

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
MESTRADO EM BIOMETRIA**

ANTÔNIO LOPES PESSOA

**UMA ABORDAGEM BAYESIANA PARA ESTUDO ESTATÍSTICO E
GEOESTATÍSTICO DE ESTIMATIVAS DE SALINIDADE DO SOLO
UTILIZANDO SENSOR DE INDUÇÃO ELETROMAGNÉTICA**

Recife / 2006

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
MESTRADO EM BIOMETRIA**

**UMA ABORDAGEM BAYESIANA PARA ESTUDO ESTATÍSTICO E
GEOESTATÍSTICO DE ESTIMATIVAS DE SALINIDADE DO SOLO
UTILIZANDO SENSOR DE INDUÇÃO ELETROMAGNÉTICA**

Dissertação apresentada ao Programa de Pós-Graduação em Biometria da Universidade Federal Rural de Pernambuco como requisito para obtenção do grau de mestre em Biometria.

Orientador: Prof. Abelardo Antônio de Assunção Montenegro, PhD
Co-orientador: Prof. Dr. Eufrázio de Souza Santos
Mestrando: Antônio Lopes Pessoa

Recife / 2006

Ficha catalográfica
Setor de Processos Técnicos da Biblioteca Central – UFRPE

P475a Pessoa, Antônio Lopes
Uma abordagem bayesiana para estudo estatístico e
geoestatístico de estimativas de salinidade do solo
utilizando sensor de indução eletromagnética / Antônio Lopes
Pessoa. -- 2006.
98 f. : il.

Orientador: Abelardo Antônio de Assunção Montenegro.
Dissertação (Mestrado em Biometria) – Universidade
Federal Rural de Pernambuco. Departamento de Estatística e
Informática.
Inclui bibliografia e anexo.

CDD 550.018 2

1. Krigagem ordinária
 2. Krigagem bayesiana
 3. Variabilidade espacial
 4. Condutividade elétrica do solo
 5. Geoestatística
 6. EM 38
 7. Indução eletromagnética
- I. Montenegro, Abelardo Antônio de Assunção
 - II. Título

ANTÔNIO LOPES PESSOA**UMA ABORDAGEM BAYESIANA PARA ESTUDO ESTATÍSTICO E
GEOESTATÍSTICO DE ESTIMATIVAS DE SALINIDADE DO SOLO
UTILIZANDO SENSOR DE INDUÇÃO ELETROMAGNÉTICA**

Dissertação julgada adequada para obtenção do título de mestre em Biometria,
defendida e aprovada por unanimidade em 24/02/2006 pela Banca Examinadora:

Orientador:

Prof. Abelardo Antônio de Assunção Montenegro, PhD.

Co-orientador:

Prof. Dr. Eufrázio de Souza Santos.

Examinadores:

Prof. José Antônio Aleixo da Silva, PhD.

Prof. Dr. Marcus Metri Corrêa.

A sabedoria não nos é dada; é preciso descobri-la por nós mesmos depois de uma viagem que ninguém nos pode poupar ou fazer por nós.

Marcel Proust

A Deus,

que me abençoou com a vida e iluminou-me na trilha do conhecimento.

Aos meus pais

Antônio Virgínio Pessoa (in memorian)

pelo incentivo e apoio em todas as minhas realizações e me transmitir perseverança e coragem de lutar.

Luiza Lopes Pessoa (in memorian)

divino exemplo de mãe, por sua dignidade, retidão de caráter, dedicação constante e incansável, o amor e carinho.

Ao meu tio Arlindo Lopes de Almeida (in memorian)

a quem sou eternamente grato pela minha formação profissional, pois, além de me proporcionar a oportunidade de estudar sempre me incentivou na busca do saber e do crescimento intelectual.

A minha esposa,

Mailde Nogueira Pessoa

pelo incentivo e apoio incondicional, amor, carinho, dedicação, companhia constante em todos os momentos.

Aos meus filhos,

Thales e Clarissa

pela felicidade que trouxeram a minha vida, incentivo e paciência em suportar minhas ausências.

DEDICO

AGRADECIMENTOS

À Universidade Federal Rural de Pernambuco – UFRPe, pela oportunidade desta qualificação.

Ao professor Dr. Abelardo Montenegro, pela orientação, pela receptividade e pela atenção dispensada e apoio imprescindível à realização deste trabalho.

Ao professor Dr. Eufrásio de Souza Santos, coordenador do curso de Pós-Graduação em Biometria e meu co-orientador, pela confiança, os ensinamentos, as sugestões e o crédito ao meu trabalho.

Aos Professores Doutores Maria Adélia, Socorro Brasileiro, Gauss Cordeiro, Paulo de Paula, Borko Stosic, José Antônio Aleixo, Cláudia e demais professores do Mestrado em Biometria pelos conhecimentos transmitidos.

Aos demais colegas do mestrado, especialmente, Adalberto, Ady Marinho, Arundo Junior, Toni Oliveira, Dâmocles, Heliovânio, Hérbetes, Luís de França, Mácio Augusto e Sérgio, pessoas que fizeram parte do convívio diário nessa empreitada, e de quem tive apoio e suporte em todos os momentos e pela amizade construída.

Aos colegas do Departamento de Tecnologia Rural da UFRPe, Adriana, Eduardo e Manoel Neto e os doutorandos Adriano Normandia, Francisco dos Santos e Veronildo de Oliveira pelo auxílio nas etapas de coleta e preparação dos dados e que sem o seu suporte muito prejudicaria o bom andamento das tarefas requeridas para a conclusão deste trabalho.

Ao secretário do mestrado Marco Antônio que atenciosamente sempre me auxiliou.

Enfim, a todos que direta ou indiretamente contribuíram com o desenvolvimento deste trabalho agradeço.

LISTA DE FIGURAS

Figura 1 – Mapa de locação dos pontos amostrais para as mensurações da condutividade elétrica do solo.

Figura 2 - Resposta relativa para o EM38 em função da profundidade (z).

Figura 3 - Vista e orientação do equipamento EM38, nas posições Horizontal (seta verde) e Vertical (seta vermelha).

Figura 4 - Representação gráfica dos parâmetros do semivariograma.

Figura 5 – Histograma com curva normal para CEaparente – vertical.

Figura 6 - Teste de normalidade de Kolmogorov-Smirnov para CEaparente – vertical.

Figura 7 – Histograma com curva normal para CEaparente – horizontal.

Figura 8 – Teste de normalidade de Kolmogorov-Smirnov para CEaparente – horizontal.

Figura 9 – Dispersão da Condutividade Elétrica na direção x, para a CEaparente-vertical.

Figura 10 - Dispersão da Condutividade Elétrica na direção y, para a CEaparente-vertical.

Figura 11 - Dispersão da Condutividade Elétrica na direção x, para a CEaparente-horizontal.

Figura 12 – Dispersão da Condutividade Elétrica na direção y, para a CEaparente-horizontal.

Figura 13 – Semivariograma gaussiano para CEaparente – vertical.

Figura 14 – Semivariograma exponencial para CEaparente – vertical.

Figura 15 – Semivariograma esférico para CEaparente – vertical.

Figura 16 – Semivariograma gaussiano para CEaparente – horizontal.

Figura 17 – Semivariograma exponencial para CEaparente – horizontal.

Figura 18 – Semivariograma esférico para CEaparente – horizontal.

Figura 19 – Histograma dos Resíduos para CEaparente – vertical.

Figura 20 – Histograma dos Resíduos para CEaparente – horizontal.

Figura 21 – Semivariograma gaussiano para CEaparente – vertical – direção 0°.

Figura 22 – Semivariograma gaussiano para CEaparente – vertical – direção 45°.

Figura 23 – Semivariograma gaussiano para CEaparente – vertical – direção 90°.

Figura 24 – Semivariograma gaussiano para CEaparente – vertical – direção 135°.

Figura 25 – Semivariograma gaussiano para CEaparente – horizontal – direção 0°.

Figura 26 - Semivariograma gaussiano para CEaparente – horizontal - 45°.

Figura 27 – Semivariograma gaussiano para CEaparente – horizontal - 90°.

Figura 28 – Semivariograma gaussiano para CEaparente – horizontal - 135°.

Figura 29 - Representação espacial e mapa de isolinhas da Krigagem para CEaparente – vertical.

Figura 30 – Representação espacial e mapa de isolinhas da Krigagem para CEaparente – horizontal.

Figura 31 – Semivariogramas teórico, experimental e dos parâmetros bayesianos da média, mediana e moda da condutividade elétrica aparente – modo vertical.

Figura 32 – Semivariogramas teórico, experimental e dos parâmetros bayesianos da média, mediana e moda da condutividade elétrica aparente – modo horizontal.

Figura 33 – Perfis das log-verossimilhanças dos parâmetros ϕ e σ^2 da condutividade elétrica aparente – modo vertical.

Figura 34 – Perfis das log-verossimilhanças dos parâmetros ϕ e σ^2 da condutividade elétrica aparente – modo horizontal.

Figura 35 – Histogramas das funções de distribuições posterioris para os parâmetros beta, σ^2 e ϕ da condutividade elétrica aparente – modo vertical.

Figura 36 – Histogramas das funções de distribuições posterioris para os parâmetros beta, σ^2 e ϕ da condutividade elétrica aparente – modo horizontal.

LISTA DE TABELAS

Tabela 1 – Estatística Descritiva das medidas de condutividade elétrica aparente.

Tabela 2 – Valores críticos dos testes KS.

Tabela 3 – Correlação entre a condutividade elétrica aparente e as coordenadas x e y .

Tabela 4 – Valores dos parâmetros dos semivariogramas ajustados.

Tabela 5 - Validação cruzada dos resíduos padronizados para a condutividade elétrica aparente.

Tabela 6 – Resultados das estimativas dos parâmetros obtidos utilizando estimadores de máxima verossimilhança de observações de condutividade elétrica aparente obtida no modo de operação vertical.

Tabela 7 – Resultados das estimativas dos parâmetros obtidos utilizando estimadores de máxima verossimilhança de observações de condutividade elétrica aparente obtida no modo de operação horizontal.

Tabela 8 - Validação cruzada dos resíduos padronizados para a condutividade elétrica aparente utilizando a estimativa de máxima verossimilhança.

SUMÁRIO

RESUMO	13
ABSTRACT	14
1. INTRODUÇÃO	15
2. REVISÃO BIBLIOGRÁFICA	18
2.1. Aspectos Gerais da Variabilidade Espacial	18
2.2. Condutividade Elétrica do Solo	19
2.3. Suficiência Amostral	21
2.4. Estudos Metodologia Geoestatística	22
2.5. Estudos da Metodologia Bayesiana	24
3. MATERIAIS E MÉTODOS	26
3.1 Localização da Área Amostral	26
3.1.1 Solos e Relevo	27
3.2 Variável Estudada	27
3.3. Equipamento Utilizado	28
3.4 Metodologia Geoestatística	29
3.4.1 Teste de Normalidade KS	30
3.4.2 Teste de Shapiro-Wilk	31
3.4.3 Suficiência Amostral	32
3.4.4 Análise de Tendências	32
3.4.4 Variáveis Regionalizadas e Função Aleatória	33
3.4.5 Momentos de uma Função Aleatória	34
3.4.6 Hipóteses de Estacionaridade	36
3.4.7 Semivariograma	38
3.4.8 Validação Cruzada	43
3.4.9 Método de Estimativa pela Máxima Verossimilhança	44
3.4.10 Critério de Informação de Akaike	48
3.5 Abordagem Bayesiana	49
3.5.1 Estimação Bayesiana	52
3.6 Softwares Utilizados	60
4. RESULTADOS E DISCUSSÃO	62
4.1 Análise Estatística Descritiva	62
4.2 Análise de Tendências	66

	12
4.3 Análise Geoestatística	69
4.4 Validação Cruzada	74
4.5 Estimadores de Máxima Verossimilhança	76
4.6 Estudo da Anisotropia	78
4.7 Estudo da Krigagem e Representação Espacial	83
4.8 Krigagem Bayesiana	84
5. CONCLUSÕES	90
REFERÊNCIAS BIBLIOGRÁFICAS	91

Pessoa, Antônio Lopes. Uma abordagem bayesiana para estudo estatístico e geoestatístico de estimativas de salinidade do solo utilizando sensor de indução eletromagnética. Orientador: Prof. Abelardo Antônio de Assunção Montenegro, PhD. Co-orientador: Prof. Dr. Eufrázio de Souza Santos.

RESUMO

Esta dissertação analisou a relação existente entre medidas de condutividade elétrica aparente de um solo aluvial da região Agreste do Estado de Pernambuco, e a sua variabilidade espacial na camada subsuperficial. A condutividade elétrica aparente do solo foi investigada através de equipamento de indução eletromagnética EM 38. As leituras efetuadas com o EM 38 foram tanto no modo vertical como no modo horizontal. As medidas obtidas em campo foram analisadas através da estatística descritiva clássica, bem como através das metodologias geoestatística e abordagem bayesiana. As análises estatísticas mostraram que os dados de condutividade elétrica aparente se ajustaram a uma distribuição normal e apresentaram uma alta variabilidade espacial para o modo de operação horizontal e uma média variabilidade espacial para o modo de operação vertical. Através da metodologia geoestatística foi construído o semivariograma experimental que, posteriormente, foi ajustado a um modelo teórico. O melhor ajuste de modelo teórico foi obtido, tanto para o modo de operação vertical como para o modo de operação horizontal, para o modelo gaussiano e para o modelo exponencial, efetuada através da validação cruzada e do Critério de Informação de Akaike. A partir da dependência espacial, foi elaborado o mapeamento das áreas salinizadas. A abordagem bayesiana focalizou a predição espacial, relacionando o método da máxima verossimilhança com as funções de distribuições prioris de cada parâmetro, considerando o grau de incerteza associado a cada uma dessas distribuições. Verificou-se que os semivariogramas ajustados não apresentaram diferenças significativas na validação da metodologia geoestatística e na abordagem bayesiana.

Pessoa, Antônio Lopes. A abordagem bayesiana para estudo estatístico e geoestatístico de estimativas de salinidade do solo utilizando sensor de indução eletromagnética. Orientador: Prof. Abelardo Antônio de Assunção Montenegro, PhD. Co-orientador: Prof. Dr. Eufrázio de Souza Santos.

ABSTRACT

This study analyzed the existing relationship among measurements of soil apparent electrical conductivity in an alluvial valley in the Agreste region of Pernambuco State and its spatial variability in the subsurface. The soil apparent electric conductivity was investigated through an electromagnetic induction EM 38 equipment. The readings have been carried out both in the vertical and horizontal modes. The measurements have been analyzed through the classic descriptive statistics as well as geostatistics and bayesian approach. The statistical analyses had shown that the data of apparent electric conductivity had adjusted to a normal distribution, presenting a high space variability for the horizontal mode and an average space variability for the way of vertical operation. In order to allow the use of the geostatistical methodology, the experimental semivariogram was constructed, and fitted to a theoretical model. Considering the spatial dependence mapping of the salinized areas have been performed. The best theoretical models for the vertical mode and for the horizontal operation were the gaussian model and the exponential model, through the crossed validation and using the Akaike's Information Criterion. The bayesian approach focused the spatial prediction relating the method of the maximum likelihood with the functions of prioris distributions for each parameter, considering the uncertainty associated to each one of these distributions. It was verified that the adjusted semivariograms had not presented significant differences in the validation of the geostatistics methodology and in the bayesian approach.

INTRODUÇÃO

O estudo da variabilidade espacial da condutividade elétrica do solo constitui um dos fatores que possibilita um manejo racional do solo e da água numa agricultura irrigada. A variabilidade espacial é entendida como variações de características dos solos dentro da taxonomia ou do mapeamento de solos.

Os diferentes tipos de solo com suas diferentes características físicas e químicas são formados de acordo com a predominância ou intensidade de atuação dos fatores responsáveis pela formação, bem como de fatores antrópicos.

A salinidade do solo é um problema que afeta o rendimento agrícola, reduzindo a produtividade das culturas não tolerantes aos excessos de sais. Esse problema é mais freqüente em regiões de clima quente e seco onde as taxas de evapotranspiração são elevadas e os índices de pluviometria baixos, como é o caso da região estudada do semi-árido do nordeste brasileiro. Dentre os vários fatores que contribuem para o processo de salinização se tem os aspectos climáticos, as condições edáficas, métodos de irrigação utilizados, drenagem insuficiente e presença de lençol freático raso. Os sais dissolvidos contidos na água de irrigação e no aquífero podem se acumular no perfil do solo ao longo do tempo (Montenegro et al., 2003).

Os solos de uma maneira geral, possuem uma grande variabilidade espacial, especialmente, os salinos e/ou os salinos-sódicos, variabilidade essa que foi investigada neste estudo.

A necessidade de compreensão da variabilidade e da dependência espacial da condutividade elétrica do solo e de outras propriedades e características correlacionadas é de fundamental importância no equacionamento das técnicas de manejo de água e solo.

A avaliação e análise dos dados para a estimativa das variáveis foram feitas através do emprego de técnicas de estatística descritiva e da metodologia geoestatística, que estudam os aspectos casuais e espaciais dos dados e se complementam.

Segundo Montenegro et al. (1999), a Geoestatística ou Teoria das Variáveis Regionalizadas, é a ciência que estuda a variabilidade de uma medida através de

um modelo probabilístico, associado à posição relativa das observações no espaço e possível correlações dessas observações em função da distância de separação entre elas.

Essas variáveis regionalizadas apresentam um duplo aspecto contraditório: primeiro revelam uma característica aleatória que se manifesta na irregularidade e na variação imprevisível de um ponto qualquer a outro no espaço; por outro lado, as características estruturais, que refletem as ligações existentes entre os pontos no espaço, não são inteiramente independentes.

As variáveis regionalizadas não podem ser tratadas apenas pela estatística clássica, tendo em vista que na estatística clássica as amostras podem ser indefinidamente repetidas. As amostras também são independentes, ou seja, as características resultadas de uma amostra não influenciam o resultado das amostras em locais vizinhos. Um estudo geoestatístico possibilita fazer inferências, baseando-se no comportamento espacial das variáveis regionalizadas estudadas.

Uma variável regionalizada manifesta a tendência de apresentar, nas suas propriedades, uma continuidade geográfica de forma que os atributos estudados em pontos vizinhos possuem valores semelhantes, que se tornam diferentes à medida que os pontos vão ficando mais distantes. Outras propriedades estudadas das variáveis regionalizadas são: localização, anisotropia e tendência.

O objetivo da análise descritiva é o de observar o comportamento geral dos dados, enquanto que a análise geoestatística tem como objetivo identificar a estrutura de variabilidade espacial dos dados. A identificação da estrutura de dependência espacial entre as observações, pode então ser feita através da função de semivariância.

A metodologia geoestatística é baseada na hipótese de que é possível inferir a partir de um único conjunto de dados experimentais ou amostrais. Essa hipótese é definida como hipótese de estacionaridade intrínseca. Assumida tal hipótese, de estacionaridade intrínseca, no caso, uma estacionaridade de segunda ordem, as medidas de semivariância mostram o grau de dependência espacial, isto é, a descrição das relações espaciais entre pontos vizinhos.

A função semivariância $\gamma(h)$ é determinada pelo método dos momentos e é definida a partir das medidas de desvios. Espera-se que os valores dos pontos

mais próximos sejam mais similares do que os valores dos pontos mais afastados, e que a diferença de valores dependam apenas da posição relativa.

A partir dos valores de semivariância, pode ser construído o semivariograma amostral para o conjunto de dados, verificando-se a hipótese de isotropia, ou seja, o mesmo comportamento de condutividade elétrica do solo nas diferentes direções espaciais. O semivariograma amostral então é ajustado a um modelo teórico e os parâmetros utilizados na interpolação de valores dos pontos não amostrados serão estimados pela técnica da krigagem.

O avanço computacional possibilitou a resolução de questões através de métodos iterativos na estimativa de parâmetros, como o método da máxima verossimilhança, possibilitando assim um critério para a escolha do melhor modelo.

O método de maximização da função de verossimilhança é utilizado em ambas formas de inferências: a clássica e a bayesiana. A aplicação da função de verossimilhança na inferência clássica difere da aplicação na inferência bayesiana, pois nesta, o parâmetro de interesse é considerado como uma variável aleatória.

A metodologia estatística bayesiana tem um importante papel na teoria e prática de simulação estocástica. Na análise de simulações estocásticas o interesse tem sido estimulado em parte, pelas pesquisas analítica e computacional, que tem possibilitado a implementação de cálculos estatísticos bayesianos. A principal diferença entre a abordagem frequentista e a abordagem bayesiana é que na estrutura de modelos bayesianos toda a incerteza é expressa em termos de distribuição de probabilidade.

A metodologia bayesiana tem como objetivo principal de estudo a obtenção da função de densidade de probabilidade posteriori do parâmetro de investigação. Com isto, a abordagem bayesiana permite uma estrutura para modelar problemas estatísticos.

Este trabalho teve como objetivo investigar os padrões de variabilidade e dependência espacial da condutividade elétrica aparente, utilizando técnicas estatísticas descritivas, metodologia geoestatística e abordagem bayesiana, visando o mapeamento da salinidade do solo como base para o manejo da irrigação e o controle da salinização da área analisada.

2. REVISÃO DE LITERATURA

Aspectos gerais da variabilidade espacial

O estudo da variabilidade espacial de algumas variáveis (características do solo, por exemplo) vem sendo objeto das preocupações de pesquisadores desde o princípio do século XX.

Segundo Vieira (2000), estudos sobre a variabilidade espacial foram iniciados no começo do século passado, foi estudada a disposição de parcelas no campo em experimentos de rendimento de variedades de milho, numa tentativa de eliminar o efeito de variações do solo. Vários estudos posteriores foram efetuados, sempre com a preocupação de caracterizar ou conhecer a variabilidade espacial do solo.

Os procedimentos usados na época baseavam-se na estatística clássica e utilizavam grandes quantidades de dados amostrais, visando caracterizar ou descrever a distribuição espacial da característica em estudo. Por estatística clássica entende-se aquela na qual se utilizam parâmetros como média e desvio padrão para representar um fenômeno e se baseia na hipótese principal de que as variações de um local para outro são aleatórias.

Os conceitos da estatística clássica e seus fundamentos podem ser encontrados em Snedecor e Cochran (1977).

Krige (1951), trabalhando com dados de concentração de ouro, concluiu que somente a informação dada pela variância seria insuficiente para explicar o fenômeno em estudo. Para tal, seria necessário levar em consideração a distância entre as observações. A partir daí surge o conceito da geoestatística que leva em consideração a localização geográfica e a dependência espacial.

Matheron (1963, 1971), baseado nas observações de Krige, desenvolveu a teoria das variáveis regionalizadas ou metodologia geoestatística. Uma variável regionalizada é uma função numérica com distribuição espacial, que varia de um ponto a outro com continuidade aparente, mas cujas variações não podem ser representadas por uma função matemática simples.

A metodologia geoestatística foi primeiramente aplicada em áreas de mineração, em seguida em estudos hidrológicos (Delhomme, 1976; Olea, 1977). Nos estudos em ciência do solo, destacam-se trabalhos desenvolvidos pelos pesquisadores Vieira e Molin (2001), e Montenegro et al (2003).

2.2 Condutividade Elétrica do Solo

Nos estudos da condutividade elétrica do solo efetuado por Bohn et al. (1982), o solo foi considerado como um recipiente truncado para solução eletrolítica, como um condutor possuindo um percurso tortuoso, ou um condutor possuindo um grande número de percursos de condução de comprimentos e seções variáveis.

Segundo Nadler e Frenkel (1980), o solo pode conduzir corrente elétrica através da água intersticial que contém eletrólitos dissolvidos e através dos cátions trocáveis que estão localizados próximo da superfície das partículas de solo carregadas, e são eletricamente móveis em vários níveis.

Rhoades e Corwin (1991) apresentam um modelo de condutividade elétrica que descreve a condutância através de três vias, atuando em paralelo:

- Condutância através de camadas alternantes de partículas do solo e entre os limites de suas soluções;
- Condutância através de soluções do solo contínuas;
- Condutância através ou em superfícies de partículas do solo em contato direto entre elas. Na ausência de sais dissolvidos na água presente na segunda via, condutividade, textura e umidade se correlacionam muito bem.

A condutividade elétrica (CE) de um solo varia geralmente com o teor de argila, o conteúdo de água e a salinidade (McNeill 1980b).

A condutividade elétrica do solo depende da solução eletrolítica existente no solo ao invés das partículas sólidas do mesmo. Normalmente, os solos secos

apresentam uma resistência à penetração alta. Sheets e Hendrickx (1995), constataram que através da medição da condutividade aparente do solo (CEa), também se podem realizar medições rápidas da umidade dos solos.

Conforme McNeil (1980a), a condutividade elétrica de solos e rochas depende da porosidade e do grau de enchimento dos poros da água. Em pesquisa realizada por Lund et al. (2001), foi constatado que a condutividade elétrica aparente do solo se correlaciona com a textura e com o tamanho das partículas do solo.

Estudos prévios mostraram que os fatores primordiais para a condutividade elétrica do solo em algumas áreas são o teor de argila (Cook & Walker, 1992); a salinidade (Rhoades et al. 1990; Lesch et al. 1995); e a água (Kachanoski et al. 1988; Kachanoski et al. 1990; Sheets e Hendrickx 1995); além dos três fatores conjuntamente, em partes diferentes de uma área de estudo (Scanlon et al. 1999).

Os íons dentro da camada eletricamente difusa de partículas de argila são capazes de conduzir eletricidade. Portanto, os minerais de argila contribuem significativamente para a condutividade elétrica do solo (Kachanoski et al. 1988).

A indução eletromagnética é uma técnica não invasiva que utiliza procedimentos para deduzir o perfil de condutividade elétrica do solo de acordo com a profundidade por meio de leituras múltiplas obtidas com o instrumento segurado a alturas variadas sobre a superfície do solo (Rhoades e Corwin, 1990). Em um estudo, Kachanoski et al. (1990) verificaram que a CEa explicava 96% da variação espacial no conteúdo de água no horizonte superior (0 - 0,5 m) do perfil de um campo de 1,8 hectares. Porém, estudos realizados por Doolittle et al. (1994) e Sudduth et al. (1995) mostraram que a profundidade de uma camada de solo superficial sobreposta a uma camada de argila poderia ser calculada precisamente com leituras efetuadas numa única altura com o sensor de indução eletromagnética EM38 (Geonics, 1999), eliminando assim, a necessidade de efetuar medidas a alturas múltiplas no processo de coleção de dados.

A medição de condutividade elétrica do solo por indução eletromagnética é um método de amostragem rápida para a coleta de informações sobre o solo e da distribuição espacial de algumas de suas propriedades.

A classificação do solo usando a condutividade elétrica aparente do solo permite uma efetiva base para delinear os atributos físicos, químicos e biológicos

do solo que são inter-relacionados e estão expressos como condições para produtividade da colheita e potencial ecológico.

Estas informações são essenciais para o monitoramento temporal da condição do solo, e para uma implementação bem-sucedida da gestão em local específico (Johnson et al., 2001).

Investigações de salinidade do solo possibilitam informações sobre o impacto de práticas agrícolas como irrigação, no campo e em micro-escala.

Triantafilis et al. (2000), consideraram que as medições eletromagnéticas têm sido muito utilizadas para estes propósitos, porque estão fortemente correlacionadas com a salinidade do solo. Por comparação com métodos elétricos diretos e o domínio do tempo de reflectometria, a não inserção física de sondas é requerida e conseqüentemente menos tempo é consumido na obtenção de uma medida em cada locação. Conseqüentemente, um grande número de medidas pode ser feito. Como exemplo de trabalhos científicos que utilizaram a indução eletromagnética para mapear a salinidade do solo, pode-se citar McKenzie et al. (1997). Em ensaios de campo, o EM 38 foi utilizado como complemento na avaliação da tolerância a sais de 26 espécies de árvores e 28 espécies de forragens e pastos.

O equipamento EM 38 é uma ferramenta precisa na preparação detalhada de mapas de salinidade do solo que possibilitem inventários precisos da extensão e severidade da salinização.

2.3 Suficiência Amostral

Um dos principais problemas que deve ser considerado em qualquer levantamento de campo, é sem dúvida a suficiência amostral, pois é através de seu cálculo que o pesquisador pode expressar suas inferências baseado no princípio de que sua amostra representa a população onde a amostragem foi executada, a um determinado nível de probabilidade e erro amostral.

Vários são os processos amostrais existentes, sendo que os mais freqüentemente utilizados são: inteiramente aleatório, estratificado, sistemático, em conglomerados, entre outros (Silva et al., 1997, Meunier et al., 2002).

O princípio básico dos processos amostrais é resultante do limite de confiança no qual a média amostral pode variar em uma amplitude de $t_{\alpha/2} \cdot s_{\bar{x}}$, em que:

$t_{\alpha/2}$ = valor do teste de t, $s_{\bar{x}}$ = o erro padrão da média. Sendo $s_{\bar{x}} = \frac{s}{\sqrt{n}}$, pode-se calcular o tamanho da amostra n no limite de confiança, sendo que os dados usados para tal são provenientes de uma amostra preliminar ou piloto, considerando um erro adotado, geralmente 10% em trabalhos de campo.

Quando o erro de amostragem proveniente da amostra piloto é menor ou igual ao erro adotado, a amostra piloto passa a ser definitiva, caso contrário, coleta-se mais observações, e se refaz o cálculo. Esse procedimento continua até quando o erro de amostragem for menor ou igual ao erro adotado.

A partir desse ponto o pesquisador pode fazer suas inferências, pois a suficiência amostral foi atingida.

2.4 Estudos da Metodologia Geoestatística

No estudo de áreas onde há um gradiente de propriedades do solo têm sido desenvolvidas e aplicadas técnicas amostrais e de análise de dados que levem em consideração a variação espacial, que muitas vezes está associada a sua formação (Webster, 1985; Goovaerts, 1994).

Segundo Cressie (1991), é de grande importância uma abordagem que leve em consideração a dependência espacial, mostrando os efeitos de autocorrelação espacial em problemas de estimação, predição e delineamento de experimentos.

Do ponto de vista matemático, o valor de uma variável (propriedade do solo), em qualquer lugar da superfície terrestre, depende da sua posição.

A geoestatística pode ser definida como um conjunto de técnicas para a solução de problemas de estimativa que envolve variáveis espaciais e tem sua base conceitual na interpretação espacial de uma variável regionalizada como realização particular de uma função aleatória.

Ribeiro Junior (1995), ressalta que a geoestatística é um conjunto de métodos que se destacam por permitir que cada observação possa ser descrita não apenas pelo seu valor, mas também por informações de sua posição, expressas em um sistema de coordenadas. Considera assim, que observações

geograficamente mais próximas tendem a apresentar valores mais similares, e que tal fato pode ser avaliado por medidas de associação.

A utilização de ferramentas geoestatísticas, além de identificar a possível existência de dependência espacial entre as observações, pode ser aplicada em mapeamento, orientação de futuras amostragens e modelagem. Se os valores próximos não são independentes entre si, pode-se usar ferramentas mais sofisticadas de interpolação e estimação de valores em lugares não medidos, a exemplo do processo de krigagem; desta forma, uma representação mais realista da área estudada pode ser obtida (Isaaks & Srivastava, 1989; Vieira, 1995).

A krigagem é um método para estimar, num contexto espacial, o valor de uma variável de interesse em um local onde não foi determinado, usando, para isso, dados obtidos em pontos adjacentes (Wackernagel, 1998).

A krigagem, ao contrário de outros métodos de interpolação, estima uma matriz de covariância espacial que permite determinar os pesos atribuídos às diferentes amostras, a redundância dos dados, a vizinhança a ser considerada no procedimento inferencial e o erro associado ao valor estimado (Goovaerts, 1997).

Existem vários algoritmos de krigagem em função do modelo adotado para estimar o conjunto de valores possíveis em relação ao atributo espacial em estudo. Um deles é o estimador de krigagem ordinária (KO) que, segundo Goovaerts (1997), permite calcular médias locais, limitando o domínio de estacionaridade da média à vizinhança local centrada no ponto a ser estimado.

Lloyd & Atkinson (2001) compararam os métodos da krigagem ordinária, krigagem associada com modelos de tendência e a krigagem por indicação nos modelos de elevação e constataram a incerteza dos valores estimados.

As variações das propriedades do solo em escala regional foram estudadas por Voltz et al. (1997) comparando a ponderação por krigagem, a ponderação em função da distância quadrática e o método dos polígonos de influência.

Estudos semelhantes foram obtidos por Burgess & Webster (1980), utilizando a técnica da krigagem para mapeamento das propriedades do solo.

A variabilidade espacial é uma característica da maioria dos fenômenos naturais, podendo-se incluir, entre eles, os parâmetros relacionados com o movimento de água no solo (Queiroz, 1995). Gonçalves et al. (1999) aplicaram a

geoestatística ao estudo da distribuição espacial de valores de armazenamento de água no solo.

A variabilidade espacial também pode ser alterada pelo manejo do solo (Cassel et al., 1986 ; Souza, 1992). Para condições de solo seco e de solo úmidos distintos comportamentos da variabilidade espacial da resistência mecânica do solo à penetração (RP) foram observados por Utset & Cid (2001).

Segundo Vieira & Molin (2001), o conhecimento da variabilidade espacial utilizando as ferramentas estatísticas adequadas é essencial para o mapeamento e determinação das unidades de gerenciamento empregados na aplicação de técnicas de agricultura de precisão. Portanto, a investigação da condutividade elétrica aparente do solo é grande utilidade da determinação das propriedades e parâmetros dos solos agrícolas.

Dentre os atributos físicos, aqueles que se relacionam com a dinâmica da água no solo, como a condutividade e difusividade hidráulica e a taxa de infiltração, exibem uma alta variabilidade espacial (Warrick & Nielsen, 1980).

2.5 Estudos da Metodologia Bayesiana

Estudos sobre informações espacialmente e facilmente interpretáveis a cerca de algumas propriedades do solo como areia, silte e conteúdo de argila através de mapas de solos contêm um importante resultado de informações herdadas do passado. A abordagem da metodologia bayesiana possui uma habilidade para produzir uma acurada estimativa de textura do solo para um mapa de solo com um limitado número de amostras adicionais.

Segundo Bogaert e D'Or (2002), as estimativas produzidas com a abordagem bayesiana são mais acuradas do que as obtidas por krigagem ordinária e os mapas de solos assim obtidos mostram os mais realísticos padrões espaciais.

Os chamados estatísticos clássicos ou frequentistas argumentavam que os métodos bayesianos só poderiam resolver problemas simples, onde a distribuição a posteriori pudesse ser descrita analiticamente. Isto exigia a escolha de distribuições a priori muitas vezes convenientes e, por isto, nem sempre representativas do verdadeiro fenômeno físico. Além disto, problemas mais complicados requereriam modelos mais complexos e, assim, o cálculo de integrais também mais complicadas (Carlin e Louis, 1996).

Com o advento de computadores mais rápidos e mais baratos, a análise Bayesiana de dados pôde avançar e eliminar as restrições sobre as distribuições prioris e adoção de modelos. Os métodos de integração de Monte Carlo tornaram-se populares entre os bayesianos, assim como os métodos iterativos de Monte Carlo, dentre eles, o Markov Chain Monte Carlo (MCMC). As técnicas de MCMC permitem a geração de cadeias de Markov, que podem ser definidas como processos descrevendo trajetórias onde quantidades sucessivas são descritas probabilisticamente de acordo com o valor de seu predecessor imediato (Gamerman, 1997).

Kitanidis (1986), considerando dados espacialmente contínuos, desenvolveu estudo sobre análise bayesiana da incerteza, na estimação de parâmetros em funções espaciais. Neste trabalho, o autor mostrou que a krigagem ordinária e a universal, principais técnicas para análise de dados espacialmente contínuos, podem ser vistas como técnicas de predição bayesiana, onde existe um desconhecimento da distribuição priori dos parâmetros da superfície de tendência (atribuição de prioris vagas para a média da superfície).

Omre (1987) explicita o termo krigagem bayesiana, discussão que continua em Omre (1989), trabalho no qual o autor utiliza o método bayesiano empírico.

Diggle et al (1998) apresentam uma revisão dos métodos de krigagem para dados gaussianos e propõem um modelo para dados não gaussianos, utilizando métodos bayesianos na estimação dos parâmetros envolvidos.

Outros trabalhos fazem uso do método da krigagem bayesiana. Host (1999), por exemplo, estudou a emissão de enxofre numa região a partir de medição em determinados locais, utilizando krigagem bayesiana.

Diggle e Ribeiro Jr. (2002) discutiram a geoestatística baseada em modelos gaussianos e, aplicando métodos bayesianos como em Kitanidis (1986) e Handcock e Stein (1993), analisaram dados de precipitação pluviométrica no Paraná.

Cowles e Zimmerman (2003) realizaram uma análise espaço-temporal dos dados de deposição de ácido, utilizando a abordagem bayesiana.

A precipitação pluviométrica no estado do Pará também foi modelada através da krigagem bayesiana, na tentativa de entender seu papel na incidência de malária no Estado (Nobre et al., 2005).

Schmidt et al. (2002) fazem uma revisão dos modelos para análise de dados de área e espacialmente contínuos, em especial aqueles utilizados na análise bayesiana.

Banerjee et al. (2000) e Gill e Swartz (2004) discutiram o problema da análise de dados de iteração espacial sob o ponto de vista bayesiano.

3. MATERIAIS E MÉTODOS

Os dados foram provenientes de medições de condutividade elétrica do solo. Foram utilizados programas computacionais tanto para a análise estatística descritiva como para a análise geoestatística e bayesiana.

3.1 Localização da Área Amostral

O estudo da salinidade do solo foi efetuado na Fazenda Nossa Senhora do Rosário propriedade situada no município de Pesqueira, região Agreste do Estado de Pernambuco, localizada na bacia hidrográfica do Rio Ipanema e situada entre as coordenadas geográficas 8° 15' e 8° 30' de Latitude Sul, 31° 45' e 37° 00' de Longitude Oeste de Greenwich e 650 m de altitude.

Segundo a classificação de KÖPPEN, o clima da região é semi-árido quente tipo estepe, com uma precipitação média anual de 730 mm (Hargreaves, 1974). A temperatura média mensal oscila em torno de 20 a 30°C.

A Figura 1 mostra, o mapa de locação dos pontos amostrais para as mensurações da condutividade elétrica do solo.

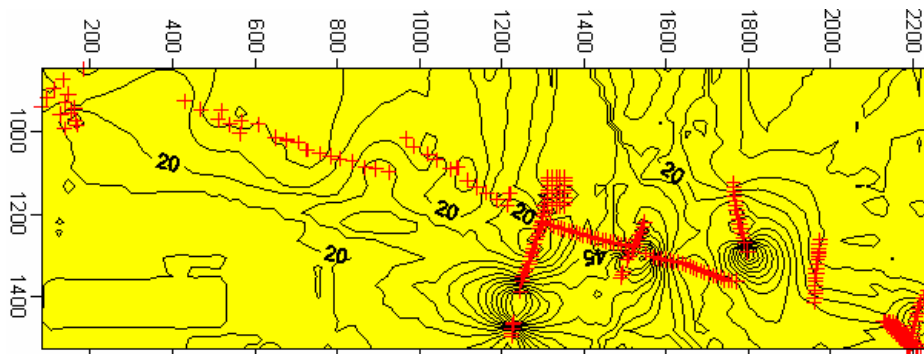


Figura 1 - Mapa de localização dos pontos amostrais para as mensurações da condutividade elétrica do solo.

Foi utilizado um esquema de amostragem referenciada seguindo uma linha de piezômetros existentes e efetuadas medições horizontais e verticais em 260 pontos. A coleta dos dados de condutividade elétrica aparente foi executada nos meses de outubro e novembro de 2004, em período seco e sem incidência de chuvas, com o aparelho EM 38 (Geonics, 1999).

3.1.1 Solos e Relevo

Os solos encontrados na região de estudo (EMBRAPA, 1997) são classificados como associações de:

- Neossolo Flúvico.
- Regosol Eutrófico com fragipan fase caatinga hipoxerófila relevo suave ondulado e ondulado.
- Solos Litólicos Eutróficos com A fraca textura arenosa e/ou média fase pedregosa e rochosa caatinga hipoxerófila relevo ondulado e ondulado.
- Afloramentos de rocha.

As medições foram realizadas prioritariamente nas áreas de solos Neossolos Flúvicos.

3.2 Variável Estudada

A condutividade elétrica (CE) é a habilidade que um meio poroso apresenta em conduzir uma corrente elétrica. É determinada por:

- Porosidade – forma e tamanho de poros, número, tamanho e forma da passagem de interconexão.
- Umidade – para os poros que estão cheios de água ou conteúdo da mistura.
- Concentração de eletrólitos – quantidade de eletrólitos dissolvidos no conteúdo da mistura que preenche os poros.
- Temperatura – estado e fase da água nos poros.

A leitura de condutividade elétrica do solo pelo aparelho de indução eletromagnética sofre contribuições de diferentes profundidades.

A condutividade elétrica aparente - CE_a é a integral da resposta relativa $\sigma(z)$ pela CE(z), dada por:

$$CE_a = \int_0^{\infty} \sigma(z) CE(z) dz \quad 3.1$$

Em que: CE(z) é a função da condutividade elétrica ao longo da profundidade z e $\sigma(z)$ é a função de ponderação para o instrumento. A integral da resposta relativa de 0 a infinito é igual a 1, ou seja:

$$\int_0^{\infty} \sigma(z).dz = 1 \quad 3.2$$

A curva resposta relativa para o aparelho EM38 (Geonics, 1999) em função da profundidade (z), conforme a Figura 4, mostra que a maior contribuição para a leitura está situada a 0,40 m da superfície para o modo vertical, e 0 (zero) para o modo horizontal (Montenegro et al., 2003).

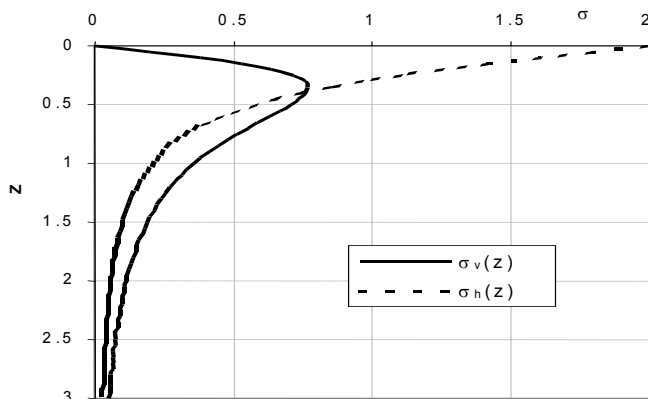


Figura 2 - Resposta relativa para o EM38 em função da profundidade (z).

3.3 Equipamento Utilizado

O instrumento utilizado nesta dissertação foi o EM38 (Geonics, 1999), sendo bem aceito atualmente na medição de condutividade elétrica do solo. O EM38 tem uma profundidade de investigação efetiva de aproximadamente 1,5 metro e possui compensação automática de temperatura. É relativamente pequeno leve e fácil de se manejar. Com esse instrumento podem-se medir condutividades de aproximadamente 0,5 a 1000mS/m. O equipamento EM38, mostrado na Figura 5, trabalha com o princípio da indução eletromagnética e pode ser instalado nas posições vertical ou horizontal. As medições foram efetuadas na superfície do solo.

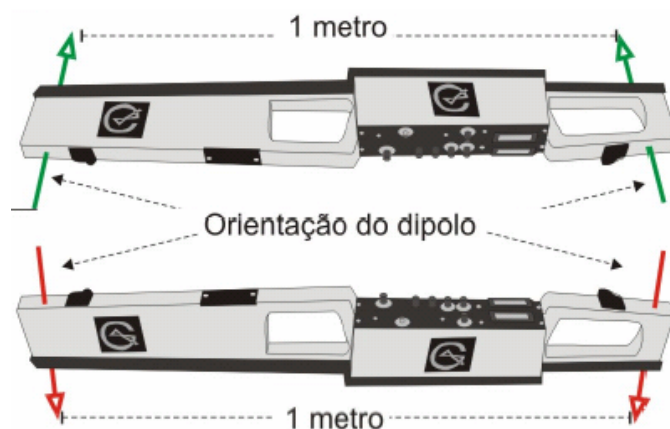


Figura 3 – Vista e orientação do equipamento EM38, nas posições Horizontal (seta verde) e Vertical (seta vermelha) (Borges e Porsani, 2003).

Os locais onde foram efetuadas as medidas de condutividade elétrica foram plotadas em um sistema de eixos cartesianos, de forma que a distância entre cada um dos pontos pudesse ser medida.

3.4 Metodologia Geoestatística

A análise de dados proposta por Fisher em 1935, em sua metodologia admite as seguintes suposições:

- a normalidade da variável;
- a homoscedasticidade (homogeneidade) de variância;
- a independência de erros.

A pressuposição de Fisher exige que, para estimar as semivariâncias, a variável a ser analisada deve ter uma distribuição aproximada normal.

Neste trabalho a normalidade da variável foi testada através de programas estatísticos, por meio de testes específicos, tais como: teste W de Shapiro & Wilk e o teste KS proposto por Kolmogorov-Smirnov. A homoscedasticidade de variância pode ser testada através do teste F máximo de Hartley.

3.4.1 Teste de Normalidade KS

Este teste de aderência avalia a concordância entre a distribuição observada da amostra e uma determinada distribuição teórica, verificando assim se os dados amostrais aproximam-se razoavelmente de uma determinada distribuição. Para isso, utiliza-se a função distribuição acumulada observada, a qual é comparada com uma teórica. Verifica-se, portanto, se a discrepância que ocorre é aleatória ou não. Para que se faça este teste, os dados devem seguir uma escala ordinal.

De acordo com Kreysig (1970), considerando-se uma distribuição teórica acumulada $F_0(x)$ e uma distribuição observada $F_1(x)$ em uma amostra de n observações, a normalidade é testada através dos seguintes passos:

Determina-se então o maior valor das diferenças (desvio máximo) entre $F_0(x)$ e $F_1(x)$ utilizando a expressão:

$$d = \max | F_0(x) - F_1(x) |$$

3.3

Em que: d é o desvio entre $F_0(x)$ e $F_1(x)$, sendo $F(x)$ a função de distribuição da população e $F_1(x)$ a função de distribuição da amostra.

Posteriormente compara-se este valor com o valor crítico obtido de tabelas. Se o valor de d obtido for maior do que o tabelado o resultado será significativo e a hipótese é rejeitada.

As vantagens do teste de Kolmogorov-Smirnov em relação a outros testes é que ele pode ser aplicado para pequenas amostras, e trata os dados individualmente, sem perder informações. Na maioria dos casos, é mais poderoso Kreysig (1970).

3.4.2 Teste de Shapiro-Wilk

O método de Shapiro-Wilk (1965), tem como resultado o valor favorecido pela estatística W , podendo variar de 0 a 1, e a probabilidade (p -valor) que descreve a hipótese da normalidade também varia de 0 a 1.

Segundo Cordeiro e Lima Neto (2004), a estatística W do teste de Shapiro-Wilk é dada por:

$$W = \frac{\left\{ \sum_{i=1}^n a_i z(i) \right\}^2}{\left\{ \sum_{i=1}^n (z_i - \bar{z})^2 \right\}} \quad 3.4$$

Em que: $z(1) \leq z(2) \leq \dots \leq z(n)$ são os dados ordenados, em ordem crescente, e os a_i 's são constantes tabuladas com níveis de significância para W .

O procedimento para aplicação do teste de normalidade Shapiro-Wilk (1965), segue os seguintes passos:

- i) Ordenam-se as observações amostrais (n) em ordem crescente;

ii) Calcula-se a soma do quadrado do erro associado à amostra

$$\left\{ \sum_{i=1}^n (z_i - \bar{z})^2 \right\};$$

iii) Calcula-se o valor de k, quando n é par, ou seja, $k = 2n$, quando n é

ímpar, $k = 2n + 1$, em seguida encontra-se o valor de: $\left\{ \sum_{i=1}^k a_i z(i) \right\}$;

Os valores de a_i 's, são obtidos em tabelas especiais. Quando o número de observações é ímpar, o valor da mediana não entra nos cálculos;

iv) Calcula-se a estatística W e compara-se o valor encontrado com o valor obtido em tabelas, caso $W < W_{\text{tab}}$; o teste rejeita a hipótese de nulidade, indicando a normalidade das observações.

O teste de normalidade de Shapiro-Wilk (1965) é comumente empregado para pequenas amostras.

3.4.3 Suficiência Amostral

O processo de amostragem utilizado foi o inteiramente aleatório, usando a seguinte fórmula (Meunier et al., 2002):

$$n = \frac{t_{\alpha/2}^2 \cdot CV^2}{E\%^2} \quad 3.5$$

em que: n = tamanho da amostra representativa da população, $t_{\alpha/2}$ = valor tabelado do teste de t, CV = coeficiente de variação, E% = erro de amostragem, considerou-se 10%.

O erro de amostragem, em porcentagem (EA%) pode ser calculado através da fórmula (Silva et al., 1997):

$$EA\% = \left(\frac{t_{\alpha/2} \cdot s_{\bar{X}}}{\bar{X}} \right) 100 \quad 3.6$$

em que: EA% = erro de amostragem, $t_{\alpha/2}$ = valor tabelado do teste de t, $s_{\bar{x}}$ = erro padrão da média, \bar{X} = média.

3.4.4 Análise de Tendências

A existência de fatores externos que possam influenciar o comportamento de uma variável mais em uma direção do que em outra é verificada na análise de tendências.

A existência de tendências é caracterizada quando o coeficiente de correlação (r), entre a variável e os eixos direcionais X e Y for superior a 70% (Landim, 1998).

As relações entre duas variáveis podem ser representadas num diagrama de dispersão, em que cada par de valores fornecerá um ponto, e utilizando-se o método de regressão linear simples, pode-se calcular a equação da reta que melhor se ajuste à nuvem de distribuição.

Algumas vezes tem-se necessidade de testar dependências cumulativas de uma variável dependente em relação a diversas variáveis independentes, tal modelo, exige então uma análise de regressão linear múltipla.

Na regressão múltipla, cada variável é isolada e mantida constante enquanto as outras restantes variam sistematicamente, sendo observados os seus efeitos sobre a variável dependente.

O processo segue por etapas até que o comportamento de todas as variáveis independentes em relação à variável dependente seja verificado. Valores do coeficiente de determinação r^2 forneceram uma medida dimensional de quantidade do ajuste do modelo de regressão múltipla aos dados. Quanto mais próximo da unidade, mais significativo o coeficiente de determinação (Landim, 1998).

3.4.5 Variáveis Regionalizadas e Função Aleatória

Em seus trabalhos com dados de mineração, Daniel Krige observou que a variância dos dados possuía uma estruturação que dependia da distância de amostragem (Vieira, 2000).

A partir desta constatação foram formulados os fundamentos teóricos da geoestatística nos trabalhos desenvolvidos por Matheron (1971), com a Teoria das Variáveis Regionalizadas.

A aplicabilidade da análise espacial de dados utilizando a geoestatística não se restringe apenas a áreas de mineração, geologia ou ciência do solo, mas, com a facilidade computacional que permite a resolução de cálculos trabalhosos, a geoestatística está bastante difundida em vários ramos de conhecimento envolvendo não apenas áreas de ciências exatas, como também áreas das ciências humanas e biológicas.

Por definição, uma variável aleatória $Z(x)$ é aquela que pode assumir uma certa quantidade de valores possíveis associados a uma determinada lei de probabilidade.

De acordo com Ribeiro Junior (1995), o conjunto de valores $z(x_i)$ para todos os pontos x_i de uma determinada região é considerado uma realização particular do conjunto das variáveis aleatórias $\{ Z(x_i) \dots \}$. Uma função aleatória $Z(x)$ é um conjunto infinito de variáveis aleatórias.

A teoria das variáveis regionalizadas pressupõe que uma variável aleatória pode ser expressa pela soma de três componentes: uma componente estrutural, associada a um valor médio constante ou a uma tendência constante; uma componente aleatória, espacialmente correlacionada e um ruído aleatório ou erro residual (Burrough, 1987).

Pode-se, então, escrever:

$$Z(x) = m(x) + w(x) + \varepsilon \quad 3.7$$

Em que: $Z(x)$ é uma função dos valores da variável resposta nas posições (x_1, x_2, \dots, x_i) , $m(x)$ é uma função determinística que descreve a componente estrutural de Z em x ; $w(x)$ é um termo estocástico, que varia localmente e depende espacialmente de $m(x)$; ε é um ruído aleatório não correlacionado, com distribuição normal com média zero e variância σ^2 .

3.4.6 Momentos de uma Função Aleatória

No estudo da geoestatística, em que as estimativas de pontos não amostrados são obtidas por meio de combinação linear, é suficiente a estimativa dos dois primeiros momentos da distribuição de $Z(x)$.

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

$$E[Z(x)] = m(x) \quad 3.8$$

Além deste, existem três momentos de segunda ordem considerados em geoestatística, conforme destacado por Braga (1990), os quais são apresentados como:

a) variância

A variância de $Z(x)$, σ^2 quando existe, pode ser definida como o momento de segunda em relação à média $m(x)$.

$$\sigma^2 = \text{Var}[Z(x)] = E\{[Z(x) - m(x)]^2\} \quad 3.9$$

b) covariância

Sejam duas variáveis aleatórias $Z(x_i)$ e $Z(x_j)$, nos pontos x_i e x_j e, conseqüentemente possuem variância finita. A expressão da covariância é dada por:

$$C(x_i, x_j) = E\{[Z(x_i) - m(x_i)][Z(x_j) - m(x_j)]\} \quad 3.10$$

A covariância deve ser definida positiva para assegurar a existência de todas as distribuições acumuladas.

c) semivariograma $\gamma(x_i, x_j)$

Pode-se definir semivariograma pela expressão:

$$\gamma(x_i, x_j) = \frac{1}{2} E \left[[Z(x_i) - Z(x_j)]^2 \right] \quad 3.11$$

Tanto a variância quanto o semivariograma são sempre positivos, enquanto que a covariância pode apresentar valores negativos.

3.4.7 Hipóteses de estacionaridade

Uma variável é estacionária se ocorrer de maneira aproximadamente homogênea, com oscilações aleatórias contínuas em torno de um valor médio. A hipótese de estacionaridade permite a repetição de um experimento mesmo que as amostras sejam observadas em pontos diferentes em relação ao experimento inicial. Com isto, as hipóteses de estacionaridade asseguram que o primeiro e o segundo momento são constantes.

a) Estacionaridade restrita

Uma função aleatória é estritamente estacionária se sua função de distribuição de probabilidade, dada por:

$$F_{Z(x_1), Z(x_2), \dots, Z(x_k)}(Z_1, Z_2, \dots, Z_k) = \Pr [Z(X_1) \leq Z_1, Z(X_2) \leq Z_2, \dots, Z(X_k) \leq Z_k] \quad 3.12$$

É invariante em relação a qualquer translação do vetor \mathbf{h} (distâncias entre as observações), ou então, se a função de distribuição do vetor aleatório:

$$[Z(x_1), Z(x_2), \dots, Z(x_k)] \quad 3.13$$

é idêntica à do vetor:

$$[Z(x_1 + h), Z(x_2 + h), \dots, Z(x_k + h)] \quad 3.14$$

para qualquer \mathbf{h} .

Embora a geoestatística empírica baseie-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.

b) Estacionaridade de primeira ordem

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

$$E [Z(x_1)] = E [Z(x_2)] = \dots = E [Z(x_i)] = E [Z(x)] = m \quad 3.15$$

c) Estacionaridade de segunda ordem

Uma hipótese de estacionaridade menos restritiva, ou seja, uma hipótese fraca de estacionaridade, utiliza o segundo momento.

Uma função aleatória $Z(x)$ é estacionária de segunda ordem se, além de atender a estacionaridade de primeira ordem, apresentar a característica de que, para cada par de valores $z(x)$ e $z(x+h)$, a covariância existe e depende apenas da distância h (Journel & Huijbregts, 1978). As equações relacionadas a esta hipótese, especificamente a esperança matemática, covariância e variância são:

$$E [Z(x)] = m, \quad \forall x \quad 3.16$$

$$C(x+h, x) = C(h) = E [Z(x+h)Z(x)] - m^2 \quad 3.17$$

$$C(h) = E [Z(x) - m][Z(x+h) - m] \quad 3.18$$

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:

$$\text{Var} [Z(x)] = E [Z(x) - m]^2 = C(0) \quad 3.19$$

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

De acordo com esta hipótese, a função variograma $2\gamma(x, h)$ depende somente do vetor de separação h (módulo e direção), e não da localização de x . Então, é possível estimar o variograma $2\gamma(h)$ com os dados disponíveis. 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:

i) o incremento entre dois pontos tem esperança nula

$$E[z(x+h) - z(x)] = 0 \quad 3.20$$

ii) para qualquer vetor h , o incremento tem uma variância finita, independente da posição x .

$$\text{Var}[z(x+h) - z(x)] = E[z(x+h) - z(x)]^2 = 2\gamma(h) \quad 3.21$$

Sendo $[2\gamma(h)]$ a definição de variograma.

A hipótese intrínseca é mais freqüentemente utilizada na geoestatística por ser mais abrangente; portanto, o semivariograma é mais empregado porque exige apenas a hipótese intrínseca, enquanto o autocorrelograma exige a estacionaridade de segunda ordem.

3.4.8 Semivariograma

O semivariograma é uma ferramenta fundamental na metodologia geoestatística, caracterizando as estruturas e correlações espaciais da variável estudada.

Suponha dois valores numéricos $z(x)$ e $z(x+h)$ para dois pontos (x) e $(x+h)$ separados pelo vetor h . A variabilidade entre estas duas quantidades é caracterizada pela função variograma $2\gamma(x,h)$, a qual é definida como a esperança da variável aleatória $[Z(x) - Z(x+h)]^2$, (Journel & Huijbregts, 1978).

A função semivariograma é por construção simétrica $[\gamma(h) = \gamma(-h)]$ e positiva $[\gamma(h) > 0]$. Em geral, $\gamma(h)$ aumenta quando h aumenta, ou seja, quanto mais afastados os pontos (x_i) e (x_i+h) , menos eles estão correlacionados.

A função semivariograma é expressa como:

$$2\gamma^*(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} [z(x) - z(x+h)]^2 \quad 3.22$$

Em que: $N(h)$ é o número de pares experimentais $[z(x_i), z(x_i+h)]$ dos dados separados pelo vetor h .

a) Parâmetros do Semivariograma

Os parâmetros do semivariograma, expressos na Figura 1, são:

- Alcance (ϕ): distância dentro da qual as amostras apresentam-se correlacionadas espacialmente.
- Patamar ($\tau^2 + \sigma^2$): é o valor do semivariograma correspondente a seu alcance (ϕ). Deste ponto em diante, considera-se que não existe mais dependência espacial entre as amostras, porque a variância da diferença entre pares de amostras ($\text{Var}[z(x) - z(x+h)]$) torna-se invariante com a distância.
- Efeito Pepita (τ^2): por definição, $\gamma(0)=0$; entretanto, na prática, à medida que h tende para 0 (zero), $\gamma(h)$ se aproxima de um valor positivo chamado Efeito Pepita.
- Contribuição ou variância estruturada (σ^2): é a diferença entre o patamar ($\tau^2 + \sigma^2$) e o Efeito Pepita (τ^2).

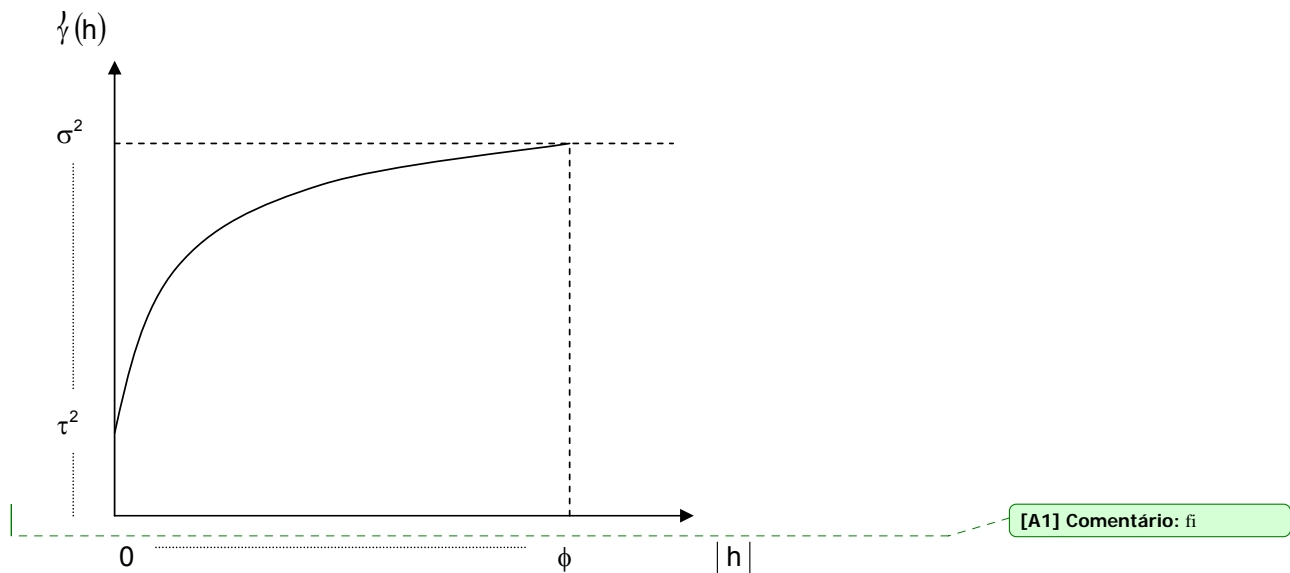


Figura 4 – Representação gráfica dos parâmetros do semivariograma.

b) Continuidade

Na definição do variograma $2\gamma(h)$, a distância “h” representa um vetor de módulo $|h|$ e direção α .

Considerando uma direção α particular e iniciando na origem $\gamma(0) = 0$, o variograma aumenta em geral com o módulo $|h|$.

A descontinuidade na origem, denominada efeito pepita, é proveniente da existência de eventuais erros devido ao processo de coleta de amostras, ao método analítico ou a correlações espaciais existentes em distâncias inferiores à das observações.

O efeito pepita reflete a variabilidade não explicada ou as variações não detectadas pela distância insuficiente da malha de amostragem (Cambardella et al., 1994).

A maneira como o variograma aumenta para pequenos valores de $|h|$ caracteriza o grau de continuidade espacial da variável estudada.

O uso do semivariograma para a análise geoestatística permite verificar a presença ou não da dependência espacial nas propriedades analisadas. Se o semivariograma, ao invés de ser crescente e dependente da distância, for

[A1] Comentário: fi

constante e igual ao patamar para qualquer distância, tem-se a ausência de dependência espacial na escala estudada, denominada de efeito pepita puro (Vieira, 2000). A independência total das amostras determina o efeito aleatório puro.

c) Anisotropia

A anisotropia é uma característica que a variabilidade ou distribuição espacial apresenta e que ocorre mais intensamente em uma direção do que em outra. Se a anisotropia observada possuir um mesmo patamar, mas, diferentes alcances para um mesmo modelo, é denominada isotropia geométrica. A anisotropia é denominada anisotropia zonal, caso a anisotropia observada possua um mesmo alcance, mas, diferentes patamares.

As convenções direcionais usadas na geoestatística para determinação da anisotropia são: 0°, 45°, 90° e 135°. Verificando-se uma similaridade entre eles, o fenômeno é considerado isotrópico, e um único modelo é suficiente para descrever a variabilidade espacial estudada.

d) Modelagem

Após a avaliação das semivariâncias experimentais, deve-se proceder ao ajuste do semivariograma experimental a um modelo teórico existente.

Os principais modelos de semivariogramas são:

i) Modelo esférico

$$\begin{aligned} \gamma(h) &= \tau^2 + \sigma^2 \left[\frac{3h}{2\phi} - \frac{1}{2} \left(\frac{h}{\phi} \right)^3 \right] && \text{para } 0 \leq h \leq \phi \\ \gamma(h) &= \tau^2 + \sigma^2 && \text{para } h > \phi \end{aligned} \quad 3.23$$

Sendo: τ^2 = efeito pepita, σ^2 = semivariância estruturada, $\tau^2 + \sigma^2$ = patamar, ϕ = alcance, h = distância.

Neste modelo, bastante empregado na modelagem geoestatística, observa-se à similaridade de valores mais próximos e a dissimilaridade à medida que aumenta a separação das amostras. A distância “ ϕ ” é o alcance corresponde à distância de influência de cada amostra, e tende a estabilizar próximo do patamar, indicando a independência das amostras.

ii) Modelo gaussiano

Este modelo pode ser expresso por:

$$\begin{aligned} \gamma(h) &= \tau^2 + \sigma^2 \left[1 - \exp\left(-\frac{h^2}{\phi^2}\right) \right] && \text{para } 0 \leq h \leq \phi \\ \gamma(h) &= \tau^2 + \sigma^2 && \text{para } h > \phi \end{aligned} \quad 3.24$$

Em que: τ^2 , σ^2 , $\tau^2 + \sigma^2$, ϕ , h têm o mesmo significado como no modelo anterior.

Neste modelo, a curva é parabólica junto à origem e indica pequena variabilidade para curtas distâncias. Trata-se de um modelo semivariográfico mais regular.

iii) Modelo exponencial

O modelo exponencial pode ser representado pela seguinte função:

$$\begin{aligned} \gamma(h) &= \tau^2 + \sigma^2 \left[1 - \exp\left(-\frac{h}{\phi}\right) \right] && \text{para } 0 \leq h \leq \phi \\ \gamma(h) &= \tau^2 + \sigma^2 && \text{para } h > \phi \end{aligned} \quad 3.25$$

Sendo que: τ^2 , σ^2 , $\tau^2 + \sigma^2$, ϕ , h têm o mesmo significado como nos modelos anteriores.

Este modelo é de fácil aplicação e o patamar só é alcançado quando “ h ” tende a infinito, sendo uma assíntota da equação de $\gamma(h)$.

Para alguns modelos o alcance (ϕ) é igual ao alcance efetivo como é o caso dos modelos esférico. Para o modelo gaussiano o alcance efetivo é igual a “3 ϕ ”, enquanto para o modelo exponencial o alcance efetivo é igual a “ 1,7 ϕ ” (Guerra, 1988).

e) Dependência espacial

Se os valores próximos não são independentes entre si, pode-se utilizar o processo de krigagem para interpolar e estimar valores em lugares não medidos (Isaaks & Srivastava, 1989).

A análise estrutural foi obtida através da análise e do ajuste do semivariograma experimental a um modelo teórico.

O uso da porcentagem (%) da semivariância do Efeito Pepita foi proposto por Trangmar et al. (1985) para determinar a dependência espacial.

Cambardella et al. (1994) propuseram o Índice de Dependência Espacial (IDE) que é dado por:

$$\text{IDE} (\%) = \frac{\tau^2}{\tau^2 + \sigma^2} 100 \quad 3.26$$

Sendo:

IDE \leq 25% - forte dependência espacial

25% < IDE < 75% - moderada dependência espacial

75% \leq IDE < 100% - fraca dependência espacial

IDE = 100% - variável independente

3.4.9 Validação Cruzada

Há vários métodos para verificação da qualidade do ajuste do modelo de semivariograma aos dados experimentais.

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 (Ribeiro Junior, 1995).

Este trabalho utilizou a metodologia dos mínimos quadrados para ajustar o modelo teórico ao semivariograma experimental, o critério para a escolha do melhor modelo foi quanto menor o valor obtido melhor o ajuste do modelo.

O processo de validação cruzada consiste dos seguintes passos:

1. Retirar do conjunto de dados um elemento qualquer.
2. Calcular através da krigagem o valor estimado do ponto retirado empregando os dados restantes.
3. Calcular o erro de estimação. O erro de estimação é calculado pela expressão:

$$\text{Erro} = \frac{(Z(x_i) - \hat{Z}(x_i))}{\sigma_i} \quad 3.27$$

Em que: $Z(x_i)$ é o valor suprimido, $\hat{Z}(x_i)$ é o valor estimado e σ_i é o desvio.

4. Repetir os passos anteriores para todos os outros elementos.
5. Análise dos erros de estimação.

Os erros de estimação requerem que as condições de não tendência e de variância mínima sejam satisfeitas. As expressões que denotam tais condições são:

$$\frac{1}{n} \sum_{i=1}^n (Z_i - \hat{Z}_i) = 0 \quad 3.28$$

$$\frac{1}{n} \sum_{i=1}^n \left(\frac{Z_i - \hat{Z}_i}{\sigma_i} \right)^2 = 1 \quad 3.29$$

A validação cruzada é uma técnica eficiente na avaliação da qualidade das estimativas da krigagem e também na escolha de modelos de semivariogramas.

3.4.10 Método de Estimativa pela Máxima Verossimilhança

A predição e avaliação de tendências no estudo de modelos lineares necessitam o conhecimento dos componentes de variâncias e da covariância das observações. Os avanços da computação e o desenvolvimento de algoritmos iterativos na estimativa de parâmetros de modelos com dependência espacial permitem solucionar tais questões através do estimador de máxima verossimilhança.

O método de máxima verossimilhança, proposto por Hartley & Rao (1967), é um método iterativo o qual consiste em maximizar a função de densidade de probabilidades, em relação aos efeitos fixos e aos componentes de variâncias dos efeitos aleatórios do modelo, fornecendo estimativa “mais verossímil” de parâmetros.

Um conjunto de observações (z_1, z_2, \dots, z_n) geradas a partir de um modelo linear gaussiano pode ser explicada pela soma de três componentes: uma componente determinística, uma componente estocástica e um erro aleatório como já apresentado na equação 2.8.

Diggle & Ribeiro Junior (2000) afirmam que ao utilizar a estimação de parâmetros de modelos baseados em verossimilhança deve-se considerar a pressuposição gaussiana, e apresentaram um modelo de regressão linear gaussiano com tendência linear, em que a equação destacada 2.8 é reescrita como:

$$Z(x_i) = \mu(x_i) + S(x_i) + \varepsilon_i \quad : \quad i = 1, \dots, n \quad 3.30$$

Em que: $Z(x_i)$ é a variável resposta; $\mu(x)$ é uma função determinística que descreve a componente estrutural de Z em x ; $S(x_i)$ é a componente estocástica que explica o comportamento espacial autocorrelacionada da variável Z e segue um processo gaussiano com variância estruturada σ^2 e função de correlação $\rho(h) = \text{Corr}\{S(x), S(x-h)\}$ e $h_{ij} = \|x_i - x_j\|$ é a distância euclidiana entre dois pontos; ε é o erro aleatório não correlacionado espacialmente, mutuamente independente e identicamente distribuídos, seguindo uma distribuição normal com média zero e variância τ^2 , ou seja, $\varepsilon \sim N(0, \tau^2)$.

Uma formulação equivalente para o modelo gaussiano é expressa por:

$$z(x_i) = \mu + S(x_i) + \varepsilon_i \quad : \quad i = 1, \dots, n \quad 3.31$$

Em que: $z(x_i)$ é o valor observado no ponto x_i ; μ é uma constante determinística na condição da hipótese intrínseca; $S(x_i)$ é a componente estocástica e ε é o erro aleatório.

Por conveniência, escreve-se o modelo denotado na equação 2.31 em forma matricial. Portanto, a componente determinística $\mu(x)$ é representada pelo modelo linear:

$$\mu(x_i) = \sum_{j=1}^k f_j(x_i) \beta_j \quad 3.32$$

Em que: \mathbf{X} é a matriz de valores das funções f_k . O conjunto de k funções (z_1, z_2, \dots, z_k) são variáveis referenciadas espacialmente que descrevem deterministicamente a variável Z a partir das coordenadas espaciais (x_1, x_2, \dots, x_k); β é o vetor de parâmetros do modelo linear.

Na definição da matriz de covariância dos dados observados (z_1, z_2, \dots, z_n), utilizando-se o modelo de dependência espacial com o vetor de parâmetros $\theta = (\tau^2, \sigma^2, \phi)$, tem-se a covariância matricialmente descrita por:

$$\text{Var}(Z) = K = \sigma^2 \Sigma + \tau^2 I \quad 3.33$$

Em que: σ^2 é a variância estruturada; Σ é a matriz de covariâncias baseada no modelo de dependência espacial; τ^2 é a variância dos erros aleatórios (efeito pepita) e I é a matriz identidade.

Segundo Diggle & Ribeiro Junior (2000), na estimação por máxima verossimilhança assume-se que os dados seguem uma distribuição multivariada gaussiana na qual a função de densidade de probabilidades da variável resposta é dada por:

$$f(\underline{z}) = \frac{1}{(2\pi)^{n/2} |\mathbf{K}|^{1/2}} e^{\left[-\frac{1}{2} (\underline{z} - \underline{u})' \mathbf{K}^{-1} (\underline{z} - \underline{u}) \right]} \quad 3.34$$

Em que: \underline{z} vetor da variável resposta; $\underline{\mu}$, vetor de média constante na condição da hipótese intrínseca; \mathbf{K} é igual a $\sigma^2 \Sigma + \tau^2 \mathbf{I}$.

A função de distribuição segue uma distribuição conjunta multivariada normal definida pela equação:

$$\underline{Z} \sim \text{MNV}(\underline{\mu}, \sigma^2 \Sigma + \tau^2 \mathbf{I}) \quad 3.35$$

Em que: $\underline{\mu}$ é a média; $\mathbf{1}$ é o vetor de valores um; σ^2 é a variância estruturada; Σ é a matriz de correlação definida conforme os parâmetros do modelo selecionado; \mathbf{I} é a matriz identidade e τ^2 é a variância do erro aleatório.

Através da função de densidade de probabilidades, pode-se escrever a função de verossimilhança, que é dada por:

$$L(\beta, \theta) = \prod f(\underline{z}) \quad 3.36$$

A maximização da função de verossimilhança é obtida através da logaritmização da função de verossimilhança.

Mardia & Marshall (1984) desenvolveram teoria estatística para a estimação de máxima verossimilhança, para o caso em que a função aleatória $Z(x)$ é gaussiana e de estacionaridade fraca. De acordo com os autores, a função de verossimilhança a ser maximizada é determinada pela equação:

$$\ln(L(\beta, \theta)) = \ln\left(\prod f(\underline{z})\right) = l(\beta, \theta) \quad 3.37$$

Tendo a seguinte forma:

$$l(\beta, \theta) \propto -\frac{1}{2} \left\{ \log |K| + (z - X\beta)^t (K)^{-1} (z - X\beta) \right\} \quad 3.38$$

A maximização da função resulta:

$$l(\beta, \tau^2, \sigma^2, \phi) \propto -\frac{1}{2} \left\{ \log |(\sigma^2 \Sigma + \tau^2 I)| + (z - X\beta)^t (\sigma^2 \Sigma + \tau^2 I)^{-1} (z - X\beta) \right\} \quad 3.39$$

Na prática, a maximização da função de verossimilhança é obtida por processos iterativos nos quais a matriz de covariância é atualizada e invertida a cada iteração.

As estimativas de máxima verossimilhança são obtidas a partir das derivadas parciais do logaritmo da função de verossimilhança, em relação aos parâmetros de interesse. Diferenciando-se parcialmente $l(\beta, \tau^2, \sigma^2, \phi)$ em relação a β e σ^2 e igualando-se as derivadas a zero, tem-se:

$$\hat{\beta} = \left(X^T (\sigma^2 \Sigma + \tau^2 I)^{-1} X \right)^{-1} X^T (\sigma^2 \Sigma + \tau^2 I)^{-1} z \quad 3.40$$

$$\hat{\sigma}^2 = \frac{1}{n} (z - X\hat{\beta})^T (\sigma^2 \Sigma + \tau^2 I)^{-1} (z - X\hat{\beta}) \quad 3.41$$

A suposição da função de distribuição do erro aleatório, $\varepsilon \sim N(0, \tau^2)$, mostra que a estimativa de máxima verossimilhança é equivalente à estimativa por mínimos quadrados.

3.4.11 Critério de Informação de Akaike

Na avaliação do desempenho dos modelos, o Critério de Informação de Akaike - Akaike's Information Criterion (AIC) tem sido bastante utilizado para diferentes estruturas de covariâncias. Akaike (1973) relacionou a discrepância,

medida existente entre o modelo verdadeiro e o modelo aproximado, com a máxima verossimilhança.

O princípio do Critério de Informação de Akaike é selecionar uma combinação de variáveis explanatórias e modelos para a função de correlação que minimize o valor do AIC. É importante observar que, em muitas situações clássicas, tais como regressão linear ou em modelos de séries temporais, o AIC não é uma condição consistente para a seleção de modelos. Ou seja, como o crescente aumento do tamanho das amostras, há uma alta probabilidade de que um modelo selecionado pelo AIC não corresponda ao verdadeiro modelo.

O AIC é definido por:

$$\text{AIC} = -2l + 2p \qquad 3.42$$

Em que: l é o logaritmo da verossimilhança e p é o número de parâmetros do modelo considerado.

O uso do critério de informação de Akaike constitui-se em importante instrumento utilizado para a escolha do melhor modelo para ajuste do semivariograma, eliminando a subjetividade que permeia o processo e influenciando diretamente na melhoria da confiança dos valores interpolados (Vendrusculo, 2003). De acordo com Webster (1985) o melhor ajuste teórico do semivariograma experimental é aquele que apresenta o menor valor de AIC.

3.5 Abordagem Bayesiana

A análise estatística clássica tem como objetivo estimar parâmetros de uma certa população e/ou testar hipóteses sobre eles. Uma população pode ser considerada um conjunto de valores, enquanto que um parâmetro pode ser uma média, uma proporção ou um coeficiente de regressão. Normalmente, uma população não pode ser inteiramente estudada, devido ao seu tamanho, dificuldade de acesso ou de medição. Assim, recorre-se a uma amostra representativa da população e toda informação sobre o parâmetro a ser estimado está contida nesta amostra. Um modelo probabilístico é então estabelecido sobre os dados amostrais, supondo-se conhecidos os parâmetros. Uma função escrita

em termos dos dados amostrais e dos parâmetros do modelo estudado é a Função de Verossimilhança.

O método empregado para a estimação de parâmetros conhecidos como Método da Máxima Verossimilhança consiste em determinar os valores dos parâmetros que maximizam a função de verossimilhança que foi estabelecida de acordo com os dados amostrais observados. Os valores dos parâmetros que tornam mais prováveis os dados coletados são chamados de estimativas de máxima verossimilhança e podem ser associados a intervalos de confiança, expressando a incerteza embutida no processo de estimação.

Por outro lado, na análise bayesiana, nem toda informação sobre o parâmetro vêm da amostra. Apenas uma parte da informação está contida na amostra, e é expressa através da função de verossimilhança. A outra parte reside num conhecimento prévio sobre o parâmetro, denominado a priori e pode ser expressa através de uma distribuição de probabilidade, denominada: “distribuição à priori do parâmetro”.

As duas partes da informação, explicitando a relação com o auxílio do Teorema de Bayes, resultam na distribuição a posteriori do parâmetro. É esta a distribuição utilizada nas inferências sobre o parâmetro. Se existe pouco ou nenhum conhecimento a priori sobre o parâmetro, costuma-se expressar este desconhecimento através da atribuição de uma distribuição a priori, conhecida como “priori vaga”, na qual a variância é grande, deixando um amplo domínio para a variação dos valores do parâmetro de interesse.

A abordagem da estatística clássica de modelos mistos apresenta algumas limitações, principalmente no que diz respeito à construção adequada de intervalos de confiança. Para contornar esse aspecto, a inferência bayesiana pode ser empregada.

A inferência bayesiana consiste em uma abordagem estatística que trata os parâmetros populacionais como variáveis aleatórias, e permite obter intervalos de confiança, denominados de intervalos de credibilidade, utilizando métodos de simulação Monte Carlo. A abordagem bayesiana tem a vantagem de incorporar informações passadas (“a priori”), enriquecendo o processo de inferência.

A análise bayesiana considera a incerteza existente sobre todos os parâmetros do modelo em sua metodologia. Assim, proporciona uma análise

descritiva completa para cada parâmetro, permitindo a construção de intervalos de credibilidade que considerem a incerteza sobre todos os parâmetros do modelo. Quando não se dispõe de informações passadas a serem levadas em conta, a metodologia bayesiana utiliza distribuições prioris não-informativas.

No contexto da inferência bayesiana, todos os parâmetros do modelo estudado, sejam fixos ou aleatórios, são igualmente considerados como variáveis aleatórias, conforme o conceito subjetivo da probabilidade.

Segundo Box & Tiao (1973), na utilização do Teorema de Bayes, informações a priori sobre os parâmetros, utilizadas em associação com os dados amostrais (representados no caso, pela função de verossimilhança), possibilitam uma inferência a posteriori dos parâmetros e a determinação de intervalos de credibilidade mais estreitos do que intervalos de confiança. Caso não existam informações a priori, a inferência bayesiana pode trabalhar com as distribuições a priori não informativas.

Na metodologia bayesiana, na maioria das vezes, as distribuições marginais são muito complexas e de difícil obtenção, mesmo aplicando técnicas usuais de integração. A obtenção da distribuição a posteriori é o grande objeto de interesse da inferência bayesiana. Para solucionar tal problema, utiliza-se comumente processo de simulações mediante iterações, como o “MCMC” (Markov Chain Monte Carlo), o algoritmo de Amostrador de Gibbs (Casella & George, 1993) e o algoritmo Metropolis Hastings (Chib & Greenberg, 1995).

O processo iterativo finaliza quando ocorre a convergência, ou seja, a série alcança um estado de estacionariedade e as distribuições condicionais estão bastante próximas das distribuições marginais.

Segundo Gamerman (1996), a utilização prática do algoritmo de Amostrador de Gibbs pode ser dificultada pela eventual complexidade dos modelos, que torna difícil a caracterização da convergência.

De acordo com Diggle & Lophaven (2004), a representação simbólica de um modelo linear clássico é expressa por:

$$[S, Y] = [S][Y | S] \quad 3.43$$

Em que tal modelo especifica a distribuição marginal da variável aleatória não observada $[S]$ e a distribuição condicional de $[Y|S]$, com o parâmetro desconhecido θ , mas, de valor fixado.

O paradigma bayesiano para modelos de parâmetros desconhecidos são também tratados como variáveis aleatórias e representados por:

$$[S, Y, \theta] = [S|\theta][Y|S, \theta][\theta] \quad 3.44$$

sendo que $[\theta]$ representa a priori de θ .

A distribuição requerida para a inferência preditiva é $[S|Y]$, a qual é denotada por:

$$[S|Y] = \int [S|Y, \theta][\theta|Y] d\theta \quad 3.45$$

A equação mostra que a distribuição preditiva bayesiana é uma distribuição preditiva $[S|Y, \theta]$ com diferentes valores de θ determinados de acordo com sua probabilidade posteriori.

3.5.1 Estimação Bayesiana

Seja um conjunto de observações (z_1, z_2, \dots, z_k) referenciadas espacialmente, que descrevem deterministicamente a variável Z a partir das coordenadas espaciais (x_1, x_2, \dots, x_k) . O conjunto das variáveis aleatórias é considerado um processo estocástico, descrito da forma:

$$\{Z(x_i): x_i \in D \subset \mathfrak{R}^2 : i=1, \dots, k\} \quad 3.46$$

Em que: $Z(x_i)$ a variável aleatória que varia continuamente em D , D é a região de estudo, x_i é a posição da variável, considerada fixa, \mathfrak{R}^2 é o espaço bidimensional.

Segundo Almeida & Ribeiro Jr. (1996), para cada ponto amostrado e referenciado x , pode-se avaliar uma determinada característica de interesse $z(x)$, realização da variável aleatória Z no ponto x .

A teoria dos modelos desenvolvida por Henderson (1975), baseada em modelos lineares que contêm efeitos fixos e aleatórios indica uma formulação equivalente para a equação 2.31, como um modelo linear gaussiano para quaisquer locações tais que $x_i \in D$, denotada por:

$$z = X\beta + S(x) + \varepsilon \quad 3.47$$

Em que: \mathbf{z} é um vetor de observações, \mathbf{X} é uma matriz de incidência dos efeitos fixos, β é um vetor de efeitos fixos, representa os efeitos das covariáveis \mathbf{X} sobre a média de \mathbf{z} , $\mathbf{S}(\mathbf{x})$ é uma matriz de incidência dos efeitos aleatórios, ε é um vetor de resíduos.

O modelo linear pode ser reescrito como:

$$z = X\beta_i + Yu_i + \varepsilon \quad 3.48$$

Em que: \mathbf{Y} é uma matriz de incidência dos efeitos aleatórios e \mathbf{u} é um vetor de efeitos aleatórios a serem preditos.

Geralmente assume-se que $\beta_i \sim N(0, \sigma_\beta^2)$, onde σ_β^2 é uma quantidade fixa conhecida, $u_i \sim N(0, \Delta\sigma^2)$, em que σ^2 é a variância e Δ é a matriz de vizinhança que descreve os coeficientes das covariâncias entre os efeitos aleatórios. Assume-se também que ε , o erro aleatório, $\varepsilon \sim N(0, I\tau^2)$, em que τ^2 é a variância residual e I é a matriz identidade.

Na especificação da metodologia bayesiana, comumente associa-se as distribuições a priori para os parâmetros $(\beta, \sigma^2, \phi, \tau^2)$. Para o parâmetro ϕ , geralmente associa-se uma priori uniforme, para os parâmetros σ^2 e τ^2 , uma

distribuição qui-quadrada inversa, enquanto o parâmetro β segue uma distribuição normal.

O parâmetro ϕ representa aquele que está de acordo com a função de correlação escolhida para o modelo. Uma priori usualmente escolhida para o parâmetro ϕ é uma distribuição gama com uma variância grande.

Seguindo o paradigma de Bayes (Box & Tiao, 1973), a distribuição posteriori de θ é dada por:

$$\text{distribuição posteriori} \propto \text{verossimilhança} \times \text{distribuição priori} \quad 3.49$$

Assumindo-se que a distribuição condicional dos dados (z) segue uma distribuição normal multivariada, em que cada componente tem a forma:

$$(z | \beta, u, \sigma^2) \sim N(X\beta + Yu, I\sigma^2) \quad 3.50$$

Então, a função de verossimilhança é representada por:

$$L(z | \beta, u, \sigma^2, \tau^2) \propto \left(\frac{1}{\sigma_o^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_o^2}(z - X\beta - Yu)^T(z - X\beta - Yu)\right\} \quad 3.51$$

Em que: σ_o^2 é a matriz de variâncias das n observações.

Segundo Sorensen (1996), na definição das distribuições a priori, admitindo-se que não haja informações prévias a respeito dos efeitos fixos, pode-se definir uma priori do tipo função constante.

Para o parâmetro β assume-se que $P(\beta) \sim \text{constante}$.

Para a variância σ^2 , pode-se considerar a priori uma distribuição qui-quadrado inversa dada por:

$$P(\sigma^2) \propto \left(\frac{1}{\sigma_u^2}\right)^{\left(\frac{\nu_u}{2} + 1\right)} \exp\left[-\frac{\nu_u S_u^2}{2\sigma_u^2}\right] \quad 3.52$$

Para a variância τ^2 , também pode-se considerar uma distribuição qui-quadrado inversa dada por:

$$P(\tau^2) \propto \left(\frac{1}{\sigma_\varepsilon^2}\right)^{\left(\frac{v_\varepsilon}{2}+1\right)} \exp\left[-\frac{v_\varepsilon s_\varepsilon^2}{2\sigma_\varepsilon^2}\right] \quad 3.53$$

Supondo-se $v_u = v_\varepsilon = -2$ e também que $s_u^2 = s_\varepsilon^2 = 0$ tem-se para as variâncias σ^2 e τ^2 funções constantes caracterizando prioris não-informativas dadas por:

Para a variância σ^2

$$\sigma^2 \propto \frac{1}{\sigma_u^2} \quad 3.54$$

Para a variância τ^2

$$\tau^2 \propto \frac{1}{\sigma_\varepsilon^2} \quad 3.55$$

Para os efeitos aleatórios, considera-se uma distribuição normal como priori dada por:

$$P(u) \propto \left(\frac{1}{\sigma_u^2}\right)^{\left(\frac{q}{2}\right)} \exp\left(-\frac{1}{2\sigma_u^2} (u' \Delta^{-1} u)\right) \quad 3.56$$

Em que: σ_u^2 é a variância dos efeitos aleatórios e Δ é a matriz de vizinhança que possui elementos que descrevem os coeficientes das covariâncias entre os efeitos aleatórios.

Aplicando o teorema de Bayes, o produto da função de verossimilhança pelas distribuições prioris de cada parâmetro simultaneamente resulta na distribuição posteriori conjunta, dada por:

$$P(u, \beta, \sigma^2, \tau^2 | z) \propto L(z | \beta, u, \sigma^2, \tau^2) P(\beta) P(\sigma^2) P(\tau^2) P(u) \quad 3.57$$

Substituindo tem-se:

$$P(u, \beta, \sigma^2, \tau^2 | z) \propto \left(\frac{1}{\sigma_\varepsilon^2}\right)^{\left(\frac{v_\varepsilon + n + 2}{2}\right)} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} [(z - X\beta - Yu)^t (z - X\beta - Yu) + v_\varepsilon s_\varepsilon^2]\right\}$$

$$\left(\frac{1}{\sigma_u^2}\right)^{\left(\frac{v_u + q + 2}{2}\right)} \exp\left\{-\frac{1}{2\sigma_u^2} [(u^t \Delta^{-1} u) + v_u s_u^2]\right\} \quad 3.58$$

A partir da distribuição posteriori conjunta, podem ser obtidas as distribuições condicionais para cada parâmetro: σ^2 , τ^2 , β e u .

Para o parâmetro σ^2 , tem-se:

$$P(\sigma^2 | u, \beta, \tau^2, z) \propto \left(\frac{1}{\sigma_u^2}\right)^{\left(\frac{v_u + q + 2}{2}\right)} \exp\left\{-\frac{1}{2\sigma_u^2} [(u^t \Delta^{-1} u) + v_u s_u^2]\right\} \quad 3.59$$

Com isso, a variância de σ^2 segue uma distribuição gama inversa, dada por:

$$(\sigma^2 | u, \beta, \tau^2, z) \sim GI\left(\frac{v_u + q}{2}; \frac{u^t \Delta^{-1} u + v_u s_u^2}{2}\right) \quad 3.60$$

Para o parâmetro τ^2 , tem-se:

$$P(\tau^2 | u, \beta, \sigma^2, z) \propto \left(\frac{1}{\sigma_\varepsilon^2}\right)^{\left(\frac{v_\varepsilon + n + 2}{2}\right)} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} [(z - X\beta - Yu)^t (z - X\beta - Yu) + v_\varepsilon s_\varepsilon^2]\right\}$$

3.61

A variância de τ^2 também segue uma distribuição gama inversa, dada por:

$$(\tau^2 | u, \beta, \sigma^2, z) \sim \text{GI} \left(\frac{\nu_\varepsilon + n}{2}, \frac{(z - X\beta - Yu)^t (z - X\beta - Yu) + \nu_\varepsilon s_\varepsilon^2}{2} \right) \quad 3.62$$

Para o parâmetro β , retornando-se à função de verossimilhança tem-se:

$$P(u, \beta, \sigma^2, \tau^2 / z) \propto \left(\frac{1}{\sigma_\varepsilon^2} \right)^{\left(\frac{\nu_\varepsilon + n + 2}{2} \right)} \exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} \left[(z - X\beta - Yu)^t (z - X\beta - Yu) + \nu_\varepsilon s_\varepsilon^2 \right] \right\}$$

$$\left(\frac{1}{\sigma_u^2} \right)^{\left(\frac{\nu_u + q + 2}{2} \right)} \exp \left\{ -\frac{1}{2\sigma_u^2} \left[(u^t \Delta^{-1} u) + \nu_u s_u^2 \right] \right\} \quad 3.63$$

Fazendo-se:

$$D = (u^t \Delta^{-1} u) + \nu_u s_u^2 \quad 3.64$$

e denotando-se B como:

$$B = (z - X\beta - Yu)^t (z - X\beta - Yu) \quad 3.65$$

$$B = \left[(z - Yu)^t - (X\beta)^t \right] \left[(z - Yu) - (X\beta) \right] \quad 3.66$$

$$B = (z - Yu)^t (z - Yu) - (z - Yu)^t (X\beta) - (X\beta)^t (z - Yu) + (X\beta)^t (X\beta) \quad 3.67$$

Definindo C como:

$$C = -(z - Yu)^t (X\beta) - (X\beta)^t (z - Yu) + (X\beta)^t (X\beta) \quad 3.68$$

$$C = -(z - Yu)^t X\beta - \beta^t X^t (z - Yu) + X^t \beta^t (X\beta) \quad 3.69$$

$$C = \left[\beta - (X^t X)^{-1} X^t (z - Y u) \right]^t (X^t X) \left[\beta - (X^t X)^{-1} X^t (z - Y u) \right] - (z - Y u)^t X (X^t X)^t X^t (z - Y u)$$

3.70

Retornando para B tem-se:

$$B = \left[(z - Y u)^t (z - Y u) \right] + C$$

3.71

Portanto para B tem-se:

$$B = \left[(z - Y u)^t (z - Y u) \right] - \left[(z - Y u)^t X (X^t X)^{-1} X^t (z - Y u) \right] +$$

$$\left[\beta - (X^t X)^t X^t (z - Y u) \right]^t (X^t X)^t \left[\beta - (X^t X)^t X^t (z - Y u) \right]$$

3.72

$$B = \left[(z - Y u)^t \left[1 - X (X^t X)^t X^t \right] (z - Y u) \right] +$$

$$\left[\beta - (X^t X)^{-1} X^t (z - Y u) \right]^t (X^t X) \left[\beta - (X^t X)^t X^t (z - Y u) \right]$$

3.73

Com isso, a distribuição posteriori condicional para o parâmetro β é dada por:

$$P(\beta | u, \sigma^2, \tau^2, z) \propto \left(\frac{1}{\sigma_\varepsilon^2} \right)^{\left(\frac{p}{2} \right)}$$

$$\exp \left\{ -\frac{1}{\sigma_\varepsilon^2} \left[\beta - (X^t X)^{-1} X^t (z - Y u) \right]^t \left(\frac{X^t X}{\sigma_\varepsilon^2} \right) \left[\beta - (X^t X)^{-1} X^t (z - Y u) \right] \right\}$$

3.74

Ou seja, o parâmetro β da distribuição condicional posteriori segue uma distribuição normal denotada por:

$$(\beta | u, \sigma^2, \tau^2, z) \sim N \left[(X^t X)^{-1} X^t (z - Y u); (X^t X)^{-1} \sigma_\varepsilon^2 \right] \quad 3.75$$

Retornando-se à função de verossimilhança, equação 3.61, para obter-se o parâmetro u , a função de verossimilhança pode ser reescrita como:

$$P(u, \beta, \sigma^2, \tau^2 / z) \propto \left(\frac{1}{\sigma_\varepsilon^2} \right)^{\left(\frac{n+2}{2} \right)} \exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} [(z - X\beta - Yu)^t (z - X\beta - Yu)] \right\}$$

$$\left(\frac{1}{\sigma_u^2} \right)^{\left(\frac{q+2}{2} \right)} \exp \left\{ -\frac{1}{2\sigma_u^2} [(u^t (\Delta^{-1}K)u)] \right\} \quad 3.76$$

Denotando:

$$K = \tau^2 / \sigma^2 \quad 3.77$$

Ou seja:

$$P(u, \beta, \sigma^2, \tau^2 / z) \propto \left(\frac{1}{\sigma_\varepsilon^2} \right)^{\left(\frac{n+2}{2} \right)} \left(\frac{1}{\sigma_u^2} \right)^{\left(\frac{q+2}{2} \right)}$$

$$\exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} \left[[(z - X\beta - Yu)^t (z - X\beta - Yu)] + [(u^t (\Delta^{-1}K)u)] \right] \right\} \quad 3.78$$

Definindo:

$$E = [(z - X\beta - Yu)^t (z - X\beta - Yu)] + [(u^t (\Delta^{-1}K)u)] \quad 3.79$$

$$E = [(z - X\beta)^t - (Yu)^t] [(z - X\beta) - (Yu)] + [(u^t (\Delta^{-1}K)u)] \quad 3.80$$

$$E = (z - X\beta)^t (z - X\beta) - (z - X\beta)^t Y u - u^t Y^t (z - X\beta) + u^t Y^t Y u + u^t (\Delta^{-1} K) u \quad 3.81$$

Denotando:

$$F = - (z - X\beta)^t Y u - u^t Y^t (z - X\beta) + u^t Y^t Y u + u^t (\Delta^{-1} K) u \quad 3.82$$

$$F = \left[u - (z^t z + \Delta^{-1} K)^{-1} z^t (z - X\beta) \right]^t (z^t z + \Delta^{-1} K) \left[u - (z^t z + \Delta^{-1} K)^{-1} z^t (z - X\beta) \right] - \left[z^t (z - X\beta) \right]^t (z^t z + \Delta^{-1} K)^{-1} \left[z^t (z - X\beta) \right] \quad 3.83$$

Retornando para E tem-se:

$$E = (z - X\beta)^t (z - X\beta) + F \quad 3.84$$

$$E = (z - X\beta)^t (z - X\beta) - \left[z^t (z - X\beta) \right]^t (z^t z + \Delta^{-1} K)^{-1} \left[z^t (z - X\beta) \right] + \left[u - (z^t z + \Delta^{-1} K)^{-1} z^t (z - X\beta) \right]^t (z^t z + \Delta^{-1} K) \left[u - (z^t z + \Delta^{-1} K)^{-1} z^t (z - X\beta) \right] \quad 3.85$$

Finalmente, a distribuição posteriori condicional para o parâmetro u é dada por:

$$P(u | \sigma^2, \tau^2, \beta, z) \propto \left(\frac{1}{\sigma_\varepsilon^2} \right)^{\left(\frac{n}{2} \right)} \exp \left\{ - \frac{1}{2 \sigma_\varepsilon^2} \left[u - (Y^t Y + \Delta^{-1} K)^{-1} Y^t (z - X\beta) \right]^t \left(\frac{Y^t Y + \Delta^{-1} K}{\sigma_\varepsilon^2} \right) \left[u - (Y^t Y + \Delta^{-1} K)^{-1} Y^t (z - X\beta) \right] \right\} \quad 3.86$$

Ou seja, o parâmetro u da distribuição condicional posteriori segue uma distribuição normal denotada por:

$$(u | \sigma^2, \tau^2, \beta, z) \sim N \left\{ (Y^t Y + \Delta^{-1} K)^{-1} Y^t (z - X\beta); (Y^t Y + \Delta^{-1} K)^{-1} \sigma_\varepsilon^2 \right\} \quad 3.87$$

3.6 Softwares Utilizados

Para as análises estatísticas descritivas, geoestatística e abordagem bayesiana foram utilizados os seguintes softwares:

1- SURFER - programa geoestatístico empregado na construção dos semivariogramas experimentais e teóricos e para a construção de mapas da krigagem em duas e três dimensões (Golden Software, 1995).

2- GeoEAS – um dos antigos softwares de geoestatística, sendo de domínio público, foi utilizado por fornecer uma melhor compreensão dos cálculos efetuados (Englund e Sparks, 1998).

3- MINITAB – programa estatístico utilizado na análise estatística descritiva, no teste de normalidade e no cálculo de correlação (Statistical Software, 2002).

4- STATISTICA – pacote estatístico também utilizado na análise estatística descritiva, no teste de normalidade e no cálculo de correlação (Statsoft, 1999).

5- GeoR – ferramenta poderosa na análise de dados espaciais, de domínio público, possui funções para modelagem de dados de áreas e visualizações dos resultados da krigagem bayesiana (Diggle e Ribeiro Jr, 2005).

4. RESULTADOS E DISCUSSÃO

4.1 Análise Estatística Descritiva

A Tabela 1 apresenta a estatística descritiva das medidas de condutividade elétrica aparente, em mS/m, obtidas em campo.

Tabela 1 – Estatísticas Descritivas das medidas de condutividade elétrica aparente.

Estatísticas	CEvert	Cehor
Média	28,19	19,32
Desvio Padrão	13,89	11,67
Variância	193,11	136,58
Coeficiente de Variação - (%)	49,29	60,49
Mediana	27	18
Mínimo	3	1
Máximo	82	57
Assimetria	1,031	0,633
Curtose	2,144	0,208
Teste KS	0,0803 p<0,10	0,0633 p>0,20
Teste W	0,9295 p<0,000	0,9447 p<0,000
Número de Observações	260	260

As medidas de dispersão estudadas foram: desvio padrão, variância, coeficiente de variação, assimetria e curtose, tanto para as leituras de condutividade elétrica aparente medidas nos locais estudados através do modo de operação vertical como para o modo horizontal. Verificou-se uma alta variabilidade espacial (CV = 60,49%) para o modo de operação horizontal e uma média variabilidade espacial (CV = 49,23%) para o modo de operação vertical. Esta heterogeneidade apresentada é própria de solos aluviais como o estudado (Warrick e Nielsen, 1980).

O tamanho da amostra representativa encontrada para condutividade elétrica aparente obtida no modo de operação vertical, nível de significância de 1% e erro de amostragem de 10%, foi 162 observações. Para o modo de operação horizontal, com o mesmo nível de significância e o mesmo erro amostral, o tamanho da amostra representativa encontrado foi 244 leituras. Verificou-se assim, que o tamanho das amostras encontradas tanto para o modo de operação vertical como para o modo de operação horizontal são inferiores ao número de leituras que foram realizadas em campo (260 leituras para cada modo de operação), caracterizando assim a suficiência amostral.

O erro de amostragem encontrado para a condutividade elétrica aparente obtida no modo de operação vertical, ao nível de significância de 1%, foi igual a 7,88%. Para o modo de operação horizontal, o erro de amostragem encontrado ao mesmo nível de significância, foi de 9,66%. Verificou-se assim, que os erros de amostragens encontrados foram inferiores ao valor recomendado para trabalhos de campo, que é igual a 10%.

A normalidade da variável estudada foi testada através dos programas estatísticos Statistica e Minitab, por meio do teste KS proposto por Kolmogorov-Smirnov e do teste W proposto por Shapiro-Wilk. Os valores críticos dos testes KS para 260 observações, com os respectivos níveis de significância, são mostrados na Tabela 2.

Tabela 2 – Valores críticos dos testes KS

Número de Observações	Nível de Significância (%)				
	$\alpha = 20$	$\alpha = 10$	$\alpha = 5$	$\alpha = 2$	$\alpha = 1$
260	0,066	0,076	0,084	0,094	0,101

A Figura 5, obtida através do pacote Statistica, apresenta o histograma de freqüências com curva normal e valores dos testes KS e W para os dados de condutividade elétrica aparente observados no modo vertical.

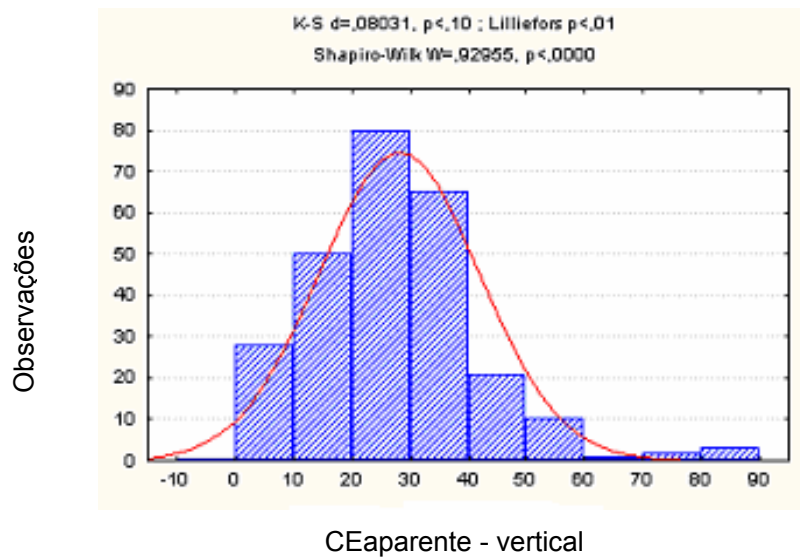


Figura 5 – Histograma com curva normal para CEaparente – vertical.

A Figura 6 apresenta o teste de normalidade de Kolmogorov-Smirnov, teste KS, obtida através do pacote estatístico Minitab, para os dados observados no modo vertical e aplicados ao nível de significância de 1%. Reforça-se a aceitação da hipótese de normalidade dos dados da condutividade elétrica vertical.

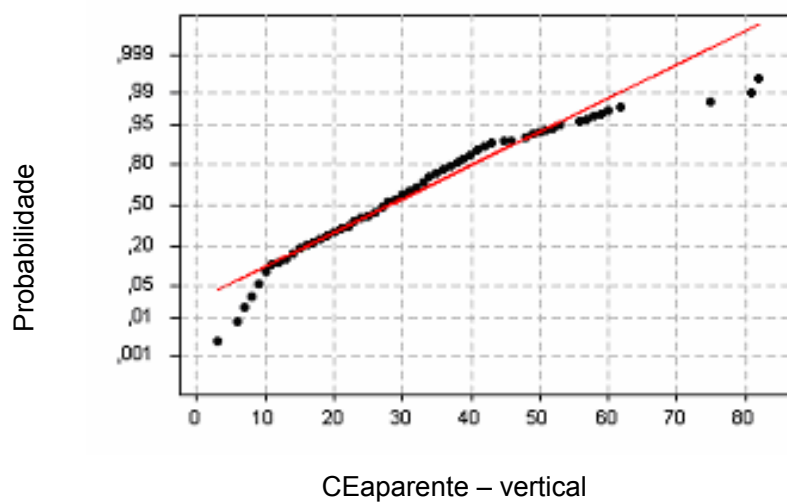


Figura 6 – Teste de normalidade de Kolmogorov-Smirnov para CEaparente – vertical.

Os gráficos dos testes de normalidade da condutividade elétrica aparente, teste KS e teste W, tanto para o modo vertical como para o modo horizontal, mostram a maioria dos valores das observações sobre a reta de tendências com poucos pontos periféricos.

A Figura 7, obtida através do pacote Statistica, permite a visualização da distribuição de freqüência dos atributos e a verificação da normalidade dos dados observados através do modo de operação horizontal.

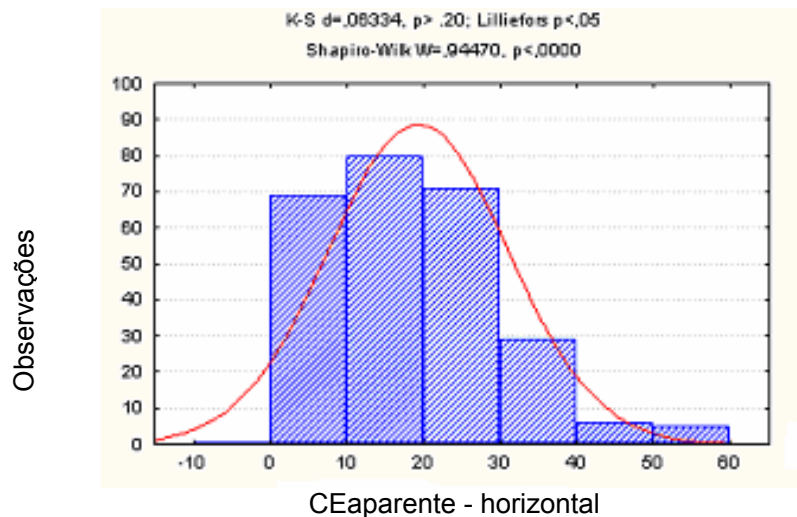


Figura 7 – Histograma com curva normal para CEaparente – horizontal.

O teste de normalidade teste KS proposto por Kolmogorov-Smirnov, apresentado para os dados obtidos no modo de operação horizontal e aplicado ao nível de significância igual a 7,8%, conforme a Figura 8, reforça a aceitação da hipótese de normalidade dos dados.

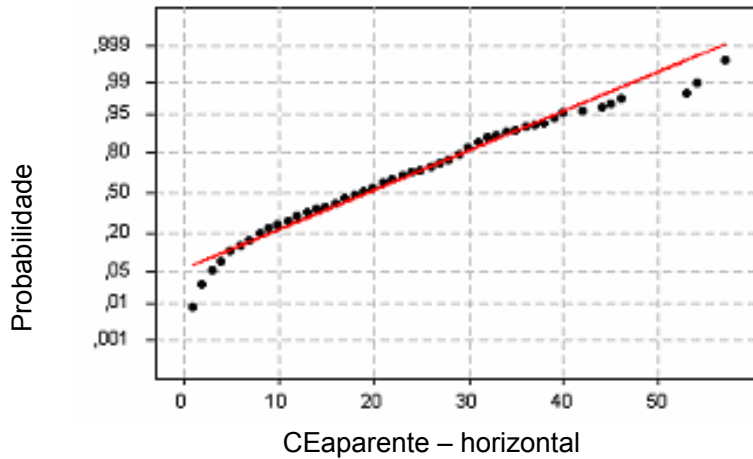


Figura 8 - Teste de normalidade de Kolmogorov-Smirnov para CEaparente – horizontal.

4.2 Análise de Tendências

A média e a variância não são suficientes para medir a continuidade espacial ou avaliar as hipóteses de estacionaridade. Tem-se ainda que avaliar a presença de tendência linear da dependência espacial, a qual indica a não existência de estacionaridade. A dispersão dos dados de condutividade elétrica aparente no modo vertical em relação às coordenadas x e y é mostrada nas Figuras 9 e 10.

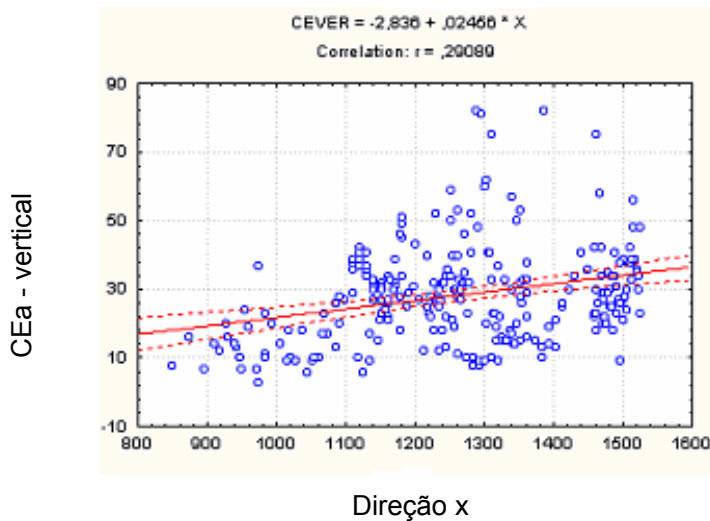


Figura 9 – Dispersão da Condutividade Elétrica na direção x, para a CEaparente-vertical .

Os dados discrepantes observados nas Figuras 9 e 10, não foram retirados do conjunto amostral e foram considerados nas análises descritivas, metodologia geoestatística e abordagem bayesiana.

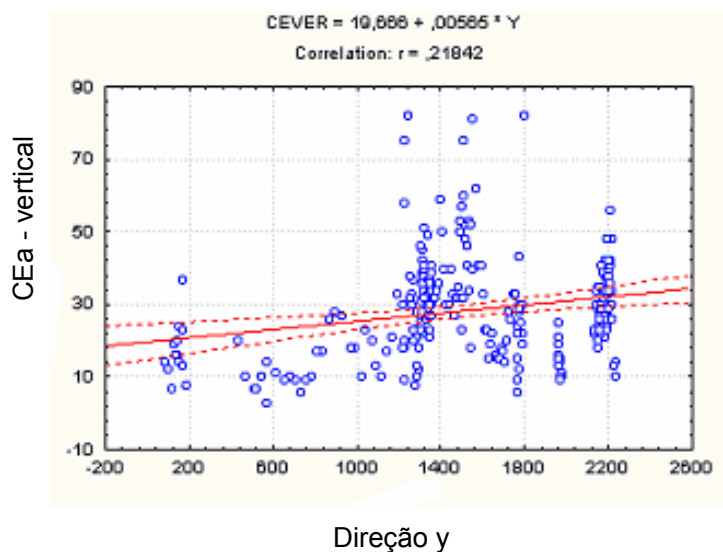


Figura 10 – Dispersão da Condutividade Elétrica na direção y, para a CEaparente-vertical .

Os dados apresentaram-se mais dispersos na direção y do que na direção x. Observou-se ainda a existência de baixas correlações entre a condutividade elétrica aparente no modo vertical e as direções x e y. Os valores para estas correlações foram, respectivamente, $r_x = 0,29$ e $r_y = 0,22$.

Analogamente, as Figuras 11 e 12 ilustram a dispersão dos dados de condutividade elétrica aparente no modo horizontal em relação às direções x e y.

Os dados apresentaram-se mais dispersos na direção y do que na direção x. Observou-se também a existência de baixas correlações entre a condutividade elétrica aparente no modo vertical e as direções x e y. Os valores para estas correlações foram, respectivamente, $r_x = 0,36$ e $r_y = 0,31$.

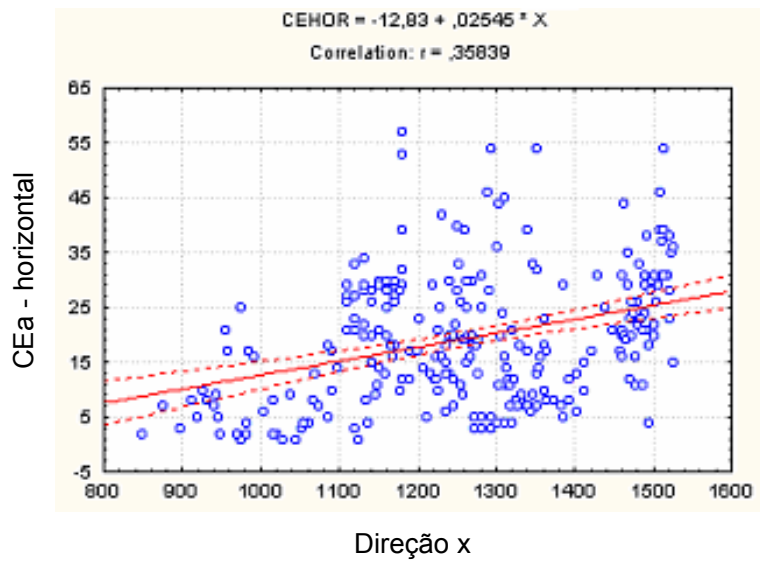


Figura 11 – Dispersão da Condutividade Elétrica na direção x, para a CEaparente-horizantal .

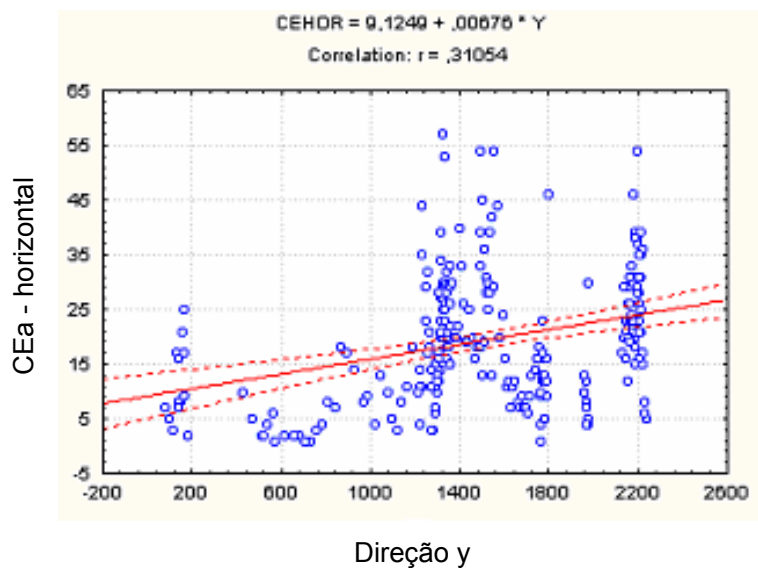


Figura 12 – Dispersão da Condutividade Elétrica na direção y, para a CEaparente-horizantal .

A Tabela 3 apresenta um resumo das correlações obtidas entre as medidas de condutividade elétrica aparente e as coordenadas x e y.

Tabela 3 – Correlação entre a condutividade elétrica aparente e as coordenadas x e y.

Correlação	CEvertical	Cehorizontal
r_x	0,291	0,358
r_y	0,218	0,311

De acordo com Landim (1998), somente os valores de correlação maiores do que 0,70 são relevantes na existência de tendência nos dados.

A verificação da tendência nos dados de condutividade elétrica aparente obtida no modo de operação vertical foi efetuada também através de uma regressão múltipla, expressa por:

$$\hat{C}E_v = -13,0 + 0,0383 X_i - 0,00476 Y_i \quad \text{com } R^2 = 0,092$$

Também foi realizada uma regressão múltipla para a verificação de tendência nos dados de condutividade elétrica aparente obtidos no modo de operação horizontal. A equação obtida é expressa por:

$$\hat{C}E_h = -14,4 + 0,0275 X_i - 0,00072 Y_i \quad \text{com } R^2 = 0,129$$

Os valores encontrados para R^2 são baixos, indicando a inexistência de tendência relevante.

4.3 Análise Geoestatística

Os modelos de semivariogramas experimentais obtidos nos modos vertical e horizontal e ajustados aos modelos matemáticos pelo pacote geoestatístico GeoEAS (England e Sparks, 1990), foram: gaussiano (Figura 13 e Figura 16), exponencial (Figura 14 e Figura 17) e esférico (Figura 15 e Figura 18). Para efeito de correções foi feita uma truncagem da distância máxima. O ajuste dos modelos aos valores estimados de acordo com as distâncias truncadas mostrou-se altamente significativo. O número de pares considerados em cada passo da

construção dos semivariogramas, incremento de distâncias, foi sempre superior ao mínimo recomendado em Journel e Huijbrets (1978), que são 30 pares.

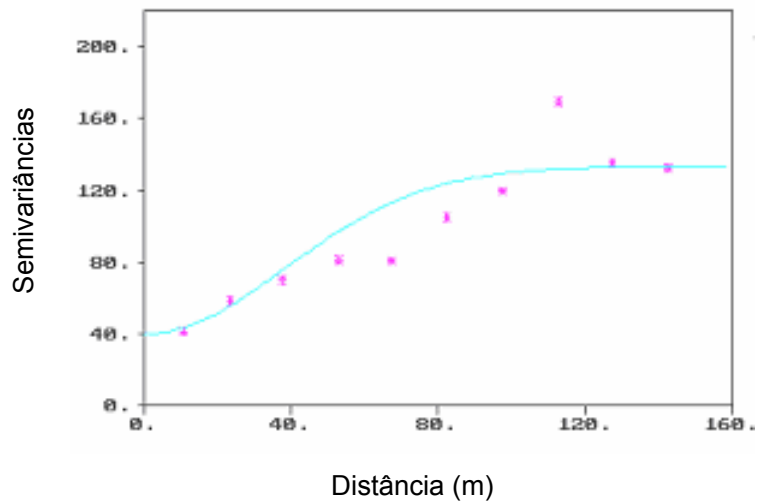


Figura 13 – Semivariograma gaussiano para CEaparente – vertical.

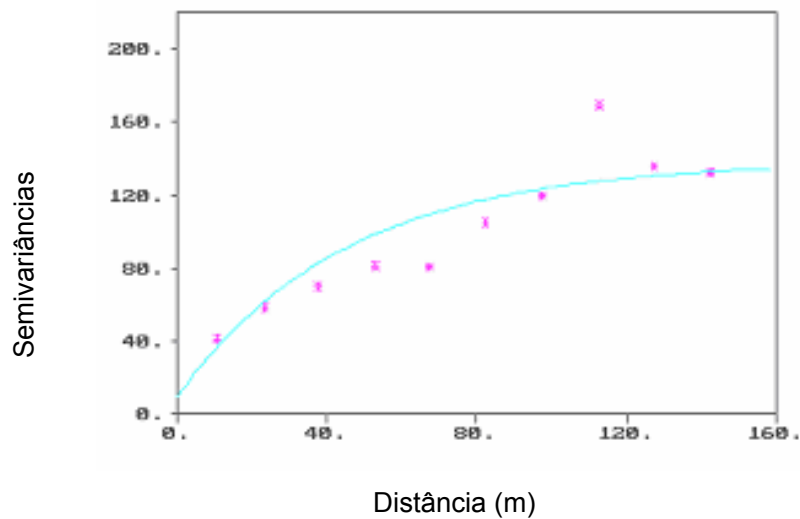


Figura 14 – Semivariograma exponencial para CEaparente – vertical.

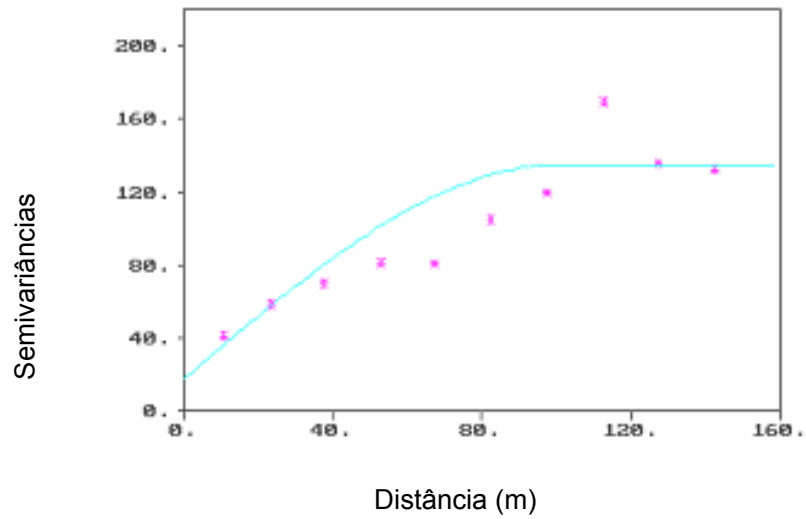


Figura 15 – Semivariograma esférico para CEparente – vertical.

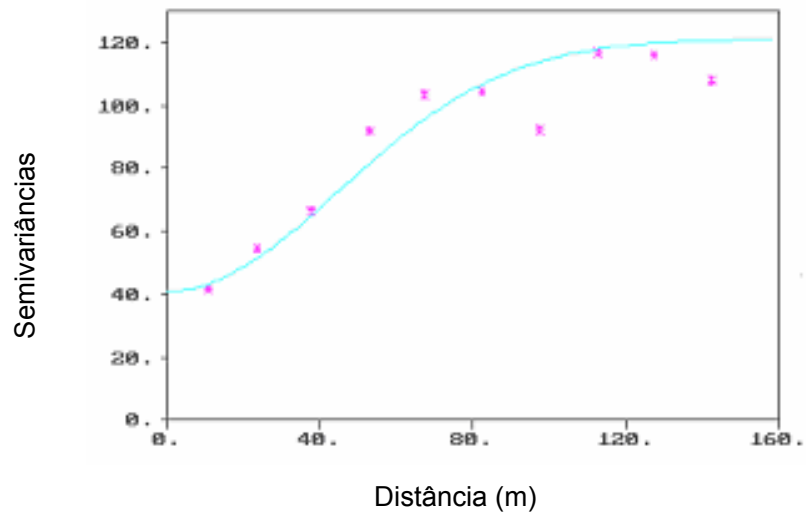


Figura 16 - Semivariograma gaussiano para CEparente – horizontal.

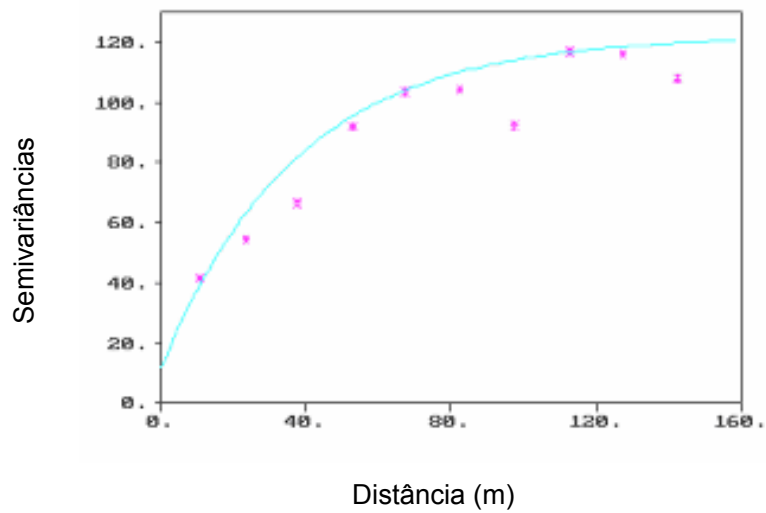


Figura 17 – Semivariograma exponencial para CEaparente – horizontal.

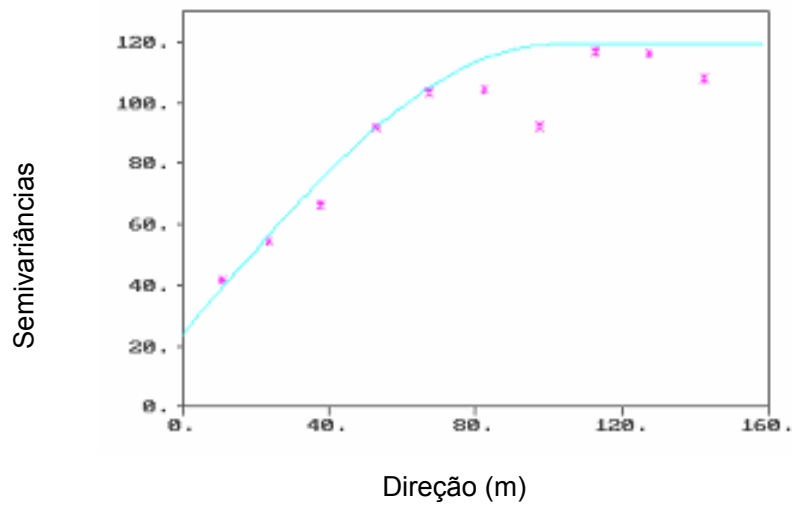


Figura 18 – Semivariograma esférico para CEaparente – horizontal.

Os parâmetros dos ajustes dos semivariogramas amostrais aos modelos teóricos estão apresentados na Tabela 4.

Tabela 4 – Valores dos parâmetros dos semivariogramas ajustados.

Semivariogramas	Modo de Operação Vertical			Modo de Operação Horizontal		
	τ^2	σ^2	ϕ	τ^2	σ^2	ϕ
Gaussiano	40	93	94	41	80	109
Exponencial	7	124	110	12	110	109
Esférico	18	117	101	24	95	101

Na análise semivariográfica a descontinuidade ou efeito pepita é atribuída às variações locais e aos erros de amostragens. Os resultados do efeito pepita (τ^2) para o ajuste ao modelo gaussiano foram maiores do que com os modelos exponencial e esférico tanto no modo de operação vertical como no modo de operação horizontal. O modelo gaussiano apresentou um menor alcance (ϕ) do que os modelos exponencial e esférico no modo de operação vertical. O patamar ($\tau^2 + \sigma^2$) apresentado pelo modelo gaussiano foi maior do que os modelos exponencial e esférico no modo de operação vertical.

A classificação proposta por Cambardella et al. (1994), considera o efeito pepita como uma porcentagem do patamar e permite comparar o grau de dependência espacial da variável estudada. Segundo esta classificação, com base na relação efeito pepita (τ^2) versus patamar ($\tau^2 + \sigma^2$), a dependência espacial considerada é representada por:

$$25\% < \frac{\tau^2}{(\tau^2 + \sigma^2)} < 75\%$$

Para o modelo gaussiano, no modo de operação vertical apresentou um alcance de dependência espacial de 94 m, efeito pepita de 40, patamar de 133 e um moderado grau de dependência espacial (IDE% = 30,1). Enquanto no modo de operação horizontal apresentou 109 m de alcance de dependência espacial, efeito pepita de 41, patamar de 121 e também moderado grau de dependência espacial (IDE % = 33,9).

4.4 Validação Cruzada

O ajuste do modelo teórico ao semivariograma amostral foi avaliado pelo método da validação cruzada.

O método da validação cruzada determinou o ajuste dos semivariogramas amostrais dos dados de condutividade elétrica nos modos vertical e horizontal.

As Figuras 19 e 20 validam os ajustes dos semivariogramas amostrais dos dados de condutividade elétrica aparente obtida no modo de operação vertical e horizontal ao modelo teórico do tipo gaussiano.

O desvio padrão e a média dos erros padronizados da validação cruzada obtidos para o modelo gaussiano no modo de operação vertical foram respectivamente iguais a 1,036 e -0,006 resultados bastante próximos da situação ideal. Em uma situação ideal, as diferenças padronizadas seguem uma distribuição normal com média nula e variância unitária. O histograma de frequência dos resíduos para a condutividade elétrica aparente no modo de operação vertical é apresentado na Figura 19.

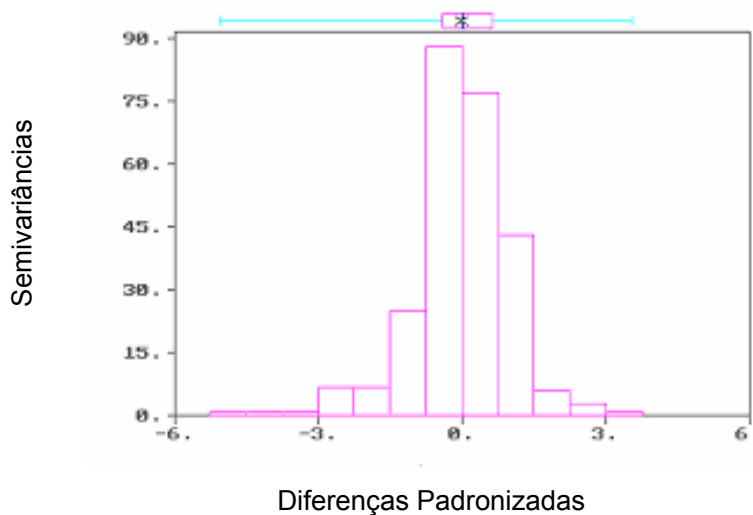


Figura 19 – Histograma dos Resíduos para CEaparente – vertical.

O histograma de frequência dos resíduos para a condutividade elétrica aparente no modo de operação horizontal apresentado na Figura 20, mostra os

valores respectivos para o desvio padrão do erro reduzido e o erro médio reduzido sendo 1,000 e -0,001, situação bastante próxima da ideal.

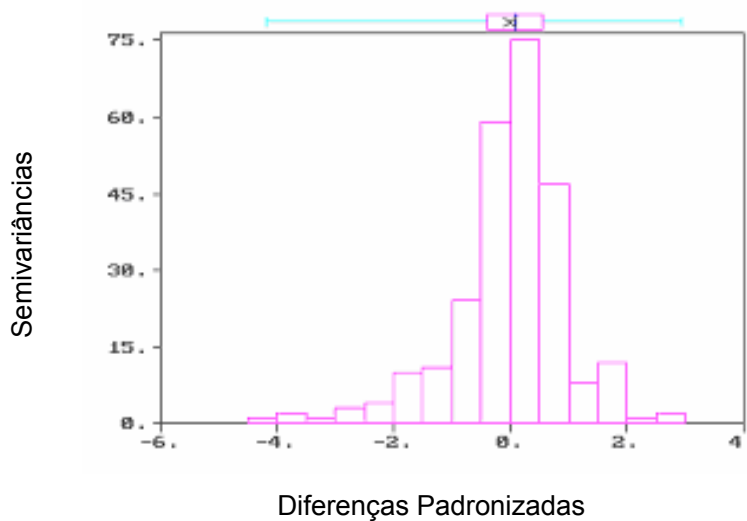


Figura 20 – Histograma dos Resíduos para CEaparente – horizontal.

A Tabela 5 mostra a estatística dos resíduos padronizados da validação cruzada para os modelos gaussianos, exponencial e esférico.

Tabela 5 - Validação cruzada dos resíduos padronizados para a condutividade elétrica aparente.

Semivariogramas	Modo Vertical		Modo Horizontal	
	Média	Desvio padrão	Média	Desvio padrão
Gaussiano	-0,001	1,138	-0,001	1,000
Exponencial	-0,001	1,145	-0,001	1,006
Esférico	-0,001	1,149	-0,001	1,004

O resumo estatístico da validação cruzada dos erros padronizados dos três modelos teóricos, ajustados “por tentativas” aos semivariogramas amostrais da variável de interesse, mostrou o modelo matemático gaussiano como o mais adequado para descrever comportamento da variabilidade espacial da condutividade elétrica aparente, tanto para o modo vertical como para o modo horizontal.

4.5 Estimadores de Máxima Verossimilhança

As análises da condutividade elétrica aparente utilizando-se o estimador de máxima verossimilhança foram baseadas nos modelos gaussiano, exponencial e esférico. O método estima os parâmetros em relação ao conjunto de dados, e não em relação aos pontos do semivariograma experimental, eliminando assim as condições de subjetividade atuantes no processo de ajuste ao semivariograma experimental. Os valores das estimativas dos parâmetros (alcance = ϕ , contribuição = σ^2 , efeito pepita = τ^2 e média = β), além do critério de informação de Akaike (AIC) e do logaritmo da função de verossimilhança foram obtidos utilizando-se o software geoR (Diggle & Ribeiro Jr, 2005).

Os resultados das estimativas de parâmetros, dos valores de AIC e os logaritmos das funções de verossimilhanças dos dados de condutividade elétrica aparente obtidos no modo de operação vertical são apresentados na Tabela 6.

Tabela 6 – Resultados das estimativas dos parâmetros obtidos utilizando estimadores de máxima verossimilhança de observações de condutividade elétrica aparente obtida no modo de operação vertical.

Parâmetros	Semivariogramas		
	gaussiano	Exponencial	Esférico
τ^2	106,70	94,10	91,28
σ^2	87,69	100,70	102,00
ϕ	20,73	22,45	39,90
β	28,25	28,13	28,27
Log verossimilhança	-1039,00	-1039,00	-1040,00
AIC	2087,00	2087,00	2088,00
Nº de Parâmetros	04	04	04

Na Tabela 6 observa-se que o valor da estimativa do efeito pepita (τ^2) foi menor para o modelo esférico. O modelo gaussiano apresentou o menor valor da estimativa do alcance (ϕ). Nas estimativas da contribuição (σ^2), o menor valor apresentado foi também para o modelo gaussiano. Os valores das estimativas do

patamar ($\tau^2 + \sigma^2$) foram maiores quando se utilizou o estimador de máxima verossimilhança, comparado com o estimador de Matheron. Pelo Critério de Informação de Akaike (AIC), o melhor ajuste foi tanto para o para o modelo gaussiano como para o exponencial.

A Tabela 7 apresenta os resultados das estimativas de parâmetros, valores de AIC e os logaritmos das funções de verossimilhanças, para o modo de operação horizontal.

Segundo a Tabela 7, os valores das estimativas do efeito pepita (τ^2) e do alcance (ϕ) foram menores para o modelo exponencial. Nas estimativas do patamar ($\tau^2 + \sigma^2$), o menor valor apresentado foi para o modelo gaussiano. Observa-se ainda que os valores das estimativas de patamar ($\tau^2 + \sigma^2$) ficaram próximos quando foi utilizado o estimador de máxima verossimilhança, comparado com o estimador clássico de Matheron.

Tabela 7 – Resultados das estimativas dos parâmetros obtidos utilizando estimadores de máxima verossimilhança de observações de condutividade elétrica aparente obtida no modo de operação horizontal.

Parâmetros	Semivariogramas		
	gaussiano	Exponencial	Esférico
τ^2	80,75	61,84	68,16
σ^2	58,09	78,26	71,09
ϕ	64,82	50,26	117,00
β	19,25	19,37	19,29
Log verossimilhança	-979,40	-978,70	-978,60
AIC	1967,00	1965,00	1965,00
Nº de Parâmetros	04	04	04

Pelo Critério de Informação de Akaike (AIC), o melhor ajuste ao semivariograma experimental ficou sendo tanto o modelo esférico como o modelo exponencial, embora, o modelo gaussiano também tenha apresentado um AIC muito próximo dos outros dois modelos.

A Tabela 8 mostra a estatística dos resíduos da validação cruzada para os modelos gaussianos, exponencial e esférico, utilizando o método da estimativa de máxima verossimilhança.

Tabela 8 - Validação cruzada dos resíduos padronizados para a condutividade elétrica aparente, utilizando a estimativa de máxima verossimilhança.

Semivariogramas	Modo Vertical		Modo Horizontal	
	Média	Desvio padrão	Média	Desvio padrão
Gaussiano	0,0014	0,9913	0,0004	0,9986
Exponencial	0,0016	0,9894	0,0011	0,9980
Esférico	0,0019	0,9877	0,0012	0,9985

O resumo estatístico da validação cruzada dos erros padronizados dos três semivariogramas experimentais ajustados pela estimativa de máxima verossimilhança aos semivariogramas teóricos da variável de interesse identificou o modelo matemático gaussiano como sendo o mais adequado para descrever a dependência espacial da condutividade elétrica aparente, tanto para o modo de operação vertical como para o modo horizontal.

4.6 Estudo da Anisotropia

A variável condutividade elétrica aparente pode apresentar um comportamento anisotrópico. Na análise do comportamento da variável estudada foi efetuado um exame dos semivariogramas ajustados ao modelo gaussiano em várias direções, tanto para o modo vertical como para o modo horizontal.

Para a análise no modo de operação vertical, foram escolhidas as direções das ordenadas, 0° e 90° (Figuras 21 e 23) e das duas diagonais, 45° e 135° (Figuras 22 e 24) com uma tolerância de 20°. Na construção dos semivariogramas direcionais, verificou-se que o número de pares dos intervalos de distância mais afastados da origem diminuiu e ficou abaixo do número recomendado por Journel e Huijbrets, (1978), 30 pares.

O comportamento dos semivariogramas apresentados para a condutividade elétrica aparente no modo vertical revelou-se isotrópico, com um crescimento

inicial monótono até alcançar uma estabilização na distância de 94m definindo um patamar de aproximadamente 133. A oscilação observada da função semivariância pode ser devido à escala de amostragem ou a presença de valores atípicos no conjunto de dados.

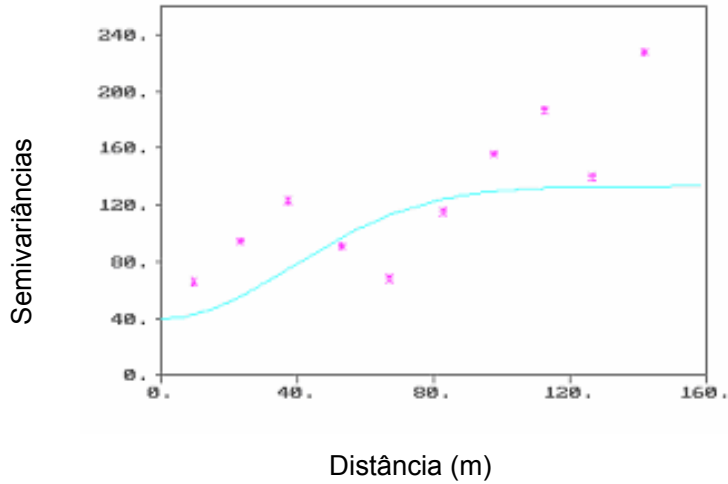


Figura 21 – Semivariograma gaussiano para CEaparente – vertical – direção 0°.

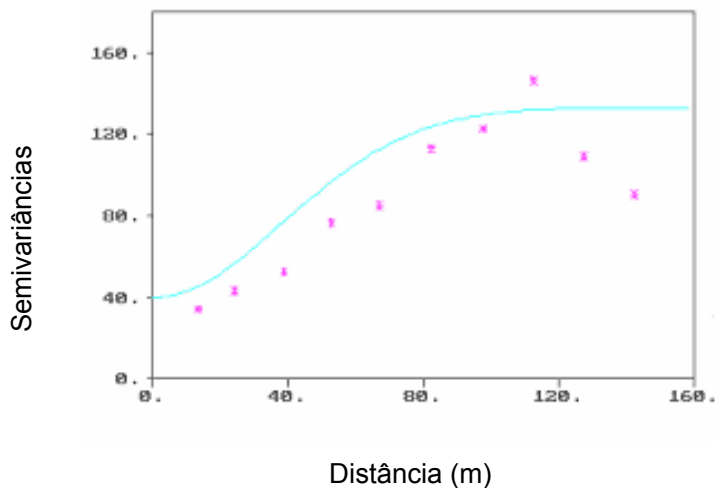


Figura 22 – Semivariograma gaussiano para CEaparente – vertical – direção 45°.

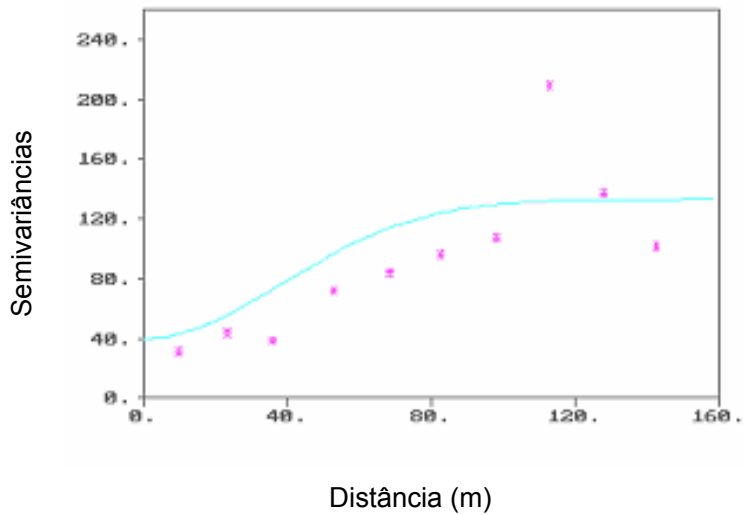


Figura 23 – Semivariograma gaussiano para CEaparente – vertical – direção 90°.

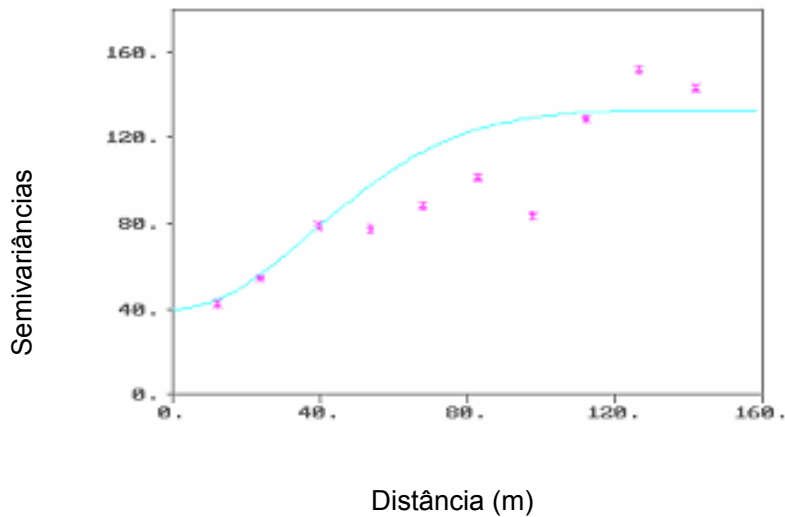


Figura 24 – Semivariograma gaussiano para CEaparente – vertical – direção 135°.

Para a análise no modo de operação horizontal, foram escolhidas as direções das ordenadas, 0° e 90° (Figuras 25 e 27) e das duas diagonais, 45° e 135° (Figuras 26 e 28) com uma tolerância de 20°. Também se verificou uma redução no número de pares dos incrementos de distância, ficando abaixo do

recomendado por Journel e Huijbrets (1978), o número de pares mais afastados da origem.

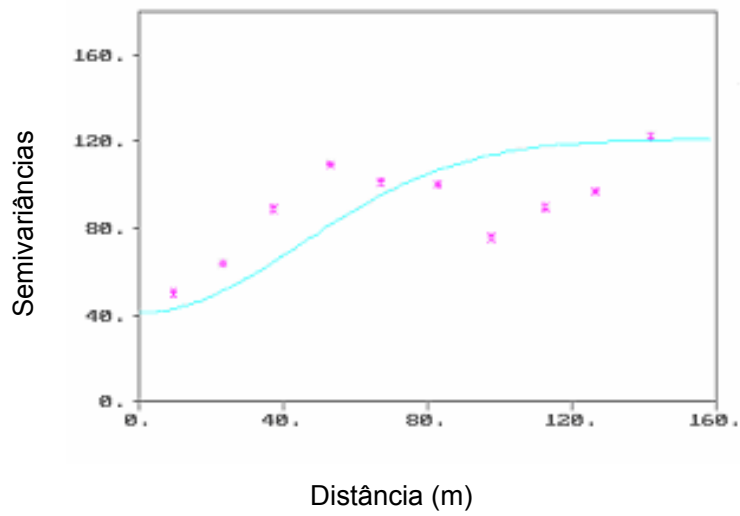


Figura 25 – Semivariograma gaussiano para CEaparente – horizontal – direção 0°.

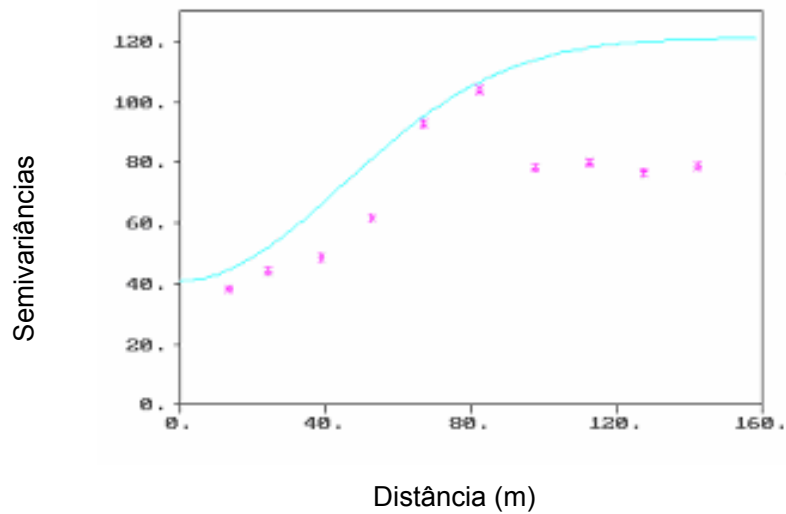


Figura 26 – Semivariograma gaussiano para CEaparente – horizontal - 45°.

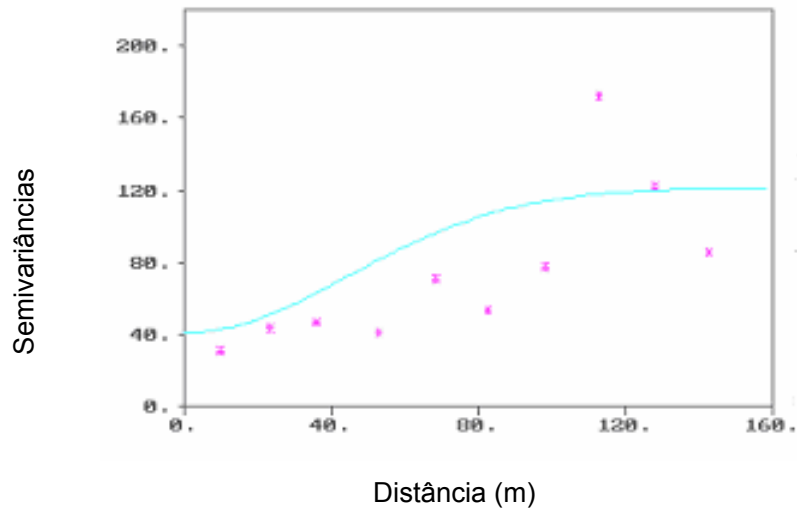


Figura 27 – Semivariograma gaussiano para CEaparente – horizontal - 90°.

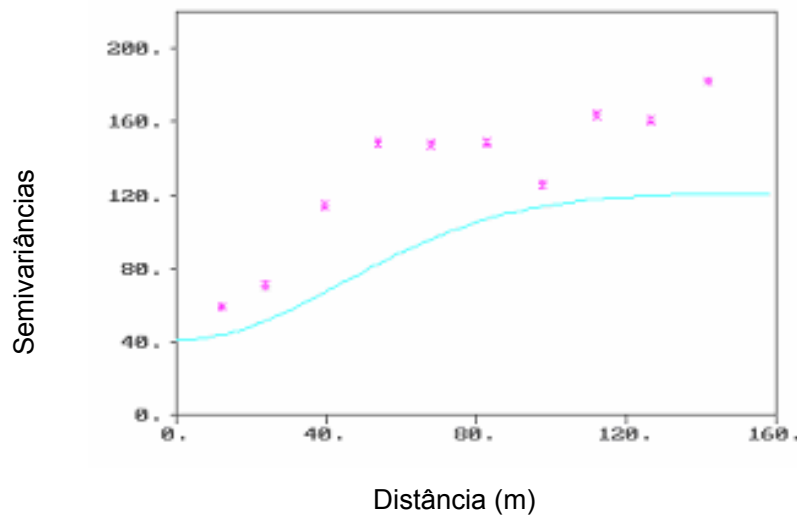


Figura 28 – Semivariograma gaussiano para CEaparente – horizontal - 135°.

O comportamento apresentado pelos semivariogramas para a condutividade elétrica aparente no modo de operação horizontal também se revelou isotrópico, com um crescimento inicial monótono até alcançar uma estabilização definindo o patamar.

4.7 Estudo da Krigagem e Representação Espacial

Conforme visto anteriormente, a krigagem é uma metodologia utilizada para estimar o valor de uma variável de interesse em um local onde não foi possível fazer a observação. É obtida através de uma interpolação, que utiliza a correlação existente entre valores dos dados obtidos em pontos máximos.

Para a aplicação da metodologia da krigagem foram considerados os modelos teóricos gaussianos ajustados ao semivariograma experimental tanto no modo vertical como no modo horizontal.

As superfícies dos valores estimados de condutividade elétrica aparente nos modos de operação vertical e horizontal representadas nas Figuras 29 e 30 apresentam as mesmas concentrações de sais das regiões estimadas pela metodologia da krigagem.

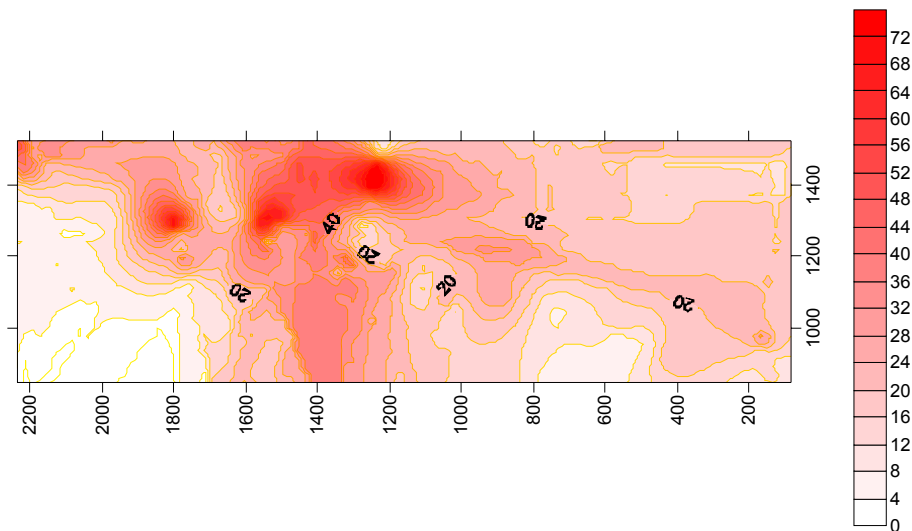


Figura 29 – Representação espacial e mapa de isolinhas da Krigagem para CE aparente – vertical.

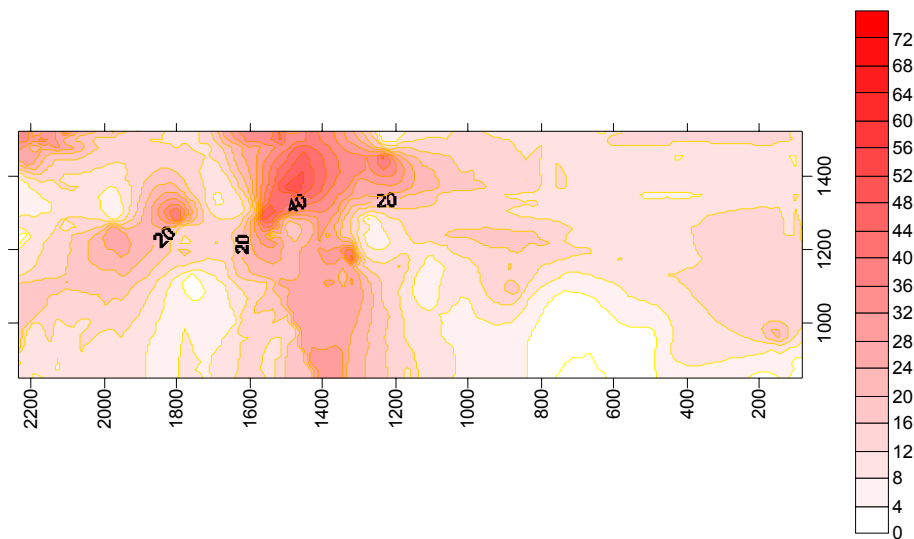


Figura 30 – Representação espacial e mapa de isolinhas da Krigagem para CE aparente – horizontal.

Os resultados da krigagem para a condutividade elétrica aparente tanto no modo de operação vertical como no modo de operação horizontal mostraram que a maior concentração de sais ocorreu na região central da área estudada.

4.8 Krigagem Bayesiana

A maior limitação imposta pela abordagem bayesiana tem sido a escolha das funções de distribuições das prioris, para o conjunto de parâmetros do modelo que podem afetar significativamente o resultado da inferência. Uma razão para isso é provocada pela natureza dos dados geoestatísticos serem espacialmente correlacionados e, mesmo ajustando completamente ao conjunto de dados, tal fato não é suficiente para influenciar a escolha da distribuição das prioris, e pode ser prejudicial para a função de verossimilhança.

Na abordagem aplicada neste trabalho, o modelo adotado ofereceu uma coerente ferramenta para a inferência geoestatística a qual permitiu uma oportunidade para a incorporação do conhecimento contextual.

Na análise bayesiana deste trabalho foi assumida as funções de distribuições para as prioris dos parâmetros (β, σ^2, ϕ) independentes. A função

de correlação assumida foi a exponencial. As medidas de condutividade elétrica aparente apresentaram uma variabilidade considerável, não havendo a necessidade de incluir outro parâmetro no modelo, além do efeito pepita (τ^2). A função de distribuição assumida para o parâmetro β foi a função de distribuição uniforme, com $p \propto 1$, que corresponde a uma distribuição priori conjugada gaussiana com extensas variâncias.

A função de distribuição priori escolhida para σ^2 é proporcional a $1/\sigma^2$, que é equivalente a forma limitada da função de distribuição qui-quadrada inversa, com grau de liberdade tendendo a zero. A função de distribuição priori escolhida para o parâmetro de correlação ϕ a uniforme.

As escolhas das funções de distribuições prioris dos parâmetros (β, σ^2, ϕ) podem ser interpretadas como desconhecimento das funções de distribuições das prioris reais para cada parâmetro.

A estimativa bayesiana do semivariograma, assim como o emprego do método da máxima verossimilhança, não é baseado no semivariograma empírico. O ajuste bayesiano não é visualmente tão perfeito como o permitido pelo método dos mínimos quadrados. Entretanto, o processo bayesiano não é contrário ao ajuste clássico. Utilizando-se os sumários das funções de distribuições posteriores (média, mediana e moda), pode-se conferir o “semivariograma bayesiano estimado” com o semivariograma empírico.

As Figuras 31 e 32 apresentam os ajustes bayesianos obtidos pela média posterior, mediana posterior e moda posterior dos parâmetros dos semivariogramas da condutividade elétrica aparente no modo vertical e no modo horizontal.

Os resultados mostram que os semivariogramas bayesianos estimados estão em concordância com os semivariogramas teóricos e amostrais.

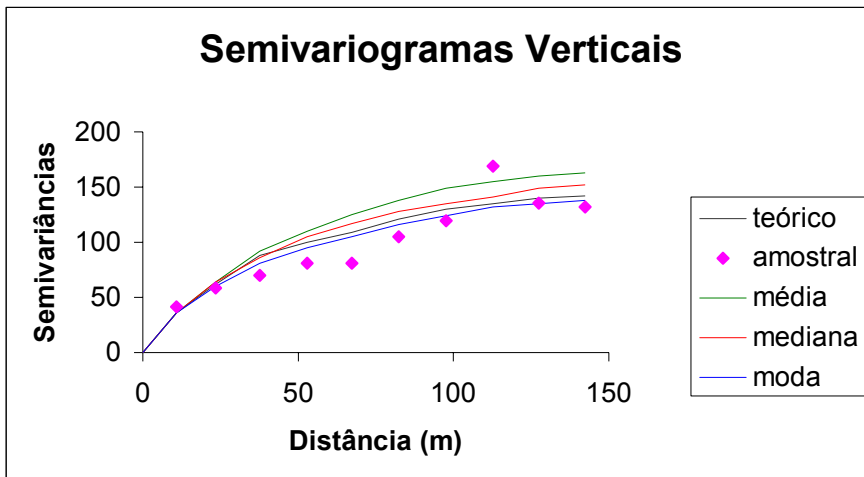


Figura 31 – Semivariogramas teórico, amostral e dos parâmetros bayesianos da média, mediana e moda da condutividade elétrica aparente – modo vertical.

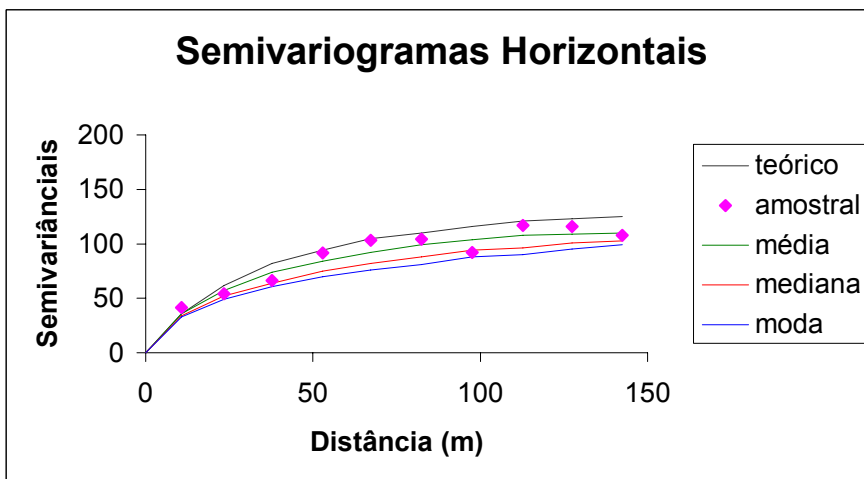


Figura 32 – Semivariogramas teórico, amostral e dos parâmetros bayesianos da média, mediana e moda da condutividade elétrica aparente – modo horizontal.

As Figuras 33 e 34 mostram os perfis da log-verossimilhanças para os parâmetros ϕ e σ^2 , das leituras de condutividade elétrica aparente obtidas no modo vertical e no modo horizontal.

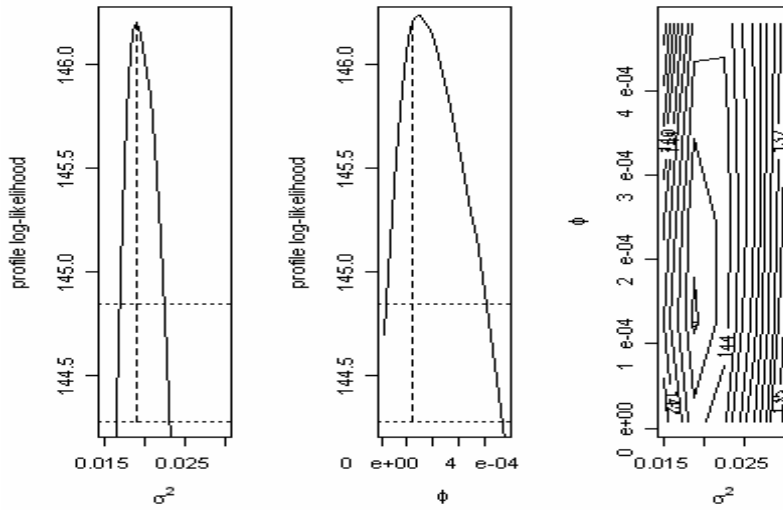


Figura 33 - Perfis das log-verossimilhanças dos parâmetros ϕ e σ^2 da condutividade elétrica aparente – modo vertical.

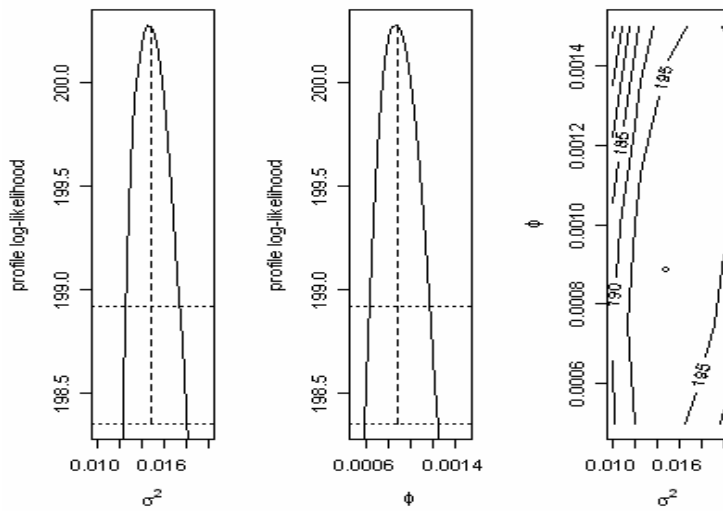


Figura 34 - Perfis das log-verossimilhanças dos parâmetros ϕ e σ^2 da condutividade elétrica aparente – modo horizontal.

As Figuras 35 e 36 apresentam os histogramas retornados da krigagem bayesiana das funções de distribuições posteriori para os parâmetros beta, σ^2 e

ϕ da condutividade elétrica aparente obtidas nos modos de operações vertical e horizontal.

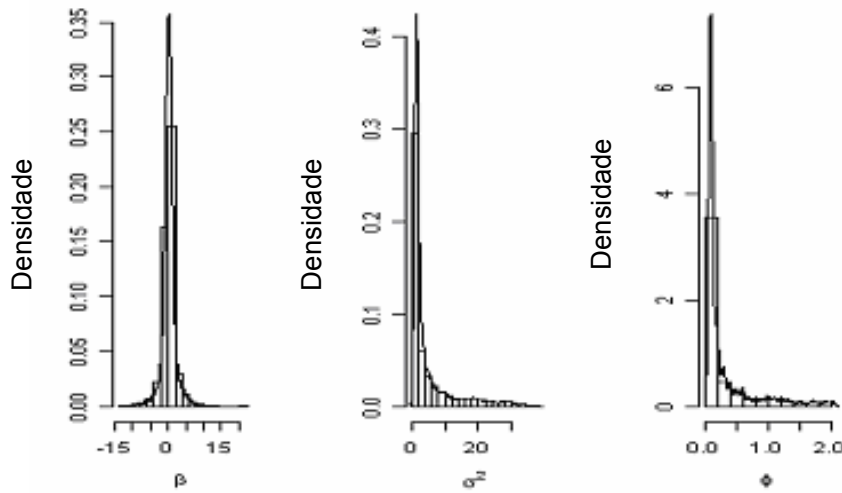


Figura 35 - Histogramas das funções de distribuições posterioris para os parâmetros beta, σ^2 e ϕ da condutividade elétrica aparente – modo vertical.

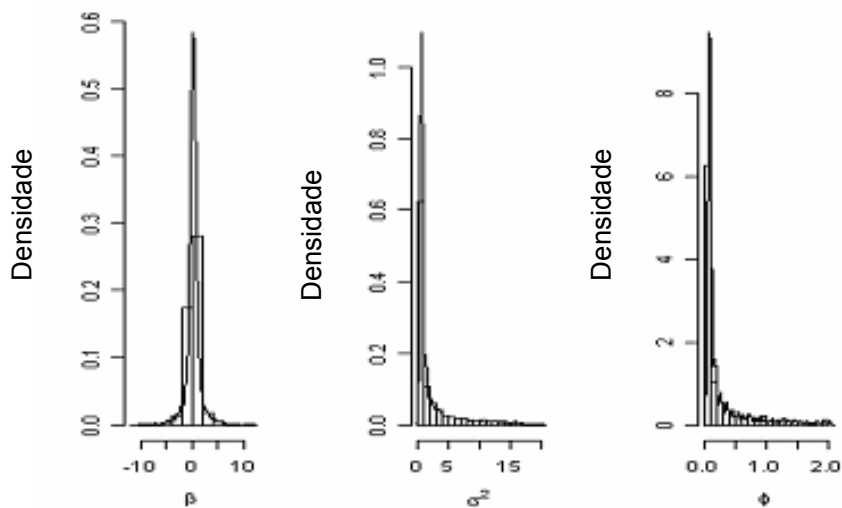


Figura 36 - Histogramas das funções de distribuições posterioris para os parâmetros beta, σ^2 e ϕ da condutividade elétrica aparente – modo horizontal.

A parte alongada da cauda visualizada nas distribuições posterioris para os parâmetros beta, σ^2 e ϕ são a contraparte bayesiana da grande assimetria

correspondente para os mesmos parâmetros observadas nos perfis das log-verossimilhanças.

Na abordagem bayesiana a determinação das funções de distribuição posterioris para a estimativa dos parâmetros depende fundamentalmente da função de máxima verossimilhança. O método da máxima verossimilhança apresenta uma superioridade em relação a outros métodos de ajuste, haja vista as fontes de subjetividades envolvidas no processo.

No processo de ajuste “a sentimento”, estão envolvidas duas fontes de subjetividades. A primeira fonte diz respeito ao número de classes de distâncias definidas para o semivariograma experimental. A outra fonte trata da incerteza por onde passa a curva do modelo entre os pontos do semivariograma experimental.

Na metodologia bayesiana com o método da máxima verossimilhança, a estimativa dos parâmetros é feita em relação ao conjunto original de dados, e não em relação aos pontos do semivariograma experimental. Desta forma, são eliminadas as fontes de subjetividade.

5. CONCLUSÕES

- a) O teste de normalidade aplicado para as leituras de condutividade elétrica aparente obtidas no modo de operação vertical como no modo horizontal, reforçam a aceitação da hipótese de normalidade dos dados.
- b) Considerando o plano de amostragem que se dispõe observa-se que não há uma anisotropia relevante no conjunto de dados, podendo assim admitir a isotropia.
- c) A verificação da tendência nos dados de condutividade elétrica aparente tanto no modo de operação vertical como no modo de operação horizontal indicou a inexistência de tendência relevante.
- d) Os semivariogramas ajustados tanto no modo de operação vertical como no modo de operação horizontal apresentaram comportamento semelhante com crescimento inicial monótono até alcançar o patamar. A escolha do melhor ajuste foi tanto para o modelo gaussiano como para o modelo exponencial.
- e) Os resultados obtidos para o comportamento espacial da condutividade elétrica aparente, indicaram um moderado grau de dependência espacial tanto no modo de operação vertical como no modo de operação horizontal.
- f) As superfícies dos valores estimados de condutividade elétrica aparente no modo de operação vertical e horizontal, através da metodologia da krigagem, apresentaram resultados semelhantes e indicou uma maior concentração de sais na região central da área estudada.
- g) Os semivariogramas ajustados apresentaram diferenças não significativas na validação da metodologia geoestatística e na abordagem bayesiana.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Akaike, H., **Information theory and na extension of maximum likelihood principle**. In Petrox, B. And Caski, F., editors, second International symposium on Information Theory, 267p, 1973.
2. Almeida, C. F. P., Ribeiro Júnior, P. J., **Estimação da distribuição espacial de retenção de água em um solo utilizando krigagem indicatriz**. Curitiba, UFPR, 1996, 38p. Relatório.
3. Banerjee, S., Gelfand, A. E.; Polasek, W. Geostatistical modelling for spatial interaction data with application to postal service performance. **Journal of Statistical Planning and Inference**, vol. 90, número 1, p. 87-105, 2000.
4. Bogaert, P., D'Or, Dimitri, Estimating soil properties from thematic soil maps: the bayesian maximum entropy approach. **Soil Science Society America Journal**, Madison, v. 66, p.1492-1500, 1991.
5. Bohn, H. L., Bem-Asher, J., Tabbara, H. S., Marwa, M., Theories and tests of electrical conductivity in soils. **Soil Science Society America Journal**, Madison, v.46, p.1143-1146, 1982.
6. Borges, W. R., Porsani, J. L., Implantação de sítio escola de geofísica no campus da USP. In: International Congress of the Brazilian Geophysical society, 2003, Rio de Janeiro, **Anais...** Rio de Janeiro: Secretaria Brasileira de Geofísica, 2003.
7. Box, G. E. P., Tiao, G. C., **Bayesian inference in stistical analysis**. New York, John Wiley & Sons, 1973, 588p.
8. Braga, L. P. V., Geoestatística e aplicações. In: SIMPÓSIO BRASILEIRO DE PROBABILIDADE E ESTATÍSTICA, 9,. 1990, São Paulo. [**Anais...**], São Paulo: IME-USP, 36p, 1990.
9. Burgess,T.M.; Webster, R. Optimal interpolation and isarithmic mapping of soil properties – I: the semi-variogram and punctual kriging. **Journal of Soil Science**, Oxford, v. 31, p. 315-331, 1980.
10. Burrough, P., **Principles of geographical information systems for land resources assessment**. Oxford, Clarendon Press, 1987, 193p.

11. Cambardella, C.A.; Moorman, T.B.; Novak, J.M.; Parkin, T.B.; Kerlen, D.L.; Turco, R.F.; Konopka, A.E. Field scale variability of soil properties in Central Iowa soils. **Soil Science Society America Journal**, Madison, vol.47, p. 1501-1511, 1994.
12. Carlin, B., Louis, T. A. **Bayes and Empirical Bayes Methods for Data Analysis**. Ed. Chapman & Hall, London, 1996, 399 p.
13. Casella, G., George, E., Explaining the gibbs sampler. **The American Statistician**, Washington, v.46, p. 167-174, 1993.
14. Cassel, D.K.; Upchurch, D.R. & Anderson, S.H. Using regionalized variables to estimate field variability of corn yield for four tillages regimes. **Soil Science Society America Journal**, Madison, v.52, p. 222-228, 1986.
15. Chib, S., Greenberg, E., Understanding the metropolis-hastings algorithm. **The American Statistician**, Washington, v.49, n. 4, p. 327-335, 1995.
16. Cook, P. G., Walker, G. R., Depth profiles of electrical conductivity from linear combinations of electromagnetic induction measurements. **Soil Science Society America Journal**, Madison, v.56, p. 1015-1022, 1992.
17. Cordeiro, G. M. e Lima Neto, E. A., **Modelos Paramétricos**. São Paulo: ABE, 2004, 246p. Trabalho apresentado ao 16º SINAPE, São Paulo, 2004.
18. Cowles, M. K. e Zimmerman, D. L. A Bayesian space-time analysis of acid deposition data combined from two monitoring networks. **Journal of Geophysical Research-Atmospheres**, v. 108, p. 299-326, 2003.
19. Cressie, N. **Statistics for spatial data**. New York, John Wiley, 1991, 900p.
20. Delhomme, J. P., Spatial variability and uncertainty in ground water flow parameters: a geostatistical approach. **Water Resources Research**, Washington, DC, v. 15, p. 269-280, 1976.
21. Diggle, P. J. et al. Model-based geostatistics. **Journal of the Royal Statistical Society Series C-Applied Statistics**, vol. 47, p. 299-326, 1998.
22. Diggle, P. J., Lophaven, S., **Bayesian geostatistical design**. Johns Hopkins University, paper 42, 2004, 17p.
23. Diggle, P. J., Ribeiro Junior, P. J., **Model based geostatistics**. Caxambu, ABE, 2000, 129p.
24. Diggle, P. J.; Ribeiro Jr, P. J. Bayesian Inference in Gaussian Model-based Geostatistics. **Geographical and Environmental Modelling**, v. 6, n. 2, p. 129 – 146, 2002.

25. Diggle, P. J.; Ribeiro Jr, P. J. **The geoR package**, Version 1.6.1, 2005, 127p.
26. Doolittle, J.A., K.A. Sudduth, N.R. Kitchen, and S.J. Indorante. Estimating depths to claypans using electromagnetic induction methods. **Journal of Soil Water Conservation**, Ankenny, v. 49, n. 6, p.572-575, 1994.
27. EMBRAPA. **Manual de Métodos de Análise de Solo**. Rio de Janeiro, 1997, 212p.
28. Englund, E. & Sparks, A. **Geostatistical Environmental Assessment Software - GeoEAS**, Version 1.2.1, Las Vegas: U. S. Environmental Protency Agency, 1990.
29. Fisher, R. A., **The design of experiments**. Oliver and Boyd, Edinburgh, 1935, 252p.
30. Gamerman, D. **Simulação estocástica via cadeias de Markov**. Caxambu: ABE, 1996, 196p.
31. Gamerman, D. **Markov Chain Monte Carlo: stochastic simulation for Bayesian inference**. Chapman and Hall, London. 1997, 245 p.
32. GEONICS – **EM 38 Ground Conductivity Meter Operating Manual**. Ontario, 1999, 69p.
33. Gill, P. S.; Swartz, T. B. Bayesian analysis of directed graphs data with applications to social networks. **Journal of the Royal Statistical Society Series C-Applied Statistics**, London, v. 53, p. 249-260, 2004.
34. Golden Software. **Surfer for Window: user's guide**. Golden, 1995.
35. Gonçalves, A.C.A.; Folegatti, M.V.; Vieira, S.R. Padrões de amostragem e intensidade de krigagem em caracterização do armazenamento de água no solo, em área irrigada por pivô central. **Revista Brasileira de Ciência do Solo**, Viçosa, MG. v. 23, n.3, p.485-495, 1999.
36. Goovaerts, P. Study of spatial relationships between two sets of variables using multivariate geostatistics. **Geoderma**, Amsterdam, v.62, p.93-107, 1994.
37. Goovaerts, P. **Geostatistics for natural resources evaluation**. New York: Oxford University Press, 1997, 481p.
38. Guerra, P. A. G. **Geoestatística operacional**. Brasília, DF: Ministério das Minas e Energia, 1988, 145p.

39. Handcock, M. S.; Stein, M. L. A Bayesian-Analysis of Kriging. **Technometrics** v. 35, n. 4, p.403-410, 1993.
40. Hargreaves, G.; **Climatic zoning for agricultural production in Northeast Brazil**. Utah State University, 1974, 6p.
41. Hartley, H. O., Rao, J. N. K., Maximum-likelihood estimation for the mixed analysis of variance model. **Biometrika**, London, v.54, p. 93-108, 1967.
42. Henderson, C. R., Best linear unbiased estimation and prediction under a selection model. **Biometrics**, Washington, v. 31, n. 3, p. 423-447, 1975.
43. Host, G. Bayesian estimation of European sulphur emissions using monitoring data and an acid deposition model. **Environmental and Ecological Statistics**, vol. 6, p. 381-399, 1999.
44. Isaaks, H.E.; Srivastava, R.M. **An introduction to applied geostatistics**. New York: Oxford University Press, 1989, 560p.
45. Johnson, C. K., Doran, J. W., Duke, R. H., Wienhold, B. J., Eskridge, K. M., Shanahan, J. F., Field-scale electrical conductivity mapping for delineating soil condition. **Soil Science Society America Journal**, Madison, v.65, p. 1829-1837, 2001.
46. Journel, A. G.; Huijbregts, C. J. **Mining geostatistics**. London: Academic, 1978, 600p.
47. Kachanoski, R. G., Gregorich, E. G., and Van Wesenbeeck, I. J.; Estimating spatial variations of soil water content using noncontacting electromagnetic inductive methods. **Canadian Journal Soil Science**, Ottawa, v. 68, p. 715–722, 1988.
48. Kachanoski, R. G., De Jong, E., and Van Wesenbeeck, I. J. Field scale patterns of soil water storage from non-contacting measurements of bulk electrical conductivity. **Canadian Journal Soil Science**, Ottawa, v.70, p. 537–541, 1990.
49. Kitanidis, P. K. Parameter Uncertainty in Estimation of Spatial Functions - Bayesian-Analysis. **Water Resources Research**, v. 22, n.4, p. 499-507, 1986.
50. Kreyszig, E.; **Introductory mathematical statistics – principles and methods**. New York, John Wiley, 1970, 470p.

51. Krige, D.G. A statistical approach to some basic mine evaluation problems on the Witwatersrand. **Journal Chemical Metall. Min. Soc. S. Afri.**, Johannesburg, v.52, p.119-139, 1951.
52. Landim, P. M. B., **Análise Estatística de Dados Geológicos**. São Paulo, 1998, 226p.
53. Lesch, S. M., Strauss, D. J., and Rhoades, J. D.; Spatial prediction of soil salinity using electromagnetic induction techniques. I. Statistical prediction models: A comparison of multiple linear regression and cokriging. **Water Resources Research**, London, v. 31, p.373–386, 1995.
54. Lloyd, C. D.; Atkinson, P. M. Assessing uncertainty in estimates with ordinary and indicator kriging. **Computers & Geosciences**, Oxford, v. 27, p. 929-937, 2001.
55. Lund, E. D., Colin, P. E., Christy, D., Drummond, P. E., Applying soil electrical conductivity technology to precision agriculture. In: European Conference on Precision Agriculture, 3., 2001, Montpellier. **Anais...**, Ecole Nationale Supérieure Agronomique, 2001, p.1089-1100.
56. Mardia, K.V., Marshall, R. J., Maximum likelihood models for residual covariance in spatial regression. **Biometrika**, v. 71, n. 1, p.135-146, 1984.
57. Matheron, G. Principles of geostatistics. **Economic Geology**, Lancaster, v. 58, p.1246-1266, 1963.
58. Matheron, G. The theory of regionalized variables and its application. **Les Cahiers du Centre de Morphologie Mathématique**, Fontainebleu, n. 5, 211p, 1971.
59. McKenzie, R. C., George, R. J., Woods, S. A., Cannon, M. E., Bennet, D. L., Use of the electromagnetic induction meter as a tool in managing salinisation. **Hydrogeology Journal**, v.5, n.1, p.37-50, 1997.
60. McNeill, J. D., **Electrical conductivity of soils and rocks**. Technical Note, TN-5, Geonics Ltda, Ontario, 1980a, 22p.
61. McNeill, J. D.; **Electromagnetic terrain conductivity measurement at low induction numbers**. Technical Note TN-6, Geonics Ltda, Ontario, 1980b, 15p.
62. Meunier, I. M. J., Silva, J.A.A. da, Ferreira, R.L.C., **Inventário florestal programa de estudo**, UFRPE, Recife, 2002, 189p.
63. Minitab. **Statistical Software for Window**, version 13.32, 2002.

64. Montenegro, A. A. A., Montenegro, S. M. G., Leal, I. G., Neto, M. C., **Variabilidade espacial da salinidade de área aluvial irrigada utilizando equipamento de indução magnética**. CONIRD, 2003, 5p.
65. Montenegro, S. M. G.; Montenegro A. A. A. e Mackay, R. Caracterização da Variabilidade Espacial de Parâmetros Hidráulicos em Solos Aluviais no Estado de Pernambuco. **RBRH**, v. 4, n. 2, p. 5-17, Abr/Jun 1999.
66. Nadler, A., Frenkel, H., Determination of soil solution electrical conductivity from bulk soil electrical conductivity measurements by the from electrode method. **Soil Science Society America Journal**, Madison, v.44, n.5, p.1216-1221, 1980.
67. Nobre, A. A.; Schmidt, A. M.; Lopes, H. F. Spatio-temporal models for mapping the incidence of malaria in Pará, **Environmetrics**, v. 16, n. 3, p. 291-304, 2005.
68. Olea, R.A. **Measuring spatial dependence with semivariograms**. Kansas Geol. Survey, University of Kansas, Lawrence, Kansas), 1977, 153p. (Series on Spatial Analysis nº 2).
69. Omre, H. Bayesian Kriging-Merging Observations and Qualified Guesses in Kriging. **Mathematical Geology**, v. 19, n. 1, p.25-39, 1987.
70. Omre, H. and K. B. Halvorsen The Bayesian Bridge between Simple and Universal Kriging. **Mathematical Geology**, v. 21, n.7, p.767-786, 1989.
71. Queiroz, J. E. **Parâmetros hidrodinâmicos de um solo de várzea para fins de drenagem subterrânea**. 1995. 167 f. Tese (Doutorado em Irrigação e Drenagem) – Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, São Paulo.
72. Reichardt, K., **Processos de transferência no sistema solo-planta-atmosfera**. 4 ed. Campinas: Fundação Cargill, 1985, 445p.
73. Rhoades, J. D., Corwin, D. L., Soil electrical conductivity: effects of soil properties and application to soil salinity appraisal. **Communications in Soil Science Plant. Analysis**, v.21, p.837-860, 1990.
74. Rhoades, J. D., Shouse, P. J., Alves, W. L., Manteghi, N. A., and Lesch, S. M. Determining soil salinity from soil electrical conductivity using different models and estimates. **Soil Science Society America Journal**, Madison, v. 54, p.46–54, 1990.

75. Rhoades, J.D. D.L. Corwin. Determining soil electrical conductivity-depth relations using an inductive electromagnetic soil conductivity meter. **Soil Science Society America Journal**, Madison, v.45, p. 255-260, 1991.
76. Ribeiro Junior, P. J. **Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo**. 1995. 99f. Dissertação de Mestrado, Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, São Paulo.
77. Scanlon, B. R., Paine, J. G., Goldsmith, R. S., Evaluation of electromagnetic induction as a reconnaissance technique to characterize unsaturated flow in an arid setting. **Ground Water**, v. 37, p.296-304, 1999.
78. Schmidt, A. M. Nobre, A. A.; Ferreira, G. S. Alguns Aspectos da Modelagem de Dados Espacialmente referenciados, **Revista Brasileira de Estatística**, v. 63, n. 220, p. 59-88, 2002.
79. Shapiro, S.S. e Wilk, M. B., An analysis of variance test for normality (complete samples), **Biometrika**, London, v.52, n. 3 e n. 4, p.591-611, 1965.
80. Sheets, K. R., and Hendrickx, J. M. H.; Noninvasive soil water content measurement using electromagnetic induction. **Water Resources Research**, Washington, DC, v. 31, p.2401–2409, 1995.
81. Silva, J. A. A. da, Campos, A. L. de, Veiga, A. F. S. L. da, Moreira, A. F. C., Marques, E. J., Estimativa da suficiência amostral para avaliar intensidade de infestação da *Diatraea SSP*, em cana de açúcar. **Pesquisa Agropecuária Brasileira**, Brasília, v. 32, n.10, p. 1003-1007, 1997.
82. Snedecor, G.W. & Cochran, W.G. **Statistical methods**. 7th Ed., Iowa State University Press, Ames. 1977, 593 p.
83. Sorensen, D., **Gibbs sampling in quantitative genetics**. Foulun, 1996, 186p.
84. Souza, L.S. **Variabilidade espacial do solo em sistemas de manejo**. 1992. 166f. Tese de Doutorado - Universidade Federal do Rio Grande do Sul, Porto Alegre,
85. Statsoft, **Statistica**, version 6, Tulsa, 1999.
86. Sudduth, K.A., N.R. Kitchen, D.F. Hughes, and S.T. Drummond. Electromagnetic induction sensing as an indicator of productivity on claypan soils. In P.C. Robert, R.H. Rust, and W.E. Larson (ed.), Proc. 2nd Intl. Conf.

- on Site-Specific Management for Agricultural Systems, 2., Madison **Proceedings**... Madison, ASA, CSSA, and SSSA, 1995, p. 671-681.
87. Trangmar, B.B., Yost, R.S., Uehara, G., Application of geostatistics to spatial studies properties. **Advances in Agronomy**, New York, v.38, p. 45-94, 1985.
88. Triantafyllis, J., Laslett, G. M., Bratney, A. B., Calibrating an electromagnetic induction instrument to measure salinity in soil under irrigated cotton. **Soil Science Society America Journal**, Madison, v.64, p. 1009-1017, 2000.
89. Utset, A.; Cid, G. Soil penetrometer resistance spatial variability in a Ferrosol at several soil moisture conditions. **Soil and Tillage Research**, Amsterdam, v.61, p.193-202, 2001.
90. Vendrusculo, L.G. Uso de Índice de desempenho e do critério de Akaike para ajuste de modelo de semivariograma. **Embrapa**, 2003, 5p. (Comunicado Técnico 58).
91. Vieira, S.R. **Curso de atualização em geoestatística**. Campinas: Instituto Agrônômico. 1995, 76p.
92. Vieira, S.R. Geoestatística em estudos de variabilidade espacial do solo. **In: Tópicos em ciência do solo**, Viçosa, MG, Sociedade Brasileira de Ciência do Solo, v.1, 2000, 54 p..
93. Vieira, S. R., Molin, J. P., Spatial variability of soil fertility for precision agriculture. In: European Conference on Precision Agriculture, 3., 2001, Montpellier, **Anais**... Montpellier: Ecole Nationale Supérieure Agronomique, 2001.
94. Voltz, M.; Lagacherie, P.; Louchart, X. Predicting soil properties over a region using sample information from a mapped reference area. **European Journal of Soil Science**, Oxford, v. 48, p. 19-30, 1997.
95. Wackernagel, H., **Multivariate Geostatistics**, Springer, Berlin, 1998, 291p.
96. Warrick, A.W., Nielsen, D.R. Spatial variability of soil physical properties in the field. In: Hillel, D. (Ed.). **Applications of Soil Physics**. New York, Academic Press, p.319-344, 1980.
97. Webster, R. Quantitative spatial analysis of soil in the field. **Advances in Soil Science**, New York, p.1-70, 1985.

DADOS CEAPARENTE

Coordenada 1	Coordenada 2	CEvert	Cehor
1518	2188	39	31
1508	2178	38	46
1498	2168	27	30
1488	2158	34	29
1478	2148	35	26
1468	2138	34	29
1458	2128	22	17
1519	2194	36	38
1509	2184	42	39
1499	2174	39	30
1489	2164	41	31
1479	2154	35	22
1469	2144	29	23
1459	2134	23	20
1520	2200	35	23
1510	2190	38	31
1500	2180	21	22
1490	2170	31	24
1480	2160	25	23
1470	2150	20	12
1460	2140	30	21
1521	2206	34	28
1511	2196	39	37
1501	2186	27	20
1491	2176	23	21
1481	2166	30	33
1471	2156	27	20
1461	2146	18	16
1522	2212	30	35
1512	2202	33	31
1502	2192	28	31
1492	2182	27	38
1482	2172	25	23
1472	2162	22	26
1462	2152	23	20
1524	2218	48	36
1514	2208	56	39
1504	2198	30	26
1494	2188	26	18
1484	2178	34	23
1474	2168	29	24
1464	2158	26	19
1180	1313	45	39
1170	1313	32	27
1160	1313	22	20
1150	1313	27	14
1140	1313	32	20
1130	1313	36	34
1120	1313	42	23
1110	1313	39	27
1180	1323	51	57
1170	1323	27	29
1160	1323	34	25
1150	1323	32	30

1140	1323	27	20
1130	1323	41	28
1120	1323	39	27
1110	1323	38	29
1180	1333	49	53
1170	1333	31	26
1160	1333	26	28
1150	1333	31	21
1140	1333	33	26
1130	1333	37	29
1120	1333	37	23
1110	1333	29	26
1180	1343	34	32
1170	1343	27	18
1160	1343	21	30
1150	1343	23	21
1140	1343	33	28
1130	1343	35	20
1120	1343	37	21
1110	1343	36	21
1180	1353	34	29
1170	1353	39	30
1160	1353	27	30
1150	1353	27	29
1140	1353	30	15
1130	1353	34	22
1120	1353	41	33
1110	1353	39	26
1098	926	27	14
1090	894	28	17
1086	868	26	18
1072	837	17	7
1066	806	17	8
1060	785	10	4
1052	757	9	3
1044	727	6	1
1026	705	9	1
1020	678	10	2
1014	651	9	2
982	608	11	2
974	568	3	1
1004	564	14	6
982	540	10	4
970	511	7	2
948	518	7	2
946	468	10	5
926	430	20	10
1178	1304	46	28
1188	1304	25	12
1198	1302	31	17
1206	1301	27	14
1216	1298	25	13
1226	1295	18	10
1234	1291	12	6
1244	1289	13	7
1254	1287	21	11
1262	1285	28	15
1272	1281	10	3

1282	1277	8	3
1292	1274	8	3
1302	1271	20	11
1312	1268	23	14
1320	1266	37	21
1330	1261	33	18
1340	1256	30	17
1350	1251	32	32
1360	1249	38	23
1386	1245	82	29
1462	1229	75	44
1466	1227	58	35
1476	1226	20	11
1486	1225	18	11
1494	1224	9	4
1150	1220	30	14
1178	1213	18	10
1164	1189	33	18
1148	1163	21	11
1134	1140	17	8
1118	1117	10	3
1086	1093	13	5
1090	1076	20	10
1068	1041	23	13
1054	1021	10	4
1036	986	18	9
1016	968	18	8
1220	1300	23	12
1223	1309	28	16
1226	1318	32	21
1228	1327	26	25
1231	1336	27	16
1234	1345	30	18
1236	1354	27	15
1239	1363	32	30
1241	1372	32	20
1244	1381	34	20
1247	1390	36	22
1250	1399	59	40
1252	1408	50	33
1254	1417	40	26
1257	1426	30	19
1259	1436	30	19
1261	1445	40	25
1264	1454	33	20
1266	1464	27	16
1268	1473	32	18
1270	1482	35	20
1273	1492	32	18
1275	1501	23	13
1277	1510	32	19
1280	1520	52	31
848	182	8	2
910	147	14	8
930	139	16	8
958	127	19	17
942	163	13	9
954	153	24	21

974	164	37	25
874	135	16	7
984	167	23	17
992	137	20	16
896	112	7	3
918	95	12	5
940	81	14	7
1352	1489	53	54
1346	1492	50	33
1340	1495	57	39
1310	1504	75	45
1300	1510	60	36
1290	1516	48	28
1280	1522	41	25
1270	1526	46	30
1260	1530	53	39
1250	1535	34	28
1240	1539	18	13
1229	1543	52	42
1218	1548	40	29
1124	1763	6	1
1135	1764	9	4
1146	1765	15	9
1157	1767	24	13
1168	1768	30	16
1179	1769	28	12
1190	1771	29	17
1201	1773	43	23
1212	1776	12	5
1223	1779	22	12
1234	1782	25	13
1245	1785	22	12
1256	1788	19	9
1267	1791	30	16
1289	1797	82	46
1262	1974	11	30
1272	1973	10	5
1282	1972	10	5
1292	1971	9	5
1301	1970	9	4
1310	1969	10	4
1320	1967	9	4
1331	1966	15	8
1342	1965	14	7
1352	1963	15	7
1362	1963	16	8
1372	1963	15	8
1382	1963	13	7
1392	1962	19	12
1402	1961	21	13
1412	1961	25	10
1384	2237	10	5
1393	2232	14	8
1402	2226	13	6
1412	2220	26	15
1421	2217	30	17
1429	2213	34	31
1438	2210	41	25

1448	2208	36	21
1459	2205	42	31
1468	2203	42	35
1477	2201	29	16
1486	2199	34	21
1496	2197	38	28
1505	2194	24	29
1514	2190	48	54
1524	2186	23	15
1294	1552	81	54
1302	1570	62	44
1306	1579	41	20
1308	1589	41	24
1310	1599	33	16
1312	1609	23	11
1314	1619	23	12
1316	1629	15	7
1318	1639	19	11
1321	1648	22	12
1325	1657	16	8
1328	1666	15	7
1332	1675	15	9
1336	1684	18	7
1340	1694	17	7
1343	1704	14	6
1347	1714	20	9
1350	1724	28	13
1353	1733	26	14
1357	1742	33	16
1360	1751	33	18
1361	1761	22	10
1362	1771	29	17