}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\mathbf{a}})^T \mathbf{Q}^{-1} (\mathbf{y} - \mathbf{X}\hat{\mathbf{a}}), \quad (35)$$

A estimativa da matriz de variância e covariância de $\hat{\mathbf{a}}$ é dada por $\text{Cov}(\hat{\mathbf{a}}) = \hat{\sigma}^2 (\mathbf{X}^T \mathbf{Q}^{-1} \mathbf{X})^{-1}$ e a variância $\hat{\sigma}^2$ por $\text{Var}(\hat{\mathbf{s}}^2) = \frac{2(\hat{\mathbf{s}}^2)^2}{n}$.

As estimativas dos parâmetros de \mathbf{e} são obtidas maximizando-se: $l(\hat{\mathbf{a}}, \hat{\sigma}^2, \mathbf{e}; \mathbf{y})$. Para maiores detalhes ver Pardo-Igúzquiza (1997).

O método de máxima verossimilhança foi também estudado por Park & Baek (2001), para estimar parâmetros de modelos com estrutura espacial. Os autores obtiveram estimativas dos parâmetros de um modelo linear espacial Gaussiano, de uma família de covariâncias exponencial potência. Além disso, desenvolveram um programa computacional em linguagem Fortran, especificamente para o modelo citado com o estimador de máxima verossimilhança. Entre muitas funções de covariâncias possíveis a expressão considerada para a família exponencial potência foi descrita pelos autores como:

$$\text{Cov}[Z(x_i), Z(x_j)] = \mathbf{S}^2 * e^{-\sum_{k=1}^d q_k |x_{ik} - x_{jk}|^{\alpha_k}} + \mathbf{S}_e^2 \delta_{ij}$$

em que, $\theta_k \geq 0$, $0 < \alpha_k \leq 2$ para todo k , σ^2_e é o efeito pepita ou *nugget* e δ_{ij} é o delta Kronecker.

Apesar de a família de covariância exponencial potência permitir muita flexibilidade na modelagem da dependência espacial, Park & Baek (2001) sugeriram que deve ser considerado o estimador de máxima verossimilhança restrita ou residual.

A motivação para o uso de métodos estatísticos fidedignos, no estudo geoestatístico foi a preocupação principal de alguns autores. São apresentados os métodos de inferência Bayesiana e a comparação entre os métodos tradicionais não-paramétricos e empíricos na avaliação da estrutura de covariância e na interpolação de dados não amostrados. Desta forma, pesquisas mais complexas, com dependência espacial em modelo misto linear generalizado, vêm sendo desenvolvidas e aplicadas em dados experimentais que apresentam concentrações de radiação nuclear e no mapeamento de doenças. A metodologia de estimação da estrutura de covariância é baseada em técnicas Bayesianas e no método de Markov Chain Monte Carlo MCMC (Diggle & Tawn, 1998; Ribeiro Junior & Diggle, 1999,2002). Os autores apresentaram também vários métodos de *krigagem* como: *krigagem* linear, *krigagem* disjuntiva e *krigagem* indicatriz.

A estimativa dos parâmetros de modelos baseados em verossimilhança produz de maneira geral estimadores não viesados e eficientes quando aplicados para grandes amostras. Diggle & Ribeiro Junior (2000), afirmaram ainda que ao utilizar estes métodos deve-se considerar a pressuposição Gaussiana e ilustraram este estudo com um modelo linear Gaussiano em que a média é definida por um modelo de regressão linear, como o descrito abaixo:

$$Y_i = \mu(x_i) + S(x_i) + \epsilon_i, i = 1, \dots, n$$

$$\mu(x_i) = \sum_{k=1}^p f_k(x_i) \beta_k,$$

em que $f_k(\cdot)$ são variáveis explanatórias referenciadas espacialmente,

$S(\cdot)$ é um processo estacionário Gaussiano com variância s^2 e função de correlação $r(\mathbf{h}, \mathbf{f})$,

$\mathbf{h} = \|x_i - x_j\|$ e \mathbf{f} = "range".

ϵ_i é o ruído ou erro, mutuamente independentes, $\epsilon_i \sim N(0, t^2)$.

2.7.2.2 Máxima Verossimilhança Restrita

No modelo misto, usando-se o método de máxima verossimilhança restrita, cada observação é dividida em duas partes independentes, uma referente aos efeitos fixos e outra aos aleatórios, de forma que a função de densidade de probabilidades é dada pela soma das funções densidades de probabilidade de cada parte referente aos efeitos aleatórios, em relação aos componentes de variância, elimina o viés resultante de graus de liberdade na estimação dos efeitos fixos do modelo (Patterson et al., 1971, Martins et al., 1993)

Nos ensaios de campo na agricultura e na pesquisa com melhoramento animal, como método de seleção, a estimativa de máxima verossimilhança restrita dos parâmetros de covariância espacial, tem sido preferida à de máxima verossimilhança (Cressie & Lahiri, 1996; Kitanidis & Shen, 1996).

Kitanidis & Shen (1996) afirmaram que nos métodos de estimação linear, (*krigagem* ordinária) e, a distribuição dos erros estimados, não têm bom desempenho, porque não são descritos adequadamente pelos valores da média e do quadrado médio. Sugerem o uso de métodos geoestatísticos não lineares como: a *krigagem* disjuntiva, a *krigagem* indicatriz e a *krigagem* probabilística.

Os autores definiram que no método de máxima verossimilhança restrita, combinações lineares especiais dos dados, chamados de incrementos generalizados, são usados no lugar das observações originais. Estes incrementos generalizados filtram as tendências e estimam somente os parâmetros de covariância, levando a um estimador não viesado.

Os incrementos generalizados foram apresentados por Pardo-Igúzquiza (1997) como:

$$\mathbf{z} = \mathbf{L}\mathbf{y} \quad (36)$$

em que, \mathbf{L} é a matriz construída a partir da matriz de projeção para as somas de quadrados residuais mínimas, sem o conhecimento dos valores verdadeiros de $\hat{\alpha}$ e $\hat{\epsilon}$ que, é apresentada na seguinte expressão:

$$\mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}', \quad (37)$$

eliminando as p linhas, entre os incrementos generalizados $\mathbf{z} = \mathbf{P}\mathbf{y}$. Estes p incrementos são linearmente dependentes dos outros e, que o posto do projetor \mathbf{P} ortogonal é dado por $m = n - p$.

Então, pelas propriedades da matriz \mathbf{P} , tem-se:

$$\mathbf{P}\mathbf{X} = 0 \Rightarrow \mathbf{P}\mathbf{y} = \mathbf{P}\mathbf{X}\hat{\mathbf{a}} + \mathbf{P}\mathbf{e} = \mathbf{P}\mathbf{e}. \quad (38)$$

Da mesma forma que para o modelo linear, Pardo-Igúzquiza (1998,a) desenvolveu o método de máxima verossimilhança para um modelo de covariograma com a *krigagem* indicatriz. Sugeriu que estimativas dos parâmetros fossem obtidas pelo método clássico não-paramétrico. Neste trabalho o autor utilizou um programa desenvolvido em Fortran (Pardo-Igúzquiza, 1997), e a expressão do estimador do semivariograma pelo método clássico com indicador é o apresentado a seguir:

$$\hat{\mathbf{g}}(\mathbf{h}; z_k) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [I(x_i + \mathbf{h}; z_k) - I(x_i; z_k)]^2$$

em que, $N(\mathbf{h})$ é o número de pares de pontos separados por uma distância \mathbf{h} , z_k é o limiar $\{z_k; k = 1, \dots, K\}$ que torna discreta a função distribuição acumulada e, I é o indicador $I(x; z_k)$.

Este método usa indicadores binários e a função de distribuição acumulada aparece como a esperança de um indicador da função aleatória inicial $Z(\mathbf{x})$; isto é

$$F(z_k) = P\{Z(x) \leq z_k\} = E\{I(x; z_k)\} = m_I$$

com

$$I(x; z_k) = \begin{cases} 1 & \text{se } Z(x) \leq z_k \text{ e} \\ 0 & \text{para os demais casos.} \end{cases}$$

A função de distribuição acumulada deve, então, ser estimada a partir de dados indicadores usando a *krigagem* indicatriz dada por:

$$\hat{F}(z_k) = \sum_{i=1}^m \ddot{e}_i(z_k) * I(x_i; z_k), \quad k=1, \dots, K.$$

Pardo-Igúzquiza (1998,b) desenvolveu novos programas para os mesmos métodos de estimação dos parâmetros e, utilizou o modelo de variograma potência. Afirmou que este modelo é amplamente usado na prática para variáveis como elevações topográficas, precipitação de chuva e ponto piezométrico. A expressão para o modelo de variograma potência é dado por:

$$\gamma(h) = \alpha * |h|^\beta$$

sendo que, $\alpha > 0$ e $\beta \in (0, 2)$ são restrições para os parâmetros na seqüência, e que $-\alpha * |h|^\beta$ é uma função definida positiva condicional admitindo um modelo variograma.

Discutiu o procedimento clássico não-paramétrico em que se estima o variograma para diferentes distâncias, usando o estimador de Matheron (1963), e concluiu que os métodos de máxima verossimilhança e máxima verossimilhança restrita são melhores estimadores dos parâmetros na forma e escala de um modelo de variograma potência do que o erro de quadrados mínimos e estimadores de quadrados mínimos em termos de viés, variância e quadrado médio do erro (Pardo-Igúzquiza, 1998,b).

Diggle & Ribeiro Junior (2000), consideraram, que o estimador de máxima verossimilhança restrita é menos viesado em pequenas amostras, embora a justificativa original foi para diferentes contextos de estimação de componentes de variância em delineamentos experimentais. Os autores comentaram ainda, que este estimador é amplamente recomendado para estimar parâmetros de modelos geoestatísticos, mas fazem uma ressalva de que é mais sensível do que o estimador de máxima verossimilhança para a falta de especificação do modelo para a média.

2.8 Perfil de Verossimilhança

Uma das formas encontradas para investigar a variabilidade dos parâmetros de vários modelos que apresentam dependência espacial foi o perfil de verossimilhança. Avalia-se a superfície do logaritmo da função de verossimilhança, a qual contém informações sobre os parâmetros do modelo focalizando-se somente sob um conjunto deles, pois a dimensão desta

superfície não permite inspeção direta. Uma propriedade deste método é que corresponde a uma transformação ortogonal do parâmetro, se existir (Diggle & Ribeiro Junior, 2000).

O modelo considerado pelos autores foi um modelo linear Gaussiano com tendência linear segundo a expressão abaixo:

$$Y_i = \mathbf{m}(x_i) + S(x_i) + \mathbf{a}_i, i = 1, \dots, n. \quad (39)$$

em que

Y_i é a variável resposta;

$$\mathbf{m}(x_i) = \sum_{k=1}^p f_k(x_i) \mathbf{b}_k \quad \text{de efeito fixo};$$

$S(x_i)$ é um processo estacionário Gaussiano com variância \mathbf{S}^2 e função de correlação $\mathbf{r}(\mathbf{h}; \mathbf{f})$, sendo \mathbf{h} a distância euclidiana entre dois pontos e, \mathbf{a}_i são independentes, $N(0, \mathbf{t}^2)$.

Isto implica que $Y \sim MVN(\hat{\mathbf{F}}\hat{\mathbf{a}}, G(\hat{\mathbf{e}}))$ tem distribuição normal multivariada.

Apresentaram ainda o logarítmo da função verossimilhança e a matriz de covariância como:

$$\begin{aligned} l(\hat{\mathbf{a}}, \hat{\mathbf{e}}) &= \frac{1}{2} \left[\log|G(\hat{\mathbf{e}})| + (\mathbf{y} - \hat{\mathbf{F}}\hat{\mathbf{a}})^T [G(\hat{\mathbf{e}})]^{-1} (\mathbf{y} - \hat{\mathbf{F}}\hat{\mathbf{a}}) \right] \\ \hat{\mathbf{e}} &= (\hat{\sigma}^2, \hat{\phi}^2, \hat{\phi}) \\ \mathbf{G}(\hat{\mathbf{e}}) &= \text{Var}(Y) = [\hat{\sigma}^2 \mathbf{I} + \hat{\phi}^2 \mathbf{R}(\hat{\phi})] \end{aligned}$$

Consideraram o modelo com parâmetros $(\hat{\mathbf{a}}, \hat{\mathbf{e}})$ e a função de verossimilhança por $L(\hat{\mathbf{a}}, \hat{\mathbf{e}})$ e a expressão do perfil de verossimilhança foi apresentado como:

$$L_p(\hat{\mathbf{a}}) = L[\hat{\mathbf{a}}, \hat{\mathbf{e}}(\hat{\mathbf{a}})] = \max_{\hat{\mathbf{e}}} [L(\hat{\mathbf{a}}, \hat{\mathbf{e}})]$$

Salientaram que o procedimento do perfil gera diretamente verossimilhança ortogonal, independente do modelo a ser parametrizado, tal que o parâmetro *nuisance* é ortogonal para o parâmetro de interesse ou não. Apresentaram os perfis de verossimilhança para

os parâmetros de covariância (σ^2, \mathbf{f}) para dois modelos: o primeiro com média constante e o segundo com tendência linear.

Dadas as considerações dos autores pode-se observar que embora demande uma metodologia mais complexa e tempo de computador para o estudo espacial, a obtenção das estimativas dos parâmetros dos modelos e as interpretações dos resultados são elaboradas com maior confiabilidade comparado aos métodos em que as estimativas dos parâmetros são obtidas através do variograma de forma subjetiva.

2.9 Critério de Informação de Akaike na comparação de modelos com dependência espacial

Em geral, na seleção de modelos, alguns autores têm utilizado o Critério de Informação de Akaike (AIC) para diferentes estruturas de covariâncias.

O AIC conforme apresentado por Xavier (2000) é baseado na teoria de decisão e pode ser definido como:

$$AIC = -2l + 2p \quad (41)$$

em que:

l é o \ln da função de verossimilhança;

p é o número de parâmetros do modelo considerado.

De acordo com este critério o melhor modelo é aquele que apresenta o menor valor de AIC.

2.10 Teste da Razão de Verossimilhanças

Um método geral para comparar modelos ajustados por máxima verossimilhança é o teste da razão de verossimilhanças. Pode também ser usado com modelos ajustados pela máxima verossimilhança restrita. Um modelo ajustado por máxima verossimilhança não deve ser comparado com outro ajustado por máxima verossimilhança restrita e vice-versa (Pinheiro & Bates, 2000).

Se denotarmos por L_2 a verossimilhança do modelo mais geral e L_1 a verossimilhança do modelo restrito tem-se $L_2 > L_1$ e, consequentemente $\log L_2 > \log L_1$.

A estatística do teste da razão de verossimilhanças é sempre positiva como a seguir:

$$2\log(L_2 / L_1) = 2[\log(L_2) - \log(L_1)] \quad (42)$$

Xavier (2000) comentou que embora este teste seja eficaz não permite que sejam comparados mais que dois modelos simultaneamente, sendo que um deles é caso especial do outro.

Esses autores salientam que deve também ser considerada a diferença entre os números de parâmetros dos modelo a ser comparados com o valor crítico do χ^2 .

Se a diferença do logarítmico das funções de verossimilhanças do segundo modelo *versus* o primeiro for menor que o valor do χ^2 tabelado o modelo com menor número de parâmetros será adequado (Pinheiro & Bates, 2000).

2.11 Método da validação cruzada

A análise geoestatística, em algumas situações, exige um certo grau de subjetividade (Ribeiro Junior, 1995). Sendo assim, para avaliar se o fenômeno foi modelado de forma adequada, alguns autores têm utilizado o método da validação cruzada, de forma que sejam obtidas informações fidedignas da amostra.

A validação cruzada consiste das seguintes fases:

- suprimir do conjunto de dados, um ponto qualquer, por exemplo, Z_i ;
- calcular pelo método da *krigagem* conforme expressão (25), o valor estimado do ponto x_i , empregando somente os dados restantes, isto é, \hat{Z}_i ;
- calcular o erro de estimação através da expressão $(Z_i - \hat{Z}_i)$ dividido pelo desvio

$$\frac{(Z_i - \hat{Z}_i)}{\sigma_i} \quad e,$$

- d) repetir as fases anteriores para todos os pontos e comprovar que:
 i) os erros de *krigagem* são sistemáticos, então pode-se afirmar que:

$$\frac{1}{n} \sum_{i=1}^n (Z_i - \hat{Z}_i) \approx 0;$$

ii) os erros são compatíveis com a variância de predição, então diz-se que:

$$\frac{1}{n} \sum_{i=1}^n \left(\frac{Z_i - \hat{Z}_i}{\sigma_i} \right)^2 \approx 1.$$

Pela validação cruzada, além de ser avaliado o modelo escolhido para o semivariograma, toda a modelagem do processo estocástico está também sendo avaliada. Isto inclui a decisão da hipótese estacionária, os estimadores utilizados, o tratamento dos valores discrepantes, o modelo de semivariograma escolhido e a decisão relativa à anisotropia (Samper & Ramirez, 1990; Ribeiro Junior, 1995; Gonçalves, 1997).

3 METODOLOGIA

3.1 Material

Os dados experimentais, que fazem parte deste trabalho, foram coletados do levantamento detalhado de solos da Estação Experimental de Campos, Fazenda Angra (Capeche et al., 1997), por pesquisadores da PESAGRO e Embrapa Solos. No estudo pedológico foram avaliadas as características morfológicas, físicas e químicas dos solos, e apresentadas também, informações referentes à distribuição geográfica. Para o estudo geoestatístico foi considerada a variável agronômica teor de Cálcio ($\text{mmol}_{\text{cdm}}^{-3}$), nas camadas de 0-20 e 20-40cm.

Esta pesquisa é parte integrante do Projeto “Planejamento de uso das terras da Estação Experimental de Campos/PESAGRO-RIO, e subsídios ao manejo de irrigação”.

3.2 Local experimental e método de amostragem

A área compreendida pelo levantamento situa-se ao Norte do estado do Rio de Janeiro, à margem esquerda do Rio Paraíba do Sul, no município de Campos dos Goytacazes, entre os paralelos $21^{\circ}44'47''$ e $41^{\circ}18'24''$ WGr.

A Estação Experimental de Campos da PESAGRO-RIO abrange uma área de 105 hectares, e faz parte de uma das áreas mais tradicionais do cultivo da cana-de-açúcar no Brasil. Esta cultura se desenvolveu em áreas denominadas de tabuleiros: Podzólico Amarelo e Latossolo Amarelo e Baixada: associações e complexos de solos Hidromórficos e Cambissolos desenvolvidos de substratos colúvio-aluvionais (Embrapa Solos, 1997). Devido às enchentes do Rio Paraíba do Sul, à esquerda da fazenda e à direita pelas enchentes das Lagoas dos Prazeres e Taquaraçu, ocorreu a deposição de sedimentos, que contribuiu para a grande variação nas propriedades avaliadas no solo associadas às adubações realizadas na área experimental.

Na Fazenda Angra, a monocultura com cana-de-açúcar foi substituída, recentemente, pelo cultivo de grãos (arroz, feijão e milho) sob sistema de irrigação por aspersão.

O objetivo principal do levantamento foi o planejamento adequado do uso dos solos possibilitando um manejo racional e sustentável, evitando o processo de erosão dando subsídios em campos experimentais na condução de pesquisas que são extrapoladas para áreas edafoclimáticas similares.

As amostras de solo foram coletadas com um trado tipo holandês, em uma malha regular, sob duas coordenadas geográficas: leste e norte, distando nas duas direções 900 metros. Esta malha foi realizada com espaçamento de 50 metros, totalizando 178 amostras, nas camadas de 0-20 e 20-40cm (Figura 4).

As amostragens regionalizadas possibilitaram mapear a direção da variabilidade das propriedades granulométricas (textura) e as propriedades químicas, permitindo mapear as produções das culturas face às diferentes situações de fertilidade do solo e tipos de solo.

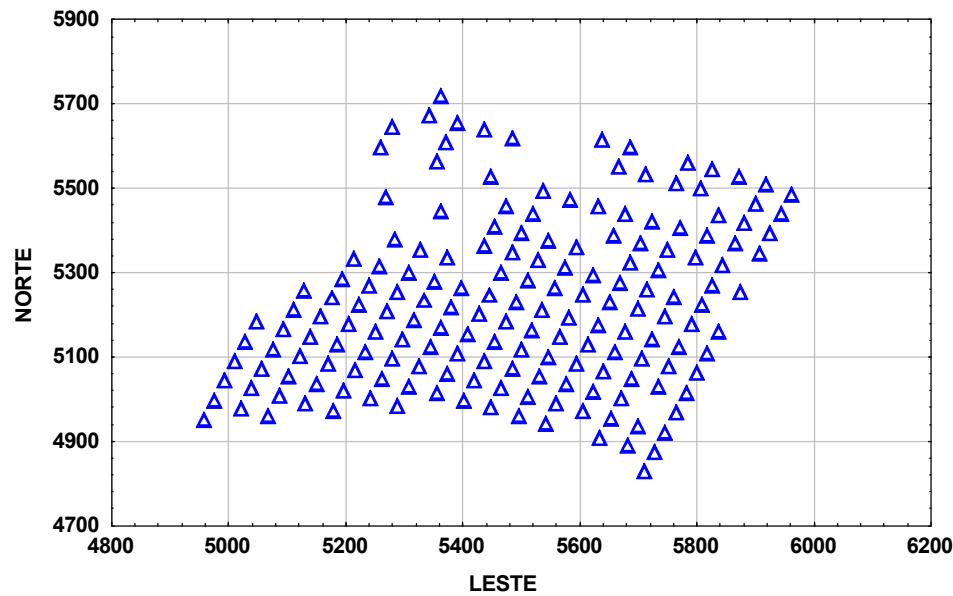


Figura 4- Mapa de amostragem em malha regular para mensurações da variável Cálcio ($\text{mmol}_{\text{c}}\text{dm}^{-3}$), nas camadas de 0-20 e 20-40cm do solo.

3.3 Métodos

O problema mais freqüente encontrado nos solos brasileiros é a grande variabilidade nos teores de micronutrientes. Em geral, em experimentos de campo, as amostragens são realizadas dentro de cada parcela aleatoriamente e, dado a variabilidade destes solos devido as manchas de fertilidade, não são detectadas diferenças estatísticas entre os tratamentos. Uma forma encontrada para detectar tal variabilidade e a existência de dependência espacial entre as parcelas, no caso de experimentação ou em levantamentos amostrais, foi a de avaliar estes nutrientes considerando a posição espacial de cada amostra, para dar subsídios aos pesquisadores na instalação de experimentos. Portanto, foi necessário avaliar diferentes estruturas de covariância identificando assim, a que melhor identificassem e interpretassem a dependência espacial do cálcio no solo. Além disso, devido à adubação realizada na área experimental, em períodos distintos, foi necessário considerar estas regiões como covariável bem como, a influência na dependência espacial na direção das coordenadas.

3.3.1 Descrição dos modelos

Considerando que as observações medidas para o teor de cálcio nos diferentes pontos amostrais apresentam dependência espacial e não existem repetições em cada ponto para predizer o que acontece na região toda por um processo de pares de distâncias. Assim, tem-se um vetor multidimensional de observações correlacionadas em duas dimensões espaciais, as quais foram analisadas por meio de um modelo linear misto, com diferentes estruturas para a matriz de covariâncias. Um modelo misto com estrutura de variabilidade espacial pode ser representado na forma matricial como $\mathbf{Y} = \mathbf{XB} + \mathbf{Z}\hat{\mathbf{a}} + \mathbf{a}$ em que: $\mathbf{Y} = (Y_1, \dots, Y_k)$ é o vetor dos valores da variável resposta nas posições (x_1, \dots, x_k) ; $\hat{\mathbf{a}}$ é o vetor de parâmetros desconhecidos associados aos efeitos fixos; $\tilde{\mathbf{a}}$ é o vetor de parâmetros desconhecidos associados aos efeitos aleatórios e, \mathbf{a} é o vetor dos erros independentes e normalmente distribuídos.

Neste trabalho, foram utilizadas as funções de covariância exponencial e a família Matérn, para $\hat{\epsilon} = 1$ e $\hat{\epsilon} = 2$, conforme apresentadas no item 3.4.7.

a) Modelo com a covariável região

Neste modelo foi incluído o efeito de regiões como covariável (modelo 1) a qual permitiu interpretar a resposta do cálcio no solo adequadamente. Desta forma, a expressão para o modelo misto é apresentado por:

$$Y_i = \hat{a}_0 + \mathbf{b}_1 \text{região2}(x_i) + \mathbf{b}_2 \text{região3}(x_i) + S(x_i) + \mathbf{e}_i : i = 1, \dots, n \quad (43)$$

em que: Y_i é valor da variável resposta na posição x_i ; \mathbf{b}_0 é uma constante; \hat{a}_1 e \hat{a}_2 são os parâmetros de regressão desconhecidos associados à covariável região sendo $-(\hat{a}_1 + \hat{a}_2)$ o efeito da região 1, \hat{a}_1 o efeito da região 2 e \hat{a}_2 o efeito da região 3; $S(x_i)$ é um processo estacionário Gaussiano com variância s^2 e função de correlação $r(\mathbf{h}; \mathbf{f})$ em que, $\mathbf{f} = "range"$ $\mathbf{h} = \|x_i - x_j\|$ e, \mathbf{e}_i é o ruído ou erro, mutuamente independentes, $\mathbf{e}_i \sim N(0, t^2)$.

b) 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 (modelo 2), 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:

$$\begin{aligned} Y_i = & \hat{a}_0 + \mathbf{b}_1 \text{região2}(x_i) + \hat{a}_2 \text{região3}(x_i) + \hat{a}_3 \text{leste}(x_i) + \\ & + \hat{a}_4 \text{norte}(x_i) + S(x_i) + \mathbf{e}_i : i = 1, \dots, n \end{aligned} \quad (44)$$

em que, além dos efeitos considerados na eq. (43) tem-se:

\hat{a}_3 é o parâmetro de regressão desconhecido associado a coordenada leste e \hat{a}_4 é o parâmetro de regressão desconhecido associado a coordenada norte.

Esses modelos podem ser escritos na forma matricial são expressos por:

$$\tilde{\mathbf{O}} = \mathbf{X}\hat{\mathbf{a}} + \mathbf{S}(x) + \tilde{\mathbf{a}} \quad (45)$$

em que: $\mathbf{Y} = (Y_1, \dots, Y_k)$ é o vetor dos valores da variável resposta nas posições (x_1, \dots, x_k) ; $\hat{\mathbf{a}}$ é o vetor de parâmetros desconhecidos associados aos efeitos fixos; $\mathbf{S}(x)$ é um processo estocástico estacionário e, $\tilde{\mathbf{a}}$ é o vetor dos erros independentes e normalmente distribuídos.

Para \mathbf{Y} é assumida distribuição normal multivariada, isto é, $MVN(\mathbf{X}\hat{\mathbf{a}}, \sigma^2 \mathbf{R}_f + \sigma^2 \mathbf{I})$ em que, \mathbf{I} é uma matriz identidade $n \times n$, \mathbf{R}_f é a matriz de covariância de dimensão $n \times n$ com (i, j) -ésimo elemento $\tilde{n}(h_{ij})$ com $h_{ij} = \|x_i - x_j\|$ a distância Euclidiana entre x_i e x_j .

Para a estrutura de covariância exponencial, a função de correlação $r(\mathbf{h})$ é expressa por:

$$r(\mathbf{h}, f) = \exp\left(-\frac{\mathbf{h}}{f}\right)$$

e para a estrutura de covariância Matérn por:

$$r(\mathbf{h}, f, k) = \left\{ 2^{k-1} \Gamma(k) \right\}^{-1} \left(\frac{\mathbf{h}}{f} \right)^k K_k \left(\frac{\mathbf{h}}{f} \right)$$

em que, K_k é a função de Bessel modificada de ordem k . No caso em que $k = 0,5$ a função de correlação Matérn reduz-se a função de correlação exponencial. Ainda, se $k \rightarrow \infty$ tem-se a função de correlação Gaussiana.

3.3.2 Métodos de estimação dos parâmetros

3.3.2.1 Máxima verossimilhança

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 de observações \mathbf{Y} conforme apresentado no item 3.3.1.

Considerando os modelos apresentados no item 3.3.1 e as estruturas de covariância exponencial e família Matérn e assumindo que $\mathbf{Y} = (Y_1, \dots, Y_k)$ segue uma distribuição multivariada Gaussiana tem-se que a função de densidade de probabilidades de \mathbf{Y} por:

$$f(\mathbf{y}; \hat{\mathbf{a}}, \hat{\mathbf{e}}) = (2\delta)^{-\frac{n}{2}} |\mathbf{O}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\hat{\mathbf{a}})^T \mathbf{O}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{a}})\right] \quad (46)$$

em que: \mathbf{y} é um vetor de observações experimentais; $\hat{\mathbf{a}}$ é o vetor de parâmetros desconhecidos associados aos efeitos fixos; $\hat{\mathbf{e}}$ são os parâmetros da matriz de covariância com $\hat{\mathbf{e}} = (\delta^2, \sigma^2, \mathbf{f})$ se a estrutura de covariância é a exponencial e $\hat{\mathbf{e}} = (\delta^2, \sigma^2, \mathbf{f}, \hat{e})$ se a estrutura de covariância é a Matérn; T é a matriz transposta e, $|\mathbf{O}|$ é o determinante da matriz de variâncias e covariâncias com dimensão $n \times n$, das n observações experimentais.

A função $f(\mathbf{y}; \hat{\mathbf{a}}, \hat{\mathbf{e}})$ com \mathbf{y} mantido fixo em seus valores observados é chamada função de verossimilhança e é expressa por:

$$L(\hat{\mathbf{a}}, \hat{\mathbf{e}}; \mathbf{y}) = (2\delta)^{-\frac{n}{2}} |\mathbf{O}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\hat{\mathbf{a}})^T \mathbf{O}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{a}})\right]. \quad (47)$$

Note que a função de verossimilhança é uma função dos parâmetros desconhecidos e não dos dados, e é interpretada como uma função que serve para medir as chances do vetor $(\hat{\mathbf{a}}, \hat{\mathbf{e}})$ explicar os dados \mathbf{y} , ou seja, consiste em encontrar os valores $(\hat{\mathbf{a}}, \hat{\mathbf{e}})$ de $(\hat{\mathbf{a}}, \hat{\mathbf{e}})$ mais prováveis de terem produzido os valores observados \mathbf{y} .

Desta forma, para $\mathbf{Y} = \mathbf{y}$, procura-se o vetor de $(\hat{\mathbf{a}}, \hat{\mathbf{e}})$ que satisfaça a seguinte expressão:

$$L[(\hat{\mathbf{a}}, \hat{\mathbf{e}}); \mathbf{y}] = f[(\mathbf{y}; \hat{\mathbf{a}}, \hat{\mathbf{e}})] = \max \{f(\mathbf{y}; \hat{\mathbf{a}}, \hat{\mathbf{e}})\} \quad \hat{\mathbf{a}}, \hat{\mathbf{e}} \in \hat{E}$$

Como existem vários valores possíveis para os parâmetros, é necessário utilizar algum mecanismo que possibilite obter os valores que maximizam a função $L(\hat{\mathbf{a}}, \hat{\mathbf{e}}; \mathbf{y})$ (Kitanidis & Shen, 1996; Xavier, 2000; Diggle & Ribeiro Junior, 2000).

Maximizar $L(\hat{\mathbf{a}}, \hat{\mathbf{e}}; \mathbf{y})$ é o mesmo que maximizar o logarítmico de $L(\hat{\mathbf{a}}, \hat{\mathbf{e}}; \mathbf{y})$, uma vez que a função $g = \ln(x)$, com $x > 0$, que é uma função estritamente monótona crescente. Assim maximizar a expressão (46) é o mesmo que maximizar o logarítmico da função de verossimilhança $L(\hat{\mathbf{a}}, \hat{\mathbf{e}}; \mathbf{y})$ a qual é dada por:

$$l(\hat{\mathbf{a}}, \hat{\mathbf{e}}; \mathbf{y}) = \ln L(\hat{\mathbf{a}}, \hat{\mathbf{e}}; \mathbf{y}) = -\frac{n}{2} \ln(2\delta) - \frac{1}{2} \ln|\mathbf{O}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\hat{\mathbf{a}})^T \mathbf{O}^{-1} (\mathbf{y} - \mathbf{X}\hat{\mathbf{a}}) \quad (48)$$

Fatorando a matriz \mathbf{O} como $\mathbf{O} = \mathbf{s}^2 \mathbf{Q}$, em que \mathbf{s}^2 é a variância e \mathbf{Q} a matriz de correlação obtém-se o logarítmico da função de verossimilhança $l(\hat{\mathbf{a}}, \hat{\mathbf{e}}, \mathbf{s}^2; \mathbf{y})$ apresentada na equação (33). Diferenciando-se parcialmente $l(\hat{\mathbf{a}}, \hat{\mathbf{e}}, \mathbf{s}^2; \mathbf{y})$ em relação a $\hat{\mathbf{a}}$ e \mathbf{s}^2 e igualando-se as derivadas a zero, obtém-se

$$\hat{\mathbf{a}} = (\mathbf{X}^T \mathbf{Q}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Q}^{-1} \mathbf{y}$$

e

$$\hat{\mathbf{s}}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\mathbf{a}})^T \mathbf{Q}^{-1} (\mathbf{y} - \mathbf{X}\hat{\mathbf{a}}),$$

$$\text{com } \text{Cov}(\hat{\mathbf{a}}) = \delta^2 (\mathbf{X}^T \mathbf{Q}^{-1} \mathbf{X})^{-1} \text{ e } \text{Var}(\hat{\mathbf{s}}^2) = \frac{2(\hat{\mathbf{s}}^2)^2}{n}.$$

As estimativas de $\hat{\mathbf{e}}$ são obtidas substituindo-se $\hat{\mathbf{a}}$ e $\hat{\mathbf{s}}^2$ no logarítmico da função de verossimilhança apresentada na equação (48). Desta forma, o vetor de estimativas de máxima verossimilhança serão obtidos por algum procedimento de maximização numérica. A subrotina geoR (Ribeiro Junior & Diggle, 2001), encontra-se disponível no programa estatístico R uma vez que, de forma analítica não é possível obter as derivadas da variância *nugget* (δ^2) e do alcance (\mathbf{f}), porque estes parâmetros estão contidas na matriz \mathbf{O}^{-1} .

3.3.2.2 Máxima verossimilhança restrita

Uma alternativa ao método de máxima verossimilhança que não leva em conta a perda de graus de liberdade devido à estimação dos efeitos fixos, é o estimador de máxima verossimilhança restrita ou residual (Patterson et al., 1971; Cressie, 1993; Xavier, 2000; Diggle & Ribeiro Junior, 2000).

O método da máxima verossimilhança residual foi utilizado inicialmente para estimar os componentes de variância na análise de variância para os efeitos fixos e aleatórios independentemente, ou seja, a função de verossimilhança é fatorada em duas partes, sendo uma delas totalmente livre dos efeitos fixos. Esta fatoração é obtida por meio de uma transformação linear nos dados originais (Patterson & Thompson, 1971; Xavier, 2000).

Neste método ao invés de trabalhar com os dados originais, são usadas transformações lineares destes, isto é, $\mathbf{Y}^* = \mathbf{PY}$ tal que \mathbf{Y}^* não depende de $\hat{\mathbf{a}}$. Estas combinações são também chamadas de incrementos generalizados conforme apresentado em (36). A matriz \mathbf{P} é construída sem o conhecimento dos valores verdadeiros de $\hat{\mathbf{a}}$ e $\hat{\mathbf{e}}$ como, por exemplo, o projetor ortogonal \mathbf{P} apresentado na eq. (37), isto é,

$$\mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

O estimador de máxima verossimilhança restrita para $\hat{\mathbf{e}}$ pode ser calculado por maximizar a seguinte expressão:

$$L^*(\hat{\mathbf{e}}) = -\frac{n-p}{2} \ln(2\delta) + \frac{1}{2} \ln(\mathbf{X}'\mathbf{X}) - \frac{1}{2} \left\{ \ln|G(\hat{\mathbf{e}})| + \ln|\mathbf{X}'[G(\hat{\mathbf{e}})]^{-1}\mathbf{X}| + \right. \\ \left. + (\mathbf{y} - \mathbf{X}\hat{\mathbf{a}})'[G(\hat{\mathbf{e}})]^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{a}}) \right\}$$

em que, $\tilde{\mathbf{a}} = \hat{\mathbf{a}}(\hat{\mathbf{e}})$ e $G(\hat{\mathbf{e}}) = \mathbf{It}^2 + \mathbf{s}^2\mathbf{R}(f)$.

De forma análoga ao estimador de máxima verossimilhança foram considerados os modelos do item 3.3.1 e as estruturas de covariância exponencial e família Matérn.

Para determinar a máxima verossimilhança restrita para os modelos estudados, considerou-se que \mathbf{R} é uma combinação linear de \mathbf{Y} , obteve-se a função de distribuição de

\mathbf{R} em função de \mathbf{Y} , as funções densidades de $\hat{\mathbf{b}}$ e $p(R\hat{\mathbf{a}})$, o projetor ortogonal \mathbf{P} , o vetor dos resíduos $\mathbf{R} = \mathbf{PY}$.

\mathbf{Y} é o vetor das observações experimentais,

\mathbf{R} é o vetor dos resíduos,

\mathbf{P} é o projetor de \mathbf{Y} no subespaço residual definido pelo método dos quadrados mínimos ordinários sendo o seu posto $n - p$, ou seja, \mathbf{P} tem $n - p$ linhas linearmente independentes.

Seja $\hat{A}[\mathbf{Y}] = \hat{A}[\mathbf{PY}]$

Pela eq. (37) podemos escrever que $\hat{A}[\mathbf{Y}] = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\hat{A}[\mathbf{Y}] = 0$

$$p(\mathbf{R} | \mathbf{y}) = |J|^{-1} \left[\frac{f(\mathbf{y} | \mathbf{e})}{f(\hat{\mathbf{a}} | \mathbf{y})} \right]$$

$$f(\hat{\mathbf{a}}) = (2\delta)^{\frac{-p}{2}} (\mathbf{X}^T \mathbf{O}^{-1} \mathbf{X})^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} [\hat{\mathbf{a}} - \mathbf{a}]^T (\mathbf{X}^T \mathbf{O}^{-1} \mathbf{X}) [\hat{\mathbf{a}} - \mathbf{a}]\right\}$$

$$p(R; y) = \frac{(\mathbf{X}^T \mathbf{X})^{-\frac{1}{2}} \exp\{[\mathbf{y} - \mathbf{X}\hat{\mathbf{a}}]^T (\mathbf{O})^{-1} [\mathbf{y} - \mathbf{X}\hat{\mathbf{a}}]\}}{(2\delta)^{\frac{n}{2}} (|\mathbf{O}|^{\frac{1}{2}}) (2\delta)^{\frac{-p}{2}} |\mathbf{X}^T \mathbf{O}^{-1} \mathbf{X}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2} [\hat{\mathbf{a}} - \mathbf{a}]^T (\mathbf{O})^{-1} |\mathbf{X}^T \mathbf{O}^{-1} \mathbf{X}| [\hat{\mathbf{a}} - \mathbf{a}]\right\}}$$

$$f(\mathbf{R}; \mathbf{y}) = \frac{\exp\left\{-\frac{1}{2} [\mathbf{y} - \mathbf{X}\hat{\mathbf{a}}]^T (\mathbf{O})^{-1} [\mathbf{y} - \mathbf{X}\hat{\mathbf{a}}] \right\} [\hat{\mathbf{a}} - \mathbf{a}]^T (\mathbf{O})^{-1} |\mathbf{X}^T \mathbf{O}^{-1} \mathbf{X}| [\hat{\mathbf{a}} - \mathbf{a}]}{(2\delta)^{\frac{-n+p}{2}} (|\mathbf{O}|^{\frac{1}{2}}) (\mathbf{X}^T \mathbf{X})^{-\frac{1}{2}} |\mathbf{X}^T \mathbf{O}^{-1} \mathbf{X}|^{\frac{1}{2}}}$$

Resolvendo-se a expressão dentro do exponencial proposta por Harville (1974) o log da verossimilhança restrita para as estruturas exponencial e esférica será:

$$\begin{aligned}
 l(\hat{\mathbf{e}}) = & -\frac{n-p}{2} \ln(2\delta) + \ln \frac{1}{2} (\mathbf{X}^T \mathbf{X}) - \ln \frac{1}{2} (\| \mathbf{O} \|) - \\
 & - \ln \frac{1}{2} \| \mathbf{X}^T \mathbf{O}^{-1} \mathbf{X} \| \left\{ \frac{1}{2} [\mathbf{y} - \mathbf{X}\hat{\mathbf{a}}]^T (\mathbf{O})^{-1} [\mathbf{y} - \mathbf{X}\hat{\mathbf{a}}] \right\}
 \end{aligned}$$

3.4 Perfil de Verossimilhança

Na investigação da variabilidade dos parâmetros dos modelos estudados que apresentam dependência espacial foi utilizado o perfil de verossimilhança e, maiores detalhes desta metodologia encontram-se no item 2.8.

Apresentaram-se os perfis de verossimilhança para os parâmetros de covariância (δ^2, f, δ^2) para os modelos com a covariável região (modelo 1) e com região e tendência linear nas coordenadas Leste e Norte (modelo 2) de acordo com as eqs. (43 e 44).

3.5 Critério de Informação de Akaike para seleção de modelos com dependência espacial

Foi utilizado para a seleção dos modelos o Critério de Informação de Akaike (AIC) para diferentes estruturas de covariâncias conforme apresentado no item 2.9, o qual é baseado na teoria de decisão e pode ser definido pela eq. (41).

Neste estudo foi considerado como melhor modelo aquele que apresentou o menor valor de AIC.

3.6 Teste da Razão de Verossimilhanças

Além do AIC foi também utilizado para comparar modelos ajustados por máxima verossimilhança o teste da razão de verossimilhanças apresentado no item 2.10 utilizando-se da eq. (42).

3.7 Método da validação cruzada

Para avaliar se a distribuição espacial do cálcio no solo foi modelado de forma adequada utilizou-se o método da validação cruzada, para que sejam obtidas informações fidedignas da amostra conforme apresentado no item 2.11.

3.8 Método de análise estatística

Os métodos de análise estatística utilizados neste trabalho, já discutidos em itens anteriores, foram:

- a) análise exploratória de dados (estatísticas descritiva e gráfica);
- b) análise exploratória espacial (gráfica);
- c) método de interpolação pela *krigagem* (*krigagem* ordinária);
- d) método paramétrico e não-paramétrico para avaliar a dependência espacial ;
- e) estimativas dos parâmetros pelos métodos de máxima verossimilhança e máxima verossimilhança restrita;
- f) avaliação da variabilidade dos parâmetros pelo perfil de verossimilhança e intervalos de confiança;
- g) ajuste dos modelos exponencial e Matérn aplicados aos dados de teor de cálcio no solo;
- h) mapas da predição e da variância de predição para os dois modelos com o melhor estimador e com as estruturas de covariâncias estudadas;
- i) critério de informação de Akaike para seleção dos modelos;
- j) método da validação cruzada para avaliar o ajuste da *krigagem*.

3.9 Programas estatísticos utilizados

Para a análise gráfica e estatística para os métodos propostos foram utilizados os seguintes os programas: GeoEAS (Englund & Sparks, 1988), SAS versão 8.0, o programa R versão 1.6.0, e a biblioteca geoR versão 1.3.7 (Ribeiro Junior & Diggle, 2000, 2001) e Statistica.

4 RESULTADOS E DISCUSSÃO

4.1 Generalidades

A compreensão de dados com dependência espacial tem sido a maior preocupação de alguns pesquisadores, principalmente, na área de Ciência do solo. Com os recursos cada vez mais escassos destinados à pesquisa, o conhecimento e interpretação das propriedades físicas e químicas do solo, limitantes para o aumento da produtividade agrícola, têm sido prioridade nos projetos de pesquisa nas diferentes culturas.

Além desta preocupação existe também o interesse desses pesquisadores em que o método estatístico de análise dos dados permita melhor interpretação dos resultados, de maneira que dê subsídios para a instalação de experimentos de campo, onde o conhecimento do solo é restrito. Na literatura brasileira, até a década de 80, os procedimentos estatísticos encontrados para analisar dados com dependência espacial, em que a posição georreferenciada é considerada, limitava-se a métodos estatísticos subjetivos, nos quais não era exigido o conhecimento da modelagem probabilística. Desta forma, a proposta de análise dos dados de cálcio no solo em duas camadas, 0-20 e 20-40cm, foi a de estudá-los baseados em modelos lineares multivariados misto e as estimativas dos parâmetros, obtidas através de estimadores de máxima verossimilhança com a subrotina geoR versão 1.37 (Anexo A).

4.2 Análise estatística

Com o objetivo de diagnosticar o comportamento espacial do teor de cálcio no solo ($\text{mmol}_c \text{dm}^{-3}$) foram realizadas as análises estatísticas descritivas e espacial as quais estão inseridas num contexto mais amplo da análise exploratória. Essas análises deram subsídios para avaliar a forma de como este teor se dispersou no espaço e permitiu verificar as hipóteses de estacionaridade (itens 2.1.6 e 2.1.6.1).

a) Análise descritiva

Observou-se pelos resultados da Tabela 1 que a variável teor de cálcio apresentou, nas duas profundidades leve assimetria e valores para a média e mediana bem próximos. Os coeficientes de variação para as duas profundidades foram, respectivamente 21,86% e 30,10%. Estes coeficientes foram aceitos como razoáveis uma vez que tais teores foram obtidos de maneira sistemática e, na região experimental existem manchas de fertilidade. Embora os valores para os coeficientes de assimetria e curtose encontrarem-se próximos de zero os resultados para os testes de aderência foram distintos para as duas profundidades (Figura 5a e 5b). Para a camada de 0-20cm a hipótese nula de normalidade da variável teor de cálcio não foi rejeitada, ou seja, a variável em estudo nesta camada pode ser estudada segundo a distribuição normal.

Tabela 1. Resultados das estatísticas descritivas para a variável teor de cálcio ($\text{mmol} \cdot \text{dm}^{-3}$), nas camadas de 0-20 e 20-40cm do solo.

Estatísticas	Cálcio 0-20cm	Cálcio 20-40cm
Número de Observações	178	178
Média	50,67	45,00
Mediana	50,50	44,00
Mínimo	21,00	21,00
Máximo	78,00	79,00
Amplitude Total	57,00	58,00
Amplitude Interquartílica	15,00	22,00
Variância	122,77	183,25
Coeficiente de Assimetria	-0,098	0,099
Coeficiente de Curtose	-0,33	-0,78
Desvio Padrão	11,08	13,54
C.V. (%)	21,86	30,10

Na camada inferior contudo, foi indicado pelos testes de Lilliefors e Shapiro-Wilk, a rejeição desta hipótese (Tabela 2, Figura 5a, 5c). Devido às diferenças nas várias metodologias utilizadas os resultados dos testes não são coincidentes com respeito à normalidade conforme apresentado pelo teste de Kolmogorov-Smirnov (Tabela 2). Observou-se no gráfico de

probabilidade normal e na distribuição de freqüências sob a curva normal (Figura 5b, 5d) para a profundidade 20-40cm uma leve assimetria que foi confirmada pelos testes de aderência, cujos resultados encontram-se na Tabela 2.

Tabela 2. Resultados dos testes de aderência para a variável teor de cálcio ($\text{mmol}_c \text{dm}^{-3}$), nas camadas de 0-20 e 20-40cm do solo.

Métodos	Cálcio 0-20		Cálcio 20-40	
	Valor do teste	Probabilidade	Valor do teste	Probabilidade
Kolmogorov-Smirnov	$d = 0,0496$	Prob > 0,20	$d = 0,07091$	Prob > 0,200
Lilliefors	-	Prob > 0,20	-	Prob < 0,050
Shapiro-Wilk	$W = 0,9946$	Prob < 0,77	$W = 0,9783$	Prob < 0,007

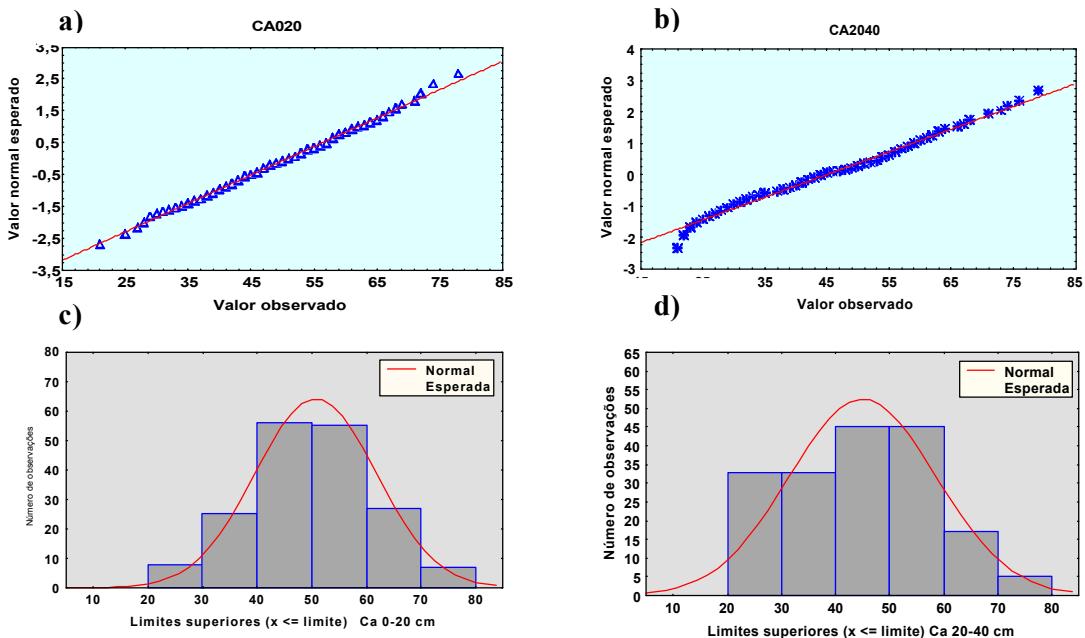


Figura 5 - Gráficos da probabilidade normal (a,b), distribuição de freqüências sob a curva normal (c,d), para o teor de cálcio ($\text{mmol}_c \text{dm}^{-3}$), nas camadas de 0-20 e 20-40cm do solo.

b) Análise exploratória espacial

Os resultados da análise descritiva fornecem uma idéia geral do comportamento da dispersão dos dados de cálcio mas, são contudo penalizados por não estarem levando em conta a posição geográfica de cada observação do teor de cálcio realizada. A média e a variância não são 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 0-20cm e, na direção leste os dados apresentaram-se mais dispersos (Figura 6a, 6b). Foi também observado a existência de correlação entre o cálcio e as coordenadas norte e leste. Os valores para estas correlações foram respectivamente $r = -0,519$ e $r = -0,0488$.

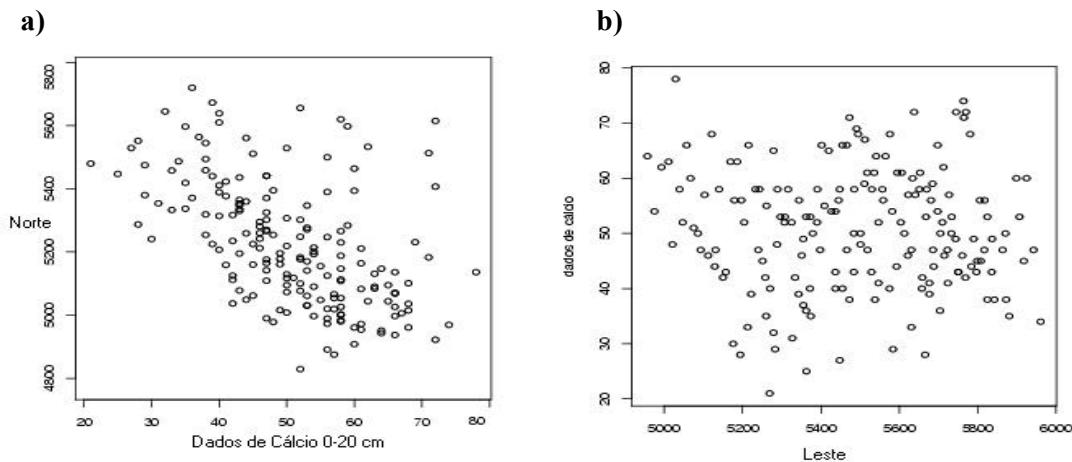


Figura 6 - Perfil do teor de cálcio na direção norte (a), e na direção leste (b), para a camada 0-20cm do solo.

Graficamente tais informações evidenciaram a presença de tendência, mas foram, no entanto, insuficientes para que as suposições de existência de estacionaridade do processo estocástico fossem aceitas para a variável teor de cálcio.

Este diagnóstico estatístico descritivo e espacial associado aos conhecimentos agronômicos da região permitiram que fosse investigado quais variáveis explanatórias deveriam ser consideradas no modelo a ser estudado, visando assim, uma melhor interpretação da fertilidade do solo e deficiência de cálcio no solo.

Analogamente, para o teor de cálcio na camada 20-40cm observou-se uma tendência linear na direção norte e, na direção leste, os dados estão mais dispersos na região (Figura 7a e 7b). Além da tendência houve observaram-se correlações de -0,482 e -0,257 entre o teor de cálcio e as coordenadas norte e leste, respectivamente.

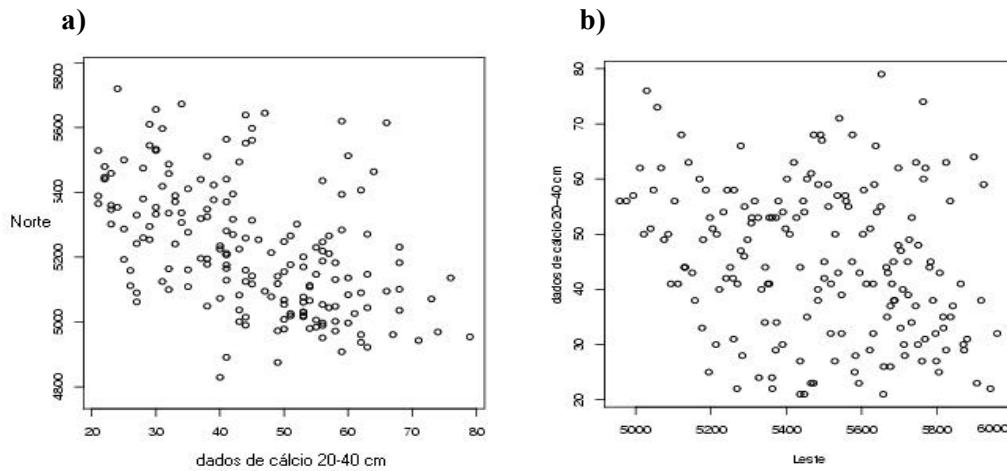


Figura 7 – Perfil do teor de cálcio na direção norte (a), e na direção leste (b), para a camada 20-40cm do solo.

Uma outra visão dos dados é aquela em que cores mais fortes são associadas aos maiores valores do teor de cálcio. Desse modo, tem-se que os teores de cálcio nas duas camadas, podem ser representados como mostrados na Figura 8. Desta figura observa-se que existe um padrão espacial distinto nas duas camadas sendo que, na camada 20-40cm, os teores de cálcio estão mais dispersos em relação aos valores mínimos e máximos. Este fato está associado à movimentação do solo com os implementos agrícolas e a influência da água das lagoas dos Prazeres, Taquaraçu e do Rio Paraíba do Sul no período das cheias.

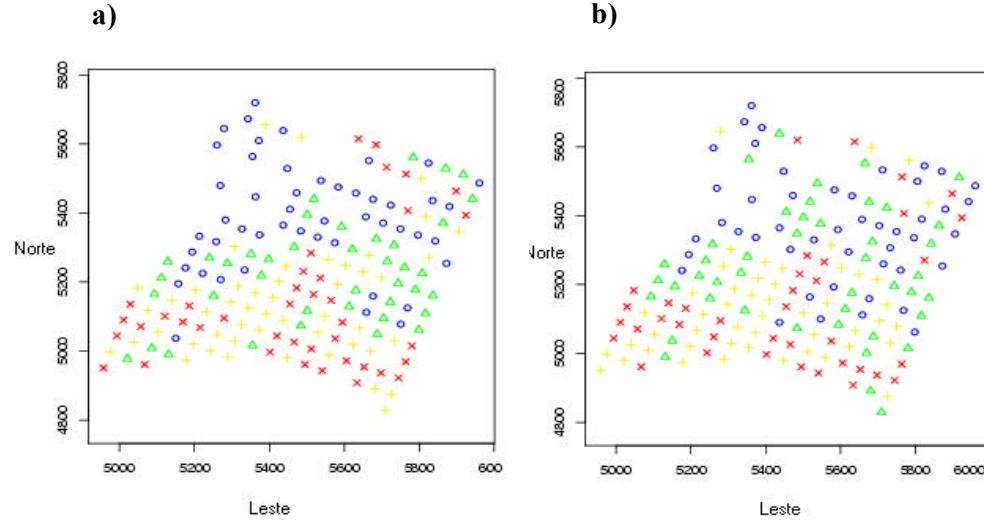


Figura 8 - Localização geográfica do teor de cálcio associando as cores fortes com a magnitude dos dados nas camadas 0-20cm (a) e 20-40cm (b) do solo.

4.3 Análise geoestatística

Na presença de correlação e tendência linear entre as coordenadas e o teor de cálcio, é importante caracterizar a dependência espacial. Sendo assim e, fundamentando-se nos resultados das análises anteriores, 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 nas estruturas de covariância exponencial e família Matérn para $\hat{e} = 1$ e $\hat{e} = 2$ que, de maneira mais simples estas estruturas serão também identificadas por modelo exponencial e modelo Matérn .

Foi verificado, também, por meio de métodos exploratórios como variogramas empíricos de nuvens (Figuras 9a, 9b, 10a e 10b) e pontuais (Figuras 9c, 9d, 10c e 10d), se algum possível valor discrepante influenciou a resposta do cálcio nas duas camadas do solo. Nas Figuras 9a, 9c, 10a e 10c encontram-se os variogramas obtidos pelo método clássico de Matheron (1963) e nas Figuras 9b, 9d, 10b e 10d foram obtidos os variogramas robustos proposto por Cressie (1993). Embora existam valores que se distanciam da massa de dados, observou-se que os mesmos não são *outliers* e sim diversos valores repetidos e muito pequenos, oriundos da região em que a adubação não foi recente. Como os variogramas empíricos são

muito erráticos em função dos diferentes números de pares que geram cada ponto do variograma e cada ponto entra várias vezes no cálculo das distâncias, um método com maior precisão foi necessário para estimar os parâmetros e identificar novos variogramas com maior precisão. Estes foram obtidos baseando-se em modelos e, seus parâmetros por estimadores de máxima verossimilhança.

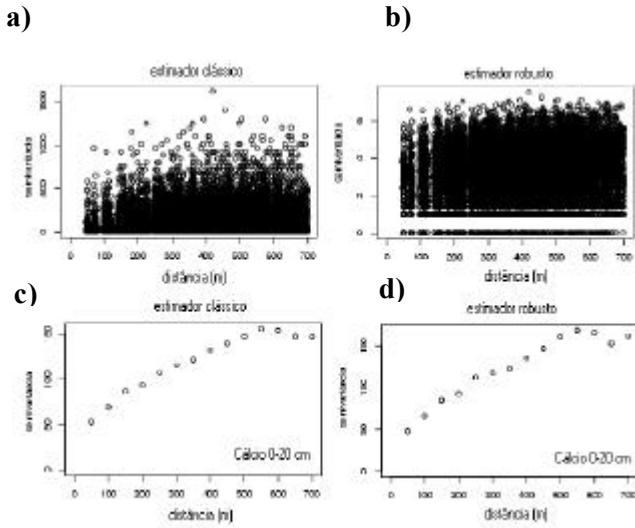


Figura 9 - Variogramas empíricos de nuvens e pontuais pelos métodos de Matheron (a e c) e de Cressie (b e d) para a variável teor de cálcio na camada de 0-20cm do solo.

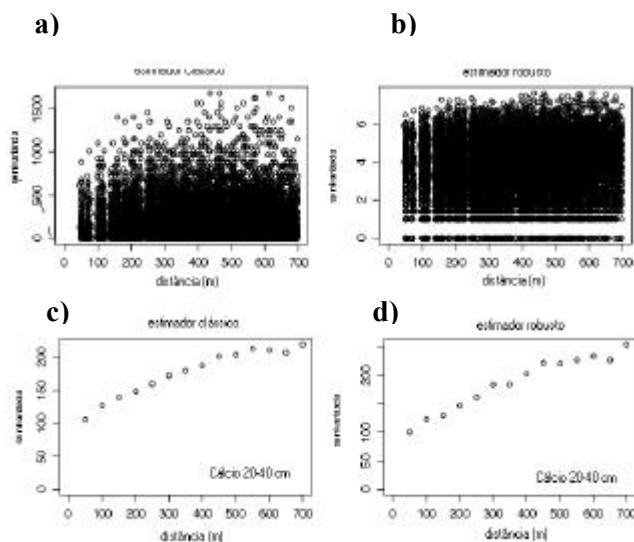


Figura 10 – Variogramas empíricos de nuvens e pontuais pelos métodos de Matheron (a e c) e de Cressie (b e d) para a variável teor de cálcio na camada de 20-40cm do solo.

Embora os variogramas pontuais das Figuras 9 e 10 não sejam precisos devido ao pares de distâncias serem correlacionados, é possível, ainda assim, identificar a existência de dependência espacial para a variável teor de cálcio sempre que ocorrer um aumento da dependência em função de uma distância \mathbf{h} até o patamar e, estabilizar-se em seguida.

Anteriormente a análise espacial, baseada em modelos foram obtidos, pelo processo empírico e, utilizando-se o pacote geoestatístico GeoEas (ENGLUND et al., 1988), os valores iniciais para os parâmetros do modelo exponencial que encontram-se na Tabela 3. Como este programa não contém a estrutura de covariância para a família Matérn, os valores iniciais para a mesma, com $\hat{\epsilon} = 1$ e $\hat{\epsilon} = 2$, foram considerados arbitrários sem qualquer determinação por programa computacional para a camada 0-20cm e, também incluídos na mesma tabela.

Tabela 3. Resultados dos valores iniciais para os parâmetros do modelo exponencial obtidos pelo programa GeoEas nas camadas do solo de 0-20 e 20-40cm.

Modelos Parâmetros	Exponencial		Família Matérn para $\hat{\epsilon} = 1$ e $\hat{\epsilon} = 2$	
	Cálcio 0-20cm	Cálcio 20-40cm	Cálcio 0-20cm	Cálcio 0-20cm
<i>Nugget</i> (δ^2)	15	75	65	65
<i>Sill</i> (σ^2)	160	145	100	100
<i>Range</i> (f)	550	600	200	200

O resultado para a variância *nugget* na camada 20-40cm do solo, com o modelo exponencial, foi maior do que a apresentada na camada superior. Isto pode estar associado à distribuição não homogênea do cálcio no solo realizada através da ação da água e do uso de implementos agrícolas no manejo dos solos.

4.4 Análise geoestatística para modelos com covariável

4.4.1 Estimadores de máxima verossimilhança e máxima verossimilhança restrita com a estrutura de covariância exponencial

As análises do teor de cálcio, na camada de 0-20cm, foram baseadas em modelos e, os valores das estimativas dos parâmetros, do critério de informação de Akaike (AIC) e do logarítmico da função de verossimilhanças, utilizando-se o estimador de máxima verossimilhança, para os modelos com a covariável região e região mais a tendência linear nas coordenadas estão apresentados na Tabela 4. Observou-se que os valores das estimativas dos efeitos *nugget* e do *range*, foram sempre menores quando incluídas as regiões mais a tendência linear nas coordenadas (modelo 2) conforme eq. (44). O range determina a distância a partir da qual um ponto da variável teor de cálcio não tem mais influência sobre a vizinhança.

Tabela 4. Resultados das estimativas dos parâmetros, critério de informação de Akaike (AIC) e logarítmico da função de verossimilhanças para os dados do teor de cálcio na camada de 0-20cm, com estrutura de covariância exponencial, utilizando-se os estimadores de máxima verossimilhança e máxima verossimilhança restrita.

Estimadores Parâmetros	Máxima verossimilhança		Máxima verossimilhança restrita	
	Região	Região + coordenadas	Região	Região + coordenadas
<i>Nugget</i> (δ^2)	3,26	0	13,07	7,35
<i>Sill</i> (δ^2)	98,70	94,77	104,52	108,70
<i>Range</i> (f)	72,61	63,02	104,19	92,57
AIC	1270,65	1271,56	1246,92	1233,06
Razão de Verossimilhanças	3,1	-	17,86	-
Log Verossimilhança	-629,33	-627,78	-617,46	-608,53
Número de Parâmetros	06	08	06	08

O *nugget* (\hat{o}^2) sempre foi menor quando da inclusão de um novo efeito no modelo, no caso a tendência linear nas coordenadas (modelo 2). No caso das estimativas para o *sill* (\hat{o}^2), no modelo 2, este valor também foi pequeno, mas muito próximo do obtido para o modelo com menor número de parâmetros (Tabela 4).

As estimativas dos parâmetros de regressão e as matrizes de variâncias e covariâncias destes, referentes aos modelos das equações (43 e 44), com os estimadores de máxima verossimilhança, máxima verossimilhança restrita e estrutura de covariância exponencial, para o teor de cálcio, na camada 0-20cm do solo, encontram-se nos Anexos B, C, D, E, H, I, J e K.

O resultado do teste da razão de verossimilhanças (Eq. 42) quando compararam-se o modelo 2 e o modelo com menor número de parâmetros (modelo 1) foi 3,1 valor este, inferior ao valor crítico de $\chi^2_{(2,0,05)} = 5,99$. Sendo assim, ao nível de significância de 5%, conclui-se que o ajuste do segundo modelo não foi significativamente melhor do que o modelo 1. Estes resultados foram também confirmados pelo critério de informação de Akaike (AIC) obtido para os dois modelos, em que os parâmetros foram estimados por máxima verossimilhança (Tabela 4). Praticamente iguais indicando que deve ser preferido o modelo apresentado na eq. (43); sendo assim, rejeita-se a hipótese alternativa (modelo 2).

De maneira análoga, e na mesma camada do solo, para os mesmos modelos e com os parâmetros estimados por máxima verossimilhança restrita observou-se que as estimativas da variância *nugget* (\hat{o}^2) foram superiores às estimativas obtidas pela máxima verossimilhança, mas no modelo com maior número de parâmetros (modelo 2) esta estimativa reduziu-se a um pouco mais que a metade. Os valores das estimativas do *sill* foi menor para o modelo com menor número de parâmetros (modelo 1) e o valor do *range* no modelo 2 foi também menor, da mesma forma daquele obtido com o estimador de máxima verossimilhança (Tabela 4).

Utilizando-se o estimador de máxima verossimilhança restrita, a diferença entre o logarítmico da função de verossimilhanças do modelo 2 e o modelo 1 foi 17,86, o qual apresentou valor superior ao valor crítico de $\chi^2_{(2,0,05)} = 5,99$. Portanto, ao nível de significância de 5%, conclui-se que o ajuste do modelo com região mais a tendência linear nas coordenadas foi significativamente melhor do que o modelo 1 e, não se rejeitou a hipótese alternativa. O modelo com maior número de parâmetros (eq. 44), utilizando-se a estrutura de covariância exponencial,

na camada de 0-20cm do solo, apresentou o menor valor de (AIC) e foi o modelo selecionado (Tabela 4).

Um aspecto importante nesta discussão é que os valores das estimativas foram sempre inferiores aos valores iniciais obtidos por um processo subjetivo. Embora obtidos estes valores a comparação entre métodos subjetivos (não-paramétricos) e os baseados em modelos (paramétricos) deve ser realizada com estudos mais complexos utilizando-se centenas ou milhares de simulações. Convém salientar que mesmo assim, os métodos paramétricos são mais confiáveis na qualidade de estimação, principalmente aqueles em que se utilizam os estimadores de máxima verossimilhança (Diggle & Ribeiro Junior, 2000).

Para o teor de cálcio na camada 20-40cm, foram obtidos os valores das estimativas dos parâmetros, do critério de informação de Akaike (AIC) e o logarítmico da função de verossimilhanças, utilizando-se o estimador de máxima verossimilhança e os modelos com a covariável região e região mais a tendência linear nas coordenadas (Tabela 5).

Os valores das estimativas dos parâmetros *nugget*, *range* e *sill* ($\hat{\sigma}^2$, \mathbf{f} e \mathbf{o}^2), foram sempre menores no modelo com a covariável região mais a tendência linear nas coordenadas (modelo 2, eq. 44) quando estes foram estimados por máxima verossimilhança.

As estimativas dos coeficientes de regressão e as matrizes de variâncias e covariâncias destes, conforme os modelos apresentados nas eqs. (43 e 44), com os estimadores de máxima verossimilhança, máxima verossimilhança restrita e estrutura de covariância exponencial, para o teor de cálcio, na camada 20-40cm do solo, encontram-se nos Anexos F, G, L, M, N e O.

Com o mesmo estimador, o resultado para o teste da razão de verossimilhanças (Eq. 42) quando compararam-se os modelos 2 com o modelo 1 foi 4,82, que é inferior ao valor crítico de $\chi^2_{(2,0,05)} = 5,99$. Assim, ao nível de significância de 5%, conclui-se que o ajuste do modelo 2 não foi significativamente melhor do que o modelo com menor número de parâmetros (modelo 1). Estes resultados foram confirmados pelo critério de informação de Akaike (AIC) obtido para os dois modelos, o qual indicou que existe preferência pelo modelo apresentado na eq. (43), sendo assim, rejeita-se a hipótese alternativa (Tabela 5).

Tabela 5. Resultados das estimativas dos parâmetros, critério de informação de Akaike (AIC) e o logarítmico da função de verossimilhanças para os dados do teor de cálcio na camada de 20-40cm, com estrutura de covariância exponencial, utilizando-se os estimadores de máxima verossimilhança e máxima verossimilhança restrita.

Estimadores	Máxima verossimilhança		Máxima verossimilhança restrita	
	Parâmetros	Região	Região + coordenadas	Região
Nugget (δ^2)	80,16	79,68	83,97	85,61
Sill (δ^2)	65,67	60,80	79,82	95,32
Range (f)	101,54	87,25	152,39	200,00
AIC	1377,31	1379,72	1352,69	1339,30
Razão de Verossimilhanças	4,82	-	26,78	-
Log Verossimilhança	-682,66	-681,86	-670,3427	-661,65
Número de Parâmetros	06	08	06	08

A estimativa do parâmetro *nugget* foi menor no modelo 2 e da mesma forma para os parâmetros *sill* e *range* quando os parâmetros foram estimados pelo estimador de máxima verossimilhança.

Com o estimador de máxima verossimilhança restrita, na camada de 20-40cm, o resultado para o teste da razão de verossimilhanças (Eq. 42) quando compararam-se os modelos 2 com o modelo 1 foi 26,78, superior ao valor crítico de $\chi^2_{(2,0,05)} = 5,99$ ao nível de 5% de significância. Foi obtido também o menor valor de AIC = 1339,50, indicando assim, que o modelo escolhido foi aquele com o maior número de parâmetros, não rejeitando-se a hipótese alternativa (Tabela 5).

Comparando-se os valores para o *range* obtidos nas duas camadas, utilizando-se o estimador de máxima verossimilhança, com os dois modelos (eqs. 43 e 44), estes indicaram que para este parâmetro houve variação de 72,61 a 101,54 com uma amplitude de 28,93 metros e, de 63,02 a 87,25 com uma amplitude de 24,23 metros (Tabelas 4 e 5).

Da mesma forma, estes valores de *range*, utilizando-se o estimador de máxima verossimilhança restrita indicaram uma variação de 104,19 a 152,39 com uma amplitude de 48,20 metros (modelo com região) e, de 92,57 a 200 metros, com uma amplitude de 107,43. Este fato indicou que nas diferentes profundidades o comportamento da dependência espacial nas variáveis do solo foram influenciadas pelas técnicas de manejo de solo ao longo dos anos. Este resultado também foi encontrado por Couto & Klamt (1999) quando estudaram o teor de manganês.

Estas variações também foram observadas para o *nugget* também chamado de variabilidade de curta escala, não explicada pelo modelo, foi maior na camada do solo 20-40cm. Isto pode estar relacionado também com os mesmos fatores citados anteriormente, bem como a interação água no solo proveniente das lagoas e do rio, no período das cheias.

Com estes resultados observou-se que existe maior variabilidade na camada 20-40cm e as estimativas dos parâmetros sempre foram grandes, com os dois modelos e estimadores, com exceção do parâmetro *nugget*. Os valores para o critério de informação de Akaike foi bem maior também quando comparados aos obtidos na camada superior. O modelo que apresentou a distância de maior dependência entre as amostras de cálcio foi o modelo exponencial, em que foi utilizado o estimador de máxima verossimilhança restrita com o valor de *range* igual a 200 metros.

Os semivariogramas ajustados e experimentais para os modelos com a covariável região e região mais tendência linear nas coordenadas, com parâmetros estimados por máxima verossimilhança e máxima verossimilhança restrita, nas duas camadas do solo, encontram-se na Figura 11a e b.

A dependência espacial está bem identificada, ou seja, na forma gráfica, para valores pequenos de **h** a semivariância é pequena e com o aumento de **h** a semivariância cresce até um ponto em que estabiliza-se, atingindo o patamar.

Observou-se maiores variâncias *nugget* ($\hat{\sigma}^2$) para os dados da camada 20-40cm e a maior variância espacial ou patamar (σ^2) foi detectada no modelo com região mais tendência linear nas coordenadas (modelo 2) e, que os parâmetros foram estimados por máxima verossimilhança restrita (Figura 11b e Tabela 5).

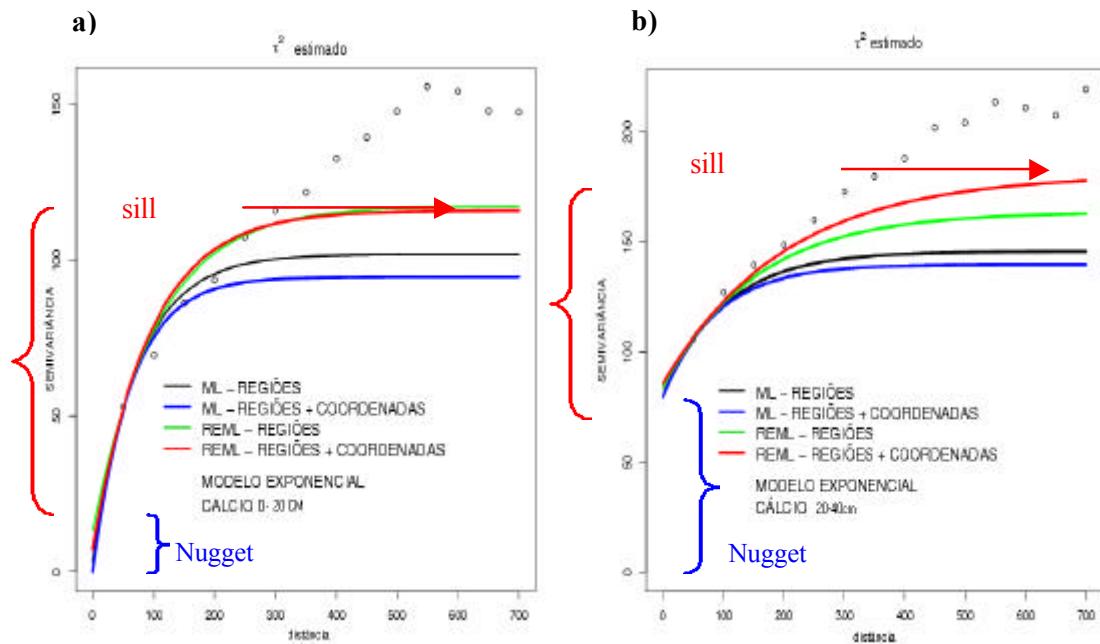


Figura 11 - Variogramas ajustados para os modelos com a covariável região e região mais tendência linear nas coordenadas. Parâmetros estimados por máxima verossimilhança e máxima verossimilhança restrita, considerando-se a estrutura de covariância exponencial, para o cálcio ($\text{mmol}_\text{C}\text{dm}^{-3}$) nas camadas de 0-20cm (a) e de 20-40cm (b) do solo.

Com estes resultados 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).

Uma forma não-paramétrica de identificar a dependência espacial para os modelos (43 e 44) foi utilizando-se do envelope simulado. Os variogramas empíricos e ajustados foram gerados a partir de 178 simulações (Figura 12a e b), para o modelo com a covariável região e a mesma covariável adicionada a tendência linear nas coordenadas (Figura 12c e d).

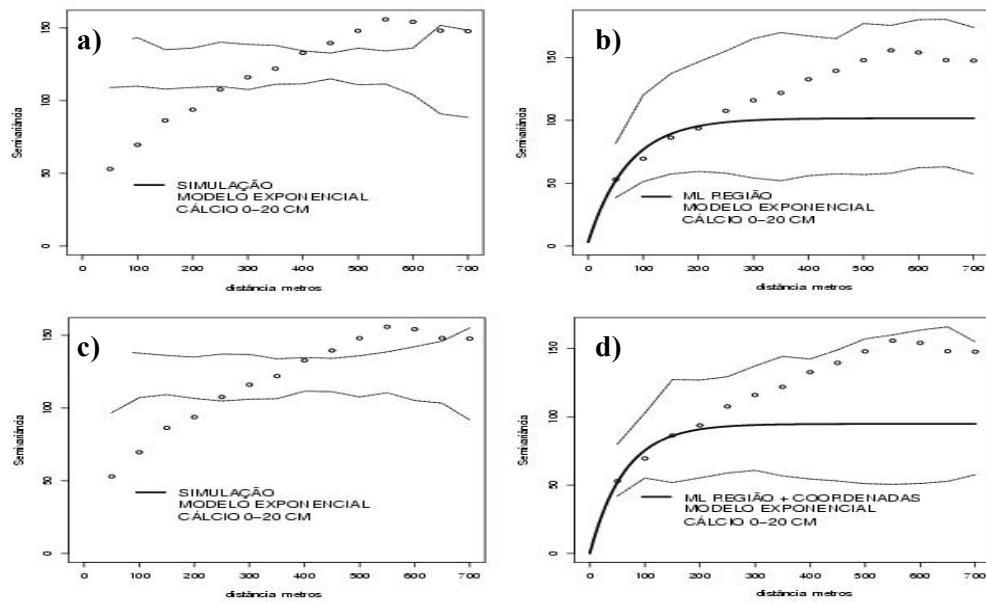


Figura 12 - Variogramas empíricos e ajustados para o modelo com a covariável região (a e b) e, com a covariável região e tendência linear nas coordenadas (c e d), para o teor de cálcio (0-20 cm). Parâmetros ajustados por máxima verossimilhança considerando-se a estrutura de covariância exponencial.

Como os gráficos ajustados encontram-se dentro do envelope pode-se afirmar que os mesmos apresentaram dependência espacial bem definida, ou seja, o grau de incerteza está entre os referidos intervalos do envelope, sendo estimados os parâmetros por máxima verossimilhança (Figura 12b e d).

Estes resultados também foram similares para os mesmos modelos com os parâmetros estimados por máxima verossimilhança restrita nas duas camadas do solo (Figuras 13, 14 e 15a,b,c e d). Observou-se para todas estas figuras que dentro de cada estimador, quando compararam-se os limites inferiores e superiores eles foram semelhantes.

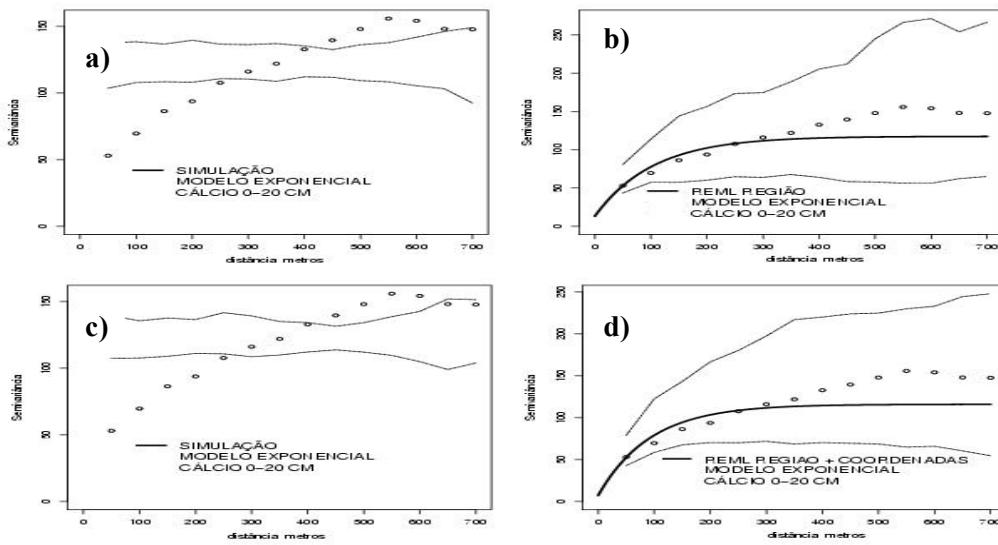


Figura 13 - Variogramas empíricos e ajustados para o modelo com a covariável região (a e b) e região mais tendência linear nas coordenadas (c e d), para o teor de cálcio na camada de (0-20 cm) do solo. Parâmetros estimados por máxima verossimilhança restrita considerando-se a estrutura de covariância exponencial.

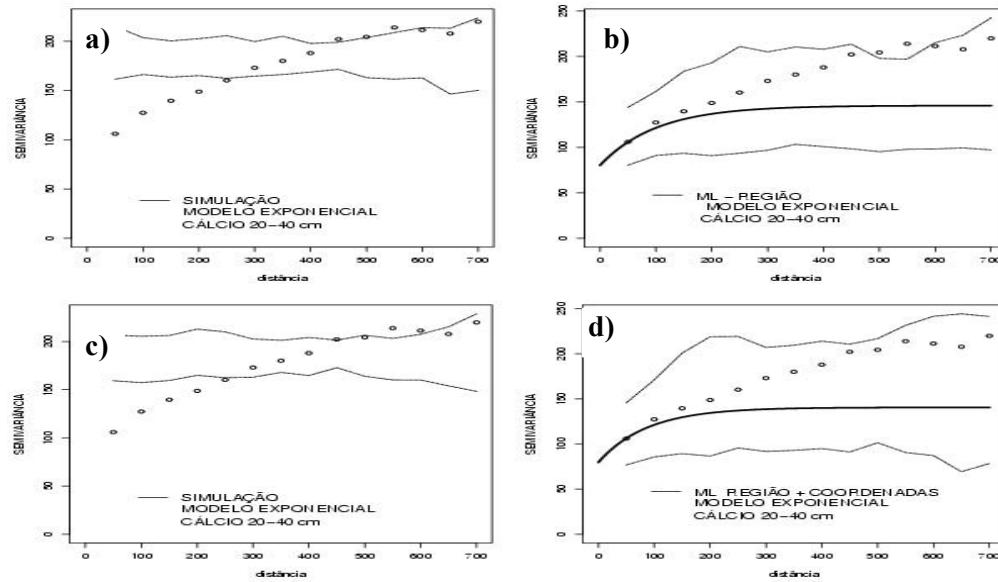


Figura 14 - Variogramas empíricos e ajustados para o modelo com a covariável região (a e b) e região mais tendência linear nas coordenadas (c e d), para o teor de cálcio na camada de (20-40cm) do solo. Parâmetros estimados por máxima verossimilhança considerando-se a estrutura de covariância exponencial.

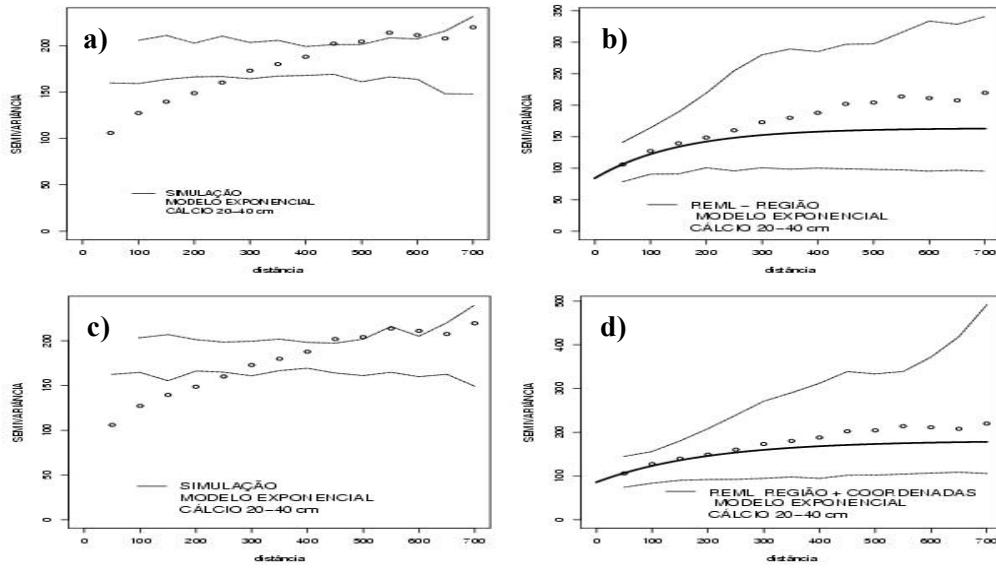


Figura 15 - Variogramas empíricos e ajustados para o modelo com a covariável região (a e b) e região mais tendência linear nas coordenadas (c e d), para o teor de cálcio na camada de (20-40cm) do solo. Parâmetros estimados por máxima verossimilhança considerando-se a estrutura de covariância exponencial.

Após identificar a existência da dependência espacial é importante inspecionar a incerteza das estimativas dos parâmetros da estrutura de covariância pelos perfis de verossimilhança. Os perfis de verossimilhanças para os efeitos *nugget*, *sill* e *range* para os modelos com a covariável região e região mais tendência linear nas coordenadas (eqs. 43 e 44) com os parâmetros estimados por máxima verossimilhança, respectivamente para as camadas do solo de 0-20cm e 20-40cm foram apresentadas nas Figuras 16 e 17. Este perfil para o parâmetro σ^2 (*sill*) ou patamar foi levemente assimétrico, o intervalo de confiança para \mathbf{f} foi amplo, apresentando maior variabilidade e, as duas linhas horizontais indicaram os níveis de significância a 10 e 5% pela distribuição de χ^2 (Figura 16a e b).

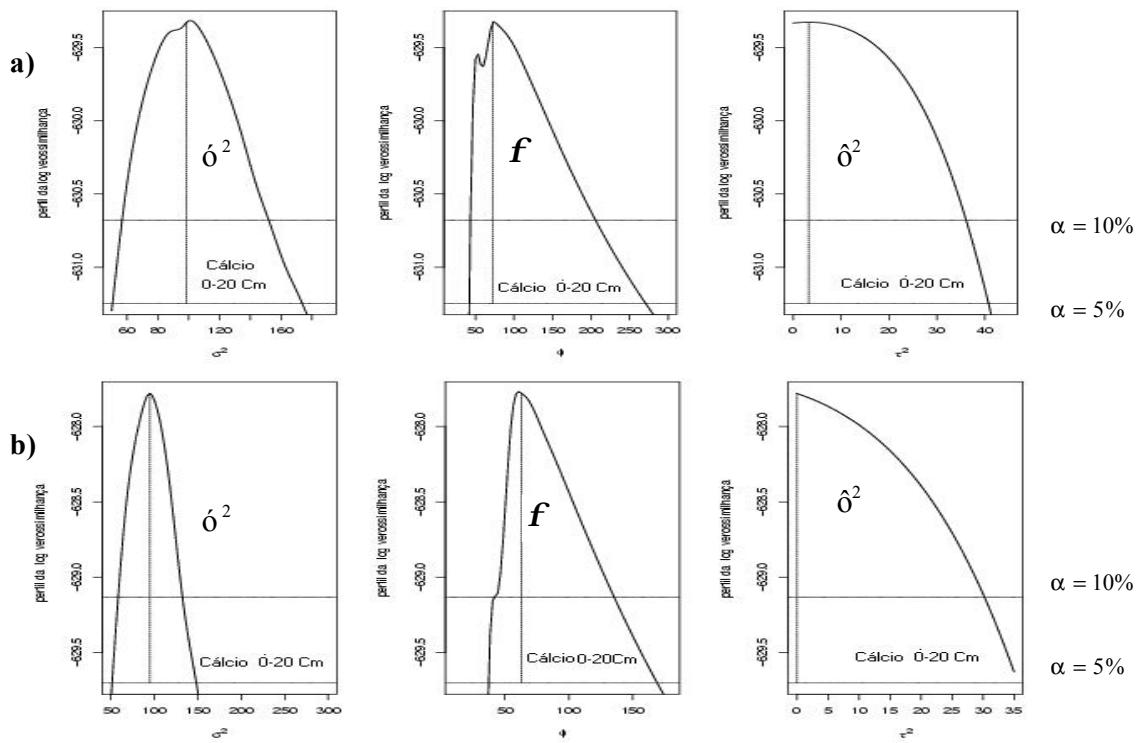


Figura 16 - Perfil de verossimilhanças para os parâmetros com estrutura de covariância exponencial, para o teor de cálcio na camada de (0-20cm) do solo. Modelo com covariável região (a) e região mais tendência linear nas coordenadas (b), com estimativa dos parâmetros ($\hat{\sigma}^2$, f , $\hat{\tau}^2$) por máxima verossimilhança.

De maneira análoga, os perfis do logarítmico da função de verossimilhanças com os mesmos parâmetros, para a variável teor de cálcio (20-40cm), também apresentaram assimetria (*nugget* e *sill*) e picos (*range*) na curva da verossimilhança os quais estão associados ao método de obtenção destes perfis para os três parâmetros (Figura 17a e b).

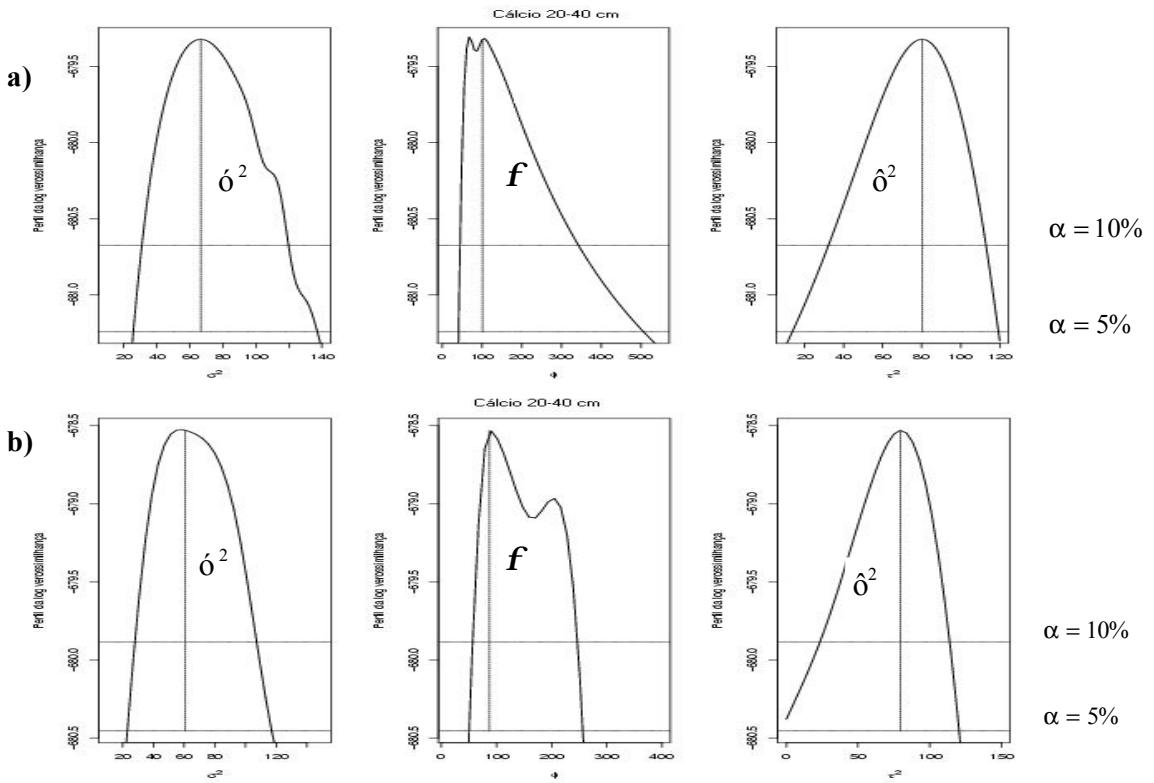


Figura 17 - Perfil de verossimilhança para os parâmetros com a estrutura de covariância exponencial, para o teor de cálcio na camada de (20-40cm) do solo. Modelo com a covariável região (a) e região mais tendência linear nas coordenadas (b) e, parâmetros (σ^2 , f , $\hat{\sigma}^2$) estimados por máxima verossimilhança.

Se existe variabilidade não se pode afirmar que tais estimativas têm alta precisão e foi sugerido realizar a superfície conjunta do perfil de verossimilhança entre os parâmetros de covariância σ^2 e ϕ (Diggle & Ribeiro Junior, 2000).

Uma vez estimados os parâmetros para os modelos com estrutura de covariância exponencial nas duas camadas do solo foi realizada a análise espacial e, os resultados das predições e variâncias das predições do teor de cálcio estão apresentadas na Figura 18a e b, respectivamente, para os modelos com a covariável região e região mais tendência linear nas coordenadas com os parâmetros estimados por máxima verossimilhança.

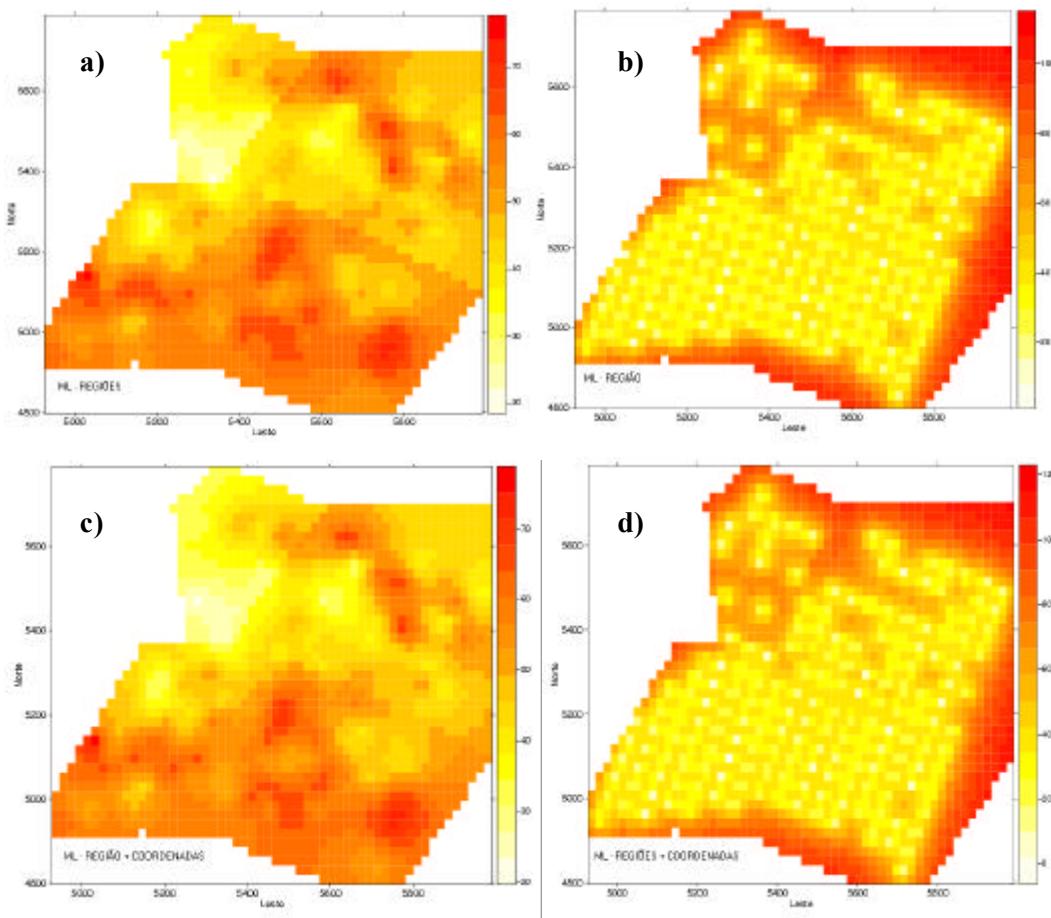


Figura 18 - Resultados da predição (a e c) e variância da predição (b e d), para o teor de cálcio (0-20cm). Modelo com a covariável região (a e b) e região mais tendência linear nas coordenadas (c e d). Parâmetros estimados por máxima verossimilhança considerando-se a estrutura de covariância exponencial.

À medida que as covariáveis foram incorporadas aos modelos pode-se observar que o padrão da predição mudou (Figura 18a e c) 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. Isto significou que as previsões cujas variâncias estão associadas são pequenas no local do estudo, mas mesmo assim, estas são similares nos modelos com região e região mais a tendência linear nas coordenadas.

A variância da predição apresentou boa confiabilidade, entretanto, nas bordas, por ausência de informações é maior, quando associada com as cores mais fortes (Figura 18b e d). Uma das justificativas para este fato 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 18b e d). De maneira geral observou-se que as variâncias estão muito próximas e que nos pontos mais claros estas variâncias são menores.

A predição obtida utilizando-se os parâmetros estimados por máxima verossimilhança restrita, com os modelos com região e região mais tendência linear nas coordenadas (modelo 2) apresentou áreas melhores definidas nos locais onde encontram-se os maiores valores de cálcio (Figura 19a e b). As variâncias das predições foram menores para o modelo com a covariável região mais a tendência linear nas coordenadas (modelo 2), mas os resultados não diferiram. Isto permitirá que sejam identificadas corretamente as regiões onde existe maior deficiência do cálcio reduzindo-se os custos na instalação de experimentos de campo.

A validação cruzada, em geral, para todos os modelos e estimadores deste estudo, indicaram que os dados do teor de cálcio nas duas camadas do solo, seguiram a distribuição normal padronizada, mas será aqui apresentada somente a figura para os resultados do melhor modelo e estimador que permitiu obter o menor logarítmico da função de verossimilhanças e menor valor do AIC dentro de cada estrutura de covariância (Figura 20).

Observou-se, ainda, que os valores preditos estão próximos da reta padrão, a distribuição dos erros positivos e negativos estão dispersos pela região de estudo e os dados da probabilidade teórica e observada encontram-se sobre a reta, indicando que o método utilizado para a predição, da krigagem ordinária foi eficiente (item 2.6 e Figura 20).

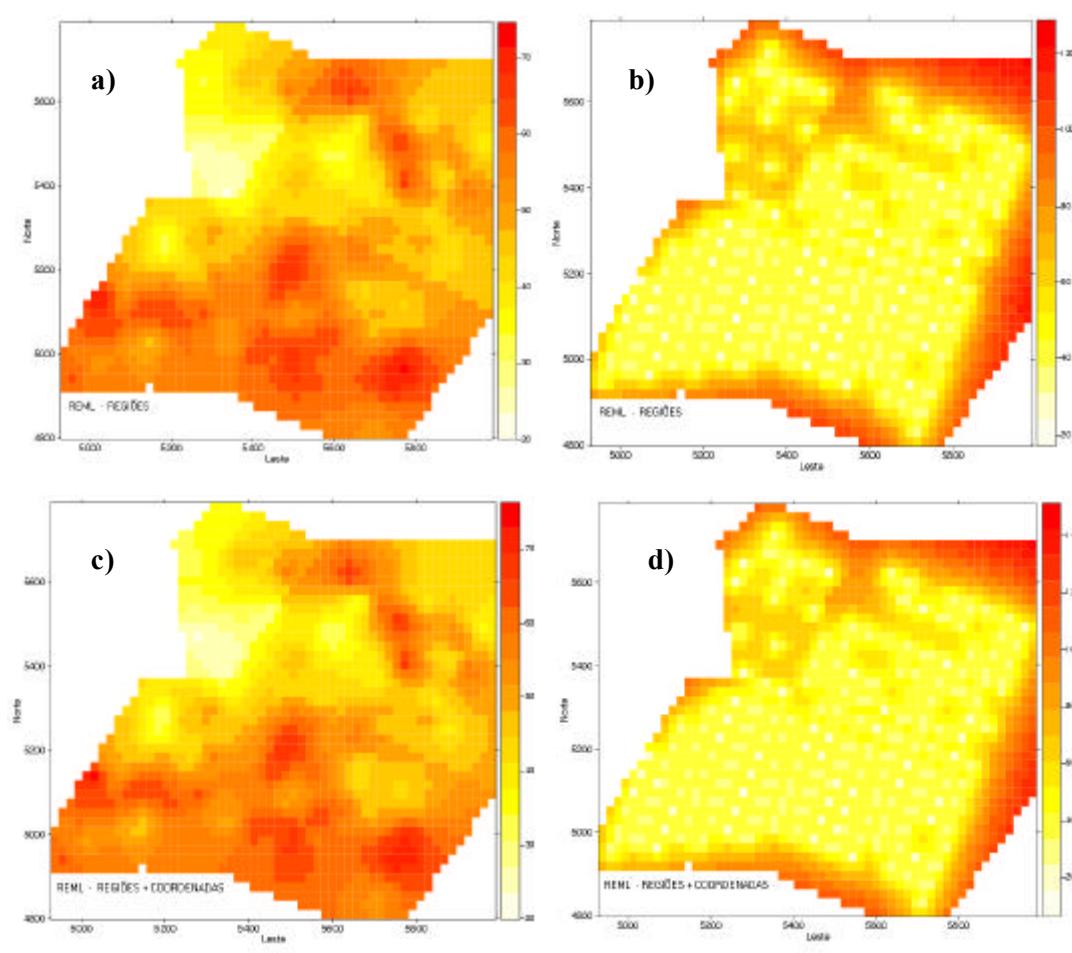


Figura 19 - Resultados da predição (a e c) e variância da predição (b e d), para o teor de cálcio (0-20 cm). Modelo com a covariável região (a e b) e região mais tendência linear nas coordenadas (c e d). Parâmetros estimados por máxima verossimilhança restrita considerando-se a estrutura de covariância exponencial.

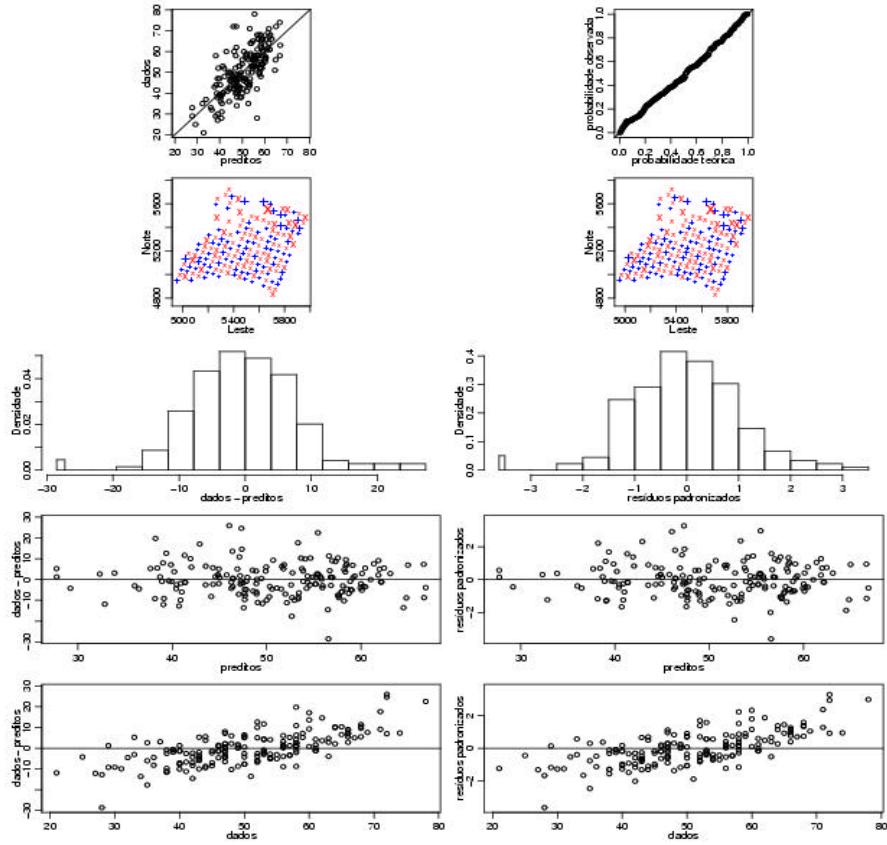


Figura 20 – Validação cruzada, para o teor de cálcio, na camada 20-40cm, utilizando-se o modelo com a covariável região mais tendência linear nas coordenadas e parâmetros estimados pelo método de máxima verossimilhança restrita considerando-se a estrutura de covariância exponencial.

Na camada 20-40cm as previsões para os modelos das eqs. (43 e 44), respectivamente, indicaram uma grande região com os maiores valores do cálcio coincidindo com a região adubada num período mais recente (Figura 21a e c). A região das variâncias das previsões indicaram que as variâncias são bem menores em quase toda região e que há boa precisão das estimativas. No modelo 2 estas respostas são similares e são apresentadas com maior nitidez, ou seja, à medida que são incorporados efeitos nos modelos e os parâmetros são estimados com métodos mais fidedignos o fenômeno em estudo pode ser melhor interpretado (Figura 21b e d).

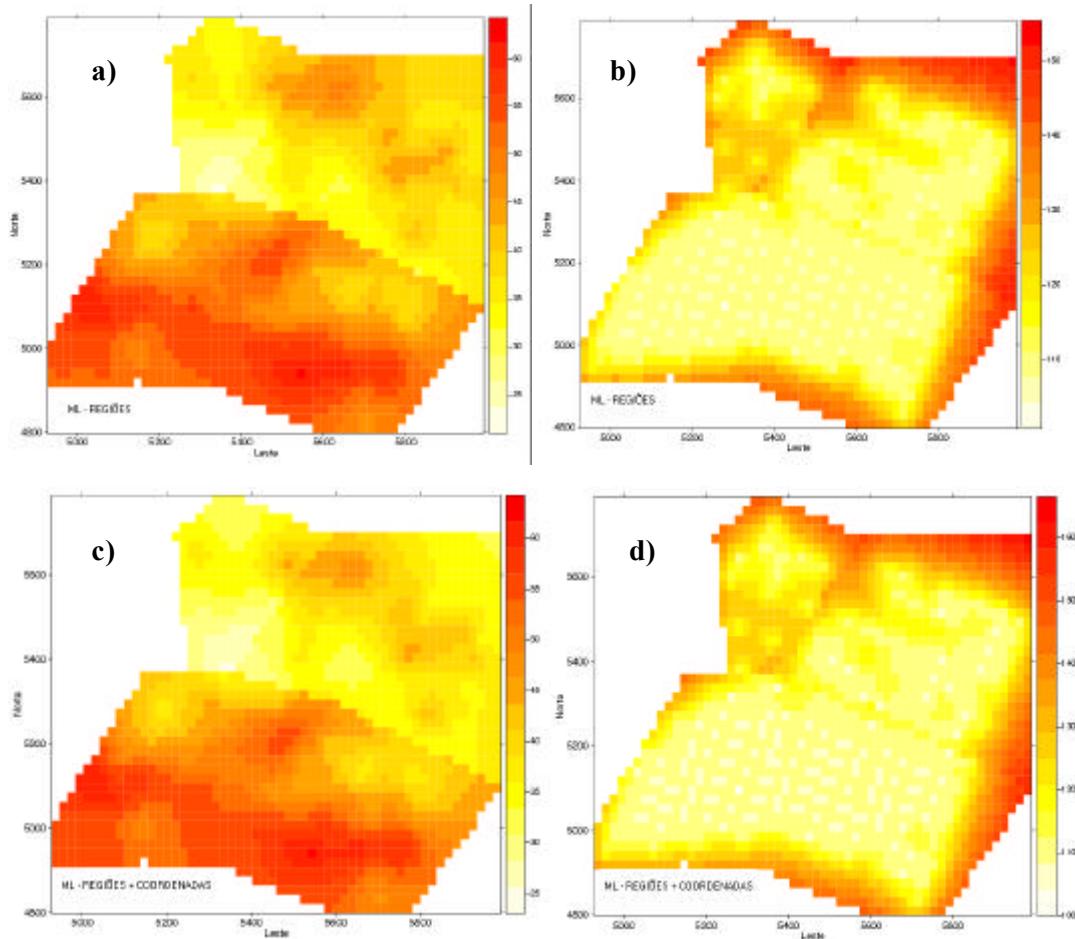


Figura 21 - Resultados da predição (a e c) e variância da predição (b e d), para a variável teor de cálcio na camada 20-40cm. Modelo com a covariável região (a e b) e região mais tendência linear nas coordenadas (c e d). Parâmetros estimados por máxima verossimilhança considerando-se a estrutura de covariância exponencial.

Análise similar foi realizada para os mesmos modelos com o estimador de máxima verossimilhança restrita (Figura 22). A evidência da qualidade da predição com o seu uso é grande, bem como modelos adequados para a interpretação dos dados de pesquisa. A qualidade deste estimador já é consagrada na literatura, isto é suas estimativas são não viesadas e só é penalizado quando aumenta-se muito o número de parâmetros, no caso de polinômios e, que o mesmo estima separadamente os parâmetros de efeitos aleatórios e fixos do modelo, conforme a teoria de modelos mistos (item 2.7.2.1). Observou-se que com este estimador com os modelos

apresentados pelas eqs. (43 e 44) as predições são bem definidas e as variâncias associadas a estas predições são em geral muito pequenas tanto para o modelo com covariável quanto para o modelo com região mais tendência linear nas coordenadas (modelo 2) (Figura 22 a,b,c e d).

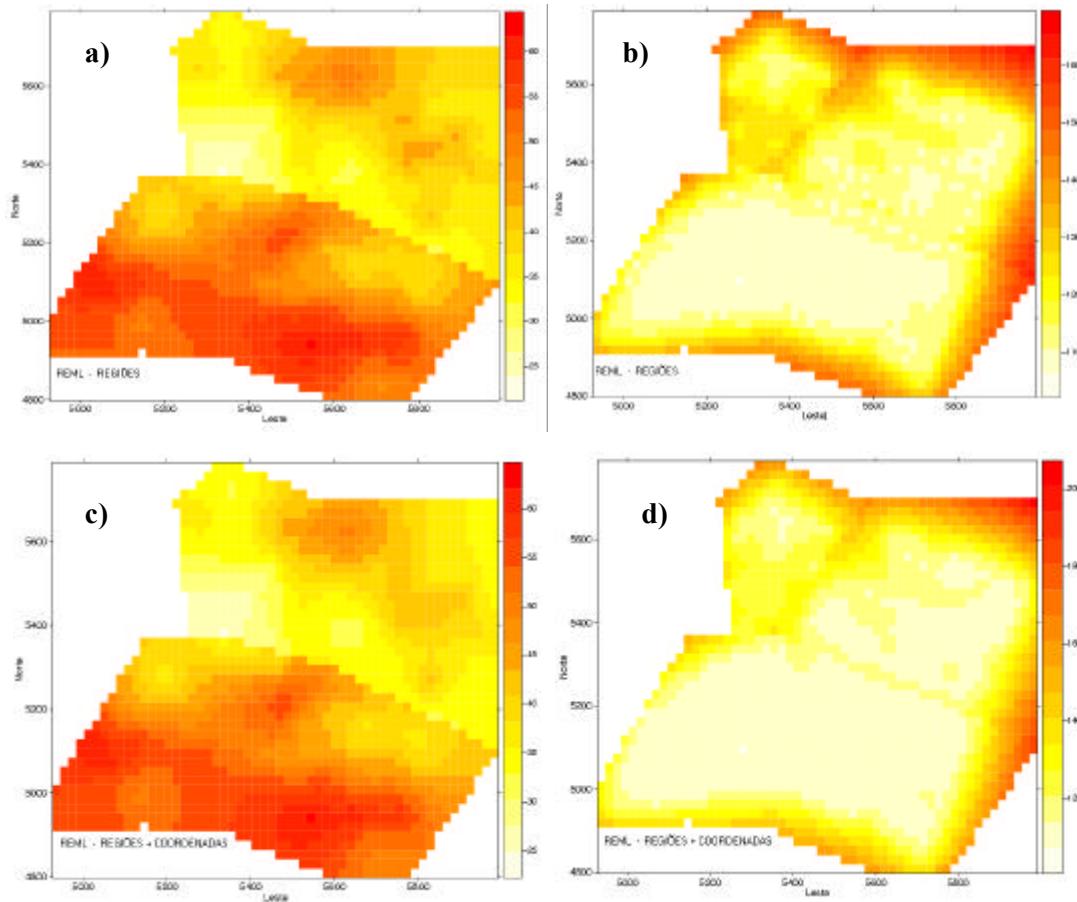


Figura 22 - Resultados da predição (a e c) e variância da predição (b e d), para o teor de cálcio (20-40cm). Modelo com a covariável região (a e b) e região mais tendência linear nas coordenadas (c e d). Parâmetros estimados por máxima verossimilhança restrita considerando-se a estrutura de covariância exponencial.

Os resultados para a validação cruzada para o teor de cálcio na camada 20-40cm foram similares aos obtidos na camada superior, para o modelo 2 e parâmetros estimados por máxima verossimilhança restrita. Este fato está diretamente associado a qualidade das estimativas obtidas e do modelo escolhido. Verificou-se que os valores preditos estão próximos da reta; a distribuição dos erros positivos e negativos está dispersa pela região de estudo e os

dados da probabilidade teórica e observada encontram-se sobre a reta, indicando que o método da krigagem ordinária utilizado para a predição foi eficiente (item 2.6 e Figura 23).

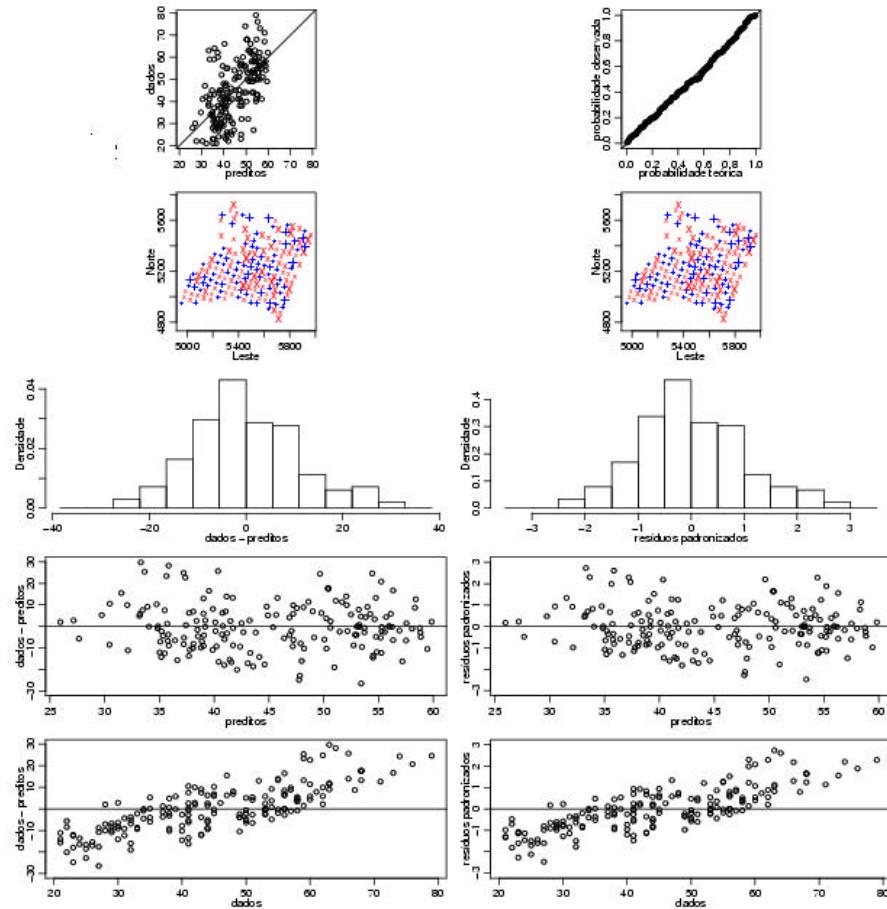


Figura 23 - Validação cruzada, para o teor de cálcio, na camada 20-40cm, utilizando-se o modelo com a covariável região mais tendência linear nas coordenadas. Parâmetros estimados pelo método de máxima verossimilhança restrita considerando-se a estrutura de covariância exponencial.

4.4.2 Estimadores de máxima verossimilhança e máxima verossimilhança restrita com a estrutura de covariância pela família Matérn para $\hat{\epsilon} = 1$ e $\hat{\epsilon} = 2$.

Fundamentando-se na análise baseada em modelos, os valores das estimativas dos parâmetros, critério de informação de Akaike (AIC) e o logarítmico da função de verossimilhanças, utilizando-se os estimadores de máxima verossimilhança e máxima verossimilhança restrita e, os modelos das eqs (43 e 44) estão apresentados nas Tabelas 6 e 7 para a estrutura de covariância Matérn para $\hat{\epsilon} = 1$ e $\hat{\epsilon} = 2$. Para esta estrutura de covariância foram utilizados somente os dados do teor de cálcio na camada de (0-20cm) do solo.

Para a estrutura de covariância Matérn para $\hat{\epsilon} = 1$, os valores das estimativas da variância *nugget* e *range* foram menores para o modelo 2 (eq. 44) estimando-se os parâmetros por máxima verossimilhança e, o valor da estimativa para o *sill* (σ^2) foi menor para o modelo com a covariável região (modelo 1), mas com pequena diferença comparada aos valores obtidos para o modelo 2 (eq. 44).

Tabela 6. Resultados das estimativas dos parâmetros, critério de informação de Akaike (AIC) e logarítmico da função de verossimilhanças para os dados do teor de cálcio na camada de 0-20cm, com estrutura de covariância Matérn com o parâmetro $\hat{\epsilon} = 1$, utilizando-se o estimador de máxima verossimilhança e máxima verossimilhança restrita.

Estimadores	Máxima verossimilhança		Máxima verossimilhança restrita		
	Parâmetros	Região	Região + coordenadas	Região	Região + coordenadas
Nugget (σ^2)	24,45	9,41	30,62	26,16	
Sill (σ^2)	76,00	83,85	82,07	84,55	
Range (f)	53,07	38,91	68,45	61,01	
AIC	1271,20	1271,74	1247,74	1233,84	
Razão de Verossimilhanças	3,45	-	17,90	-	
Log Verossimilhança	-629,60	-627,87	-617,87	-608,92	
Número de Parâmetros	06	08	06	08	

Comparando-se os valores para o *range*, estimados por máxima verossimilhança com a estrutura de covariância Matérn com $\hat{e} = 1$ indicaram que a dependência máxima variou de 38,91 a 53,07 metros, com uma amplitude de 14,16 metros para a camada 0-20cm. Conclui-se que a dependência espacial máxima estabilizou-se somente a distância de 53 metros quando usou-se o modelo com menor número de parâmetros (modelo 1, eq. 43) e para o caso do modelo 2 a aleatorização das amostras ocorreram acima de 39 metros.

O resultado para o teste da razão de verossimilhanças (Eq. 42) quando foram comparados o modelo 2 com o modelo com menor número de parâmetros (modelo 1) foi 3,45 o qual apresentou valor inferior ao valor crítico de $\chi^2_{(2,0,05)} = 5,99$. Desta forma, com o nível de significância de 5%, conclui-se que o ajuste do modelo 2 não foi significativamente melhor do que o modelo com menor número de parâmetros (Tabela 6). Estes resultados também foram coincidentes pelo critério de informação de Akaike (AIC) obtidos para os dois modelos. Sendo assim, baseando-se nesses resultados, rejeitou-se a hipótese alternativa e existe preferência pelo modelo apresentado na eq. (43).

Com o estimador de máxima verossimilhança restrita e estrutura de covariância Matérn com $\hat{e} = 1$, também os valores das estimativas para o *nugget* e *range* foram menores para o modelo 2 (eq. 44) e o valor da estimativa para o *sill* (σ^2) foi menor para o modelo com a covariável região (modelo 1), mas com pequena diferença comparada aos valores obtidos para o modelo com região mais a tendência linear nas coordenadas (modelo 2). A dependência espacial máxima nos dois modelos variou de 61,01 a 68,45 metros (Tabela 6).

O resultado para o teste da razão de verossimilhanças (Eq. 42), para a estrutura de covariância Matérn $\hat{e} = 1$, quando compararam-se os modelos das eqs. (43 e 44) foi 17,90. Este resultado foi confirmado pelo critério de informação de Akaike em que obteve-se o menor valor deste critério com o modelo com a covariável região mais a tendência linear nas coordenadas (eq. 44). Portanto, não se rejeita a hipótese alternativa e o modelo escolhido foi o modelo 2 (Tabela 6).

Na estrutura de covariância da família Matérn para $\hat{e} = 2$, usando-se o estimador de máxima verossimilhança, na camada de 0-20cm do solo, os valores das estimativas do *nugget* (σ^2) e do *range* (*f*) foram menores para o modelo 2 (eq. 44) e o valor da variação total (*sill*) foi menor para o modelo com menor número de parâmetros (modelo 1). A distância máxima da dependência espacial nos dois modelos variou de 64,25 a 69,72 metros (Tabela 7).

O valor do teste da razão de verossimilhanças (Eq. 42) quando comparados os modelos com região mais a tendência linear nas coordenadas (modelo 2) e com o menor número de parâmetros (modelo 1) foi 3,57. Este valor foi inferior ao valor crítico de $\chi^2_{(2,0,05)} = 5,99$. Sendo assim, ao nível de significância de 5%, conclui-se que o ajuste do modelo 2 não foi significativamente melhor do que o modelo com menor número de parâmetros (modelo 1). Estas respostas foram confirmadas pelo critério de informação de Akaike (AIC) obtido para os modelos das eqs (43 e 44), em que os valores do AIC e o logarítmico da verossimilhança são praticamente os mesmos. Desta forma, rejeitou-se a hipótese alternativa e o modelo com o menor número de parâmetros (modelo 1) foi selecionado quando utilizou-se o estimador de máxima verossimilhança e a estrutura de covariância Matérn para $\hat{e} = 2$ (Tabela 7).

Tabela 7. Resultados das estimativas dos parâmetros, critério de informação de Akaike (AIC) e logarítmico da função de verossimilhanças para os dados do teor de cálcio na camada de 0-20cm, com estrutura de covariância Matérn com $\hat{e} = 2$, utilizando-se o estimador de máxima verossimilhança e máxima verossimilhança restrita.

Estimadores Parâmetros	Máxima verossimilhança		Máxima verossimilhança restrita	
	Região	Região + coordenadas	Região	Região + coordenadas
Nugget (δ^2)	34,77	22,71	38,91	35,48
Sill (σ^2)	64,25	69,72	70,06	71,61
Range (f)	37,56	28,10	45,36	40,97
AIC	1271,91	1272,34	1248,69	1234,77
Razão de Verossimilhanças	3,57	-	17,92	-
Log Verossimilhança	-629,95	-628,17	-618,34	-609,38
Número de Parâmetros	06	08	06	08

As estimativas dos coeficientes de regressão e matrizes de variâncias e covariâncias, para os modelos com a covariável região e região mais tendência linear nas

coordenadas, com os dois estimadores estudados, para a estrutura de covariância Matérn para $\hat{\epsilon} = 1$ e $\hat{\epsilon} = 2$, com o teor de cálcio, na camada de 0-20cm do solo, estão apresentados nos Anexos B, C, D, E, P, Q, R, S, T, U, V e X.

Os valores obtidos para o *range* quando utilizado o estimador de máxima verossimilhança restrita indicou que a dependência máxima variou de 40,97 a 45,36, com uma amplitude de 4,39 metros, para a estrutura de covariância Matérn com $\hat{\epsilon} = 2$, na camada 0-20cm do solo. As estimativas dos parâmetros *nugget* e *range* foram menores para o modelo 2. A variância total (*sill*) nos dois modelos foram similares (Tabela 7).

Com o estimador de máxima verossimilhança restrita, o resultado do teste da razão de verossimilhanças (Eq. 42), considerando-se a estrutura de covariância Matérn $\hat{\epsilon} = 2$, para os modelos apresentados nas eqs. (43 e 44) foi de 17,92. Obteve-se para o modelo com a covariável região e tendência linear nas coordenadas (eq. 44), os menores valores de AIC e do logarítmico da função de verossimilhanças. Portanto, não rejeitou-se a hipótese alternativa e o modelo escolhido foi o modelo com a covariável região mais a tendência linear nas coordenadas (modelo 2).

Comparando-se os resultados obtidos para a estrutura de covariância da família Matérn para $\hat{\epsilon} = 1$ e $\hat{\epsilon} = 2$, observou-se que com o estimador de máxima verossimilhança, a variância total (*sill*) foi maior para $\hat{\epsilon} = 1$, obtendo-se o maior valor para o modelo 2. À medida que aumentou o grau do parâmetro de suavização do processo estocástico ($\hat{\epsilon}$) decresceu a variância total (*sill*), bem como a distância de maior dependência (*f*) entre as amostras e aumentou-se a variância *nugget* (Tabelas 6 e 7). Com os resultados dessas tabelas, foram construídos os semivariogramas ajustados e experimentais e bem identificadas a dependência espacial conforme apresentada na Figura 24a e b.

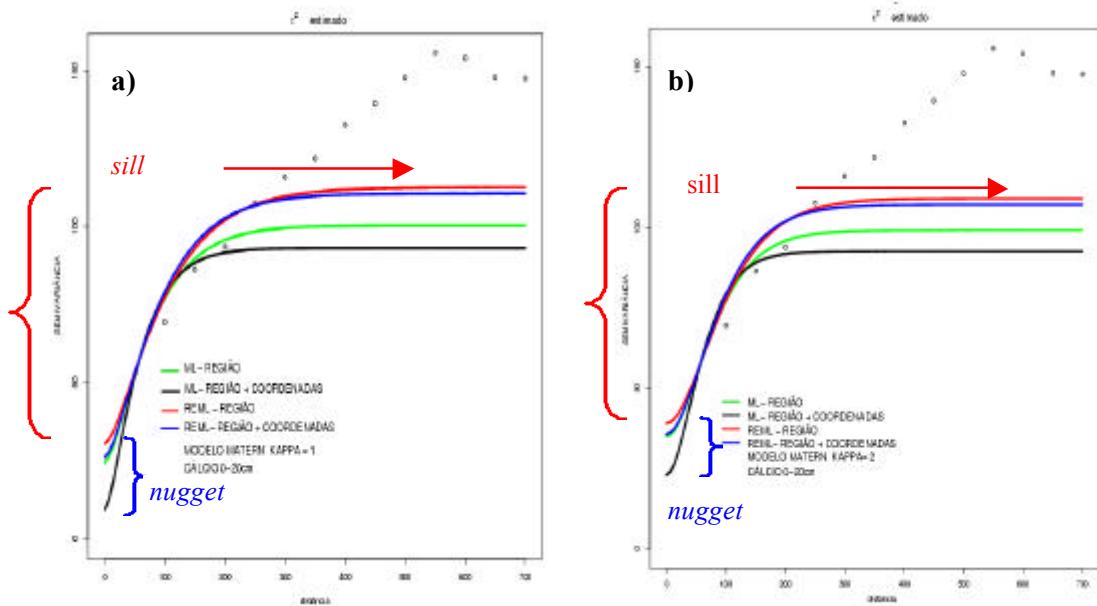


Figura 24 – Variogramas ajustados para o teor de cálcio ($\text{mmol}_\text{C} \text{dm}^{-3}$) para os modelos com covariável região e região mais a tendência linear nas coordenadas. Parâmetros estimados por máxima verossimilhança e máxima verossimilhança restrita, com estrutura de covariância Matérn na camada 0-20cm para $\hat{\epsilon} = 1$ (a) e $\hat{\epsilon} = 2$ (b).

Na identificação da dependência espacial foi usado um método não-paramétrico considerando-se os modelos (43 e 44), pelo envelope simulado, com a variável teor de cálcio na camada 0-20cm, para os dois parâmetros da estrutura de Matérn e, nas duas situações, pode-se confirmar que existe dependência espacial para os dados de cálcio utilizando-se os estimadores de máxima verossimilhança e máxima verossimilhança restrita (Figuras 25 e 26). Os variogramas empíricos gerados a partir de 178 simulações para o modelo com a covariável região e o segundo, a mesma covariável adicionada a tendência linear das coordenadas (norte, leste) são apresentados nas mesmas figuras.

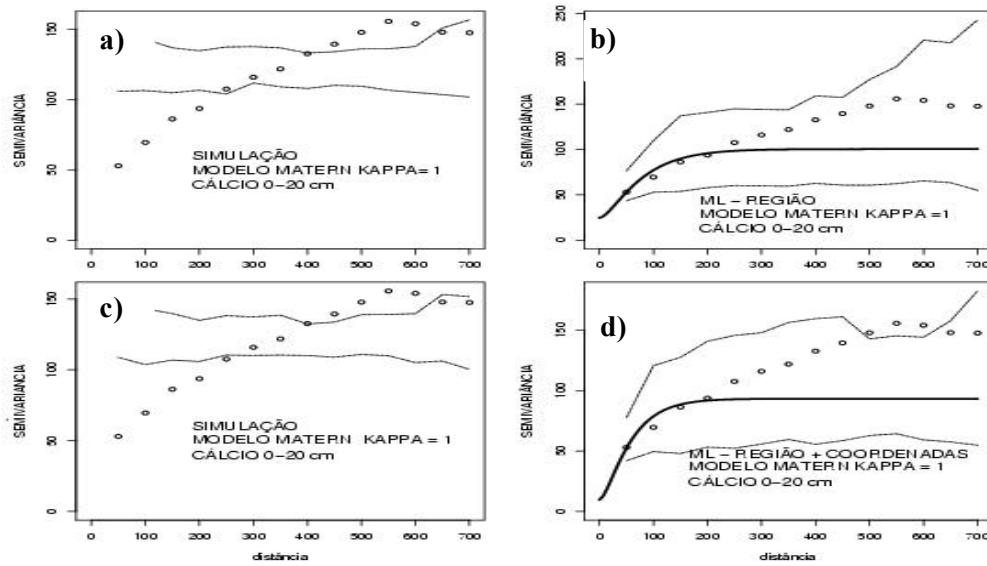


Figura 25 - Variogramas empíricos e ajustados para o modelo com a covariável região (a e b) e, região mais tendência linear nas coordenadas (c e d), para o teor de cálcio na camada de (0-20 cm) do solo. Parâmetros estimados por máxima verossimilhança considerando-se a estrutura de covariância Matérn com $\hat{\epsilon} = 1$.

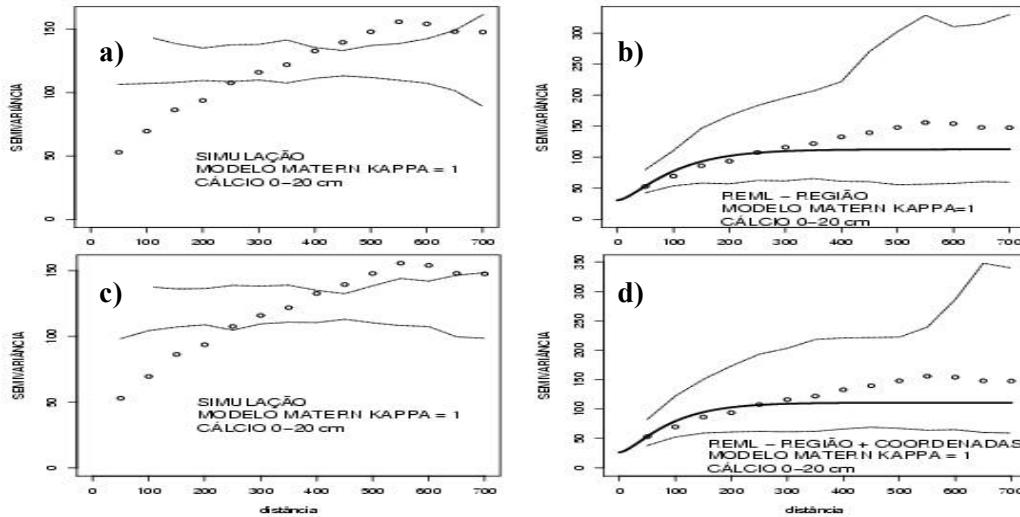


Figura 26 - Variogramas empíricos e ajustados para o modelo com a covariável região (a e b) e, região mais tendência linear nas coordenadas (c e d), para o teor de cálcio na camada (0-20cm) do solo. Parâmetros estimados por máxima verossimilhança restrita considerando-se a estrutura de covariância Matérn com $\hat{\epsilon} = 1$.

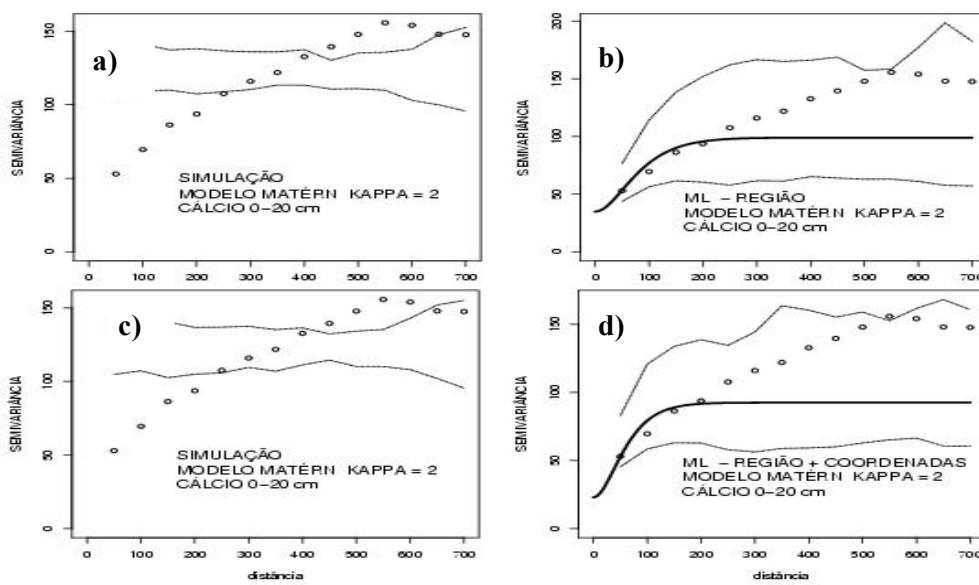


Figura 27 - Variogramas empíricos e ajustados para o modelo com covariável região (a e b) e, região mais tendência linear nas coordenadas (c e d), para o teor de cálcio na camada de (0-20cm). Parâmetros estimados por máxima verossimilhança considerando-se a estrutura de covariância Matérñ com $\hat{\epsilon} = 2$.

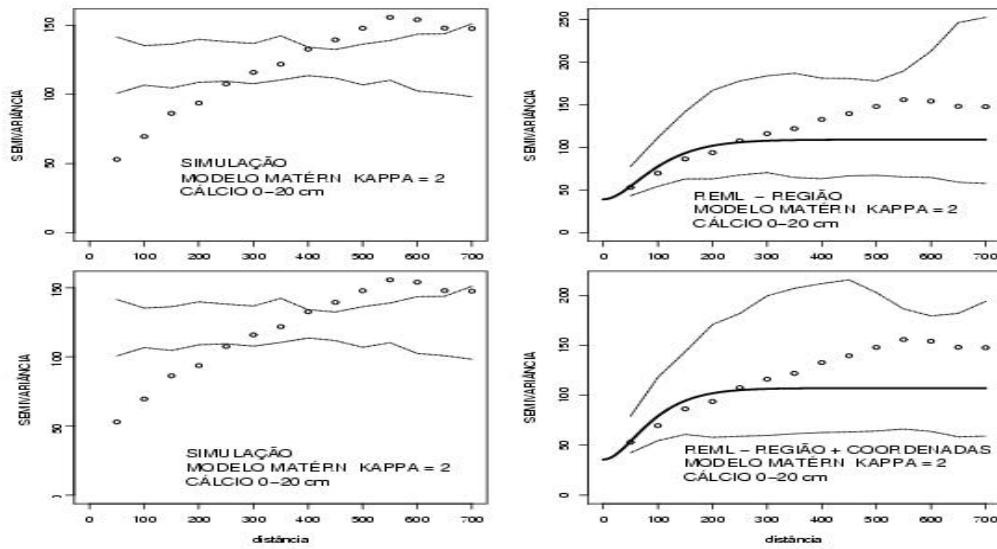


Figura 28. Variogramas empíricos e ajustados para o modelo com a covariável região (a e b) e, região mais tendência linear nas coordenadas (c e d), para o teor de cálcio na camada de (0-20cm). Parâmetros estimados por máxima verossimilhança restrita considerando-se a estrutura de covariância Matérñ com $\hat{\epsilon} = 2$.

Nos dois tipos de variogramas foram obtidas as estimativas dos parâmetros, considerando-se a estrutura de covariância da família Matérn para $\hat{\epsilon} = 1$ e $\hat{\epsilon} = 2$ e os métodos de máxima verossimilhança (ML) e máxima verossimilhança restrita (REML) (Figuras 27a e b e 28a e b).

Observou-se que os variogramas experimentais e ajustados encontram-se dentro dos limites inferior e superior do envelope e apresentaram continuidade espacial para todos os modelos, ou seja, o grau de incerteza está entre os referidos intervalos do envelope (Figuras 27a e b e 28a e b).

Identificada a dependência espacial é importante inspecionar a incerteza das estimativas dos parâmetros da estrutura de covariância. Desta forma, foram realizados os perfis de verossimilhanças para os efeitos *nugget*, *sill* e *range* para os modelos conforme eqs. (43 e 44) estimados pelo estimador de máxima verossimilhança (Figuras 29 e 30), respectivamente para $\hat{\epsilon} = 1$ e $\hat{\epsilon} = 2$ na camada do solo de 0-20cm.

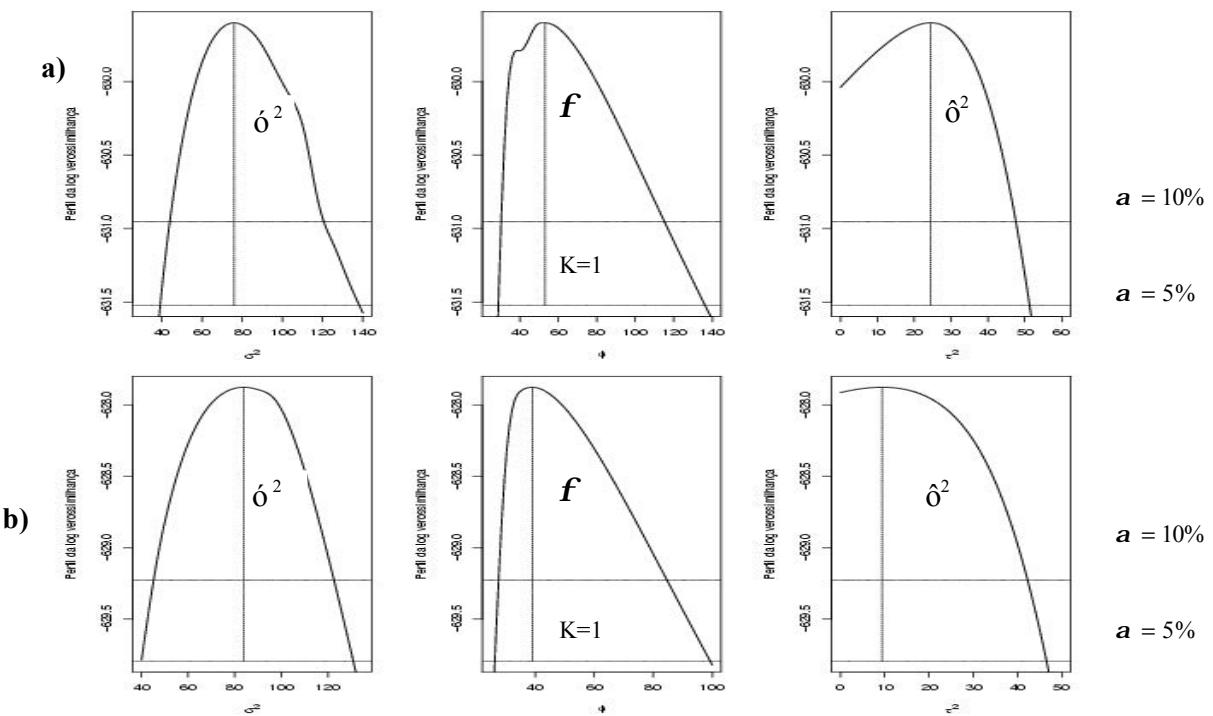


Figura 29 – Perfil de verossimilhança para os parâmetros do modelo Matérn com $\hat{\epsilon} = 1$, para a variável teor de cálcio (0-20cm). Modelo com a covariável região (a) e, região mais tendência linear nas coordenadas (b) com estimativa dos parâmetros (σ^2, f, τ^2) por máxima verossimilhança.

O perfil de verossimilhanças para o *sill* e *range* foram assimétricos quando utilizou-se a família Matérn para $\hat{\epsilon} = 1$ com os modelos apresentados nas eqs. (43 e 44). O intervalo de confiança foi amplo para o alcance e para o patamar obteve-se maior variabilidade. As linhas horizontais indicaram os níveis de significância a 10 e 5% pela distribuição de χ^2 (Figura 29a). No modelo com a covariável região e tendência linear nas coordenadas a variância espacial (σ^2) apresentou-se simétrica e, para o parâmetro *range* foi assimétrico com intervalo de confiança amplo (Figura 29b).

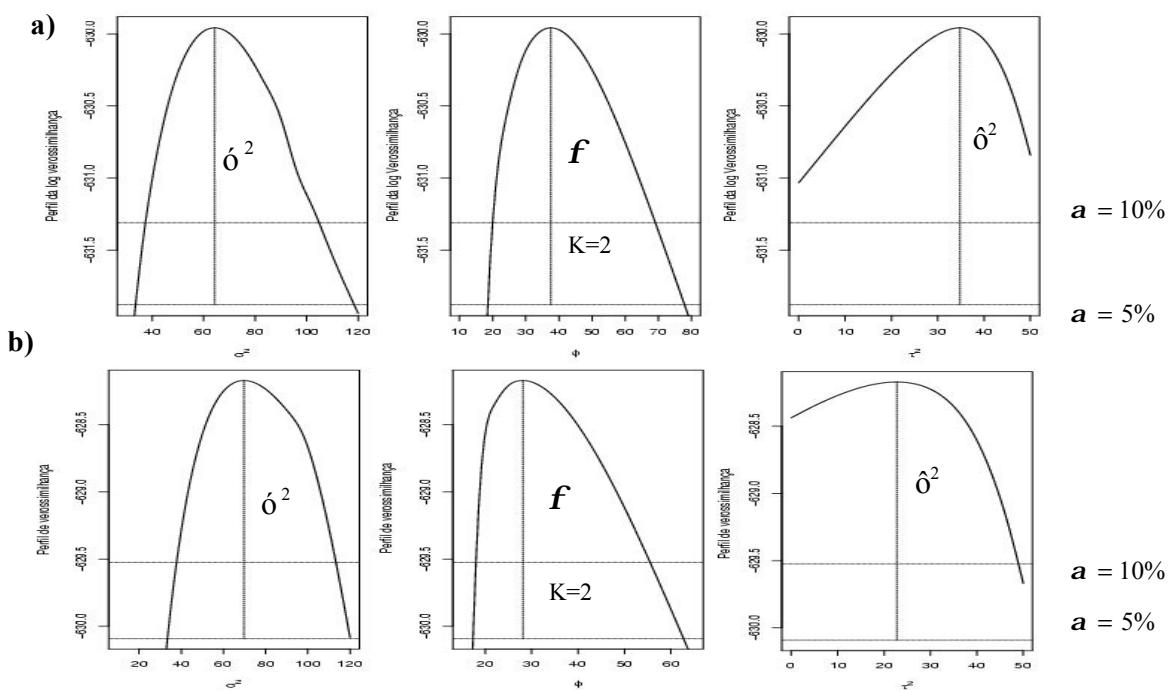


Figura 30 - Perfil de verossimilhanças para os parâmetros do modelo Matérn com $\hat{\epsilon} = 2$, para a variável teor de cálcio (0-20cm). Modelo com a covariável região (a) e, região mais tendência linear nas coordenadas (b) com estimativa dos parâmetros ($\sigma^2, f, \hat{\sigma}^2$) por máxima verossimilhança.

Comparando-se os perfis de verossimilhanças para a estrutura de covariância Matérn, os parâmetros de covariâncias (σ^2, ϕ) foram mais simétricos quando utilizou-se o modelo com a covariável região mais a tendência linear nas coordenadas (modelo 2) e o parâmetro $\hat{\epsilon} = 2$, ou seja, a variabilidade foi menor com este parâmetro (Figura 30).

Em geral, os perfis de verossimilhanças para a estrutura de covariância Matérn indicaram pequena variabilidade se comparado a estrutura exponencial. Se existem grandes variabilidades foi sugerido realizar a superfície conjunta do perfil de verossimilhança entre os dois parâmetros de covariância (δ^2, f) (Diggle & Ribeiro Junior, 2000).

O objetivo maior da geoestatística na análise de dados com dependência espacial é a partir de um método de interpolação de variância mínima e não viesado fazer previsões de pontos não amostrados na região de estudo e com estruturas que suavize a variabilidade espacial (Figuras 31 a 34). Com os resultados apresentados nessas figuras serão discutidos as previsões e variâncias das previsões para a estrutura de covariância da família Matérn com os parâmetros $\hat{\epsilon} = 1$ e $\hat{\epsilon} = 2$.

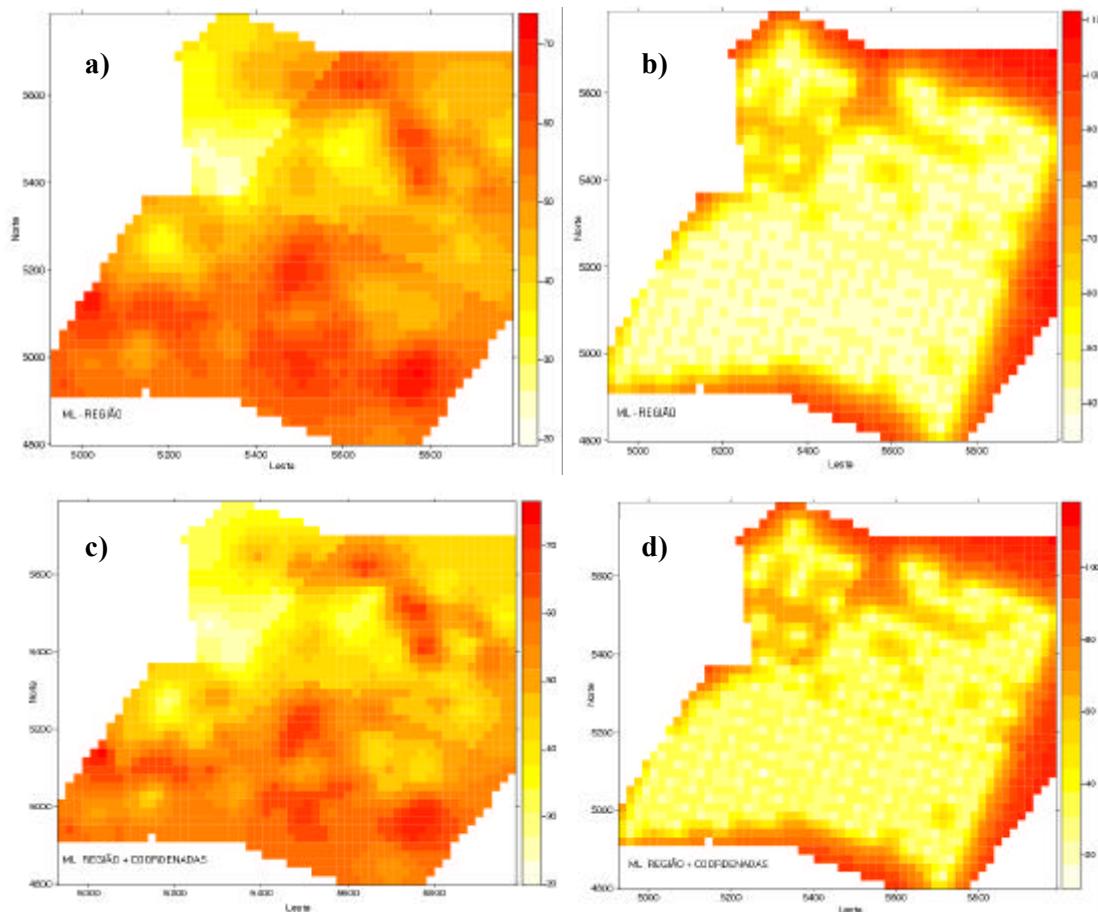


Figura 31 - Resultados da previsão (a e c) e variância da previsão (b e d), para o teor de cálcio (0-20cm). Modelo com a covariável região (a e b) e região mais tendência linear nas coordenadas (c e d). Parâmetros estimados por máxima verossimilhança para a estrutura de covariância Matérn para $\hat{\epsilon} = 1$.

Considerando-se a estrutura Matérn serão analisados os modelos das eqs. (43 e 44) com os parâmetros estimados pelos estimadores de máxima verossimilhança e máxima verossimilhança restrita com a variável teor de cálcio na camada 0-20cm conforme descrito em 2.7.2. Utilizando-se o estimador de máxima verossimilhança para estimar os parâmetros dos modelos apresentados nas equações 43 e 44, com $\hat{\epsilon} = 1$ detectou-se uma região maior da predição onde localizaram-se os maiores valores dos teores cálcio e as variâncias das previsões menores (Figura 31a e b). No caso do modelo 2, estas regiões, tanto para as previsões quanto para suas variâncias, apresentaram-se mais homogêneas (Figura 31c e d).

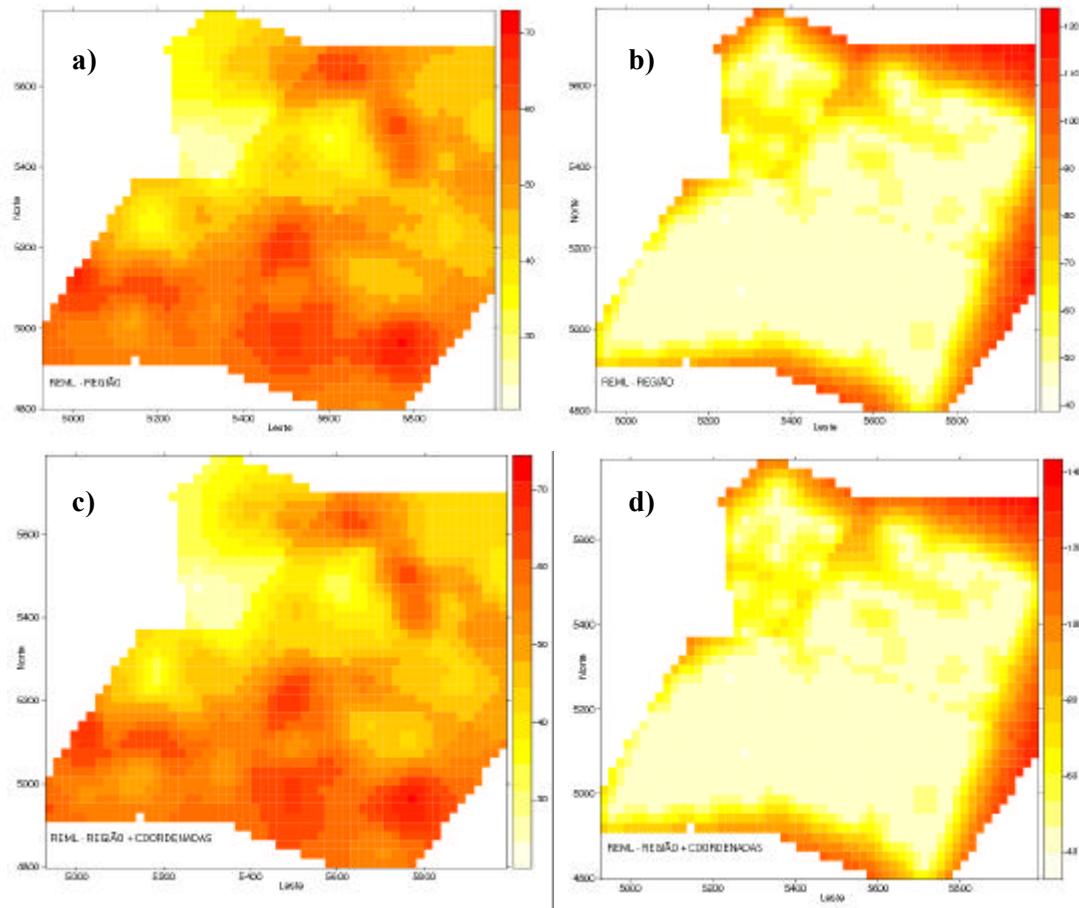


Figura 32 - Resultados da predição (a e c) e variância da predição (b e d), para o teor de cálcio (0-20cm). Modelo com a covariável região (a e b) e região mais tendência linear nas coordenadas (c e d). Parâmetros estimados por máxima verossimilhança restrita considerando-se a estrutura de covariância Matérn para $\hat{\epsilon} = 1$.

A predição com o estimador de máxima verossimilhança restrita para a estrutura de covariância Matérn com $\hat{e} = 1$ apresentou-se com maior suavização na região para a variável teor de cálcio (0-20cm) e, com menor efeito *nugget* comparando-se à estrutura de covariância exponencial principalmente, com o modelo 2 (Figura 32c).

As variâncias das predições foram mais abrangentes e um maior número de variâncias pequenas foram detectadas e, estas estão associadas às cores mais claras. Quanto mais clara é a cor da região menor é a variância e, em consequência, pode-se obter maior precisão das estimativas dos parâmetros (Figura 32a,b,c e d).

De maneira análoga, foram obtidas as predições e variâncias das predições, com os modelos apresentados nas eqs. (43 e 44) utilizando-se a estrutura de covariância de Matérn com $\hat{e} = 2$ e parâmetros estimados por máxima verossimilhança (Figura 33).

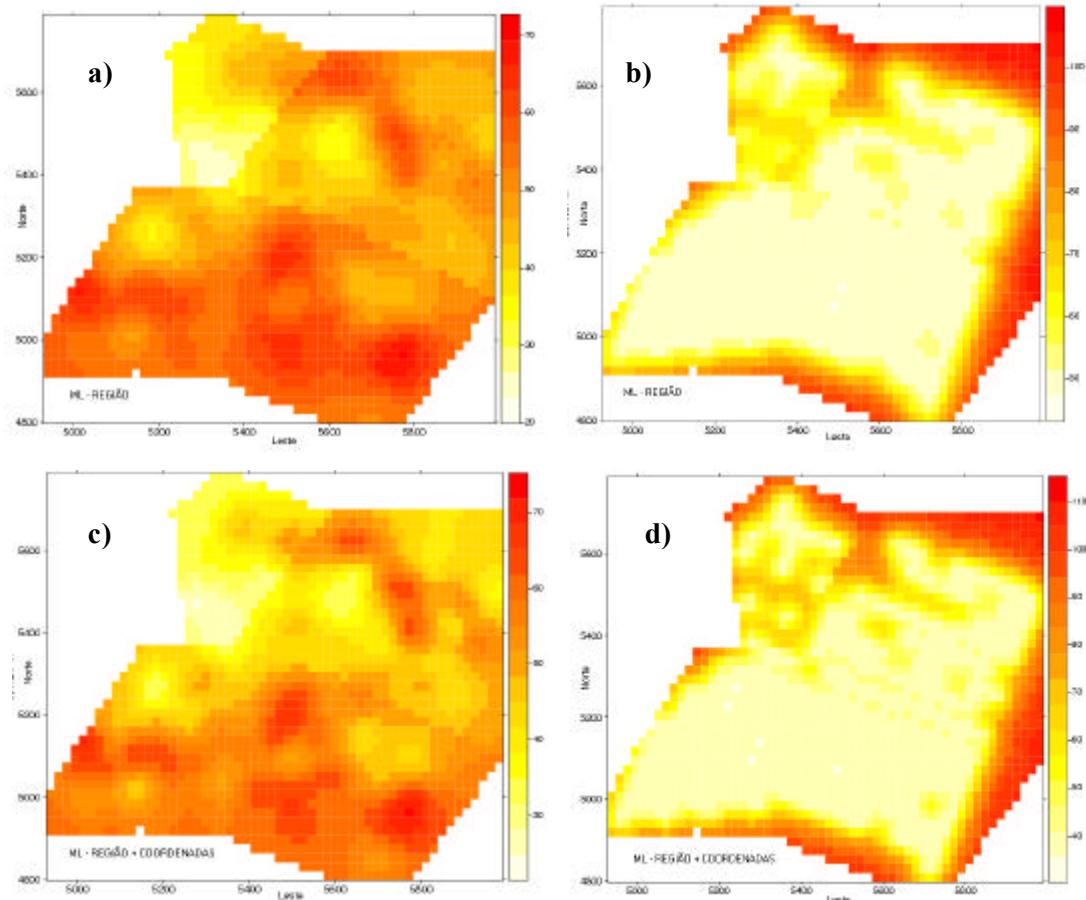


Figura 33 - Resultados da predição (a e c) e variância da predição (b e d), para o teor de cálcio (0-20cm). Modelo com a covariável região (a e b) e região mais a tendência linear nas coordenadas (c e d). Parâmetros estimados por máxima verossimilhança considerando-se a estrutura de covariância Matérn para $\hat{e} = 2$.

Considerando-se o modelo com menor número de parâmetros (modelo 1) e o estimador de máxima verossimilhança (Figura 33a), as regiões onde localizam-se os maiores valores para a variável teor de cálcio não estão bem definidas quando comparadas ao modelo 2 (Figura 33c). Da mesma forma, ocorreu com as variâncias das previsões (Figura 33b e 33d), nos dois modelos para $\hat{e} = 2$, mas se compararmos os mesmos resultados para $\hat{e} = 1$, o número de variâncias das previsões para este último parâmetro foi menor do que para $\hat{e} = 2$.

Os resultados obtidos da previsão e variância da previsão com a estrutura de covariância Matérn para $\hat{e} = 2$, para o estimador de máxima verossimilhança restrita, com os modelos com covariável região e região mais tendência linear nas coordenadas encontram-se na Figura 34.

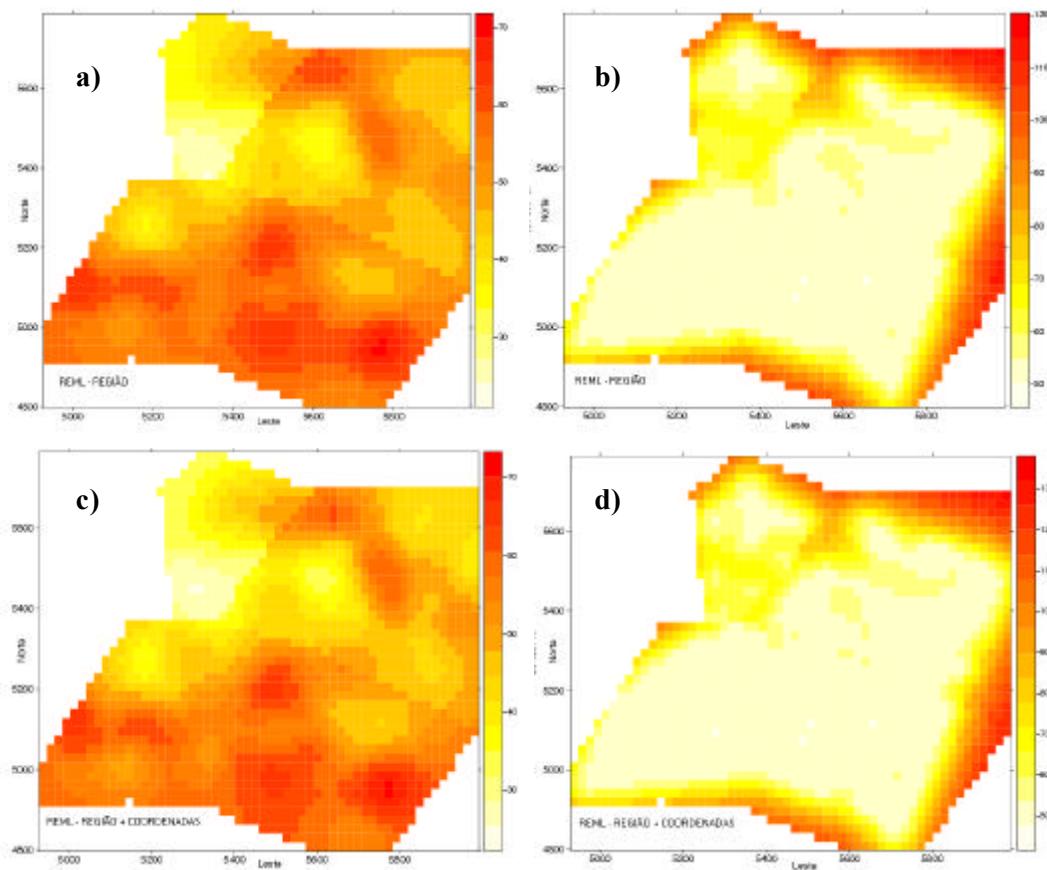


Figura 34 - Resultados da previsão (a e c) e variância da previsão (b e d), para o teor de cálcio (0-20cm). Modelo com a covariável região (a e b) e região mais tendência linear nas coordenadas (c e d). Parâmetros estimados por máxima verossimilhança restrita considerando-se a estrutura de covariância Matérn para $\hat{e} = 2$.

Observou-se que os resultados obtidos para as variâncias totais (*sill*) utilizando-se a estrutura de covariância Matérn foi maior do que a encontrada com a estrutura exponencial, mas a variância da predição para a primeira estrutura foi bem menor e mais abrangente para a região de estudo apresentada. 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 34b e d). Este fato ocorreu porque 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.

Em geral, observou-se que as variâncias estão muito próximas e que nos pontos mais claros estas variâncias são menores (Figura 34b e 34d). No caso em que utilizou-se o modelo 2 a área com cores mais claras foi maior do que aquela em que ajustou-se o modelo com a covariável região (modelo 1), sendo assim, obteve-se maior precisão quando os parâmetros foram estimados por máxima verossimilhança restrita e estrutura de covariância Matérn com $\hat{c} = 2$ (Figura 34d).

Da mesma forma que com a estrutura de covariância exponencial, a validação cruzada para a estrutura de covariância Matérn com o modelo com a covariável região mais a tendência linear nas coordenadas (modelo 2) indicou que os dados do teor de cálcio na camada superficial do solo, seguiu distribuição normal padronizada. Apresentou-se somente as figuras para os resultados do melhor modelo e estimador que permitiu obter o menor logarítmico da função de verossimilhança e menor valor do AIC dentro de cada estrutura de covariância. O modelo escolhido foi com a covariável região e tendência linear nas coordenadas, e estimador de máxima verossimilhança restrita (Figuras 35 e 36).

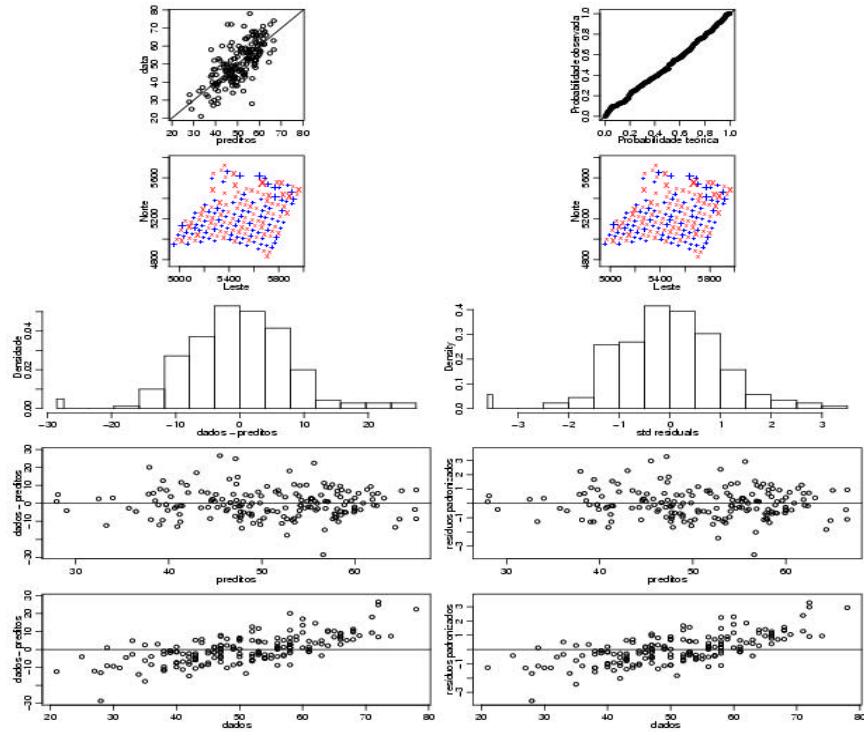


Figura 35 - Validação cruzada, para o teor de cálcio, na camada 0-20cm, utilizando-se o modelo com covariável região e tendência linear nas coordenadas e parâmetros estimados por máxima verossimilhança restrita para a estrutura de covariância Matérn para o parâmetro $\hat{e} = 1$.

Os valores preditos estão próximos da reta; a distribuição dos erros positivos e negativos está dispersa pela região de estudo e os dados das probabilidades teórica e observada encontram-se sobre a reta, indicando que o método utilizado para a predição, pela krigagem ordinária foi eficiente (item 2.6 e Figuras 35 e 36).

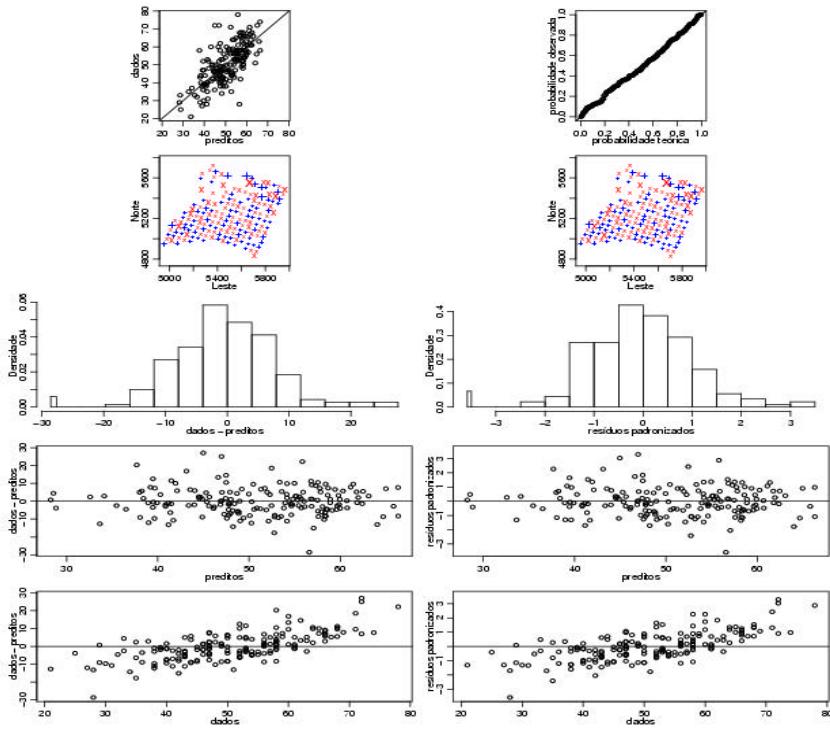


Figura 36 – Validação cruzada, para o teor de cálcio, na camada 0-20cm, utilizando-se o modelo com covariável região e tendência linear nas coordenadas e parâmetros ajustados pelo método de máxima verossimilhança restrita para a estrutura de covariância Matérn para o parâmetro $\hat{\epsilon} = 2$.

5 CONCLUSÕES

Face as metodologias empregadas neste trabalho e os resultados obtidos, pode-se concluir que:

- a) foi possível identificar a dependência espacial da variável cálcio nas diferentes camadas do solo com os diferentes métodos geoestatísticos utilizados;
- b) as análises exploratórias descritivas e espaciais foram importantes na identificação e escolha do modelo que melhor descreveu a distribuição espacial do cálcio no solo;
- c) quando os parâmetros foram estimados por máxima verossimilhança e máxima verossimilhança restrita, utilizando-se as estruturas de covariância exponencial e família Matérn com $\hat{\epsilon} = 1$ e $\hat{\epsilon} = 2$, na camada de 0-20cm do solo, os modelos selecionados foram respectivamente, incluindo-se a covariável região (modelo 1) e região mais a tendência linear nas coordenadas (modelo 2);
- d) quando os parâmetros foram estimados por máxima verossimilhança e máxima verossimilhança restrita, utilizando-se a estrutura de covariância exponencial na camada de 20-40cm do solo os modelos selecionados foram respectivamente, incluindo-se a covariável região e região mais tendência linear nas coordenadas;
- e) a estrutura de covariância Matérn foi mais precisa ao detectar as menores variâncias das previsões comparadas aquelas obtidas com a estrutura de covariância exponencial;

- f) o estimador de máxima verossimilhança não foi eficiente para detectar diferenças entre os modelos com a covariável região e o modelo com a covariável região mais a tendência linear nas coordenadas (modelo 2);
- g) com o estimador de máxima verossimilhança restrita obtiveram-se estimativas dos parâmetros confiáveis bem como, as menores variâncias de predições na região de estudo, menores valores para o critério de informação de Akaike e para o logarítmico da função de verossimilhança;
- h) a variabilidade espacial do cálcio foi maior na camada 20-40cm;
- i) o critério de informação de Akaike foi similar ao teste da razão da verossimilhança para selecionar os modelos estudados dentro de cada estimador utilizado, estrutura de covariância e camada do solo e,
- j) pelos métodos aplicados considerando a posição geográfica foi possível indicar com maior precisão os locais da distribuição do cálcio, dando subsídios para instalação de futuros experimentos com custos reduzidos.

ANEXOS

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\nu} = 1$ ou $\hat{\nu} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```

## Carregando a geoR
##
library(geoR)
par.ori <- par(no.readonly=TRUE)
##
## Lendo os dados
## Modelo Matern-2 n=178 observações Cálcio na camada 0-20cm Ca20
##
## Dados do teor de cálcio estão na coluna 5 e as covariáveis nas colunas 3 e 4
##
ca20 <- read.geodata("c:\\usr\\cristina\\programas\\hori2ec2.R", head=T, data.col=5,
covar.col=3:4)
borda1 <- read.table("sub1.R", head=T)[,2:3]
borda2 <- read.table("sub2.R", head=T)[,2:3]
borda3 <- read.table("sub3.R", head=T)[,2:3]
reg <- read.table("reg.R", head=F)
x1 <- ca20$coords[,1]
x2 <- ca20$coords[,2]
##
##Gráfico separando os dados de acordo com os quantis
##
postscript ("c:\\\\Emate2-ca20\\\\mat20-1.ps", horizontal = FALSE,
height=6, width=6, pointsize=10)
plot(ca20)
points(ca20, pt.div= "quintile", xlab = "leste", ylab ="norte")
cloud10 <- variog(ca20, option = "cloud", max.dist=700)
cloud20 <- variog(ca20, option = "cloud", estimator.type = "modulus", max.dist=700)
bin1 <- variog(ca20, uvec = seq(0,700, l=15))
bin20 <- variog(ca20, uvec = seq(0,700, l=15), estimator.type = "modulus")
##
## Realizando variogramas com estimador clássico e estimador de Hawkins e Cressie
##
par(mfrow=c(2,2))
plot(cloud10, main = "classical estimator")
plot(bin1, main = "classical estimator")

```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\epsilon} = 1$ ou $\hat{\epsilon} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```

plot(cloud20, main = "modulus estimator")
plot(bin20, main = "modulus estimator")
par(mfrow=c(1,2))
bin10 <- variog(ca20, uvec = seq(0,700, l=15), bin.cloud = T)
bin20 <- variog(ca20, uvec = seq(0,700, l=15), estimator.type = "modulus", bin.cloud
=T)
plot(bin10, bin.cloud = T, main ="classical estimator")
plot(bin20, bin.cloud = T, main ="modulus estimator")
bin10 <- variog(ca20, uvec =seq(0,700, l=15))
dev.off()
##
##Variograma empírico considerando no primeiro modelo a covariável região
##Variograma empírico com covariável região e tendência linear nas coordenadas
##
postscript("c:\\\\Emate2-ca20\\\\mat20-3.ps",horizontal=F, height=5, width=8, pointsize=8)
par(mfrow=c(1,3))
plot(variog(ca20, trend = ~as.factor(ca20$covar[,2]), max.dist=700))
plot(variog(ca20, trend = ~as.factor(ca20$covar[,2])+ca20$coords, max.dist=700))
dev.off()
##
##Modelo com a covariável região e com nugget estimado
##Ajustados por máxima verossimilhança e máxima verossimilhança restrita
##
ml1a <- likfit(ca20, trend = ~as.factor(ca20$covar[,2]), ini=c(100,200), nug=65, meth
="ML", kappa=2, cov.model="matern" )
reml1a <- likfit(ca20, trend =~as.factor(ca20$covar[,2]), ini=c(100,200), nug=65, meth
="REML", kappa=2, cov.model="matern")
##
##Modelo com a covariável região com tendência linear nas coordenadas e parâmetros
##ajustados por Máxima verossimilhança e máxima verossimilhança restrita
## considerando nugget estimado
##
ml4 <- likfit(ca20,trend=~as.factor(ca20$covar[,2])+ca20$coords, ini=c(100,200),
nug=65,meth= "ML", kappa=2, cov.model="matern")

```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\epsilon} = 1$ ou $\hat{\epsilon} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```

reml4<-likfit(ca20,trend=~as.factor(ca20$covar[,2])+ca20$coords, ini=c(100,200),
nug=65, meth = "RML", kappa=2, cov.model="matern")
##
##Plotando modelos ajustados contra variogramas empíricos com covariável região
##e região mais tendência linear com parâmetros obtidos por máxima
##verossimilhança e máxima verossimilhança restrita
##Estrutura de covariância Matérn com k=2
##
postscript("c:\\\\Emate2-ca20\\\\mat20-4.ps", horizontal=F, height=8, width=8,
pointsize=8)
par(mfrow =c(1,1))
plot(bin10, main = expression(paste("estimated", tau^2)))
lines(ml1a , col ="green", lwd=3, max.dist=700)
lines(ml4, lwd=3, max.dist=700)
lines(rem11a, col ="red", lwd=3, max.dist=700)
lines(rem14,col="blue", lwd=3, max.dist=700)
legend(80,50, bty="n", legend=c("ML- REGIÃO","ML- REGIÃO +
COORDENADAS","REML - REGIÃO","REML- REGIÃO +
COORDENADAS","MODELO MATERN KAPPA= 2","CÁLCIO 0-20cm"),
lty=c(1,1,1,1,-1,-1), lwd=c(2,2,2,2,2,2), col=c(3,1,2,4,8,8), cex=1.3)
dev.off()
##
##Gráficos para identificar dependência espacial por meio de envelopes simulados
## Envelopes simulados e parâmetros estimados por máxima verossimilhança
## e máxima verossimilhança restrita com os modelos com a covariável região
## e região mais tendência linear nas coordenadas
##Estrutura de covariância com o Modelo Matérn com k=2 para o teor de Cálcio 0-
##20cm
##
postscript("c:\\\\Emate2-ca20\\\\mat20-5.ps", horizontal =F, onefile =T, height=7, width=7,
pointsize=8)
par(mfrow=c(2,2))
env.mc <- variog.mc.env(ca20, obj.var=bin10, nsim=178, save.sim =TRUE)
env.model <- variog.model.env(ca20,obj.var=bin10, nsim=178, save.sim =TRUE,
model=ml1a)
plot(bin10, envelope=env.mc)

```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\epsilon} = 1$ ou $\hat{\epsilon} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```

legend(50, 60, bty="n", legend= c("SIMULAÇÃO", "MODELO MATÉRN KAPPA = 2", "CÁLCIO 0-20 cm"), lty=c(1,1,1), lwd=c(2,2,2), col=c(-1,-1,-1), cex=1.3)
plot(bin10, envelope=env.model)
lines.variomodel(list(nugget=34.78 , cov.pars=c(64.26,37.57), max.dist=700, kappa=2,
cov.model="matern"), lwd=2)
legend(50, 60, bty="n", legend= c("ML - REGIÃO", "MODELO MATÉRN KAPPA = 2", "CÁLCIO 0-20 cm"), lty=c(1,1,1), lwd=c(1,1,1), col=c(-1,-1,-1), cex=1.3)
env.mc <- variog.mc.env(ca20, obj.var=bin10, nsim=178, save.sim =TRUE)
env.model <- variog.model.env(ca20,obj.var=bin10, nsim=178, save.sim =TRUE,
model=ml4)
plot(bin10, envelope=env.mc)
legend(50, 60, bty="n", legend= c("SIMULAÇÃO", "MODELO MATÉRN KAPPA = 2", "CÁLCIO 0-20 cm"), lty=c(1,1,1), lwd=c(2,2,2), col=c(-1,-1,-1), cex=1.3)
plot(bin10, envelope=env.model)
lines.variomodel(list(nugget=22.72, cov.pars=c(69.73,28.11), max.dist=700, kappa=2,
cov.model="matern"), lwd=2)
legend(50, 60, bty="n", legend= c("ML - REGIÃO + COORDENADAS", "MODELO MATÉRN KAPPA = 2", "CÁLCIO 0-20 cm"), lty=c(2,2,2), lwd=c(1,1,1), col=c(-1,-1,-1), cex=1.3)
env.mc <- variog.mc.env(ca20, obj.var=bin10, nsim=178, save.sim =TRUE)
dev.off()
##
postscript("c:\\\\Emate2-ca20\\\\mat20-6.ps", horizontal =F, onefile =T, height=7, width=7,
pointsize=8)
par(mfrow=c(2,2))
env.mc <- variog.mc.env(ca20, obj.var=bin10, nsim=178, save.sim =TRUE)
env.model <- variog.model.env(ca20,obj.var=bin10, nsim=178, save.sim =TRUE,
model=reml1a)
plot(bin10, envelope=env.mc)
legend(50, 60, bty="n", legend= c("SIMULAÇÃO", "MODELO MATÉRN KAPPA = 2", "CÁLCIO 0-20 cm"), lty=c(1,1,1), lwd=c(2,2,2), col=c(-1,-1,-1), cex=1.3)
plot(bin10, envelope=env.model)
lines.variomodel(list(nugget=38.92, cov.pars=c(70.06,45.36), max.dist=700, kappa=2,
cov.model="matern"), lwd=2)

```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\nu} = 1$ ou $\hat{\nu} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```

legend(50, 60, bty="n", legend= c("REML - REGIÃO", "MODELO MATÉRN KAPPA = 2", "CÁLCIO 0-20 cm"), lty=c(1,1,1), lwd=c(2,2,2), col=c(-1,-1,-1), cex=1.3)
env.model <- variog.model.env(ca20,obj.var=bin10, nsim=178, save.sim =TRUE,
model=reml4)
plot(bin10, envelope=env.mc)
legend(60, 40, bty="n", legend= c("SIMULAÇÃO", "MODELO MATÉRN KAPPA = 2", "CÁLCIO 0-20 cm"), lty=c(1,1,1), lwd=c(2,2,2), col=c(-1,-1,-1), cex=1.3)
plot(bin10, envelope=env.model)
lines.variomodel(list(nugget= 35.49, cov.pars=c(71.61,40.97), max.dist=700, kappa=2,
cov.model="matern"), lwd=2)
legend(30, 60, bty="n", legend= c("REML - REGIÃO + COORDENADAS",
"MODELO MATÉRN KAPPA = 2", "CÁLCIO 0-20 cm"), lty=c(1,1,1), lwd=c(1,1,1),
col=c(-1,-1,-1), cex=1.3)
dev.off()
##
##Perfil de verossimilhança para avaliar variabilidade dos parâmetros
##Modelo com covariável região e parâmetros ajustados por máxima verossimilhança
##
postscript("c:\\\\Emate2-ca20\\\\mat20-7.ps", horizontal =F, onefile =T, height=5, width=8,
pointsize=12)
prof2 <- proflik(ml1a, geodata=ca20, sill.val=seq(10,140, l=16), range.val=seq(25,140,
l=16),
nugget.val=seq(0,50, l=16), uni.only =T)
par(mfrow=c(1,3))
plot(prof2, nlevels=16)
dev.off()
##
##Perfil de verossimilhança para avaliar variabilidade dos parâmetros
##Modelo com covariável região e parâmetros ajustados por máxima verossimilhança
##Restrita
##
postscript("c:\\\\Emate2-ca20\\\\mat20-8.ps", horizontal=F, onefile=T, height=5, width=8,
pointsize=12)

```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\epsilon} = 1$ ou $\hat{\epsilon} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```

prof4 <- proflik(rem11a, geodata=ca20, sill.val=seq(81,81.3, l=4), range.val=seq(80,82,
l=4),
nugget.val=seq(38,40, l=4), uni.only =T)
par(mfrow=c(1,3))
plot(prof4, nlevels=16)
dev.off()
##
##Perfil de verossimilhança para avaliar variabilidade dos parâmetros
##Modelo com covariável região e tendência linear nas coordenadas
## e parâmetros ajustados por máxima verossimilhança
##
postscript("c:\\\\Emate2-ca20\\\\mat20-9.ps", horizontal=F, onefile=T, height=5, width=8,
pointsize=12)
prof5 <- proflik(ml4, geodata=ca20, sill.val=seq(40,140, l=16), range.val=seq(10,100,
l=16),
nugget.val=seq(0,50, l=16), uni.only=T)
par(mfrow=c(1,3))
plot(prof5, nlevels=16)
dev.off()
##
##Perfil de verossimilhança para avaliar variabilidade dos parâmetros
##Modelo com covariável região e tendência linear nas coordenadas
##e parâmetros ajustados por máxima verossimilhança Restrita
##
postscript("c:\\\\Emate2-ca20\\\\mat20-10.ps", horizontal=F, onefile=T, height=5, width=8,
pointsize=12)
prof6 <- proflik(rem14, geodata=ca20, sill.val=seq(10,100, l=16), range.val=seq(55,70,
l=16),
nugget.val=seq(24,28, l=16), uni.only =T)
par(mfrow=c(1,3))
plot(prof6, nlevels=16)
dev.off()
##
##Validação cruzada para o modelo com covariável região e parâmetros ajustados
##por máxima verossimilhança e com estrutura de covariância Matérn
##

```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\epsilon} = 1$ ou $\hat{\epsilon} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```

postscript("c:\\\\Emate2-ca20\\\\mat20-11.ps",horizontal=F, onefile=T, height=8, width=8,
pointsize=10)
xv.ml1a <- xvalid(ca20, model=ml1a)
par(mfcol = c(5,2), mar=c(3,3,.5,.5), mgp=c(1.5,.7,0))
plot(xv.ml1a, ca20)
par(par.ori)
dev.off()
##
##Validação cruzada para o modelo com covariável região e tendência linear nas
##coordenadas com e parâmetros ajustados por máxima verossimilhança e com
##estrutura de covariância Matérn para k=2
##
postscript("c:\\\\Emate2-ca20\\\\mat20-12.ps",horizontal=F, onefile=T, height=8, width=8,
pointsize=10)
xv.ml4 <- xvalid(ca20, model=ml4)
par(mfcol = c(5,2), mar=c(3,3,.5,.5), mgp=c(1.5,.7,0))
plot(xv.ml4, ca20)
par(par.ori)
dev.off()
##
##Validação cruzada para o modelo com covariável região e tendência linear nas
##coordenadas com e parâmetros ajustados por máxima verossimilhança restrita
##e com estrutura de covariância Matérn para k=2
##
postscript("c:\\\\Emate2-ca20\\\\mat20-13.ps",horizontal=F, onefile=T, height=8,
width=8,pointsize=10)
xv.reml4 <- xvalid(ca20, model=reml4)
par(mfcol = c(5,2), mar=c(3,3,.5,.5), mgp=c(1.5,.7,0))
plot(xv.reml4, ca20)
par(par.ori)
dev.off()
##
##Cálculo dos valores do Critério de Informação de Akaike e Critério Bayesiano
##Parâmetros: variância do ruído (nugget), variância espacial (sill) e alcance (range)
##ML1a – Modelo com região e parâmetros estimados por máxima verossimilhança
##ML4 - Modelo com região e tendência nas coordenadas e parâmetros estimados

```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\nu} = 1$ ou $\hat{\nu} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```

##      por máxima verossimilhança
##
ml.res1 <- data.frame(npars = c(ml1a$npars, ml4$npars), logL= c(ml1a$loglik,
ml4$loglik), AIC = c(ml1a$AIC, ml4$AIC), BIC =c(ml1a$BIC, ml4$BIC), nugget=c(
ml1a$nugget,ml4$nugget), sill=c(ml1a$cov.pars[1], ml4$cov.pars[1]),
range=c(ml1a$cov.pars[2], ml4$cov.pars[2]))
rownames(ml.res1) <- paste("modelo", 1:2)
ml.res1
##
##Cálculo dos valores do Critério de Informação de Akaike e Critério Bayesiano
##Parâmetros: variância do ruído (nugget), variância espacial (sill) e alcance (range)
##REML1a – Modelo com região e parâmetros estimados por máxima
##verossimilhança Restrita
##REML4 - Modelo com região e tendência nas coordenadas e parâmetros
##estimados por máxima verossimilhança Restrita
##
reml.res1 <- data.frame(npars = c(reml1a$npars, reml4$npars), logL= c( reml1a$loglik,
reml4$loglik), AIC=c(reml1a$AIC, reml4$AIC), BIC=c(reml1a$BIC, reml4$BIC),
nugget=c(reml1a$nugget, reml4$nugget), sill=c(reml1a$cov.pars[1],reml4$cov.pars[1]),
range=c(reml1a$cov.pars[2], reml4$cov.pars[2]))
rownames(reml.res1) <- paste("modelo", 1:2)
reml.res1
##
##Avaliação dos variogramas de acordo com a Anisotropia em diferentes direções
##e numa mesma direção avaliando a isotropia com estrutura de covariância Matérn e
##k=2
##
postscript ("c:\\\\Emate2-ca20\\\\mat20-14.ps", horizontal=F, height=5, width=5,
pointsize=10)
ca20.v2 <- variog4(ca20, max.dist=700, trend =~as.factor(ca20$covar[,2]))
plot(ca20.v2, omn=T, legend=F)
legend(20,50, legend= c("única direção", "0", "45", "90", "135"), lty=c(1,2,3,4,5),
lwd=c(1,1,1,1,1), col=c(1,2,3,4,5))
plot(ca20.v2, omn=T, same=F)
dev.off()

```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\epsilon} = 1$ ou $\hat{\epsilon} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```
##
##Avaliação dos variogramas de acordo com a Anisotropia em diferentes direções
##e numa mesma direção avaliando a isotropia com estrutura de covariância Matérn e
k=2
##considerando o modelo com a covariável região
##
postscript ("c:\\Emate2-ca20\\mat20-14a.ps", horizontal=F, height=5, width=5,
pointsize=10)
ca20.v2 <- variog4(ca20, max.dist=700, trend =~as.factor(ca20$covar[,2]))
plot(ca20.v2, omn=T)
plot(ca20.v2, omn=T, same=F)
dev.off()
##
## Opções de gráficos avaliando a anisotropia com a covariável região e tendência
## linear nas coordenadas para o teor de cálcio na camada 0-20 cm nas 4 direções e
## numa única direção
#
postscript ("c:\\Emate2-ca20\\mat20-15.ps", horizontal=F, height=5, width=5,
pointsize=10)
ca20.v3 <- variog4(ca20, max.dist=700, trend
=~as.factor(ca20$covar[,2])+ca20$coords)
plot(ca20.v3, omn=T, legend=F)
legend(20,50, legend= c("única direção", "0", "45", "90", "135"), lty=c(1,2,3,4,5),
lwd=c(1,1,1,1,1), col=c(1,2,3,4,5))
plot(ca20.v3, omn=T, same=F)
dev.off()
##
postscript ("c:\\Emate2-ca20\\mat20-15a.ps", horizontal=F, height=5, width=5,
pointsize=10)
ca20.v3 <- variog4(ca20, max.dist=700, trend
=~as.factor(ca20$covar[,2])+ca20$coords)
plot(ca20.v3, omn=T)
plot(ca20.v3, omn=T, same=F)
dev.off()
##
## Lendo os dados
```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\nu} = 1$ ou $\hat{\nu} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```

## Modelo Matérn para k=2 n=178 observações Profundidade 0-20cm
##Definição das coordenadas das regiões de acordo com a adubação da área
##experimental
##
ca20 <- read.geodata("hori2ec2.R", head=T, data.col=5, covar.col=3:4)
names(ca20)
##
sub1 <- matrix(c( 5590,5340,5220,5250,5240,5350,5450,5510,5590,
5690,5800,5700,5370,5370,5500,5600,5690), nc=2)
sub1
sub2 <- matrix(c(5990,5590,5350,5450,5510,5590,5800,5990,
5100,5300,5370,5500,5600,5690,5690,5690), nc=2)
sub2
sub3 <- matrix(c(5990,5590,5350,5150,4920,4920,5150,5350,5590,5780,
5100,5300,5370,5370,5000,4900,4920,4900,4800,4800),nc=2)
sub3
##
## Mapa da krigagem utilizando o modelo ML1a com covariável Região
##Estrutura de covariância Matérn com estimador de Máxima verossimilhança
##com kappa=2
##
postscript("c:\\\\Emate2-ca20\\\\mat20-16.ps",horizontal=FALSE, onefile=TRUE,
height=8, width=8, pointsize=10)
points(ca20)
polygon(sub1)
polygon(sub2)
polygon(sub3)
dev.off()
##
## Predição espacial dos dados de Cálcio na profundidade 0-20cm Com estrutura de
##covariância Matérn com kappa=2
##Modelo com a covariável região e tendência linear nas coordenadas
## Definindo pontos da região 1 a serem preditos- Região menos utilizada para a
##pesquisa
## Função de krigagem para a região 1 e com estimativas dos parâmetros pelo ##método
## de Máxima Verossimilhança no modelo com a covariável região ML1a

```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\nu} = 1$ ou $\hat{\nu} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```
##
postscript("c:\\\\Emate2-ca20\\\\mat20-17.ps",horizontal=F, onefile=T, height=6, width=6,
pointsize=10)
points(ca20)
polygon(sub1)
polygon(sub2)
polygon(sub3)
##
plot(ca20)
points(ca20)
polygon(borda1)
polygon(borda2)
polygon(borda3)
polygon(reg, lwd=2)
dev.off()
##
##Fazendo predição somente na região 1 - Estrutura de covariância Exponencial
## Modelo com a covariável Região e tendência linear nas coordenadas
## Estimador de Máxima Verossimilhança ML1a
##
postscript("c:\\\\Emate2-ca20\\\\mat20-18.ps",horizontal=F, onefile=T, height=8, width=8,
pointsize=10)
loci0 <- expand.grid(seq(4500,6000,l=80), seq(4500,5800, l=60))
points(ca20)
polygon(borda1)
polygon(borda2)
polygon(borda3)
dev.off()
##
##Predizendo somente na Região 1 nos locais onde desejo fazer predição
##
postscript("c:\\\\Emate2-ca20\\\\mat20-19.ps",horizontal=F, onefile=T, height=8, width=8,
pointsize=10)
points(ca20)
polygon(borda1)
polygon(borda2)
```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\nu} = 1$ ou $\hat{\nu} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```

polygon(borda3)
loci1 <- polygrid(seq(4500,6000,l=80), seq(4500,5800, l=60), borda1)
points(loci1, pch="+", col="red")
dev.off()
##
## definindo o trend.l (região + coordenadas dos pontos a serem preditos)
##
postscript("c:\\\\Emate2-ca20\\\\mat20-20.ps",horizontal=F, onefile=T, height=8, width=8,
pointsize=10)
t.11 <- list(covariates = factor(rep(1, nrow(loci1)), levels=1:3), coords=as.matrix(loci1))
kc41 <- krige.control(obj.model=ml1a,
                      trend.d =~as.factor(ca20$covar[,2]),
                      trend.l =~as.factor(t.11$covar))
kr41 <- krige.conv(ca20, loc=loci1, krige=kc41)

image(kr41, border=borda1, loc=loci0, col=gray(seq(1,0, l=21)))
dev.off()
##
## Fazendo predição em todas as Regiões de uma vez
##
postscript("c:\\\\Emate2-ca20\\\\mat20-21.ps",horizontal=F, onefile=T, height=8, width=8,
pointsize=10)
points(ca20)
polygon(borda1)
polygon(borda2)
polygon(borda3)
##
loci1 <- polygrid(seq(4500,6000,l=80), seq(4500,5800, l=60), borda1)
points(loci1, pch="+", col="red")
loci2 <- polygrid(seq(4500,6000,l=80), seq(4500,5800, l=60), borda2)
points(loci2, pch="+", col="blue")
loci3 <- polygrid(seq(4500,6000,l=80), seq(4500,5800, l=60), borda3)
points(loci3, pch="+", col="green")
##
loci.reg <- as.matrix(rbind(loci1,loci2,loci3))

```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\nu} = 1$ ou $\hat{\nu} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```
##
## definindo o trend.l (região + coordenadas dos pontos a serem preditos)
t.l.reg<- list(covariates = factor(c(rep(1, nrow(loci1)),rep(2, nrow(loci2)),rep(3,
nrow(loci3)))), coords=loci.reg)
##
kc42 <- krige.control(obj.model=ml1a,
                      trend.d =~as.factor(ca20$covar[,2]),
                      trend.l =~as.factor(t.l.reg$covar))
kr42 <- krige.conv(ca20, loc=t.l.reg$coords, krige=kc42)
dev.off()
##
postscript("c:\\\\Emate2-ca20\\\\mat20-22.ps",horizontal=F, onefile=T, height=8, width=8,
pointsize=10)
require(lattice)
lset()
levelplot(kr42$pred ~ loci.reg[,1] * loci.reg[,2], asp="xy",
col.reg=heat.colors(30)[30:1])
dev.off()
##
postscript("c:\\\\Emate2-ca20\\\\mat20-22A.ps",horizontal=F, onefile=T, height=8,
width=8, pointsize=10)
require(lattice)
lset()
levelplot(kr42$pred ~ loci.reg[,1] * loci.reg[,2], asp="xy", col.reg=gray(seq(1,0, l=20)))
dev.off()
##
postscript("c:\\\\Emate2-ca20\\\\mat20-22B.ps",horizontal=F, onefile=T, height=8,
width=8, pointsize=10)
require(lattice)
lset()
levelplot(kr42$krige.var ~ loci.reg[,1] * loci.reg[,2], asp="xy",
col.reg=heat.colors(30)[30:1])
dev.off()
##
```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\nu} = 1$ ou $\hat{\nu} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```

postscript("c:\\\\Emate2-ca20\\\\mat20-22C.ps",horizontal=F, onefile=T, height=8,
width=8, pointsize=10)
require(lattice)
lset()
levelplot(kr42$krige.var ~ loci.reg[,1] * loci.reg[,2], asp="xy", col.reg=gray(seq(1,0,
l=30)))
dev.off()
##
postscript("c:\\\\Emate2-ca20\\\\mat20-23.ps",horizontal=FALSE, onefile=TRUE,
height=6, width=6, pointsize=10)
points(ca20)
polygon(sub1)
polygon(sub2)
polygon(sub3)
##
plot(ca20)
points(ca20)
polygon(borda1)
polygon(borda2)
polygon(borda3)
polygon(reg, lwd=2)
dev.off()
##
##Fazendo predição somente na região 1 - Estrutura de covariância Matérn kappa=2
## Modelo com a covariável Região e tendência linear nas coordenadas
## Estimador de Máxima Verossimilhança ML4
##
postscript("c:\\\\Emate2-ca20\\\\mat20-24.ps",horizontal=F, onefile=T, height=8, width=8,
pointsize=10)
loci0 <- expand.grid(seq(4500,6000,l=80), seq(4500,5800, l=60))
points(ca20)
polygon(borda1)
polygon(borda2)
polygon(borda3)
dev.off()
##

```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\nu} = 1$ ou $\hat{\nu} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```

##Predizendo somente na Região 1 nos locais onde desejo fazer predição
##
postscript("c:\\\\Emate2-ca20\\\\mat20-25.ps",horizontal=F, onefile=T, height=8, width=8,
pointsize=10)
points(ca20)
polygon(borda1)
polygon(borda2)
polygon(borda3)
loci1 <- polygrid(seq(4500,6000,l=80), seq(4500,5800, l=60), borda1)
points(loci1, pch="+", col="red")
dev.off()
##
## definindo o trend.l (região + coordenadas dos pontos a serem preditos)
##
postscript("c:\\\\Emate2-ca20\\\\mat20-26.ps",horizontal=F, onefile=T, height=8, width=8,
pointsize=10)
t.l1 <- list(covariates = factor(rep(1, nrow(loci1)), levels=1:3), coords=as.matrix(loci1))
kc43 <- krige.control(obj.model=ml4,
                      trend.d =~as.factor(ca20$covar[,2])+ca20$coords,
                      trend.l =~as.factor(t.l1$covar)+t.l1$coords)
kr43 <- krige.conv(ca20, loc=loci1, krige=kc43)

image(kr43, border=borda1, loc=loci0, col=gray(seq(1,0, l=21)))
dev.off()
##
## Fazendo predição em todas as Regiões de uma vez
##
postscript("c:\\\\Emate2-ca20\\\\mat20-27.ps",horizontal=F, onefile=T, height=8, width=8,
pointsize=10)
points(ca20)
polygon(borda1)
polygon(borda2)
polygon(borda3)
##
loci1 <- polygrid(seq(4500,6000,l=80), seq(4500,5800, l=60), borda1)

```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\epsilon} = 1$ ou $\hat{\epsilon} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```

points(loci1, pch="+", col="red")
loci2 <- polygrid(seq(4500,6000,l=80), seq(4500,5800, l=60), borda2)
points(loci2, pch="+", col="blue")
loci3 <- polygrid(seq(4500,6000,l=80), seq(4500,5800, l=60), borda3)
points(loci3, pch="+", col="green")
##
loci.reg <- as.matrix(rbind(loci1,loci2,loci3))
##
## definindo o trend.l (região + coordenadas dos pontos a serem preditos)
t.l.reg <- list(covariates = factor(c(rep(1, nrow(loci1)),rep(2, nrow(loci2)),rep(3,
nrow(loci3)))), coords=loci.reg)
##
kc44 <- krige.control(obj.model=ml4,
                      trend.d =~as.factor(ca20$covar[,2])+ca20$coords,
                      trend.l =~as.factor(t.l.reg$covar)+t.l.reg$coords)
kr44 <- krige.conv(ca20, loc=t.l.reg$coords, krige=kc44)
dev.off()
##
postscript("c:\\\\Emate2-ca20\\\\mat20-28.ps",horizontal=F, onefile=T, height=8, width=8,
pointsize=10)
require(lattice)
lset()
levelplot(kr44$pred ~ loci.reg[,1] * loci.reg[,2], asp="xy",
col.reg=heat.colors(20)[20:1])
dev.off()
##
postscript("c:\\\\Emate2-ca20\\\\mat20-28A.ps",horizontal=F, onefile=T, height=8,
width=8, pointsize=10)
require(lattice)
lset()
levelplot(kr44$pred ~ loci.reg[,1] * loci.reg[,2], asp="xy", col.reg=gray(seq(1,0, l=20)))
dev.off()
##
postscript("c:\\\\Emate2-ca20\\\\mat20-28B.ps",horizontal=F, onefile=T, height=8,
width=8, pointsize=10)

```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\nu} = 1$ ou $\hat{\nu} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```

require(lattice)
lset()
levelplot(kr44$krige.var ~ loci.reg[,1] * loci.reg[,2], asp="xy",
           col.reg=heat.colors(20)[20:1])
dev.off()
##
postscript("c:\\\\Emate2-ca20\\\\mat20-28C.ps",horizontal=F, onefile=T, height=8,
           width=8, pointsize=10)
require(lattice)
lset()
levelplot(kr44$krige.var ~ loci.reg[,1] * loci.reg[,2], asp="xy", col.reg=gray(seq(1,0,
           l=30)))
dev.off()
##
## Mapa da krigagem utilizando o modelo ML1a com a covariável Região
##Estrutura de covariância Esférica com estimador ML
##
postscript("c:\\\\Emate2-ca20\\\\mat20-29.ps",horizontal=F, onefile=T, height=8, width=8,
           pointsize=10)
t.l1 <- list(covariates = factor(rep(1, nrow(loci1))), levels=1:3), coords=as.matrix(loci1))
kc45 <- krige.control(obj.model=reml1a,
                       trend.d =~as.factor(ca20$covar[,2]),
                       trend.l =~as.factor(t.l1$covar))
kr45 <- krige.conv(ca20, loc=loci1, krige=kc45)
image(kr45, border=borda1, loc=loci0, col=gray(seq(1,0, l=21)))
dev.off()
##
## Fazendo predição em todas as Regiões de uma vez
##
postscript("c:\\\\Emate2-ca20\\\\mat20-30.ps",horizontal=F, onefile=T, height=8, width=8,
           pointsize=10)
points(ca20)
polygon(borda1)
polygon(borda2)
polygon(borda3)

```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\epsilon} = 1$ ou $\hat{\epsilon} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```
##
loci1 <- polygrid(seq(4500,6000,l=80), seq(4500,5800, l=60), borda1)
points(loci1, pch="+", col="red")
loci2 <- polygrid(seq(4500,6000,l=80), seq(4500,5800, l=60), borda2)
points(loci2, pch="+", col="blue")
loci3 <- polygrid(seq(4500,6000,l=80), seq(4500,5800, l=60), borda3)
points(loci3, pch="+", col="green")
##
loci.reg <- as.matrix(rbind(loci1,loci2,loci3))
##
## definindo o trend.l (região + coordenadas dos pontos a serem preditos)
t.l.reg<- list(covariates = factor(c(rep(1, nrow(loci1)),rep(2, nrow(loci2)),rep(3,
nrow(loci3)))), coords=loci.reg)
##
kc46 <- krige.control(obj.model=rem11a,
                      trend.d =~as.factor(ca20$covar[,2]),
                      trend.l =~as.factor(t.l.reg$covar))
kr46 <- krige.conv(ca20, loc=t.l.reg$coords, krige=kc46)
dev.off()
##
postscript("c:\\\\Emate2-ca20\\\\mat20-31.ps",horizontal=F, onefile=T, height=8, width=8,
pointsize=10)
require(lattice)
lset()
levelplot(kr46$pred ~ loci.reg[,1] * loci.reg[,2], asp="xy",
col.reg=heat.colors(20)[20:1])
dev.off()
##
postscript("c:\\\\Emate2-ca20\\\\mat20-31A.ps",horizontal=F, onefile=T, height=8,
width=8, pointsize=10)
require(lattice)
lset()
levelplot(kr46$pred ~ loci.reg[,1] * loci.reg[,2], asp="xy", col.reg=gray(seq(1,0, l=20)))
dev.off()
##
```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\nu} = 1$ ou $\hat{\nu} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```

postscript("c:\\\\Emate2-ca20\\\\mat20-31B.ps",horizontal=F, onefile=T, height=8,
width=8, pointsize=10)
require(lattice)
lset()
levelplot(kr46$krige.var ~ loci.reg[,1] * loci.reg[,2], asp="xy",
col.reg=heat.colors(30)[30:1])
dev.off()
##
postscript("c:\\\\Emate2-ca20\\\\mat20-31C.ps",horizontal=F, onefile=T, height=8,
width=8, pointsize=10)
require(lattice)
lset()
levelplot(kr46$krige.var ~ loci.reg[,1] * loci.reg[,2], asp="xy", col.reg=gray(seq(1,0,
l=30)))
dev.off()
##
postscript("c:\\\\Emate2-ca20\\\\mat20-32.ps",horizontal=FALSE, onefile=TRUE,
height=6, width=6, pointsize=10)
points(ca20)
polygon(sub1)
polygon(sub2)
polygon(sub3)
##
plot(ca20)
points(ca20)
polygon(borda1)
polygon(borda2)
polygon(borda3)
polygon(reg, lwd=2)
dev.off()
##
##Fazendo predição somente na região 1 - Estrutura de covariância Esférica
## Modelo com a covariável Região e tendência linear nas coordenadas REML4
## Estimador de Máxima Verossimilhança REML
##
```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\nu} = 1$ ou $\hat{\nu} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```

postscript("c:\\\\Emate2-ca20\\\\mat20-33.ps",horizontal=F, onefile=T, height=8, width=8,
pointsize=10)
loci0 <- expand.grid(seq(4500,6000,l=80), seq(4500,5800, l=60))
points(ca20)
polygon(borda1)
polygon(borda2)
polygon(borda3)
dev.off()
##
##Predizendo somente na Região 1 nos locais onde desejo fazer predição
##
postscript("c:\\\\Emate2-ca20\\\\mat20-34.ps",horizontal=F, onefile=T, height=8, width=8,
pointsize=10)
points(ca20)
polygon(borda1)
polygon(borda2)
polygon(borda3)
loci1 <- polygrid(seq(4500,6000,l=80), seq(4500,5800, l=60), borda1)
points(loci1, pch="+", col="red")
dev.off()
##
## definindo o trend.l (região + coordenadas dos pontos a serem preditos)
##
postscript("c:\\\\Emate2-ca20\\\\mat20-35.ps",horizontal=F, onefile=T, height=8, width=8,
pointsize=10)
t.l1 <- list(covariates = factor(rep(1, nrow(loci1)), levels=1:3), coords=as.matrix(loci1))
kc47 <- krige.control(obj.model=reml4,
trend.d =~as.factor(ca20$covar[,2])+ca20$coords,
trend.l =~as.factor(t.l1$covar)+t.l1$coords)
kr47 <- krige.conv(ca20, loc=loci1, krige=kc47)
image(kr47, border=borda1, loc=loci0, col=gray(seq(1,0, l=21)))
dev.off()
##
## Fazendo predição em todas as Regiões de uma vez
##

```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\epsilon} = 1$ ou $\hat{\epsilon} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```

postscript("c:\\Emate2-ca20\\mat20-36.ps",horizontal=F, onefile=T, height=8, width=8,
pointsize=10)
points(ca20)
polygon(borda1)
polygon(borda2)
polygon(borda3)
##
loci1 <- polygrid(seq(4500,6000,l=80), seq(4500,5800, l=60), borda1)
points(loci1, pch="+", col="red")
loci2 <- polygrid(seq(4500,6000,l=80), seq(4500,5800, l=60), borda2)
points(loci2, pch="+", col="blue")
loci3 <- polygrid(seq(4500,6000,l=80), seq(4500,5800, l=60), borda3)
points(loci3, pch="+", col="green")
##
loci.reg <- as.matrix(rbind(loci1,loci2,loci3))
##
## definindo o trend.1 (região + coordenadas dos pontos a serem preditos)
t.l.reg <- list(covariates = factor(c(rep(1, nrow(loci1)),rep(2, nrow(loci2)),rep(3,
nrow(loci3)))), coords=loci.reg)
##
kc48 <- krige.control(obj.model=reml4,
trend.d =~as.factor(ca20$covar[,2])+ca20$coords,
trend.l =~as.factor(t.l.reg$covar)+t.l.reg$coords)
kr48 <- krige.conv(ca20, loc=t.l.reg$coords, krige=kc48)
dev.off()
##
postscript("c:\\Emate2-ca20\\mat20-37.ps",horizontal=F, onefile=T, height=8, width=8,
pointsize=10)
require(lattice)
lset()
levelplot(kr48$pred ~ loci.reg[,1] * loci.reg[,2], asp="xy",
col.reg=heat.colors(20)[20:1])
dev.off()
##
postscript("c:\\Emate2-ca20\\mat20-37A.ps",horizontal=F, onefile=T, height=8,
width=8, pointsize=10)

```

Anexo A - Subrotina geoR para obtenção da análise geoestatística e gráfica para os dados da variável teor de cálcio nas camadas de 0-20cm, utilizando-se a estrutura de covariância Matérn com $\hat{\nu} = 1$ ou $\hat{\nu} = 2$, permitindo adaptação para outras estruturas de covariâncias.

```
require(lattice)
lset()
levelplot(kr48$pred ~ loci.reg[,1] * loci.reg[,2], asp="xy", col.reg=gray(seq(1,0, l=20)))
dev.off()
##
postscript("c:\\\\Emate2-ca20\\\\mat20-37B.ps",horizontal=F, onefile=T, height=8,
width=8, pointsize=10)
require(lattice)
lset()
levelplot(kr48$krige.var ~ loci.reg[,1] * loci.reg[,2], asp="xy",
col.reg=heat.colors(30)[30:1])
dev.off()
##
postscript("c:\\\\Emate2-ca20\\\\mat20-37C.ps",horizontal=F, onefile=T, height=8,
width=8, pointsize=10)
require(lattice)
lset()
levelplot(kr48$krige.var ~ loci.reg[,1] * loci.reg[,2], asp="xy", col.reg=gray(seq(1,0,
l=30)))
dev.off()
```

Anexo B. Estimativas dos parâmetros obtidas por máxima verossimilhança com o modelo com a covariável região para a variável teor de cálcio, na camada 0-20cm, do solo, com as estruturas de covariâncias exponencial ($\hat{e} = 0,5$) e Matérn para $\hat{e} = 1$ e $\hat{e} = 2$.

\hat{e}	\hat{a}_0	\hat{a}_1	\hat{a}_2	\hat{o}^2	\hat{f}	\hat{o}^2	$\log \hat{L}$
0,5	39,71	8,03	13,81	98,70	72,61	3,26	-629,326
1	39,24	8,41	14,32	76,00	53,07	24,45	-629,600
2	38,79	8,78	14,83	64,25	37,56	34,77	-629,957

Anexo C. Estimativas dos parâmetros obtidas por máxima verossimilhança com o modelo com a covariável região mais tendência linear nas coordenadas para a variável teor de cálcio na camada 0-20cm, do solo, com as estruturas de covariâncias exponencial ($\hat{e} = 0,5$) e Matérn para $\hat{e} = 1$ e $\hat{e} = 2$.

\hat{e}	\hat{a}_0	\hat{a}_1	\hat{a}_2	\hat{a}_3	\hat{a}_4	\hat{o}^2	\hat{f}	\hat{o}^2	$\log \hat{L}$
0,5	121,90	6,02	9,20	0,0024	-0,017	94,77	63,02	0,00	-627,779
1	125,24	6,16	9,27	0,0026	-0,018	83,85	38,91	9,41	-627,874
2	124,99	6,30	9,46	0,0026	-0,0185	69,72	28,10	22,71	-628,170

Anexo D. Estimativas dos parâmetros obtidas por máxima verossimilhança restrita com o modelo com a covariável região para a variável teor de cálcio, na camada 0-20cm do solo, com as estruturas de covariâncias exponencial ($\hat{e} = 0,5$) e Matérn para $\hat{e} = 1$ e $\hat{e} = 2$.

\hat{e}	\hat{a}_0	\hat{a}_1	\hat{a}_2	\hat{o}^2	\hat{f}	\hat{o}^2	$\log \hat{L}$
0,5	41,19	7,05	12,08	104,52	104,19	13,07	-617,460
1	40,38	7,70	12,94	82,07	68,45	30,62	-617,871
2	39,65	8,30	13,74	70,06	45,36	38,91	-618,347

Anexo E. Estimativas dos parâmetros obtidas por máxima verossimilhança restrita com o modelo com a covariável região mais tendência linear nas coordenadas para a variável teor de cálcio, na camada de 0-20cm do solo, com as estruturas de covariâncias exponencial ($\hat{e} = 0,5$) e Matérn para $\hat{e} = 1$ e $\hat{e} = 2$.

\hat{e}	\hat{a}_0	\hat{a}_1	\hat{a}_2	\hat{a}_3	\hat{a}_4	$\hat{\sigma}^2$	\hat{f}	$\hat{\sigma}^2$	$\log \hat{L}$
0,5	125,08	5,34	8,40	0,0015	-0,0171	108,70	92,57	7,35	-608,530
1	124,70	5,75	8,78	0,0018	-0,0174	84,55	61,01	26,16	-608,920
2	123,75	6,14	9,17	0,0019	-0,0174	71,61	40,97	35,48	-609,38

Anexo F. Estimativas dos parâmetros obtidas por máxima verossimilhança com o modelo com a covariável região e região mais tendência linear nas coordenadas, para variável teor de cálcio, na camada 20-40cm do solo, com a estrutura de covariância exponencial.

Modelos	\hat{a}_0	\hat{a}_1	\hat{a}_2	\hat{a}_3	\hat{a}_4	$\hat{\sigma}^2$	\hat{f}	$\hat{\sigma}^2$	$\log \hat{L}$
Região	35,69	2,94	13,86	-	-	65,67	101,54	80,16	-682,66
Tendência	129,28	3,44	10,86	-0,0062	-0,0111	60,80	87,25	79,68	-681,86

Anexo G. Estimativas dos parâmetros obtidas por máxima verossimilhança restrita com o modelo com a covariável região e região mais tendência linear nas coordenadas, para a variável teor de cálcio, na camada 20-40cm do solo, com a estrutura de covariância exponencial.

Modelos	\hat{a}_0	\hat{a}_1	\hat{a}_2	\hat{a}_3	\hat{a}_4	$\hat{\sigma}^2$	\hat{f}	$\hat{\sigma}^2$	$\log \hat{L}$
Região	36,88	2,33	12,60	-	-	79,87	152,39	83,97	-670,34
Tendência	121,08	2,22	10,35	-0,0072	-0,0082	95,32	200,00	85,61	-661,65

Anexo H. Matriz de variâncias e covariâncias para a estrutura de covariância exponencial, na camada do solo de 0-20cm, incluindo no modelo a covariável região e parâmetros estimados por máxima verossimilhança.

$$M_1 = \begin{bmatrix} 15,68671 & -13,86193 & -13,80741 \\ -13,86193 & 20,91124 & 14,34958 \\ -13,80741 & 14,34958 & 17,56388 \end{bmatrix}$$

Anexo I. Matriz de variâncias e covariâncias para a estrutura de covariância exponencial, na camada do solo de 0-20cm, incluindo no modelo a covariável região e tendência linear nas coordenadas e, com os parâmetros estimados por máxima verossimilhança.

$$M_2 = \begin{bmatrix} 3876,25192 & -5,95159448 & -140,115238600 & -0,2396322000 & -0,47143920000 \\ -5,95159 & 23,27782392 & 17,228329420 & -0,0135737800 & 0,01221444000 \\ -140,11523 & 17,22832942 & 24,587040000 & -0,0054613010 & 0,02881134000 \\ -0,23963 & -0,01357378 & -0,005461301 & 0,0000527500 & -0,00000816420 \\ -0,47143 & 0,01221444 & 0,028811340 & -0,0000081642 & 0,00009448496 \end{bmatrix}$$

Anexo J. Matriz de variâncias e covariâncias para a estrutura de covariância exponencial, na camada do solo de 0-20cm, incluindo no modelo a covariável região com os parâmetros estimados por máxima verossimilhança restrita.

$$M_3 = \begin{bmatrix} 21,10844 & -16,56750 & -16,65468 \\ -16,56750 & 25,15409 & 17,48448 \\ -16,65468 & 17,48448 & 21,48378 \end{bmatrix}$$

Anexo K. Matriz de variâncias e Covariâncias para a estrutura de covariância exponencial, na camada do solo de 0-20cm, incluindo no modelo a covariável região e tendência linear nas coordenadas com os parâmetros estimados por máxima verossimilhança restrita.

$$M_4 = \begin{bmatrix} 5553,99036\,6500 & -7,30368528 & -143,640500\,000 & -0,38436740\,000 & -0,63887310\,000 \\ -7,30368530\,0 & 28,9433736\,3 & 21,2752900\,00 & -0,01671839\,000 & 0,01507007\,000 \\ -143,640542\,800 & 21,2752949\,4 & 28,3674400\,00 & -0,00774022\,800 & 0,03129963\,000 \\ -0,38436740\,0 & -0,01671839 & -0,00774022\,8 & 0,00008661\,161 & -0,00001554\,208 \\ -0,63887310\,0 & 0,01507007 & 0,03129963\,0 & -0,00001554\,208 & 0,00013366\,690 \end{bmatrix}$$

Anexo L. Matriz de variâncias e covariâncias para a estrutura de covariância exponencial, na camada do solo de 20-40cm, incluindo no modelo a covariável região e parâmetros estimados por máxima verossimilhança.

$$M_5 = \begin{bmatrix} 21,25512 & 18,77943 & 19,19586 \\ -18,77943 & 27,75912 & 19,42258 \\ -19,19586 & 19,42258 & 24,29149 \end{bmatrix}$$

Anexo M. Matriz de variâncias e covariâncias para a estrutura de covariância exponencial, na camada do solo de 20-40cm, incluindo no modelo a covariável região e tendência linear nas coordenadas e, com os parâmetros estimados por máxima verossimilhança.

$$M_6 = \begin{bmatrix} 5119,8164621 & -19,74551261 & -223,73060000 & -0,2948289000\,00 & -0,6405692000\,00 \\ -19,7455124 & 31,54741786 & 24,35452000 & -0,0171402200\,00 & 0,0173182100\,00 \\ -223,7305975 & 24,35451673 & 36,94434000 & -0,0058216490\,00 & 0,0433465400\,00 \\ -0,2948289 & -0,01714022 & -0,00582164 & 0,0000062588\,00 & -0,0000076992\,95 \\ -0,6405692 & 0,01731821 & 0,04334654 & -0,0000076992\,95 & 0,0001243476\,00 \end{bmatrix}$$

Anexo N. Matriz de variâncias e covariâncias para a estrutura de covariância exponencial, na camada do solo de 20-40cm, incluindo no modelo a covariável região com os parâmetros estimados por máxima verossimilhança restrita.

$$M_7 = \begin{bmatrix} 29,34711 & -22,48269 & -23,41364 \\ -22,48269 & 33,82193 & 23,82692 \\ -23,41364 & 23,82692 & 30,40982 \end{bmatrix}$$

Anexo O. Matriz de variâncias e covariâncias para a estrutura de covariância exponencial, na camada do solo de 20-40cm, incluindo no modelo a covariável região e tendência linear nas coordenadas e, com os parâmetros estimados por máxima verossimilhança restrita.

$$M_8 = \begin{bmatrix} 9360,8018391 & -26,35235622 & -223,40006304 & -0,6758964000 & -1,0488600000 \\ -26,3523562 & 42,73603924 & 32,33557049 & -0,0231656100 & 0,0236188100 \\ -223,4000631 & 32,33557049 & 44,49198619 & -0,0105491300 & 0,0472203900 \\ -0,6758964 & -0,02316561 & -0,01054913 & 0,0001492548 & -0,0000247466 \\ -1,0488602 & 0,02361881 & 0,04722039 & -0,0000247466 & 0,0002193039 \end{bmatrix}$$

Anexo P. Matriz de variâncias e covariâncias para a estrutura de covariância Matérn com $\hat{\epsilon} = 1$, na camada do solo de 0-20cm, incluindo no modelo a covariável região e parâmetros estimados por máxima verossimilhança.

$$M_9 = \begin{bmatrix} 14,91580 & -13,50201 & -13,41038 \\ -13,50201 & 20,26679 & 13,92627 \\ -13,41038 & 13,92627 & 16,97253 \end{bmatrix}$$

Anexo Q. Matriz de variâncias e covariâncias para a estrutura de covariância Matérn com $\hat{\epsilon} = 1$, na camada do solo de 0-20cm, incluindo no modelo a covariável região e tendência linear nas coordenadas e, com os parâmetros estimados por máxima verossimilhança.

$$M_{10} = \begin{bmatrix} 3513,7522182 & -5,18432581 & -141,5388000 & -0,2055771000 & -0,4379339000 \\ -5,1843258 & 21,60065109 & 16,01910000 & -0,0127106100 & 0,0113696700 \\ -141,5387514 & 16,01910432 & 23,56268000 & -0,0047525930 & 0,0284954500 \\ -0,2055771 & -0,01271061 & -0,004752593 & 0,0000443745 & -0,0000060255 \\ -0,4379339 & 0,01136967 & 0,028495450 & -0,0000060255 & 0,0000859400 \end{bmatrix}$$

Anexo R. Matriz de variâncias e covariâncias para a estrutura de covariância Matérn com $\hat{\epsilon} = 1$, na camada do solo de 0-20cm, incluindo no modelo a covariável região com os parâmetros estimados por máxima verossimilhança restrita.

$$M_{11} = \begin{bmatrix} 18,94682 & -15,93323 & -15,92793 \\ -15,93323 & 24,01090 & 16,67466 \\ -15,92793 & 16,67466 & 20,36884 \end{bmatrix}$$

Anexo S. Matriz de variâncias e Covariâncias para a estrutura de covariância Matérn com $\hat{\epsilon} = 1$, na camada do solo de 0-20cm, incluindo no modelo a covariável região e tendência linear nas coordenadas com os parâmetros estimados por máxima verossimilhança restrita.

$$M_{12} = \begin{bmatrix} 4944,3715620 & -7,258946120 & -146,763400000 & -0,3277024000 & -0,5817530000 \\ -7,2589461 & 27,843840290 & 20,578790000 & -0,0161759100 & 0,0146025500 \\ -146,7634213 & 20,578789630 & 27,900970000 & -0,0072742850 & 0,0314657000 \\ -0,3277024 & -0,016175910 & -0,007274285 & 0,0000737911 & -0,0000130327 \\ -0,5817530 & 0,014602550 & 0,031465700 & -0,0000130327 & 0,0001201594 \end{bmatrix}$$

Anexo T. Matriz de variâncias e covariâncias para a estrutura de covariância Matérn com $\hat{\epsilon} = 2$, na camada do solo de 0-20cm, incluindo no modelo a covariável região e parâmetros estimados por máxima verossimilhança.

$$M_{13} = \begin{bmatrix} 14,26188 & -13,14296 & -13,02950 \\ -13,14296 & 19,62758 & 13,51784 \\ -13,02950 & 13,51784 & 16,41571 \end{bmatrix}$$

Anexo U. Matriz de variâncias e covariâncias para a estrutura de covariância Matérn com $\hat{\epsilon} = 2$, na camada do solo de 0-20cm, incluindo no modelo a covariável região e tendência linear nas coordenadas e, com os parâmetros estimados por máxima verossimilhança.

$$M_{14} = \begin{bmatrix} 3399,78802350 & -5,65297223 & -142,167988340 & -0,1948284000 & -0,4273810000 \\ -5,65297223 & 20,98472079 & 15,636340000 & -0,0123327200 & 0,0111326900 \\ -142,16788340 & 15,63633834 & 23,259630000 & -0,0044934920 & 0,0283902500 \\ -0,1948284 & -0,012332720 & -0,004493492 & 0,0000417264 & -0,0000053547 \\ -0,4273810 & 0,011132690 & 0,028390250 & -0,0000053547 & 0,0000832502 \end{bmatrix}$$

Anexo V. Matriz de variâncias e covariâncias para a estrutura de covariância Matérn com $\hat{\epsilon} = 2$, na camada do solo de 0-20cm, incluindo no modelo a covariável região com os parâmetros estimados por máxima verossimilhança restrita.

$$M_{15} = \begin{bmatrix} 17,42000 & -15,28525 & -15,22310 \\ -15,28525 & 22,88782 & 15,89884 \\ -15,22310 & 15,89884 & 19,33474 \end{bmatrix}$$

Anexo X. Matriz de variâncias e covariâncias para a estrutura de covariância Matérn com $\hat{\epsilon} = 2$, na camada do solo de 0-20cm, incluindo no modelo a covariável região e tendência linear nas coordenadas com os parâmetros estimados por máxima verossimilhança restrita.

$$M_{16} = \begin{bmatrix} 4558,4128915 & -7,68976552 & -148,916800000 & -0,29183730000 & -0,5455308000 \\ -7,6897655 & 26,84716003 & 19,965260000 & -0,01562414000 & 0,0142098600 \\ -148,9168142 & 19,96525790 & 27,475740000 & -0,00684933900 & 0,0314901500 \\ -0,2918373 & -0,01562414 & -0,006849339 & 0,00006548314 & -0,0000112646 \\ -0,5455308 & 0,01420986 & 0,031490150 & -0,00001126460 & 0,0001114234 \end{bmatrix}$$

REFERÊNCIAS BIBLIOGRÁFICAS

- ATKINSON, P.M.; LLOYD, C.D., Mapping precipitation in Switzerland with ordinary and indicator kriging. **Journal of Geographic Information and Decision Analysis**, v.2, n.2, p.72-86, 1998.
- BRAGA, L.P.V. Geoestatística e aplicações. In: SIMPÓSIO BRASILEIRO DE PROBABILIDADE E ESTATÍSTICA, 9., São Paulo: IME-USP, 1990. 36p..
- CÂMARA, G.; MONTEIRO, A .M.; FUKS, S.D.; FELGUEIRAS, C. **Análise espacial de dados geográficos.** 2. Ed. S. José dos Campos: INPE, 2001.
- CAPECHE, C.L.; MACEDO, J.R.; MANZATTO, H.R.H.; SILVA, E.F. Caracterização pedológica da fazenda Angra - PESAGRO/RIO - Estação experimental de Campos (RJ). (compact disc). In: CONGRESSO BRASILEIRO DE CIÊNCIA DO SOLO. 26., Informação, globalização, uso do solo; Rio de Janeiro, 1997. **trabalhos.** Rio de Janeiro: Embrapa/SBCS, 1997.
- COUTO, E.G.; KLAMT, E. Variabilidade espacial de micronutrientes em solo sob pivô central no sul do Estado de Mato Grosso. **Pesquisa agropecuária brasileira**, v.34, n.12, p.2321-2329, 1999.
- CRESSIE, N. A. C. **Statistics for spatial data.** New York: John Wiley, 1993. 900p.

- CRESSIE, N.; LAHIRI, S.N. Asymptotics for REML estimation of spatial covariance parameters. **Journal of Statistical Planning and Inference**, v.50, p.327-341. 1996.
- DAVID, MICHEL. Geostatistical ore reserve estimation. New York, Elsevier Scientific, 1977. 363p.
- DIGGLE, P.J.; TAWN, J. A. Model-based geostatistics. **Applied Statistical** v.47, pt 3, p.299-350, 1998.
- DIGGLE, P.J.; RIBEIRO JÚNIOR, P.J. Bayesian inference in Gaussian model-based geostatistics. **Geographical and Environmental Modelling**, v.6, n. 2, p.131-148, 2002.
- DIGGLE, P.J.; RIBEIRO JÚNIOR, P.J. **Model based geostatistics**. Caxambu: ABE, 2000. 129p.
- EMBRAPA. Centro Nacional de Pesquisa de Solos. (Rio de Janeiro, RJ). **Manual de métodos de análise de solo**. 2. ed. Rio de Janeiro, 1997. 212p. (EMBRAPA-CNPS. Documentos;1).
- ENGLUND, E.; SPARKS, A. **Geo-EAS (Geostatistical Environmental Assessment Software)**. Las Vegas: U.S. Environmental Protection Agency, 1988. 1v. (EPA/600/4-88/033a).
- FARIAS, P.R.S. Distribuição espacial do nematóide *reniforme* com o uso da geoestatística, em um solo com rotação de culturas. Jaboticabal, 1999. 109p. Tese (Doutorado)- Universidade Estadual Paulista Faculdade de Ciências Agronômicas, Universidade Estadual Paulista “Júlio de Mesquita Filho”.
- FELGUEIRAS, C.A. Modelagem ambiental com tratamento de incertezas em sistemas de informação geográfica: O paradigma geoestatístico por indicação. São José dos Campos, 1999. 181p. Tese (Doutorado)-Ministério da Ciência e Tecnologia, Instituto Nacional de Pesquisas Espaciais - INPE.
- 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. Piracicaba, 1997. 118p. Tese-(Doutorado)-Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.
- GRONDONA, M. O.; CRESSIE, N. Using spatial considerations in the analysis of experiments. **Technometrics**, v. 33, n.4, p.381-391, 1991.
- GUERRA, P.A.G. **Geoestatística operacional**. Brasília: Ministério das Minas e Energia, 1988. 145p.
- HAMLETT, J.M.; HORTON, R.; CRESSIE, N. A. C. Resistant and exploratory techniques for use in semivariogram analysis. **Soil Science Society of American Journal**, v. 50, p.868-875, 1986.
- HARTLEY, H.O.; RAO, J.N.K. Maximum-likelihood estimation for the mixed analysis of variance model. **Biometrika**, v.54, p.93-108, 1967.
- HARVILLE, D. A. Bayesian inference for variance components using only error contrasts. **Biometrika**, v. 61, p.383-385, 1974.
- HOAGLIN, D.C.; MOSTELLER, F.; TUKEY, J.W. **Análise exploratória de dados: técnicas robustas**. Lisboa: Ed. Salamandra, 1992. 446p.
- ISAAKS, E.H.; SRIVASTAVA, R.M. **An introduction to applied geostatistics**. New York: Oxford University, 1989. 561p.
- JOURNEL, A.G.; HUIJBREGTS, C.J. **Mining geostatistics**, Centre de Geoestatistique, Fontainebleau, France, New York: Centre de Geoestatistique, Academic Press, 1978. 600p.
- KITANIDIS, P.K. **Introduction to geostatistics applications to hydrology**. Cambridge: University Press, 1997. 249p.

- KITANIDIS, P.K.; SHEN, K.F. Geostatistical interpolation of chemical concentration. **Advances in Water Resources**, v.19, n.6, p.369-78, 1996.
- LANDIM, P.M.B. **Introdução à geoestatística**. Rio Claro: UNESP, 1998, 226p.
- LIBARDI, P.L.; MANFRON, P.A.; MORAES, S.O.; TUON, R.L. Variabilidade da umidade gravimétrica de um solo hidromórfico. **Revista Brasileira de Ciência do Solo**, v.20, n.2, p.1-12, 1996.
- MARTINS, E.N.; LOPES, P.S.; SILVA, M. A.; REGAZZI, A. J. Modelo linear misto. Viçosa: Universidade Federal de Viçosa, Departamento de Zootecnia, 1993. 46p.
- MATHERON, G. Principles of Geostatistics. **Economic Geology**, v.58, p.1246-1266. 1963.
- MATHERON, G. The theory of regionalized variables and its applications. **Les cahiers du CMM**, n.5, 211p, 1971.
- McBRATNEY, A. B.; WEBSTER, R. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. **Journal of Soil Science**, v.37, p.617-639. 1986.
- MENK, J.R.F. Programa de microcomputador (PGeoEAS) para converter arquivo de dados ao formato do pacote geoestatístico Geo-EAS. **Sociedade Brasileira de Ciência do Solo**, v.15, n. 3 p.102-119, 1990.
- OLIVEIRA, M.C.N.; CORRÊA-FERREIRA, B.S. PROTEÇÃO DE PLANTAS. Diagnóstico exploratório na análise de dados entomológicos: efeito do diflubenzuron no consumo alimentar de *Anticarsia gemmatalis* (Hübner). **Anais da Sociedade Entomológica do Brasil**, v.27, n.4, p.627-637, 1998.
- PANNATIER, Y. **VARIOWIN**: software for spatial data analysis in 2D. New York: Springer-Verlag, 1996, 90p.

PARDO-IGÚZQUIZA, E. MLREML: A computer program for the inference of spatial covariance parameters by maximum likelihood and restricted maximum likelihood. **Computers & Geosciences**. v.23, n.2. p.153-162, 1997.

PARDO-IGÚZQUIZA, E. Inference of spatial indicator covariance parameters by maximum likelihood using MLREML. **Computers & Geosciences**. v.24, n.5. p.453-464, 1998a.

PARDO-IGÚZQUIZA, E. MLREML4: A program for the inference of the power variogram model by maximum likelihood and restricted maximum likelihood. **Computers & Geosciences**. v.24, n.6. p.537-543, 1998b.

PARK, J.S.; BAEK, J. Efficient computation of maximum likelihood estimators in a spatial linear model with power exponential covariogram. **Computers & Geosciences**. v.27 , p.1-7, 2001.

PATTERSON, H.D., THOMPSON,R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, 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. **Journal of Statistical Planning and Inference**, v.91, p.499-509, 2000.

REICHARDT, K; VIEIRA, S.R.; LIBARDI, P.L. Variabilidade espacial de solos e experimentação de campo. **Revista Brasileira de Ciência do Solo**, v.10, n.1, p1-6, 1986.

RIBEIRO JUNIOR, P.J. *Krigagem: A estimação geoestatística*. Piracicaba: Esalq, 1992. 47p.

RIBEIRO JUNIOR, P.J. Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo. Piracicaba, 1995. 99p. Dissertação (Mestrado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.

RIBEIRO JUNIOR, P.J.; DIGGLE, P.J. Bayesian inference in Gaussian model-based geostatistics. Lancaster: Dep. of Mathematics and Statistics, University Lancaster, UK, 1999. 45p. (Technical report, ST-99-08)

RIBEIRO JUNIOR, P.J.; DIGGLE, P.J. **geoR**: A package for geostatistical analysis. Lancaster : R News, 2001. 4p.

ROSSI, R.E. Geostatistical tools for modeling and interpreting ecological spatial dependence. **Ecological Monographs**, v.62, n.2, p.277-314, 1992

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.

SOARES, A. **Geoestatística para ciência da terra e do ambiente**. Lisboa: Instituto Superior Técnico, 2000. 206p.

TRANGMAR, B. B.; YOST, R. S.; UEHARA, G. Application of geostatistics to spatial studies of soil properties. **Advanced in Agronomy**, v.38, p.45-94, 1985.

VALENTE, J.M.G.P. Geomatématica. Lições de Geoestatística. Ouro Preto: Fundação Gorceix, 1989. v.3, p.511-713.

VIEIRA, S. R. **Geoestatística**. Curso de extensão universitária. Botucatu: UNESP, 1996. 71p.

VIEIRA, S.R.; NIELSEN, D.R.; BIGGAR, J.W. Spatial variability of field-measured infiltration rate. **Soil Science Society of American Journal** v.45, n.3, p.1040-1048, 1981.

VIEIRA, S.R.; HATFIELD, J.L.; NIELSEN, D.R.; BIGGAR, J.W. Geostatistical theory and application to variability of some agronomical properties. **Hilgardia**, v.51, n.3, p.1-75, 1983.

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

YATES, F. The comparative advantages of systematic and randomized arrangements in the design of agricultural and biological experiments. **Biometrika**, v. 30, p.444-466, 1938.

ZAMBOTI, J.L. Mapas pluviométricos médios mensais no verão para o estado do Paraná, utilizando métodos geoestatísticos. Botucatu, 2001. 117p. Tese (Doutorado) – Universidade Estadual Paulista Faculdade de Ciências Agronômicas, Universidade Estadual Paulista “Júlio de Mesquita Filho”.

APÊNDICE

APÊNDICE 1. Planta topográfica planialtimétrica na escala 1:1000, divida em três regiões, correspondentes aos períodos de adubação do solo.

