**APLICAÇÃO DE MÉTODOS GEOESTATÍSTICOS PARA
IDENTIFICAR A MAGNITUDE E A ESTRUTURA DA
VARIABILIDADE ESPACIAL DE VARIÁVEIS FÍSICAS DO SOLO**

GLAUCY DA CONCEIÇÃO ORTIZ

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

PIRACICABA
Estado de São Paulo - Brasil
Fevereiro - 2002

**APLICAÇÃO DE MÉTODOS GEOESTATÍSTICOS PARA
IDENTIFICAR A MAGNITUDE E A ESTRUTURA DA
VARIABILIDADE ESPACIAL DE VARIÁVEIS FÍSICAS DO SOLO**

GLAUCY DA CONCEIÇÃO ORTIZ

Engenheira Agrônoma

Orientadora: Profa. Dra. **SÔNIA MARIA DE STEFANO PIEDADE**

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

PIRACICABA

Estado de São Paulo - Brasil

Fevereiro - 2002

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

Ortiz, Glaucy da Conceição

Aplicação de métodos geoestatístico para identificar a magnitude e a estrutura da variabilidade espacial de variáveis físicas do solo / Glaucy da Conceição Ortiz. - Piracicaba, 2002.

75 p. : il.

Dissertação (mestrado) - Escola Superior de Agricultura Luiz de Queiroz, 2002.

Bibliografia.

1. Densidade do solo 2. Física do solo 3. Geoestatística 4. Relação solo-planta 5. Umidade do solo 6. Variabilidade espacial I. Título

CDD 631.43

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

A Deus

Autor da vida.

Ao meu querido papai,

Tertuliano Ortiz,

a minha amada mamãe,

Nora Ney da Conceição Ortiz,

E aos meus irmãos **Carlos, Douglas e Márcia** que
juntamente com suas famílias me motivam valorizando
minhas decisões profissionais.

Dedico

AGRADECIMENTOS

A Deus

“ Bendize, ó minha alma, ao Senhor, e tudo o que há em mim bendiga o seu santo nome.

Bendize, ó minha alma, ao Senhor, e não te esqueças de nenhum de seus benefícios...”

Salmos 103: 1 e 2.

Sempre temos os que agradecer... Pelo fôlego de vida, Dom de Deus, oportunidade única de viver melhor a cada dia evitando os erros do dia anterior e valorizando as pessoas a nossa volta que direta ou indiretamente ajudam em nossa jornada, por isso os meus agradecimentos:

A professora Dra. Sônia Maria De Stefano Piedade, pela orientação, atenção e confiança que muito me ajudou no término deste trabalho.

Aos professores doutores Roseli Aperecida Leandro, Antonio Augusto Franco Gracia e José Eduardo Corrente, também a Rosa Maria Alves e Solange de Assis Paes Sabadin e demais professores e funcionários do Departamento de Ciências Exatas da ESALQ - USP pelo apoio, carinho, respeito e amizade.

Aos professores doutores Antônio Carlos Andrade Gonçalves, Klaus Reichardt e Sergio Oliveira Moraes, docentes da ESALQ/USP, pela ajuda profissional no desenvolvimento deste trabalho.

Ao professor Dr. Luiz Carlos Ferreira de Souza, docente do Departamento de Ciências Agrárias da UFMS, ao Grupo PET-AGRONOMIA/UFMS e a professora Dra. Denise Santana Garcia da UFU pelo apoio e suporte constantes nesta etapa acadêmica.

À CAPES, pela bolsa de estudos concedida.

À amiga Maria Cristina Neves de Oliveira e demais colegas da pós-graduação, pela convivência agradável tornando momentos difíceis em momentos superáveis.

A todos que de uma maneira ou de outra me ajudaram a concluir mais uma etapa em minha vida fica aqui o meu muito obrigada!

SUMÁRIO

	Página
LISTA DE FIGURAS	vii
LISTA DE TABELAS	x
RESUMO	xi
SUMMARY	xiii
1 INTRODUÇÃO	1
2 REVISÃO DE LITERATURA	5
2.1 Modelagem probabilística e estacionaridade	8
2.2 Descrição e modelagem da estrutura de variabilidade espacial	14
2.2.1 Semivariância e semivariograma	14
2.2.2 Autocorrelação e autovariograma	22
2.3 Métodos de estimação do semivariograma	24
2.4 A estimação geoestatística	25
2.4.1 Krigagem ordinária pontual	26
2.5 Validação cruzada	29
3 MATERIAL E MÉTODOS	30
3.1 Localização e delineamento experimental	30
3.2 Análise estatística	31
3.2.1 Análises exploratórias descritivas	32

	vi
3.2.2 Análise exploratória espacial	34
3.2.3 Análise geoestatística	35
4 RESULTADOS E DISCUSSÃO	39
4.1 Análise exploratória descritiva	39
4.1.1 Análise exploratória espacial	43
4.2 Análise geoestatística	48
4.2.1 Análise semivariográfica	48
4.2.2 Estudo anisotrópico	51
4.2.3 Autocorrelograma	56
4.2.4 Interpolação por krigagem	58
4.2.5 Validação cruzada	61
5 CONCLUSÕES	69
REFERÊNCIAS BIBLIOGRÁFICAS	71

LISTA DE FIGURAS

	Página
1 Comportamento na vizinhança da origem de quatro tipos de semivariogramas.	16
2 Semivariograma com efeito pepita.	17
3 Principais modelos semivariográficos teóricos.	19
4 Aspecto do semivariograma com total independência entre as amostras. .	21
5 Tipos de autocorrelogramas.	23
6 Área experimental com a localização amostral dos 250 pontos.	31
7 Histograma de frequência (a) “box-plot” e (b) gráfico de probabilidade normal (c) para os dados de densidade do solo ($kg.m^{-3}$).	41
8 “Box-plot” para os dados de umidade do solo nas tensões 9,8kPa (Cc) e 1,5MPa (Pmp).	42
9 Histogramas de frequência (a) e gráficos de probabilidade normal (b) para os dados de umidade do solo nas tensões 9,8kPa (Cc) e 1,5MPa (Pmp), respectivamente.	42
10 Superfície dos valores (a) e dispersão das medidas (b) de densidade do solo ($kg.m^{-3}$) amostradas nos sentidos das SO-NE e NO-SE.	44
11 Gráficos de dispersão das médias e variâncias dos valores de densidade do solo ($kg.m^{-3}$) amostradas nos sentidos das linhas (a) e colunas (b). . . .	45
12 “Postplot” dos dados de densidade do solo ($kg.m^{-3}$).	45
13 Superfície de valores (a) e dispersão dos dados nas linhas e colunas (b) da umidade do solo na tensão 9,8kPa (Cc).	46

14	Superfície de valores (a) e dispersão dos dados nas linhas e colunas (b) da umidade do solo na tensão 1,5MPa (Pmp).	47
15	“Postplot” dos dados de umidade do solo nas tensões 9,8kPa (a) e 1,5MPa (b).	47
16	Semivariograma para densidade do solo, ajustado ao modelo gaussiano com os seguintes parâmetros: $C_0 = 3400[kg.m^{-3}]^2$; $a = 17m$ e $C = 800[kg.m^{-3}]^2$	49
17	Semivariogramas para umidade do solo, ajustados às tensões 9,8kPa (a) e 1,5MPa (b).	50
18	Semivariogramas da densidade do solo $(kg.m^{-3})^2$ a 90° (a), 45° (b) e -45° (c), respectivamente, na direção das ordenadas.	52
19	Semivariogramas da umidade do solo na tensão 9,8kPa (Cc) a 90° (a) na direção das ordenadas, 45° (b) e -45° (c) nas duas diagonais, respectivamente.	54
20	Semivariogramas da umidade do solo na tensão 1,5MPa (Pmp) a 90° (a) na direção das ordenadas, 45° (b) e -45° (c) nas duas diagonais, respectivamente.	55
21	Autocorrelograma para os dados de densidade do solo $(kg.m^{-3})$	56
22	Autocorrelograma para os dados de umidade do solo nas tensões 9,8kPa (a) e 1,5MPa (b).	57
23	Área interpolada para densidade do solo $(kg.m^{-3})$ pelo método de <i>krigagem</i> ordinária pontual considerando o modelo gaussiano.	59
24	Superfície de valores de densidade do solo $(kg.m^{-3})$, interpolada pelo método de <i>krigagem</i>	59
25	Áreas interpoladas para umidade do solo em ambas tensões 9,8kPa (a) e 1,5MPa (b) pelo método de <i>krigagem</i> ordinária pontual.	60
26	Superfícies de valores de umidade do solo nas tensões 9,8kPa (a) e 1,5MPa (b), interpoladas pelo método de <i>krigagem</i>	61

27	Mapa de erros dos valores observados em relação aos estimados ajustado ao modelo gaussiano com 16 vizinhos.	62
28	Gráficos de dispersão das estimativas de <i>krigagem</i> em relação aos valores observados (a) e as diferenças padronizadas (b) na validação cruzada para a variável densidade do solo ($kg.m^{-3}$) no modelo gaussiano.	63
29	Histograma de frequência dos erros padronizados da densidade do solo ($kg.m^{-3}$), ajustado ao modelo gaussiano com 16 vizinhos.	64
30	Gráficos de erros dos valores observados em relação aos estimados ajustados aos modelos esférico referente a tensão 9,8kPa (a) e gaussiano referente a tensão 1,5MPa (b).	65
31	Gráficos de dispersão das estimativas de <i>krigagem</i> em relação aos valores observados e as diferenças padronizadas na validação cruzada para a umidade do solo (Pa) nas tensões 9,8kPa (a) e 1,5MPa (b) nos modelos esférico e gaussiano, respectivamente.	66
32	Histogramas de frequência dos erros padronizados da umidade do solo nas tensões 9,8kPa (a) e 1,5MPa (b).	68

LISTA DE TABELAS

Página

1	Análise estatística descritiva para as variáveis densidade do solo (DS) em $kg.m^{-3}$, umidade do solo às tensões 9,8kPa (Cc) e de 1,5MPa (Pmp). . . .	40
2	Modelos de semivariogramas ajustados e seus respectivos parâmetros. . . .	49
3	Modelos de semivariogramas ajustados para os dados de umidade do solo nas tensões 9,8kPa (Cc) e 1,5MPa (Pmp) com seus respectivos parâmetros. . . .	51
4	Resumo estatístico da validação cruzada para os erros padronizados de três modelos ajustados ao semivariograma experimental da variável densidade do solo.	62
5	Resumo estatístico da validação cruzada para o modelo gaussiano com 16 vizinhos.	63
6	Resumo estatístico da validação cruzada para os erros padronizados de três modelos ajustados ao semivariograma experimental da variável umidade do solo em diferentes tensões 9,8kPa (Cc) e 1,5MPa (Pmp).	65
7	Resumo estatístico da validação cruzada para o modelo esférico com 16 vizinhos.	67
8	Resumo estatístico da validação cruzada para o modelo gaussiano com 16 vizinhos.	67

APLICAÇÃO DE MÉTODOS GEOESTATÍSTICOS PARA IDENTIFICAR A MAGNITUDE E A ESTRUTURA DA VARIABILIDADE ESPACIAL DE VARIÁVEIS FÍSICAS DO SOLO

Autora: GLAUCY DA CONCEIÇÃO ORTIZ

Orientadora: Profa. Dra. SÔNIA MARIA DE STEFANO PIEDADE

RESUMO

O estudo da dependência espacial de variáveis do solo ou da planta, através da teoria das variáveis regionalizadas ou geoestatística, permite a interpretação e a projeção dos resultados com base na estrutura da sua variabilidade natural. Com o conhecimento das posições relativas das amostragens ou medidas feitas a campo, a variabilidade espacial passa a ser utilizada como um fator positivo identificando interações importantes nas conclusões experimentais. O estudo da variabilidade de um solo, além de caracterizar uma região, pode indicar o número e a distribuição de amostras a serem retiradas possibilitando ainda maior detalhamento da área e dos resultados. A base para a adoção dos procedimentos geoestatísticos e sua justificativa é o entendimento e o tratamento da variável em questão. Para aplicação desses procedimentos é fundamental a compreensão dos instrumentos e conceitos

utilizados, tais como: semivariograma, autocorrelação, hipótese de estacionaridade, efeito pepita, patamar, alcance, anisotropia, *krigagem* e outros. Desta maneira o presente trabalho apresenta e discute conceitos e métodos geoestatísticos, analisando o comportamento de dados reais de duas variáveis físicas do solo: densidade do solo e umidade do solo nas tensões da capacidade de campo e ponto de murcha permanente, de modo a avaliar a magnitude da variabilidade espacial, as características qualitativas ligadas à estrutura do fenômeno natural que elas representam e a estrutura de correlação existente entre valores tomados em dois pontos adjacentes no espaço. Para tanto, foram obtidas estimativas das estatísticas, gráficos e mapas de dispersão e superfície que descrevessem espacial o comportamento da variável em estudo. Os resultados demonstram a capacidade dos métodos em informar e ajudar a compreender o comportamento dos fenômenos fornecendo informações para o planejamento e análise estatística futuras.

APPLICATION OF GEOESTATISTICAL METHODS TO IDENTIFY THE MAGNITUDE AND STRUCTURE OF SPATIAL VARIABILITY OF PHYSICAL VARIABLE OF THE SOIL

Author: GLAUCY DA CONCEIÇÃO ORTIZ

Adviser: Profa. Dra. SÔNIA MARIA DE STEFANO PIEDADE

SUMMARY

The study of the spatial dependence of parameters of soil or plant, through the theory of regionalized variables or geostatistics, can allow the interpretation and the projection of the results based on the structure of its natural variability. Knowing the relative positions of the samplings or field measures, the spatial variability can be used as a positive factor identifying important interactions in the experimental conclusions. The study of the soil variability, besides area characterization can indicate the number and the distribution of samples that should be removed and facilitate larger details of the area and the results. The adoption of geostatistical procedures and its justification is based on the understanding and the treatment of the cited variable. To apply these procedures it is important to know the instruments and concepts used, such as: semivariogram, autocorrelation,

estationarity hypothesis, effect nugget, sill, range, anisotropy, kriging and others. Thereby this work presents and discusses geostatistical concepts and methods, analyzing the behavior of true data of two physical variables of the soil: density of the soil and soil-water content in the tensions of field capacity and point of permanent withered, to evaluate the magnitude of the spatial variability, the qualitative characteristics linked to the structure of the natural phenomenon that they represent and the structure of existent correlation among values taken in two adjacent points in the space. For such analysis statistics estimates, graphs and maps of dispersion and surface, that described spatially the behavior of the variable in study were obtained. The results demonstrate the capacity of the methods in informing and helping to understand the behavior of the phenomenons supplying information for the planning and future statistical analysis.

1 INTRODUÇÃO

O estudo da variabilidade espacial das variáveis físico-químicas do solo constitui, entre os exemplos de aplicação, um dos fatores que possibilita o manejo do solo e da água numa agricultura irrigada.

A variabilidade espacial é entendida como variação das propriedades do solo dentro das unidades taxonômicas e/ou do mapeamento de solos. A mesma variabilidade espacial pode ser agrupada em duas categorias: variabilidade sistemática e variabilidade ao acaso.

A variabilidade sistemática é uma mudança gradual ou acentuada das propriedades do solo como função da paisagem, aspectos morfológicos, fatores de formação e/ou do próprio manejo do solo pelo homem. Enquanto que a variabilidade ao acaso constitui-se nas mudanças das propriedades do solo que não podem ser relacionadas a uma causa conhecida podendo ser decorrência da litologia diferencial, intensidade do intemperismo, adição diferencial, erosão, fatores biológicos, hidrologia diferencial, erros analíticos e de amostragem etc. (Wilding & Drees, 1983).

A variabilidade espacial tem sido uma das preocupações de pesquisadores praticamente desde o início do século. Smith (1910) considerou a disposição de parcelas no campo, em experimentos de rendimento de variedades de milho, uma tentativa de eliminar o efeito de variações do solo. Waynick (1918), ressaltou a importância da variabilidade espacial no estudo em Ciência do Solo.

Os procedimentos usados na época baseavam-se na estatística clássica ou de Fisher, que se utilizava de parâmetros como média e desvio padrão para representar um fenômeno, assumindo a hipótese principal de que as variações de um

local para outro são aleatórias.

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

A teoria das variáveis regionalizadas, também denominada geoestatística por sua aplicação nos estudos de Geologia e Mineralogia, foi um importante passo no conhecimento humano. Segundo Guerra (1988), entende-se como variáveis regionalizadas, V.R., uma função que varia de um lugar a outro no espaço com certa aparência de continuidade. São variáveis cujos valores estão relacionados de algum modo com a posição espacial que ocupam.

A continuidade atribuída às V.R. se manifesta sobretudo pela tendência de tomarem valores mais próximos, quanto menos afastados estejam dois pontos amostrados.

As peculiaridades dessas variáveis ferem o pressuposto básico da técnica de estatística clássica, ou seja, a independência entre as observações. Uma medida tomada em um ponto guarda relações de dependência com medidas tomadas em pontos adjacentes sugerindo uma estrutura de correlação.

Essas variáveis portam consigo um duplo aspecto contraditório, por um lado revelam uma característica aleatória que se manifesta na irregularidade e variação imprevisível de um ponto para outro e por outro lado, manifestam-se nas V.R., características estruturantes ou espaciais que refletem ligações existentes entre os pontos do espaço, ou seja, os valores numéricos observados não são inteiramente independentes.

As variáveis regionalizadas também não podem ser tratadas pela estatística clássica visto que: (i) nesta as amostras podem ser indefinidamente repetidas; (ii) as amostras são independentes, o resultado de uma amostra não pode ser influenciado pelo resultado daquelas amostras precedentes.

Estas imposições da estatística clássica, em geral, não são respeitadas pelas variáveis regionalizadas, pois quando se tira uma amostra num determinado ponto s o teor da referida amostra é um valor único, fisicamente determinado, sendo muito difícil a repetição dessa experiência.

Se fosse retirada outra amostra num ponto muitíssimo próximo do s considerado, poder-se-ia dizer que a condição (i) estaria satisfeita, porém neste caso não se estaria respeitando o condicionante (ii).

Segundo Guerra (1988), a partir das V.R., representadas na prática por certa quantidade de dados numéricos disponíveis, se obtêm as informações sobre as características qualitativas ligadas à estrutura do fenômeno natural que elas representam, tais características são: a localização, a continuidade e a anisotropia.

Considerando ainda que os fenômenos naturais não podem ser tratados como fenômenos aleatórios, faz-se necessário utilizar ferramentas matemáticas que permitam estudar as duas características essenciais das V.R., ou seja, aspecto aleatório e aspecto espacial.

Deste modo a geoestatística se propõe estudar dois objetivos principais: ser capaz de extrair da aparente desordem dos dados disponíveis, uma imagem da variabilidade dos mesmos e uma medida da correlação existente entre os valores tomados em dois pontos do espaço, separados por um certo vetor $\mathbf{h}$ em que $|\mathbf{h}| = h$, as quais são objetivos da análise estrutural, que é feita por meio do semivariograma; e ser capaz de medir a precisão de toda predição ou estimativa através de dados fragmentados, ou seja, há necessidade de uma teoria de estimativa, que é obtida através da *krigagem*.

O estudo da dependência espacial de propriedades do solo ou da planta, através da teoria das variáveis regionalizadas ou geoestatística, permite a interpretação e a projeção dos resultados com base na estrutura da sua variabilidade natural. Por exemplo, permite separar nas respostas das culturas a proporção referente ao tratamento ou manejo, daquela correspondente às diferenças de solos entre pontos de um mesmo campo.

Outra influência da dependência espacial está relacionada à informação dos graus de liberdade, pois, em experimentação agronômica se a hipótese de independência entre as observações não é verdadeira então, a informação dos graus de liberdade é falsa e os resultados da análise de variância não são confiáveis, devido também a redundância de informação entre os pontos dependentes.

O estudo geoestatístico apresenta um valor econômico, pois com o conhecimento da magnitude da dependência dos dados é possível planejar de maneira mais eficaz a implantação de malhas de amostragem.

Assim, com o conhecimento das posições relativas das amostragens ou medidas feitas a campo, a variabilidade espacial passa a ser utilizada como um fator positivo identificando interações importantes.

O objetivo do presente trabalho é aplicar a metodologia geoestatística afim de se verificar em uma área considerada homogênea para experimentação agrícola:

- a magnitude da variabilidade espacial de propriedades físicas do solo como densidade e umidade;
- as características qualitativas ligadas à estrutura do fenômeno natural que elas representam, tais como: a localização, a continuidade espacial e a anisotropia;
- a estrutura de correlação existente entre valores tomados em dois pontos adjacentes no espaço.

2 REVISÃO DE LITERATURA

Estudos que se utilizam da geoestatística para caracterizar a variabilidade de solos têm se tornado cada vez mais frequentes e são encontrados na literatura de Ciências do Solo de forma sistemática. O conhecimento da variabilidade de um solo, além de caracterizar uma região, pode indicar o número e a distribuição de amostras a serem retiradas, possibilitando ainda melhor detalhamento da área e dos resultados.

Diversos trabalhos mostram resultados obtidos por pesquisadores na aplicação da técnica em Ciências do Solo, como o de Reichardt (1984) e Vieira et al. (1983) em que estabeleceram os semivariogramas, covariogramas e autocorrelogramas para identificar a dependência entre uma amostra e outra. Reichardt et al. (1986) em revisão sobre variabilidade de solos relataram o semivariograma, autocorrelograma e correlograma cruzado como ferramentas que vêm sendo utilizadas em estudos no Brasil e no exterior.

Outros autores apresentaram a geoestatística como ferramenta para o estudo espacial de propriedades do solo e determinação do número de amostras a serem retiradas de uma área (Webster & Burgess, 1984; Warrick et al., 1986; Trangmar et al., 1987 e Silva, 1988).

Todos os trabalhos envolvendo conceitos geoestatísticos, em Ciência do Solo, apresentaram maior desenvolvimento na década de 80 apontando como principal preocupação o estudo da variabilidade espacial de propriedades físicas e químicas do solo, destacando-se Burgess & Webster (1980) e McBratney et al. (1986).

Segundo Warrick & Nielsen (1980), uma primeira aproximação para o

estudo da variabilidade espacial foi utilizar-se das estatísticas valor médio e desvio padrão, observando como se comportou a distribuição de frequência e se apresentou um comportamento simétrico. Outros autores como Libardi et al. (1996) e Ribeiro Jr. (1995) denominaram este procedimento como análise exploratória dos dados, em que foi possível obter uma primeira idéia das medidas de posição e variabilidade do conjunto em estudo, acompanhada da análise dos gráficos de ramos e folhas e “box-plot”, para verificação da adequação das medidas obtidas em face de problemas de valores periféricos, “outliers”.

Anterior a qualquer aplicação de estatísticas espaciais associadas à geoestatística deve-se realizar a análise exploratória através de histogramas, gráficos de dispersão e outros que permitam identificar “outliers”, forma de distribuição, escolha da análise a ser aplicada, decisão do tipo de estacionaridade que pode ser assumida, caracterização da variabilidade e outros.

Segundo Hoaglin et al. (1992), a análise exploratória dos conjuntos de dados pode facilitar o entendimento de como os processos geradores definiram este conjunto indicando a forma da distribuição, os padrões espaciais e contaminações no conjunto de dados, auxiliando na decisão das hipóteses a serem assumidas e na análise a ser desenvolvida.

Um outro exemplo de aplicação destas técnicas é encontrado em Hamlett et al. (1986), onde técnicas resistentes e exploratórias foram utilizadas para verificar se as condições restritivas relativas à estacionaridade não foram violadas. Cressie (1988) e Isaaks & Srivastava (1989) sugerem outras técnicas de análise exploratória que podem auxiliar na decisão do tipo de estacionaridade assumida.

Conforme Ribeiro Jr. (1995), a decisão a ser tomada é sempre subjetiva tendo como subsídios a interpretação das ferramentas de análise exploratória, procedimentos estes não condenáveis, desde que sejam claramente documentados.

No entanto, estas ferramentas da análise exploratória não são suficientes para descrever a variabilidade espacial, pois não consideram as posições das amostras no campo fazendo-se necessário a utilização de ferramentas que consideram

a localização espacial de cada amostra no campo. Segundo Gonçalves et al. (1999) as ferramentas geoestatísticas prestam-se para esta situação, pois permitem não só avaliar e modelar a estrutura de dependência espacial, tornando possível o mapeamento da propriedade na área por meio de interpolação geoestatística (*krigagem*), mas também avaliar e descrever a correlação espacial existente entre propriedades do solo em estudo.

A geoestatística, ou teoria das variáveis regionalizadas, pressupõe que a variação de uma variável pode ser expressa pela soma de três componentes, ou seja, um componente estrutural associado a um valor médio constante ou a uma tendência constante, um componente aleatório espacialmente correlacionado e um ruído aleatório ou erro residual, isto é,

$$Z(s) = T(s) + W(s) + e(s),$$

sendo:

$T(s)$ uma função determinística que descreve a componente estrutural de Z em s ;
 $W(s)$ um termo estocástico, que varia localmente e depende espacialmente de $T(s)$;
 $e(s)$ um ruído aleatório não correlacionado normalmente distribuído, com média zero e variância σ^2 .

Na prática, não é possível conhecer as funções de distribuição de $Z(s)$ em detalhe, por esse motivo, utilizam-se alguns pressupostos sobre o comportamento das funções. Normalmente, assume-se que as variáveis são descritas por funções conhecidas, do tipo gaussiano e lognormal, distribuições essas consideradas como limite de todas as funções de distribuições contínuas.

Não existem repetições das amostras nos pontos amostrados. Tais repetições não são mesmo possíveis para variáveis obtidas a partir de métodos extrativos. Assim, não é possível obter a função de distribuição no ponto e nem os momentos das variáveis $Z(s)$.

Uma forma de solucionar este problema é utilizar os dados em todo o campo para estimar a função de distribuição das variáveis aleatórias, isto é, assumir

que o comportamento da função de distribuição local é idêntico ao da função de distribuição global. Essa abordagem nada mais é do que a aplicação da hipótese de ergodicidade.

A base para a adoção dos procedimentos geoestatísticos e sua justificativa é o entendimento e o tratamento da variável em questão. Para aplicação desses procedimentos é fundamental a compreensão dos instrumentos e conceitos utilizados, tais como: modelagem probabilística e estacionaridade, descrição e modelagem da estrutura de variabilidade espacial e verificação da distribuição espacial para cada uma das variáveis obtendo mapas que representam os valores das propriedades em estudo na área.

2.1 Modelagem probabilística e estacionaridade

Considerando que em certos pontos de uma região qualquer sejam realizadas amostras de uma variável mencionada, tem-se um conjunto de dados espacialmente distribuídos, ou seja, medidas de um atributo acompanhadas de suas respectivas coordenadas. Com a informação da localização de cada amostra é possível calcular a distância euclidiana entre os pontos observados.

Segundo Almeida & Ribeiro Jr. (1996), para cada ponto amostrado, posição s_i , pode-se avaliar uma determinada característica $z(s_i)$, realização da variável aleatória Z no ponto s_i . O conjunto destas variáveis aleatórias é considerado um processo estocástico, descrito da forma:

$$Z(s_i) : s_i \in \mathbf{D} \subset \mathbb{R}^d$$

sendo:

$Z(s_i)$ a variável aleatória que varia continuamente em $\mathbf{D}$;

s_i a posição da variável, considerada fixa;

$\mathbf{D}$ é a região em estudo;

$\mathbb{R}^d$ é o espaço d -dimensional ($d=1, 2, 3$ ou 4).

Para $d=1$ os dados estão em uma transeção, com $d=2$ em um plano, $d=3$ em um volume que somando-se a variável tempo tem-se a dimensão 4.

O conjunto de valores $z(s_i)$ para todos os pontos s_i no domínio $\mathbf{D}$, isto é, a variável regionalizada $Z(s_i)$, pode ser considerada uma particular realização do conjunto de variáveis aleatórias $Z(s_i) : s_i \in \mathbf{D} \subset \mathbb{R}^d$. Este conjunto de variáveis aleatórias é chamado função aleatória e pode ser denotado por $Z(s)$ (Ribeiro Jr., 1995).

Esta definição de função aleatória expressa os aspectos de estrutura e aleatoriedade da variável regionalizada, pois: (i) localmente, no ponto s_i , $Z(s_i)$ é uma variável aleatória; (ii) $Z(s)$ é também uma função aleatória no sentido de que, para cada par de pontos s_i e s_{i+h} as correspondentes variáveis aleatórias $Z(s_i)$ e $Z(s_{i+h})$ não são, em geral independentes, mas são relatadas pela correlação, expressando a estrutura espacial da variável regionalizada inicial $Z(s_i)$.

Portanto, um tratamento para variáveis que apresentam aspectos aleatórios e espaciais, ditas regionalizadas, deve em uma análise estrutural, avaliar a variabilidade e correlação entre pontos fornecendo estimativas associadas a uma avaliação de sua qualidade.

A teoria das variáveis regionalizadas é uma proposta para problemas dessa natureza, onde os aspectos de localização, continuidade e anisotropia da variável regionalizada são levados em consideração.

A variável regionalizada é definida dentro de uma determinada região no espaço onde a variável é suscetível de tomar valores definidos e no interior do qual se estuda sua variação. Segundo Valente (1989), na prática a V.R. não é medida em pontos do espaço, conforme conceituados pela geometria euclidiana, mas sim em suportes de dimensões finitas e perfeitamente determinadas. Por exemplo, para um teor o suporte é o volume da amostra retirada e analisada.

É de fundamental importância compreender que caso mude o suporte obtém-se outra regionalização diferente da primeira, até sob o ponto de vista da

estatística clássica.

A variável regionalizada que se apresenta com uma continuidade mais ou menos estável em sua variação espacial pode ser expressa através de uma certa flutuação entre os valores de amostras vizinhas. Esta flutuação reflete o grau de dependência que existe entre um valor e outro.

Apesar da complexidade das flutuações verificadas para as diversas V.R., em geral, existe uma continuidade em média.

$$\lim_{s \rightarrow s_0} E\{Z(s) - Z(s_0)\} = 0.$$

Algumas variáveis com um caráter geométrico marcado dependente de uma formação, refletem um alto grau de dependência entre um valor e seu adjacente, como é o caso da espessura de uma camada de carvão (Valente, 1989).

Existem circunstâncias, onde esta continuidade é pouco definida e não pode ser confirmada, tal como acontece com minerais erráticos. Então neste caso surge o denominado efeito pepita ($C_0 \neq 0$), definido da forma abaixo:

$$\lim_{s \rightarrow s_0} E[\{Z(s) - Z(s_0)\}]^2 = C_0 \neq 0.$$

Por último existem casos onde os valores representativos não possuem dependência entre eles, ou seja, são totalmente independentes tratando-se neste caso de uma variável aleatória (Valente, 1989). O estudo dessa continuidade nos leva a considerar a variável aleatória como um caso particular de variável regionalizada.

E por fim, o aspecto de anisotropia que conforme Kitanidis (1997) refere-se a existência de uma direção privilegiada ao longo da qual os valores não variam de forma significativa, este fenômeno é também denominado de zonalidade.

Já que as variações da variável regionalizada ocorrem de maneira mais acentuada segundo determinadas direções, o estudo anisotrópico para algumas características, pode assumir um valor econômico, pois poder-se-á planejar de maneira mais eficaz a implantação de malhas de amostragens.

O duplo aspecto aleatório-estruturado das variáveis regionalizadas

pode ser interpretado com apoio na teoria das funções aleatórias, F.A. Uma F.A. nada mais é que a variável aleatória a uma infinidade de componentes.

No entanto, duas características principais limitam a modelagem de fenômenos naturais, espacialmente distribuídos, por variáveis aleatórias. Uma característica é a impossibilidade de repetir indefinidamente o experimento e a outra refere-se à dependência das observações. Visto que o resultado da amostragem para cada variável aleatória é composto de uma única realização em cada ponto e, portanto de cada variável, tem-se a impossibilidade de fazer qualquer tipo de inferência sobre este processo.

Por este motivo o estudo das variáveis regionalizadas exige que algum tipo de estacionaridade condizente com o problema em questão, deva ser assumida de forma que possibilite a estimação de pelo menos os dois primeiros momentos da distribuição da variável aleatória, que em geral estão relacionados com parâmetros de interesse como a média, a correlação, a covariância e a semivariância.

A função aleatória é estacionária quando sua lei de distribuição de probabilidades é invariante por translação, isto é, a F.A. $Z(s_1), Z(s_2), \dots, Z(s_n)$ possui a mesma lei de distribuição de probabilidades que $Z(s_{1+h}), Z(s_{2+h}), \dots, Z(s_{n+h})$. Nestas condições tem-se:

$$E\{Z(s)\} = m(s) = E\{Z(s+h)\} = m.$$

Pois sendo $Z(s)$ uma F.A. estacionária, a lei é invariante por translação e temos então $m(s)$ para qualquer vetor $\mathbf{h}$, sendo $m(s)$ é uma constante m independente do ponto de apoio s (Valente, 1989).

Pode-se então, neste caso, supor $m = 0$, para o que basta considerar a F.A. $Z(s) - m$, dita centrada, em vez de $Z(s)$.

Segundo Isaaks & Srisvastava (1989) e Cressie (1991), a variável regionalizada, $Z(s)$, deve satisfazer as hipóteses de estacionaridade. A forma de estacionaridade usualmente assumida na análise geoestatística é a chamada hipótese intrínseca ou estacionaridade fraca e é expressa totalmente em termos das diferenças,

$[Z(s_{i+h}) - Z(s_i)]$, das variáveis regionalizadas:

(i) $E\{Z(s_{i+h}) - Z(s_i)\} = 0$, para qualquer posição no espaço s_i e vetor separador $\mathbf{h}$, estabelecendo que $Z(s_i)$ é constante.

(ii) $E\{[Z(s_{i+h}) - Z(s_i)]^2\} = 2\gamma(h)$, para qualquer s_i e $\mathbf{h}$.

A expressão (ii) que define o valor da semivariância, $\gamma(h)$, que deve ser independente da posição dos pontos, sendo função apenas da distância entre eles.

Esta hipótese pode ser considerada um tipo de estacionaridade dos incrementos, que é formulada sob a variável aleatória

$$G(h) = [Z(s_{i+h}) - Z(s_i)],$$

a qual representa diferenças entre as variáveis separadas pela distância h .

A hipótese de estacionaridade de segunda ordem ou estacionaridade forte, requer:

(i) $E\{Z(s_i)\} = \mu_s$, para qualquer s , isto é, uma função aleatória independente da posição s_i , $Z(s_i)$ é constante como na hipótese intrínseca.

(ii) $E\{[Z(s_{i+h}) - \mu_s] \cdot [Z(s_i) - \mu_s]\} = cov(h)$, para qualquer s_i .

Existe uma função covariância para cada par de valores $Z(s_{i+h})$ e $Z(s_i)$ que depende apenas do vetor separador $\mathbf{h}$ o que implica que a variância, σ^2 , de $Z(s_i)$ não depende de s .

As ferramentas geoestatísticas utilizadas no estudo da variabilidade espacial, tais como semivariograma e autocorrelograma, que estão em 2.2.1 e 2.2.2, são funções principalmente de distâncias especificadas. Desta maneira a vizinhança usada na estimativa torna-se um ponto importante.

Vários são os métodos que podem ser usados para a determinação da vizinhança na estimativa. Tais métodos devem considerar o ganho de precisão em relação ao aumento de tempo de computação (Vieria, 1996). A seguir, são apresentados os métodos de vizinhança com algumas de suas vantagens e desvantagens.

Vizinhança única: no procedimento de vizinhança única todos os valores medidos são considerados vizinhos e serão utilizados na estimativa. No entanto, os pesos associados aos vizinhos separados por distâncias maiores que o alcance do

semivariograma não devem ter contribuição significativa no valor estimado. Para utilizar a vizinhança única é necessário que o semivariograma seja definido até a maior distância existente na área em estudo. A grande vantagem deste método está em que uma vez invertida a matriz de coeficientes, as estimativas podem ser feitas para qualquer espaçamento com pequeno consumo de tempo e processamento de computação.

Distância constante: o método de distância constante seleciona para cada ponto estimado uma vizinhança dentro de um círculo de raio especificado. Por consequência, nos cantos de uma área retangular tem-se um quarto de círculo com um quarto do número de vizinhos. Uma vantagem imediata na utilização deste método é o conhecimento exato que se tem da distância na qual os vizinhos para estimativa são procurados. Entretanto, o número de vizinhos pode mudar ao longo do campo fazendo com que o tamanho do sistema matricial seja variável. Isto pode ser um problema se exceder o valor usado na dimensão das matrizes do programa computacional utilizado.

Vizinhos constantes: o número de vizinhos constantes é um método bastante utilizado pois mantém constante o número de vizinhos em qualquer posição do campo. Os vizinhos são procurados, inicialmente, dentro de um raio inicial, se o número de vizinhos encontrados for menor que o limite especificado, a distância é incrementada, e o processo se reinicia. Caso contrário, se o número de vizinhos encontrados for maior que o limite, apenas o número especificado mais próximo será utilizado.

Procedimento de quadrantes: o procedimento de quadrantes, que consiste em utilizar um número especificado de vizinhos em cada quadrante ao redor do valor a ser estimado, fazendo com que a estimativa receba contribuição semelhante em número, de todas as direções. As desvantagens deste método estão nas estimativas dos pontos pertencentes aos cantos e extremidades do campo e também no desconhecimento da distância em que se localizam os vizinhos. Por estas razões o método dos quadrantes apresenta mais desvantagens que vantagens.

2.2 Descrição e modelagem da estrutura de variabilidade espacial

Assumida a estacionaridade, hipótese intrínseca, e considerando que a associação das variáveis em pontos distintos é maior a medida que mais próximos estes pontos estejam um do outro, a etapa seguinte consiste em descrever e modelar estas relações entre distâncias e associação espacial.

Várias medidas se prestam para a descrição desta relação, tais como a autocovariância e autocorrelação, usuais na análise de séries temporais. Em geoestatística a medida normalmente utilizada é a semivariância. Ao contrário da covariância e correlação, a semivariância é uma medida de dissimilaridade, ou seja, o valor da semivariância é maior à medida que as variáveis estão menos associadas.

2.2.1 Semivariância e semivariograma

A semivariância exige uma hipótese de estacionaridade menos restritiva em relação as outras medidas como a covariância, que exige estacionaridade de segunda ordem. Esta exigência menor no modelo de função aleatória não tem consequências na maioria dos casos práticos (Deutsch & Journel, 1992). Por este motivo, a semivariância pode ser utilizada em um maior número de situações.

A função semivariância é definida a partir dos desvios das medidas (d) em relação à reta 1:1 do gráfico de dispersão de $Z(s_i)$ em relação a $Z(s_{i+h})$.

$$[Z(s_{i+h}) - Z(s_i)]^2 = d^2 + d^2$$

$$d^2 = \frac{1}{2}[Z(s_{i+h}) - Z(s_i)]^2. \quad (1)$$

Considerando $\bar{Z} = \frac{\sum_{i=1}^n Z(s_i)}{n}$ como sendo o valor médio da variável em estudo e o erro médio em relação a média como sendo $\bar{e}(s_i) = \frac{\sum_{i=1}^n [Z(s_i) - \bar{Z}]}{n}$, em que n corresponde ao número de realizações da variável regionalizada $Z(s_i)$. Entretanto,

como o valor de $\bar{e}(s_i)$ é zero, considere o quadrado do erro $[e(s_i)]^2 = [Z(s_i) - \bar{Z}]^2$. Então o quadrado médio do desvio deve ser:

$$QMR = \frac{\sum_{i=1}^n d^2}{N(h)}, \quad (2)$$

em que $N(h)$ corresponde ao número de pares de observações separadas pelo mesmo h . Substituindo o valor da equação 1 em 2 tem-se segundo Journel & Huijbregts (1978):

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^n [Z(s_{i+h}) - Z(s_i)]^2, \quad (3)$$

quando a hipótese intrínseca é mantida, o valor de (3) é um estimador, pelo método dos momentos, sem tendências de $\gamma(h)$.

Com o semivariograma espera-se que valores mais próximos sejam mais similares do que aqueles mais afastados e que a diferença entre dois valores dependa apenas da sua posição relativa.

Na forma gráfica, a função semivariância ou o semivariograma, apresenta em condições ideais, o seguinte comportamento: (i) para um $h=0$, a função semivariância é zero ou não mostra variância; (ii) com o aumento de h , $\gamma(h)$ cresce gradualmente e aproxima-se de uma constante C , atingindo o patamar ou “sill”.

A dependência das observações é avaliada pela estrutura do semivariograma. A distância na qual estas tornam-se independentes é chamada de alcance ou raio de influência, a .

O crescimento mais ou menos rápido do semivariograma representa, com efeito, a maneira mais ou menos rápida com a qual é diminuída a influência de uma amostra sobre as zonas mais afastadas. A função $G(h)$ pode representar as características estruturais específicas de uma variável regionalizada analisando o comportamento na origem, a anisotropia e o fenômeno de transição.

A continuidade de uma V.R. está ligada ao comportamento do semivariograma nas vizinhanças da origem. Segundo Guerra (1988) podem-se distinguir quatro tipos (Figura 1):

Figura 1 - Comportamento na vizinhança da origem de quatro tipos de semivariogramas.

Uma variável que tenha continuidade estrita ou alta continuidade, por exemplo, uma variável com caráter puramente geométrico, graficamente apresenta um comportamento parabólico na origem (Figura 1a). Uma variável que possui continuidade “média” apresenta um semivariograma contínuo na origem porém apresentando uma tangente oblíqua (Figura 1b). Outra variável que apresenta uma tangente vertical na origem, ou seja, aquela em que $G(h) = C_0$, quando h tende a zero, terá um $G(h)$ como na Figura 1c. Este tipo de descontinuidade recebe o nome de efeito pepita (“nugget effect”, “random variance”). O efeito pepita, C_0 , representa as variações locais ou a pequena escala, como erros de análise, amostragem, etc. Finalmente uma variável que apresenta efeito pepita puro, ou seja, um fenômeno totalmente aleatório (Figura 1d).

Dada a relação dependência da função semivariância com a distância h e sendo h um vetor, o semivariograma deve ser construído segundo distintas direções no espaço, visando estudar como se deforma a função semivariograma ao mudar a direção de $\mathbf{h}$, de modo a colocar em evidência a anisotropia existente no fenômeno.

A anisotropia refere-se à existência de uma direção privilegiada ao longo da qual os valores da variável regionalizada não variam de forma significativa. Segundo Vieira (1996), a maioria das variáveis em Ciência do Solo pode ter um comportamento anisotrópico, aconselhando examinar semivariogramas para várias

direções, tais como: 0° na direção das abscissas; 90° na direção das ordenadas; 45° e -45° nas duas diagonais. Pode suceder, contudo, que $\gamma(h)$ não dependa da direção, isto é, calculado em diferentes direções forneça o mesmo semivariograma experimental. Neste caso, diz-se que o fenômeno regionalizado é isotrópico.

Segundo Vieira et al. (1983), diferentes tipos de erros aleatórios ou sistemáticos, na obtenção das amostras e medidas das variáveis em intervalos inferiores ao de amostragem, geram uma descontinuidade na origem, denominada de efeito pepita (“nugget effect”) e indicada por C_0 . O patamar, então, é $C + C_0$ e um exemplo pode ser visto na Figura 2.

Figura 2 - Semivariograma com efeito pepita.

O valor a (Figura 2) é denominado alcance e corresponde à idéia de zona de influência de uma amostra. Ele marca a distância a partir da qual um ponto da variável em estudo não tem mais influência sobre o ponto vizinho, marcando assim, o início da zona de pura aleatoriedade. Este tipo de situação refere-se ao fenômeno de transição que se caracteriza por semivariograma cujo crescimento tende até um patamar $C + C_0$. O patamar (“sill”, “palier”, soleira) $C + C_0$, marca o valor da variância total da variável em estudo, C é denominada variância de dispersão espacial e representa as diferenças espaciais entre os valores de uma variável tomada em dois pontos separados por distâncias cada vez maiores.

Com dados de precipitação e lâmina de irrigação, além de potenciais da solução, Dourado Neto (1989) observa a variabilidade espacial dos dados, ajustando

modelos de semivariograma linear, esférico, exponencial e de Gauss em várias direções para a constatação da anisotropia na região.

Segundo Ribeiro Júnior (1995) um modelo que é ajustado por tentativas, bem como seus parâmetros sobrepostos aos pontos estimados, não tem rigor e deve ser utilizado algum procedimento que possibilite encontrar, com boa aproximação, informações sobre a amostra. Em 2.5, utilizar-se-á a validação cruzada.

Segundo Guerra (1988), existem três tipos de semivariogramas a saber: semivariograma observado ou experimental, aquele obtido a partir da amostragem realizada, portanto o único conhecido; o semivariograma verdadeiro, é o semivariograma real do fenômeno em estudo e é sempre desconhecido. E o semivariograma teórico é um semivariograma de referência.

Dos modelos teóricos existentes, ressaltam-se: o esférico, o exponencial e o de Gauss ou gaussiano, os quais são modelos com patamar; semivariogramas linear, potência e o logarítmico ou de De Wijs são modelos sem patamar. Uma vez escolhido um semivariograma teórico serão utilizadas suas propriedades para a análise semivariográfica.

O modelo esférico ou de Matheron é bastante utilizado em geoestatística e representa um fenômeno de transição. Sua expressão matemática é:

$$\gamma(h) = \begin{cases} C[\frac{3h}{2a} - \frac{1}{2}\frac{h^3}{a^3}] & \text{se } h \leq a \\ C & h > a \end{cases}.$$

Inicia-se na origem e vai crescendo à medida que aumenta a separação das amostras, mostrando maior similaridade de valores mais próximos. Quando não há a presença do efeito pepita (C_0) tem-se $C = \sigma^2$ (variância *a priori* finita ou patamar) e a a amplitude ou alcance.

Para $h > a$ não há correlação entre $Z(s_i)$ e $Z(s_{i+h})$ e $\gamma(h)$ alcança o patamar. A distância a corresponderá a distância de influência de cada amostra. A distância a , chamada de alcance tende a se estabilizar por volta do patamar (Figura 3A), indicando uma total aleatoriedade ou independência das amostras. Isto sugere que para distâncias inferiores ao alcance adotam-se os procedimentos geoestatísticos

e, em caso contrário, as técnicas da estatística clássica.

Figura 3 - Principais modelos semivariográficos teóricos.

Havendo efeito pepita tem-se:

$$\gamma(h) = \begin{cases} C_0 + C\left[\frac{3h}{2a} - \frac{1}{2}\frac{h^3}{a^3}\right] & \text{se } h \leq a \\ C_0 + C & h > a \end{cases},$$

em que a representação genérica está na Figura 3B, onde a é a amplitude e $C_0 + C$ é o patamar. Observando a tangente na origem o modelo assume uma posição particular tratando de um modelo semivariográfico com comportamento linear na origem e o ponto onde a referida tangente corta o patamar representa $\frac{2}{3}a$ facilitando o cálculo automático.

Segundo Valente (1989), o modelo exponencial representa o fenômeno de transição, embora o patamar só seja alcançado para $h = \infty$, conforme consta na Figura 3C. Quando não há efeito pepita sua equação é:

$$\gamma(h) = C[1 - e^{-\frac{h}{a}}],$$

em que a amplitude a é determinada pelo encontro da tangente na origem com o patamar C igual à variância a priori, variância finita, sendo um modelo com comportamento linear na origem, .

Quando existe efeito pepita C_0 a sua equação fica:

$$\gamma(h) = C_0 + C[1 - e^{-\frac{h}{a}}],$$

e o patamar passa a ser $C_0 + C$. O modelo exponencial é um modelo de fácil aplicação devido a sua simples equação matemática. Note-se que o patamar é verdadeiramente uma assíntota da equação de $\gamma(h)$.

Segundo Landim (1998), um modelo semivariográfico que também representa um fenômeno de transição, a curva é parabólica junto à origem e não linear como os modelos anteriores, é o modelo de Gauss. A tangente nesse ponto é horizontal o que indica pequena variabilidade para curtas distâncias, genericamente representado na Figura 3D. Quando não há efeito pepita a sua equação é:

$$\gamma(h) = C[1 - e^{-\frac{h^2}{a^2}}],$$

tratando-se de um modelo semivariográfico mais regular possível, visto admitir tangente horizontal na origem. O seu patamar C e a sua amplitude a são determinados como no semivariograma esférico e têm exatamente o mesmo significado.

Quando existe efeito pepita (C_0) a sua equação fica:

$$\gamma(h) = C_0 + C[1 - e^{-\frac{h^2}{a^2}}],$$

em que o patamar passa a ser $C_0 + C$. A existência do efeito pepita neste modelo só admite-se pela presença de erros humanos na amostragem, uma vez que os fenômenos representados por este modelo, são os mais regulares possíveis.

Um exemplo de modelo sem patamar é o modelo Linear. É o modelo mais simples, sua equação é:

$$\gamma(h) = p(h)^\theta,$$

em que p é determinado com o valor de $\gamma(1)$ (Kitanidis, 1997). Genericamente representa-se conforme Figura 3E e pode-se dizer que é a reta que se obtém unindo o ponto experimental obtido para $h = 1$ com a origem $h = 0$.

A principal observação para este modelo refere-se ao seu crescimento monótono e o fenômeno é não estacionário. Algumas situações que não apresentam patamar definido são consideradas de infinita capacidade de dispersão, ajustando-se aos modelos que consideram este fato, como por exemplo um modelo linear.

No caso de total independência das amostras, tem-se o chamado efeito aleatório puro, e um semivariograma da Figura 4.

Figura 4 - Aspecto do semivariograma com total independência entre as amostras.

2.2.2 Autocorrelação e autovariograma

A função autocorrelação expressa a correlação linear entre uma série espacial ou temporal e ela mesma, quantificando a mudança de forte dependência para a situação de independência dos valores de um determinado parâmetro.

Na utilização da função autocorrelação os parâmetros são modelados por uma função qualquer, com propriedades suficientes para garantir a sua existência. Como tal função pode depender da área ou volume da região delimitada a sua validade não pode ser satisfatoriamente testada, embora seja requerida uma forma de estacionaridade.

A função autocorrelação tem certas vantagens em relação a outros métodos, por apresentar valores normalizados para o intervalo -1 a 1, o que facilita a interpretação dos resultados.

Para a caracterização da função autocorrelação é necessário que os valores de $Z(s_i)$ e $Z(s_{i+h})$ satisfaçam a hipótese de estacionaridade de segunda ordem. Nestas condições o coeficiente de autocorrelação é

$$\rho(h) = \frac{cov[Z(s_i), Z(s_{i+h})]}{\sqrt{var[Z(s_i)]} \sqrt{var[Z(s_{i+h})]}}, \quad (4)$$

em que $\rho(h)$, cov e var, são respectivamente, a autocorrelação, a covariância e as variâncias populacionais. A estimativa de $\rho(h)$ usando dados de amostragem é representada por $r(h)$. A covariância é dada pela equação (5)

$$cov(h) = \frac{1}{N} \sum_{i=1}^{N-h} [Z(s_i) - \bar{Z}][Z(s_{i+h}) - \bar{Z}]. \quad (5)$$

Se especificar $h = 0$, tem-se para os pares de $[Z(s_i), Z(s_i)]$ um coeficiente de correlação $r_0 = 1$. Para pares $[Z(s_i), Z(s_{i+1})]$ a distância $h = 1$ o valor obtido é r_1 . Acrescentando a distância entre os vizinhos para $h = 2$ dos pares $[Z(s_i), Z(s_{i+2})]$ tem-se o valor r_2 e assim por diante.

O autocorrelograma é um gráfico de $r(h)$ em função da distância h . O

valor máximo é 1 para $h = 0$ e tende a decrescer com o aumento de h . Quando $h = 1$ e os pontos $s_1, s_2, \dots, s_n$ são igualmente espaçados, a correlação é feita de s_1 com s_2 , s_2 com s_3 , ..., s_{n-1} com s_n . Para $h = 2$, a correlação é feita de s_1 com s_3 , s_2 com s_4 , ..., s_{n-2} com s_n .

O autocorrelograma pode ser um instrumento para determinar o espaçamento entre parcelas no qual a correlação entre vizinhos, ou pontos adjacentes, é eliminada e não há mais a interferência de uma parcela sobre a outra. Neste contexto, o alcance é definido pelo intervalo no qual as amostras espaciais mantêm relação de dependência, ou podem ser representadas por uma única amostra.

Três casos de autocorrelogramas estão apresentados na Figura 5. O primeiro caso (a) é o mais típico, com $r(h)$ decrescendo gradualmente até nivelar-se em torno de zero, neste caso o autocorrelograma é interpretado como espacialmente dependente. No caso (b), o valor de $r(h)$ declina rapidamente dentro da primeira distância (*lag*), a partir do que, mostra a independência entre as amostras. E por fim, no terceiro caso (c), o coeficiente de autocorrelação diminui lentamente indicando independência entre as amostras a longo alcance.

Figura 5 - Tipos de autocorrelogramas.

Considerando a definição de autocorrelograma, para $h = 0$, $r(h)$ é sempre igual a 1 e à medida que a distância h aumenta, os valores de $r(h)$ tendem a decrescer para $r(h)$ igual a zero. A correlação entre duas variáveis $Z(s_i)$ e $Z(s_{i+h})$ desaparece quando h se torna muito grande. Na prática pode-se dizer que $C(h) = 0$,

se $|\mathbf{h}| \geq a$. Esta distância a é chamada de amplitude ou alcance e representa o ponto a partir do qual passamos do estado onde existe uma correlação espacial $|\mathbf{h}| < a$ para o estado onde não existe tal correlação, $|\mathbf{h}| \geq a$.

2.3 Métodos de estimação do semivariograma

Existem diversos métodos de ajuste de modelos semivariográficos, McBratney & Webster (1986) comentam o ajuste de semivariogramas abordando os métodos de mínimos quadrados (ordinários, ponderados e generalizados) e de máxima verossimilhança. Segundo os autores, o método de mínimos quadrados ponderados tem melhor realismo e eficiência computacional.

O ajuste por mínimos quadrados ordinários e ajuste empírico também chamado de ajuste “a sentimento”, são bastantes comentados na literatura (Journel & Huijbregts, 1978; Isaaks & Srivastava, 1989 e Cressie, 1991).

O método de mínimos quadrados ordinário é citado por Ribeiro Júnior (1995) como inadequado porque os pontos da nuvem de dados correspondem as estimativas provenientes de diferentes pares de pontos e estas estimativas não possuem a mesma precisão e não são independentes.

No método “a sentimento” escolhe-se um modelo e seus parâmetros e este é sobreposto aos pontos estimados, a partir de uma nuvem de pontos obtém-se a estimação dos parâmetros do semivariograma através de uma avaliação visual. Alguns autores como McBratney & Webster (1986) e Ribeiro Júnior (1995) consideram o método pouco rigoroso e não fidedigno.

Apesar das críticas, diversos textos geoestatísticos clássicos, como o de Journel & Huijbregts (1979), adotam este método fazendo ressalva ao método de ajuste automático. Ribeiro Júnior (1995) recomenda a utilização do ajuste empírico como chute inicial para estimar os parâmetros do modelo de interesse e dar continuidade com outros métodos de melhor embasamento estatístico, o quais não são de fácil implementação computacional.

2.4 A estimação geoestatística

Muitas das vezes além de um modelo de dependência espacial existe o interesse em se obter valores em pontos não amostrados ou obter uma malha de pontos interpolados que possibilite a visualização do comportamento da variável na região através de mapas ou gráficos de superfície.

O interesse de se detalhar uma área sugere a idéia de um método de interpolação que forneça estimativas de valores em pontos ou médias em sub-regiões e que os estimadores utilizados sejam não viesados e de variância mínima. Segundo Ribeiro Jr. (1995) a acurácia de uma estimativa depende dentre vários fatores de:

- (i) número de pontos avaliados e a qualidade da informação obtida de cada um deles, refere-se ao método de medida, sua precisão e homogeneidade;
- (ii) posição dos pontos no campo deve ser arranjada de maneira que abranja a maior área possível evitando concentrações em sub-regiões;
- (iii) continuidade espacial de variáveis interpoladas, considera que quantidades com suaves variações podem ser melhor interpoladas que quantidades com flutuações mais erráticas.

A proposta geoestatística de interpolação é conhecida como *krigagem* e foi desenvolvida por Matheron em homenagem ao engenheiro de minas D.G.Krige, o qual em trabalhos com reserva de ouro na África do Sul introduziu o uso do processo de médias móveis para estimação em depósitos minerais. A *krigagem* é um processo que estima valores de variáveis espacialmente distribuídas a partir de valores adjacentes interdependentes considerando todos os principais fatores necessários para se ter uma boa acurácia.

O estimador de *krigagem* equação (6)

$$\hat{Z}(s_0) = \sum_{i=1}^{N(h)} \lambda_i Z(s_i), \quad (6)$$

é um interpolador BLUE (“Best Linear Unbiase Estimator”) desenvolvido a partir

de dados referentes à uma variável de natureza regionalizada. O estimador é dito linear pois a estimação é obtida combinando os valores amostrados e ponderando os pontos amostrais para que se obtenha o menor erro quadrático médio.

Pode ser aplicado para previsão do valor de uma variável em um ponto específico dentro de um campo geométrico, *krigagem* pontual; para o cálculo do valor médio de uma variável regionalizada em uma determinada região, *krigagem* em blocos; e na estimação de “drift”, tendência local de uma variável de interesse de modo similar à análise de superfície de tendência (Valente, 1989).

Existem diversos tipos de *krigagem*: a simples, ordinária, universal, indicadora, probabilística e outras. Neste trabalho, nos restringiremos a *krigagem* do tipo ordinária, com estimativas pontuais.

2.4.1 Krigagem ordinária pontual

O estimador de *krigagem* é uma combinação linear dos pontos conhecidos, em que os ponderadores são proporcionais às distâncias euclidianas entre o ponto a ser estimado e os demais pontos conhecidos, incorporando a estrutura de variabilidade na região de estimação.

As distâncias consideradas não são somente as distâncias euclidianas, mas também, as distâncias entre vizinhos. Considere um fenômeno isotrópico e a estimação de um ponto qualquer, em que os vizinhos a este ponto estão igualmente espaçados e não agrupados, nesta situação os pesos destes vizinhos são semelhantes.

Agora considere a situação em que os vizinhos ao ponto a ser estimado estão distanciados igualmente, existindo um agrupamento de alguns valores. Nesta situação é razoável que os pesos de cada valor agrupado seja menor que do valor não agrupado, mesmo que a distância entre estes pontos e o ponto estimado seja a mesma, pois os dados agrupados trazem informações redundantes de uma mesma região.

Esta característica do estimador é denominada de “declustering”, pois

o preditor considera uma medida de associação entre os pontos e suas respectivas vizinhanças. Geralmente, a medida de associação utilizada é a semivariância, por isso da necessidade do semivariograma e do modelo ajustado.

Se o modelo ajustado para o semivariograma é correto e não apresenta erros de medidas deve-se então, determinar os valores dos pesos, λ_i , que assegurem as propriedades de mínima variância e não tendenciosidade.

Assumida a hipótese intrínseca, para a não tendenciosidade ser assegurada deve-se ter:

$$\begin{aligned} E\{\hat{Z}(s_0) - Z(s_0)\} &= 0 \\ E\{\hat{Z}(s_0) - Z(s_0)\} &= E\left\{\sum_{i=1}^{N(h)} \lambda_i Z(s_i) - Z(s_0)\right\} = \\ &= \sum_{i=1}^{N(h)} \lambda_i E\{Z(s_i)\} - E\{Z(s_0)\} = 0. \end{aligned}$$

Assumindo que $E\{Z(s_i)\}$ é constante para todo ponto s_i , igual a um valor m , tem-se:

$$m \sum_{i=1}^{N(h)} \lambda_i - m = 0 \Rightarrow \sum_{i=1}^{N(h)} \lambda_i = 1.$$

Esta restrição garante a não tendenciosidade do estimador.

Os coeficientes de estimação são obtidos minimizando o erro de estimação $[\hat{Z}(s_0) - Z(s_0)]$, em que a variância de estimação é dada por:

$$Var[\hat{Z}(s_0) - Z(s_0)] = E\{[\hat{Z}(s_0) - Z(s_0)]^2\} - \{E[\hat{Z}(s_0) - Z(s_0)]\}^2,$$

em que o primeiro e o segundo termos correspondem, respectivamente, ao erro quadrático médio e ao viés. Para o estimador não viciado, o erro quadrático médio é igual à variância, ou seja, o viés é nulo.

Substituindo a equação do estimador de *krigagem* na expressão da variância tem-se:

$$Var[\hat{Z}(s_0) - Z(s_0)] = - \sum_i \sum_j \lambda_i \lambda_j \gamma(s_i, s_j) + 2 \sum_i \lambda_i \gamma(s_i, s_j).$$

Conforme Calvete & Ramirez (1990) e Ribeiro Júnior (1995) inserindo o multiplicador de Lagrange, o qual permite solução do sistema de equações de *krigagem*:

$$Var[\hat{Z}(s_0) - Z(s_0)] = - \sum_i \sum_j \lambda_i \lambda_j \gamma(s_i, s_j) + 2 \sum_i \lambda_i \gamma(s_i, s_0) - 2\mu \left(\sum_i \lambda_i \right).$$

A variância quando minimizada, por meio do cálculo das derivadas de primeira ordem em relação aos parâmetros, sujeita a restrição $\sum_{i=1}^{N(h)} \lambda_i = 1$ e igualada a zero, resulta em um sistema de equações de *krigagem* do tipo:

$$\begin{cases} \sum_{i=1}^{N(h)} \lambda_i \gamma(s_i, s_j) + \mu = \gamma(s_i, s_0) \\ \sum_{i=1}^{N(h)} \lambda_i = 1 \end{cases}.$$

Resolvendo o sistema tem-se a estimativa:

$$\hat{Z}(s_0) = \sum_{i=1}^{N(h)} \lambda_i Z(s_i),$$

em que o estimador $\hat{Z}(s_0)$ é BLUE.

O sistema de *krigagem* com solução única será assegurado somente se a matriz de covariâncias for positiva definida, indicando assim que o seu determinante deve ser positivo ou zero. Se em lugar de utilizar a covariância, for usada a semivariância na matriz, o modelo de semivariograma deve ser negativo semidefinido condicional - NSDC (Gonçalves, 1997).

Através da *krigagem* ordinária tem-se $\hat{Z}(s_0)$, que é uma estimativa do valor esperado de $Z(s_0)$ ou $E\{Z(s_0)\}$. Por estimar uma média, o processo de *krigagem* implica numa suavização dos valores preditos para a região em estudo não fornecendo uma estimativa da dispersão destas variáveis, pois a variância de *krigagem* avalia apenas a configuração da vizinhança.

Repetindo o processo de *krigagem* ordinária em vários pontos de modo a formar uma malha fina é possível obter um mapa das estimativas na região estudada facilitando a interpretação quanto ao comportamento espacial da variável.

2.5 Validação cruzada

O preditor de *krigagem*, como visto anteriormente, considera que o modelo do semivariograma é correto e que não há erros de medida. A partir do pressuposto ele é obtido de forma a atender duas qualidades desejáveis em um preditor a não tendenciosidade e a variância mínima. Mas como avaliar se o modelo escolhido está correto?

A validação cruzada apresenta-se como uma ferramenta adequada, pois a idéia é: se o fenômeno foi satisfatoriamente modelado é possível reproduzir, com boa aproximação, informações da amostra (Vieira, 1996).

O procedimento da validação cruzada consiste em retirar um ponto do conjunto original de dados e estimá-lo a partir do restante. Utilizando a *krigagem*, é também estimada a variância da estimativa retornando o ponto ao conjunto estima-se o próximo ponto repetindo o processo para todos os pontos amostrados (Ribeiro Jr, 1995).

Em cada ponto tem-se o erro de estimação e padronizando-o pelo desvio padrão da estimativa espera-se, conforme Gonçalves (1997), que os erros deduzidos de *krigagem* tenham média nula e variância unitária. Os valores das estatísticas obtidas a partir dos erros e dos gráficos mostram o comportamento das estimativas e são utilizados na avaliação.

Interessante observar que a validação cruzada não avalia simplesmente o modelo de semivariograma escolhido, mas toda a modelagem do processo estocástico em questão, isto é, são colocados à prova a decisão de estacionaridade, os estimadores utilizados, o tratamento dos dados atípicos, o modelo de semivariograma adotado e a decisão relativa a anisotropia.

Nem sempre é fácil identificar a causa dos problemas apresentados nos resultados da validação cruzada, mas com a união da experiência, conhecimento dos problemas em questão e análise minuciosa dos diversos recursos descritivos, em especial os gráficos, pode-se identificar as origens dos problemas.

3 MATERIAL E MÉTODOS

Os dados são provenientes de um conjunto de medidas de densidade e umidade do solo em diferentes tensões descritos por Moraes (1991), com o propósito de ilustrar a aplicação de algumas técnicas geoestatística sem se preocupar com a discussão de aspectos específicos de física do solo.

Os programas computacionais utilizados são: Geo-EAS, “Geostatistical Enviromental Assessment Software” de Englund & Sparks (1988) e Variowin (Pannatier, 1996) para a análise geoestatística; Statistica e Surfer para os cálculos estatísticos descritivos e obtenção dos gráficos de superfícies, respectivamente.

3.1 Localização e delineamento experimental

O experimento refere-se a uma área no campo experimental do Departamento de Ciências Exatas do Campus da ESALQ/USP, no município de Piracicaba/SP. A área experimental possui um relevo suave, com solo classificado como Nitossolo Vermelho distrófico latossólico (EMBRAPA, 1999), com dimensões de 125m por 50m, totalizando uma área de $6250m^2$. O tipo de amostragem utilizada foi a amostragem sistemática com 250 pontos. As amostras do solo foram coletadas a uma profundidade de 25cm, com espaçamento regular de 5m, resultando em uma malha de pontos amostrados de 25 linhas por 10 colunas (Figura 6).

Em cada ponto foram amostradas medidas das variáveis: densidade do solo ($kg.m^{-3}$), umidade às tensões 9,8kPa, relacionado a tensão na capacidade de

campo, e 1,5MPa que corresponde a tensão do ponto de murcha permanente.

Figura 6 - Área experimental com a localização amostral dos 250 pontos.

3.2 Análise estatística

O ponto de partida para a análise dos dados espacialmente distribuídos foi a divisão da análise em duas etapas: a primeira consistiu em uma descrição geral, sem considerar as posições das amostras, para a identificação de tendências e pontos discrepantes foram utilizadas as análises exploratórias descritivas e gráficas.

A outra etapa considerou as posições das amostras, em que as análises semivariográficas amostral e modelada e autocorrelográfica foram aplicadas para estudar a magnitude e a estrutura da variabilidade espacial das variáveis com avaliação do ajuste do modelo de semivariograma por meio da validação cruzada. Sendo possível, então, interpolações de áreas não amostradas através da *krigagem* e confecção de mapa de superfície de valores.

3.2.1 Análises exploratórias descritivas

A decisão quanto ao tipo de estacionaridade deve ser assumida é uma afirmação ou suposição a respeito do comportamento das variáveis aleatórias e depende do entendimento e conhecimento do fenômeno em estudo. Apesar da hipótese de estacionaridade não ser testada, a decisão de assumir algum tipo de estacionaridade pode ser orientada por uma análise exploratória adequada dos dados. A falta de estacionaridade pode ser tanto na média quanto na variância. Isto justifica a elaboração dos gráficos que trazem informações não só de medidas de posição mas também de dispersão.

As técnicas descritivas exploratórias devem envolver todas as diversas fases da análise, e não serem utilizadas apenas inicialmente. Análise exploratória através de gráficos de dispersão, “box-plot”, histogramas e gráfico de probabilidade normal foi utilizada para identificação de tendências, possíveis pontos discrepantes, forma da distribuição, escolha da análise a ser aplicada, decisão do tipo de estacionaridade que pode ser assumida e caracterização da variabilidade.

O histograma consiste em retângulos com base nas faixas de valores da variável e com área igual à frequência relativa da faixa respectiva, desta forma, é possível obter informações a respeito de como os dados se organizam e ainda tem-se a relação com o gráfico da função de probabilidade. O histograma também pode ser utilizado no cálculo da mediana, que é o valor da variável que divide o conjunto de dados ordenados em dois grupos.

A definição de mediana pode ser estendida para situações em que dividi-se o conjunto de dados em mais de dois sub-grupos. Um caso especial é quando o conjunto é dividido em quatro sub-grupos iguais obtendo com isso mais dois valores, em que, 25% das observações estão abaixo de um deles e 75% estão abaixo do outro, estes valores são denominados de primeiro quartil e terceiro quartil, respectivamente.

A representação gráfica envolvendo os quartis é o “box-plot”, que é

uma “caixa” com o nível superior dado pelo terceiro quartil e o nível inferior pelo primeiro quartil. O sinal no interior da “caixa” corresponde a mediana e segmentos de retas são construídos da “caixa” até os valores máximo e mínimo, que não são necessariamente observações discrepantes. A representação gráfica por “box-plot” informa também a variabilidade, simetria dos dados e identifica descritivamente, diferenças nos comportamentos de grupos de variáveis.

Existem diversos critérios para decidir se uma observação pode ser considerada discrepante ou não, um deles, o qual é utilizado neste trabalho, é o critério dos limites inferior e superior da variável, em que valores fora destes limites são considerados discrepantes.

O limite inferior é calculado subtraindo-se do quartil inferior uma vez e meio a amplitude interquartílica, que é a diferença entre os quartis superior e o inferior. Ao passo, que o limite superior é obtido somando-se ao quartil superior uma vez e meio a amplitude interquartílica.

Dados atípicos podem ser identificados nas análises descritivas necessitando com isso maior detalhamento da origem do valor, se o dado é consequência de um erro grosseiro, erro de anotação, leitura, entre outros, deve ser corrigido se possível, caso contrário eliminado. Caso o valor atípico não seja devido a erros grosseiros é necessário buscar explicações para seu comportamento estudando os impactos nos resultados das análises na presença ou ausência deles.

A localização dos dados atípicos na região em estudo deve ser considerada, pois quando os valores discrepantes ocorrem de forma concentrada, pode indicar sub-regiões com comportamentos diferenciados da área total. Ou quando ocorre de forma dispersa pode se tratar de um fenômeno com elevada variabilidade. Mostrando-se importante a análise descritiva espacial dos dados em estudo.

Outras informações utilizadas na análise são obtidas através do gráfico de probabilidade normal e dos testes de aderência que aqui são os de Lilliefors que é específico para testar se os dados seguem distribuição normal e o de Shapiro-Wilk que pode ser utilizado para testar qualquer outra distribuição de probabilidade

apresentando maior rigor para testar normalidade quando comparado ao de Lilliefors.

A análise exploratória por meio das ferramentas citadas anteriormente não tem preocupação com a distribuição de probabilidade exata das variáveis em estudo, mas verifica se seria razoável estudar os dados através da distribuição normal, admitida como não discrepante da verdadeira distribuição que é desconhecida.

3.2.2 Análise exploratória espacial

Esta análise é mais detalhada, pois ao se estudar o comportamento da variável, as posições das amostras foram consideradas em algum momento da análise. O “postplot”, gráfico de superfície de valores e outras ferramentas que consideram as posições das amostras foram técnicas utilizadas na análise exploratória espacial.

O “postplot” é um gráfico construído pelo programa Geo-EAS em que os dados são representados nos pontos onde foram amostrados. Para cada valor foi associado um símbolo indicando a qual classe ele pertence. Este gráfico constituiu-se em um instrumento de verificação de tendências associadas às posições espaciais. Concentrações de símbolos iguais podem indicar sub-regiões e variações gradativas ao longo de alguma direção indicam a existência de tendências.

Para atender à hipótese intrínseca não devem haver padrões neste gráfico. O “postplot” fornece uma idéia inicial da continuidade do fenômeno, identificando regiões com concentração de amostras de elevados ou baixos valores que sugerem a divisão da área em sub-regiões.

Outra ferramenta é o gráfico de superfície de valores que apresenta a superfície das medidas da variável definida pela amostragem. A análise da superfície é grosseira e fornece apenas uma idéia do comportamento da variável na região. Uma superfície mais detalhada é obtida após a construção por interpolações de uma malha mais fina que a amostrada.

Gráficos dos valores amostrais em relação as linhas ou colunas são métodos alternativos para identificar tendência ou sub-regiões.

Outra ferramenta para a investigação de dados espacialmente atípicos é a análise do efeito proporcional, que é o aumento da variância proporcionalmente ao aumento da média. Na natureza este efeito proporcional é observado em fenômenos com distribuições assimétricas, lognormal por exemplo.

A análise do efeito proporcional consiste em construir gráficos de dispersões da média e da variância no sentido das linhas e colunas. Em uma situação que não apresenta tendência, tem-se uma expectativa dos pontos em torno da média. Para o caso de dispersão que varia no espaço a expectativa dos pontos afasta-se da média gradualmente.

3.2.3 Análise geoestatística

Na geoestatística a análise do semivariograma é uma etapa importante, pois o modelo de semivariograma escolhido é a interpretação da estrutura de correlação espacial a ser utilizada nos procedimentos inferenciais da *krigagem*. A análise completa do semivariograma compreende os seguintes passos:

- Levantamento do semivariograma experimental: o cálculo do valor do semivariograma em cada distância é realizado utilizando os dados amostrais da variável regionalizada $Z(s_i)$ na equação (3), que é o estimador do semivariograma. O semivariograma experimental, quando apresenta dependência espacial, é uma curva irregular com flutuações que crescem com os valores de $\mathbf{h}$. Os últimos pontos deste gráfico têm menor significância estatística, pois envolvem poucos pares de pontos.

A curva experimental obtida é na realidade um estimador do verdadeiro semivariograma desconhecido, este possui propriedades matemáticas precisas. Portanto fazendo-se necessário o ajuste de um modelo teórico que sirva de base para os cálculos posteriores;

- Ajuste a uma família de modelos de semivariogramas: o modelo ajustado ao semivariograma experimental está relacionado com o comportamento do semivariograma na origem e a existência ou não de um patamar. Ao modelo une-se o

efeito pepita por extrapolação da curva.

A adequação de um modelo teórico é fundamental, pois a partir deste modelo serão feitas inferências em relação ao semivariograma verdadeiro;

- Validação do modelo a ser utilizado nos procedimentos da *krigagem*: a adequada modelagem da estrutura de dependência espacial valida qualquer outro procedimento de inferência e interpolação. Além de ser um forte subsídio para decisões práticas sobre o fenômeno em estudo.

Conhecido o semivariograma da variável em estudo e havendo dependência espacial entre as amostras, o próximo passo, é obter informações de pontos não amostrados no campo através do método de interpolação denominado de *krigagem*.

Os métodos de *krigagem* são métodos de interpolação que procuram minimizar o erro da estimação; na realidade, o erro médio de estimação é nulo. O problema que se coloca normalmente é o de estimar o valor de uma variável em locais não amostrados, $Z(s_0)$, a partir dos valores de locais amostrados, $Z(s_i)$. O estimador de *krigagem* $\hat{Z}(s_0)$, é também um estimador linear que considera a organização espacial da variável:

$$\hat{Z}(s_0) = \sum_{i=1}^{N(h)} \lambda_i Z(s_i),$$

em que λ_i é o ponderador da distância e $\sum_{i=1}^{N(h)} \lambda_i = 1$.

A *krigagem* é um método exato e não viesado, isto é, os valores nos locais amostrados são reproduzidos e o erro médio de estimação é nulo.

A análise geoestatística, em algumas situações, exige um certo grau de subjetividade, sendo assim, para avaliar se o fenômeno foi modelado de forma adequada, alguns autores têm utilizado o método da validação cruzada.

Seja $Z(s)$ uma função aleatória de segunda ordem com semivariograma $\gamma(h)$ e função de covariância $C(h)$. Sejam $z_1, z_2, \dots, z_n$ os valores de $Z(s)$ em n pontos. A validação cruzada consiste em suprimir o i -ésimo valor medido z_i e estimá-lo a partir do restante dos dados. O valor estimado neste ponto, é calculado pela *krigagem*

ordinária. Em seguida, retorna-se o ponto ao conjunto de dados e retira-se o segundo e assim sucessivamente para os demais pontos amostrados, de forma que para cada ponto é possível obter o erro de estimação pela distribuição normal padrão, com média zero e variância unitária.

É importante ressaltar que a análise geoestatística não se trata de um método contínuo em uma direção. A análise geoestatística consiste em ir e voltar, refazer e comparar antes de qualquer decisão definitiva. Várias decisões são tomadas ao longo do processo, ou seja, os resultados intermediários e finais não são obtidos de forma única e as técnicas descritivas e exploratórias devem estar em todas as fases do processo de análise e não somente na fase inicial.

A autocorrelação expressa a correlação linear entre uma série espacial, ou temporal, e a mesma série num intervalo do espaço, ou do tempo. Procurando definir a distância de separação entre observações além da qual não existe correlação espacial entre os pares de valores.

O autocorrelograma, ferramenta gráfica da função autocorrelação, é mais sensível que o semivariograma no sentido que exige uma hipótese de estacionaridade mais forte a de segunda ordem. As amostras de cada variável física são espacialmente independentes se todos os valores de $r(h)$ estão próximos de zero.

Os valores de $r(h)$ para as distâncias iniciais de cada autorrelograma são ajustados pelo método dos quadrados mínimos da equação

$$r(h) = r_0 \exp \frac{-h}{\kappa} \quad r(h) > 0, \quad (7)$$

em que $r_0 = 1$, h é a distância e o autocorrelograma é diminuído para e^{-1} que corresponde a distância onde $h = \kappa$. Para $r_0 = 1$, a magnitude de κ é calculada por

$$\sum_{i=1}^{n-1} [r_i(h) - \exp \frac{h}{\kappa}] = 0,$$

em que n é o número de observações.

Para estimar o comprimento ou alcance da autocorrelação, ou seja, a distância além da qual as observações são independentes umas das outras adotou-se

dois critérios: o primeiro, a distância foi obtida visualmente lendo-se o valor aproximado de h para a qual o correlograma aproxima-se de zero, neste caso esse valor deve ter relação próxima com o alcance do semivariograma.

O outro critério, é o conceito de escala integral (E.I.), em que o seu valor representa a distância dentro da qual as observações são correlacionadas (Vauclin et al., 1982), ou seja, a área sob a função $r(h)$. Em uma escala unidimensional esta área corresponde a

$$E.I. = \int_0^{\infty} r(h)dh. \quad (8)$$

Aplicando a equação (7) em (8) obtém-se:

$$E.I. = r_0\kappa. \quad (9)$$

O valor da E.I. (9) não corresponde ao rigor conceitual da equação (8), visto que a equação (7) implica na função autocorrelação zero no infinito.

4 RESULTADOS E DISCUSSÃO

4.1 Análise exploratória descritiva

As estatísticas descritivas para densidade do solo ($kg.m^{-3}$) e umidade à base de massa nas tensões de 9,81kPa referente à capacidade de campo (Cc) e 1,5MPa que corresponde à tensão 1,5MPa (Pmp) estão apresentados na Tabela 1.

A média e mediana apresentaram entre si valores próximos para as três variáveis. Os coeficientes de simetria e curtose da variável umidade do solo em ambas as tensões foram valores próximos de zero, indicando uma distribuição mais simétrica.

Para os dados de densidade do solo o coeficiente de curtose foi superior a uma unidade (1,37) o que indica um comportamento leptocúrtico da curva de distribuição, ou seja, a assimetria pode não ter sido afetada por valores extremos não existindo a necessidade de substituí-la por outra medida de tendência central (Webster, 1977).

Os valores do mínimo e do máximo para a variável densidade do solo estão fora do intervalo dado pelos limites inferior e superior, sugerindo possíveis candidatos a valores discrepantes (Tabela 1).

Para as estatísticas da umidade do solo, não se observou-se medidas atípicas e a distribuição dos dados apresentou-se com boa simetria, em que a média e a mediana aproximaram-se de uma constante.

Observando os valores do CV %, segundo Warrick & Nielsen (1980) pode-se classificar as propriedades do solo como tendo baixa, média e alta variação.

Tabela 1. Análise estatística descritiva para as variáveis densidade do solo (DS) em $kg.m^{-3}$, umidade do solo às tensões 9,8kPa (Cc) e de 1,5MPa (Pmp).

	DS	Cc	Pmp
Nº de observações	250	250	250
Média	1490,4	26,24	19,51
Mediana	1490	26,49	19,62
Somatório	372600	6558,97	4876,55
Mínimo	1270	20,60	14,72
Máximo	1730	31,39	23,54
Quartil Inferior	1450	24,52	17,66
Quartil Superior	1530	27,47	20,60
Amplitude total	460	10,79	8,83
Amplitude interquartílica	80	2,94	2,94
Variância	5060,88	3,99	3,93
Desvio padrão	71,14	2,00	1,98
Simetria	0,37	-0,07	-0,05
Curtose	1,37	-0,25	-0,78
Limite Inferior	1330	20,11	13,24
Limite Superior	1650	31,88	25,02
C.V.(%)	4,77	7,61	10,16
Lilliefors	($p < 0,01$)	($p < 0,01$)	($p < 0,01$)
Shapiro-Wilk	0,9718($p < 0,01$)	0,9656($p < 0,00$)	0,9448($p < 0,00$)

De maneira geral, os resultados da Tabela 1 apresentaram, para todas as variáveis estudadas, pequena magnitude de variação com coeficientes de variação menores ou aproximadamente iguais a 10% em que a menor dispersão das medidas em torno da média (CV=4,77%) foi observada para a variável densidade do solo.

A variável umidade do solo em ambas as tensões apresentou baixa

variação ($CV < 12\%$), o aumento do coeficiente de variação da tensão 9,8kPa para 1,5MPa não foi acompanhado por um aumento da amplitude total e desvio padrão. Isto indica que para caracterizar a dispersão o parâmetro estatístico não deve ser único.

Os resultados dos teste de Shapiro-Wilk e Lilliefors foram semelhantes entre si para todas as variáveis, rejeitando a hipótese de nulidade em favor da hipótese alternativa, de que os dados não seguem uma distribuição normal. No entanto, observando graficamente os valores das variáveis em estudo (Figuras 7 e 8), a distribuição normal de probabilidade acaba por ser adotado na prática, isto ocorre porque os testes de aderência foram rigorosos ao testarem a distribuição normal.

Figura 7 - Histograma de frequência (a) “box-plot” e (b) gráfico de probabilidade normal (c) para os dos dados de densidade do solo ($kg.m^{-3}$).

O histograma de frequência e o gráfico “box-plot” apresentaram os valores da média e mediana coincidindo e foram deslocados do centro da distribuição, indicando uma assimetria positiva (Figuras 7, 8 e 9a).

Figura 8 - “Box-plot” para os dados de umidade do solo nas tensões 9,8kPa (Cc) e 1,5MPa (Pmp).

Figura 9 - Histogramas de frequência (a) e gráficos de probabilidade normal (b) para os dados de umidade do solo nas tensões 9,8kPa (Cc) e 1,5MPa (Pmp), respectivamente.

O gráfico de probabilidade normal para os dados de densidade do solo demonstrou que a maioria dos dados estavam sobre a reta com alguns pontos afastados, o que indica que não deve ser grande o número de valores periféricos candidatos a “outliers” (Figura 7b). Para os valores de umidade do solo em ambas as tensões não se observou valores que se afastassem sensivelmente da reta de probabilidade normal (Figura 9b).

4.1.1 Análise exploratória espacial

Para melhor entendimento, os resultados da densidade e umidade do solo foram discutidos separadamente.

- Densidade do Solo

Observando a superfície dos valores da variável densidade do solo, observou-se uma aglomeração das medidas para valores superiores a 1600 kg.m^{-3} em uma região distinta localizada à direita superior do gráfico (Figura 10a).

Este fato pode estar associado a remoção de camadas superficiais do solo nesta parte do terreno, expondo o horizonte mais profundo e compactado levando a valores maiores de densidade do solo. Na Figura 10b observou-se pequenas concentrações de valores maiores que 1650 kg.m^{-3} , a qual representa o limite superior, nas primeiras linhas (SO-NE) e nas últimas colunas (NO-SE).

A dispersão nas linhas e colunas das medidas de densidade do solo (Figura 10b) concordou com a análise gráfica da superfície de valores da Figura 10a, onde os valores maiores que o limite superior foram encontrados nas primeiras linhas das ordenadas e últimas colunas das abscissas, confirmando área de maior densidade do solo.

Conforme Moraes (1991) os valores discrepantes de densidade do solo não são resultados de erros de leitura ou sistemáticos, mas devido a localização da extremidade da área estar próxima a uma estrada com tráfego de automóveis. E sendo o fenômeno de dependência a curta escala existiu interferência da compactação

Figura 10 - Superfície dos valores (a) e dispersão das medidas (b) de densidade do solo ($kg.m^{-3}$) amostradas nos sentidos das SO-NE e NO-SE.

da estrada sobre as medidas mais próximas à estrada, portanto, se justifica a presença das medidas mais densas as quais devem ser incluídas no conjunto para a análise.

Devido a análise estatística descritiva apontar um comportamento um pouco assimétrico do conjunto de dados de densidade do solo, a existência do efeito proporcional é investigada por meio da Figura 11, em que as medidas de posição maiores corresponderam, ainda que em poucos pontos, as medidas de dispersão mais elevadas. A Figura 11 não apresentou o efeito proporcional nítido, podendo assumir um comportamento estabilizado em torno de uma constante com alguns pontos mais afastados.

O “postplot” dos valores de densidade do solo (Figura 12), exceto pela concentração que ocorre à direita do gráfico, a qual já havia sido comentada nas figuras anteriores, pareceu estar de acordo com a hipótese intrínseca não apresentando tendências.

Figura 11 - Gráficos de dispersão das médias e variâncias dos valores de densidade do solo ($kg.m^{-3}$) amostradas nos sentidos das linhas (a) e colunas (b).

Figura 12 - “Postplot” dos dados de densidade do solo ($kg.m^{-3}$).

- Umidade do Solo

A superfície de valores da variável umidade do solo, apresenta uma diminuição nos valores na região central e à direita desta região as medidas de umidade tendem a aumentar. Este comportamento foi semelhante em ambas as tensões.

A dispersão nas linhas e colunas das medidas de umidade do solo em ambas as tensões (Figuras 13b e 14b) concordam e detalham melhor a análise gráfica da superfície de valores, revelando as posições dos maiores medidas de umidade nas regiões mais afastadas do centro das figuras.

Figura 13 - Superfície de valores (a) e dispersão dos dados nas linhas e colunas (b) da umidade do solo na tensão 9,8kPa (Cc).

O comportamento na direção NO-SE para a variável umidade do solo na tensão 9,8kPa descreveu claramente a região central comentada anteriormente, em que há aumento nos valores da umidade nas posições mais externas (Figura 13b).

A variável umidade do solo na tensão 1,5MPa (Figura 14b) apresentou uma uniformização da dispersão, principalmente, na direção SO-NE e na região central da direção NO-SE a diminuição nos valores de umidade é menos acentuada.

Os “postplots” dos valores de umidade do solo (Figura 15), exceto pela concentração que ocorre na parte inferior a direita do gráfico, a qual já havia sido comentada nas figuras anteriores, parece estar de acordo com a hipótese intrínseca não apresentando tendências.

A caracterização dos “postplot” nas tensões estudadas foi a mesma, com a ressalva que para a aplicação da maior tensão 1,5MPa, tem-se menor quantidade de água no solo.

Figura 14 - Superfície de valores (a) e dispersão dos dados nas linhas e colunas (b) da umidade do solo na tensão 1,5MPa (Pmp).

Figura 15 - “Postplot” dos dados de umidade do solo nas tensões 9,8kPa (a) e 1,5MPa (b).

Isso é fato e associado às regiões de maior adensamento do solo descritas pela Figura 12, fica fácil de se observar sub-regiões na Figura 15. Por exemplo, a posição direita inferior que é comum às duas tensões. Apesar da presença de outras sub-regiões, central e a esquerda superior dos gráficos, não existiram tendências muito evidentes no comportamento do fenômeno em estudo.

4.2 Análise geoestatística

4.2.1 Análise semivariográfica

Após o cálculo do semivariograma experimental pode-se tentar um ajuste sobre a nuvem de pontos, obtendo-se um semivariograma modelado, deve-se observar que cada ponto do semivariograma representa a medida de dessemelhança entre vizinhos próximos.

- Densidade do Solo

Inicialmente, considerou-se um comportamento isotrópico sendo avaliado para 0° na direção das abscissas.

Para os dados da densidade do solo o modelo que melhor se ajustou foi o gaussiano (Figura 16) dado pela equação

$$\gamma(h) = C_0 + C[1 - e^{-\frac{h^2}{a^2}}].$$

Os parâmetros do modelo foram efeito pepita ($C_0 = 3400[kg.m^{-3}]^2$), patamar ($C = 800[kg.m^{-3}]^2$) e alcance (a) de 17 metros (Figura 16). Ou seja, existe uma estrutura de dependência espacial entre as amostras até 17 metros.

Outros modelos ajustados com seus respectivos parâmetros podem ser observados na Tabela 2.

O menor valor do efeito pepita, $C_0 = 2650$, foi observado para o modelo exponencial, ou seja, foi modelo exponencial que mais explicou a variabilidade espacial. O modelo que descreveu a distância de maior dependência entre as amostras foi o esférico com alcance de 20 metros.

Figura 16 - Semivariograma para densidade do solo, ajustado ao modelo gaussiano com os seguintes parâmetros: $C_0 = 3400[kg.m^{-3}]^2$; $a = 17m$ e $C = 800[kg.m^{-3}]^2$.

Tabela 2. Modelos de semivariogramas ajustados e seus respectivos parâmetros.

	Exponencial	Esférico	Gaussiano
Número de <i>lags</i>	10	10	10
Efeito pepita - C_0 ($kg.m^{-3}$) ²	2650	3250	3400
Patamar - C ($kg.m^{-3}$) ²	1550	900	800
Alcance (m)	17	20	17

- Umidade do Solo

Considerando, ainda, um comportamento isotrópico sendo avaliado para 0° na direção das abscissas. Os modelos esférico e gaussiano foram os que melhores se ajustaram aos dados de umidade do solo nas tensões 9,8kPa e 1,5MPa, respectivamente.

Os parâmetros do modelo esférico, para a tensão de 9,8kPa, foram efeito pepita ($C_0 = 2,1$), patamar ($C = 1,5$) e alcance de 22 metros. Para o modelo gaussiano, na tensão 1,5MPa, os parâmetros foram efeito pepita ($C_0 = 1,6$), patamar ($C = 2,5$) e alcance de 19 metros (Figura 17).

Figura 17 - Semivariogramas para umidade do solo, ajustados às tensões 9,8kPa (a) e 1,5MPa (b).

Observou-se que a variável umidade do solo na tensão de 9,8kPa apresentou com continuidade espacial mais acentuada quando comparado a densidade do solo, visto que, o alcance da primeira variável ($a = 22m$) é maior que da segunda ($a = 17m$) fato que é confirmado, pois a variável densidade do solo é uma propriedade física com variação à curta escala (Moraes, 1991).

Outros modelos ajustados aos dados da variável umidade do solo, com seus respectivos parâmetros, estão descritos na Tabela 3. Observe que os modelos exponencial e gaussiano para a tensão de 9,8kPa, e o modelo exponencial para a tensão de 1,5MPa apresentaram menores valores do efeito pepita, mas que no entanto, não foram os mais adequados para descreverem a variabilidade espacial, significando que outras estatísticas devem ser consideradas na escolha de um modelo teórico.

Tabela 3. Modelos de semivariogramas ajustados para os dados de umidade do solo nas tensões 9,8kPa (Cc) e 1,5MPa (Pmp) com seus respectivos parâmetros.

	Cc			Pmp		
	Exp.	Esf.	Gauss	Exp.	Esf	Gauss
Nº de <i>lags</i>	10	10	10	10	10	10
Efeito pepita - $C_0 (Pa)^2$	1,8	2,1	1,8	0	2,1	1,6
Patamar - $C(Pa)^2$	2,0	1,5	1,9	4,2	2,1	2,5
Alcance (<i>m</i>)	30	22	28	15	17	19

4.2.2 Estudo anisotrópico

- Densidade do Solo

Visto que variáveis do solo podem apresentar um comportamento anisotrópico, um exame dos semivariogramas para outras direções, 90° na direção das ordenadas; 45° e -45° nas duas diagonais, estão apresentados na Figura 18.

Para o cálculo do semivariograma em 90° na direção das ordenadas, observou-se uma dependência espacial dentro das primeiras distâncias, a partir daí houve estabilização do semivariograma até a distância de 40 metros, então, o semivariograma começa a apresentar um comportamento monótono crescente sem patamar definido revelando que nesta direção a escala de amostragem não mostra toda a dependência espacial do fenômeno na direção das ordenadas (Figura 18a).

O comportamento do semivariograma na diagonal 45° é totalmente atípico, com um crescimento inicial, a partir do qual decresce e posteriormente volta a crescer. Esta oscilação da função semivariância, segundo Trangmar et al. (1987), pode ser devida a quantidade de pares de pontos envolvidos no cálculo da função ou a presença de valores atípicos no conjunto de dados. No caso, as duas hipóteses podem ter colaborado para o comportamento não estruturado do semivariograma

(Figura 18b).

Figura 18 - Semiariogramas da densidade do solo $(kg.m^{-3})^2$ a 90° (a), 45° (b) e -45° (c), respectivamente, na direção das ordenadas.

Com relação a diagonal -45° observou-se um crescimento constante da função semivariograma, não apresentando patamar definido, fato que pode estar associado a escala de amostragem, que não foi suficiente para descrever toda a

estrutura de dependência espacial na área nesta direção (Figura 18c).

- Umidade do Solo

O comportamento anisotrópico é comum em fenômenos naturais, fato que não foi diferente para a variável umidade do solo. A anisotropia dos semivariogramas da umidade do solo nas tensões de 9,8kPa e 1,5MPa nas direções, 90° na direção das ordenadas; 45° e -45° nas duas diagonais, são apresentados nas Figuras 19 e 20.

Para a umidade do solo na tensão 9,8kPa na direção de 90° das ordenadas, observou-se dependência amostral em uma escala maior, aproximadamente 40 metros, e houve estabilização da semivariância em torno do valor de 3,5 (Figura 19a).

Para o semivariograma da diagonal 45° observou-se uma dependência espacial até os trinta metros aproximadamente, a partir daí o semivariograma decresce estabilizando-se em torno de um patamar igual a três. Revelando que nesta direção existe um comportamento anisotrópico que também deve ser considerado no modelo (Figura 19b).

A diagonal -45° é a direção que apresenta melhor estrutura do semivariograma experimental, observando crescimento da função semivariograma com patamar definido. Caracterizando bem a dependência espacial entre as amostras e demonstrando que a escala de amostragem foi adequada para descrever a estrutura de dependência espacial na área nesta direção (Figura 19c).

Para o cálculo do semivariograma da umidade do solo na tensão de 1.5MPa na direção de 90° das ordenadas, observou-se dependência espacial com um efeito pepita muito pequeno, próximo de uma unidade que cresceu rapidamente e estabilizou-se em torno de um patamar igual quatro. O alcance da dependência espacial foi em torno de 20 metros. Estes valores descreveram um fenômeno de variáveis regionalizadas que se tornam diferentes em uma escala muito curta (Figura 20a).

Figura 19 - Semivariogramas da umidade do solo na tensão 9,8kPa (Cc) a 90° (a) na direção das ordenadas, 45° (b) e -45° (c) nas duas diagonais, respectivamente.

Fato que não foi observado quando se estudou o comportamento do semivariograma da diagonal 45° . Existe uma dependência espacial até os 20 metros, a partir daí o semivariograma estabilizou-se em torno de um patamar aproximada-

mente igual ao valor do efeito pepita 3,8. O comportamento anisotrópico revelado nesta direção não foi tão bem estruturado quanto o observado na direção anterior (Figura 20b).

Figura 20 - Semivariogramas da umidade do solo na tensão 1,5MPa (Pmp) a 90° (a) na direção das ordenadas, 45° (b) e -45° (c) nas duas diagonais, respectivamente.

Para a diagonal -45° observou-se um comportamento semelhante a direção de 90° , a qual apresentou um semivariograma experimental bem estruturado. O crescimento da função semivariograma para um patamar definido foi rápido nas primeiras distâncias estabilizando-se em torno do valor igual a quatro (Figura 20c).

4.2.3 Autocorrelograma

- Densidade do Solo

O autocorrelograma dos dados de densidade do solo da Figura 21, apresentou um decréscimo da correlação entre as observações, a medida que a distância entre os pares das observações aumentou, mostrando dependência espacial das amostras até aproximadamente a distância visualmente determinada, correspondente à cerca de 20 metros.

Figura 21 - Autocorrelograma para os dados de densidade do solo ($kg.m^{-3}$)

Fato que concorda com o caso (A) da Figura 5, em que o valor de $r(h)$ declina gradativamente nos primeiros *lags* igualando a zero e a partir do que, mostra a independência entre as amostras.

O alcance é aproximadamente igual ao já obtido pelo semivariograma, mostrando que função de autocorrelação para a variável densidade do solo apresentou relação com o comportamento da semivariância e existindo a função de correlação pode-se assumir uma hipótese de estacionaridade mais forte, a de segunda ordem.

Quando se estimou o valor do alcance da autocorrelação (κ) pelo critério da escala integral foi observado uma sensível diminuição na distância dentro da qual as observações são correlacionadas, ou seja, por este critério a distância a partir da qual as observações tornam-se independentes uma das outras corresponde a 2,44 metros.

- Umidade do Solo

Para os dados de umidade do solo em diferentes tensões (Figura 22), houve um decréscimo da correlação entre as observações e a medida que a distância entre os pares das observações aumentou, observando dependência espacial das amostras até aproximadamente o décimo *lag*, na distância correspondente a aproximadamente 50 metros, em ambas as tensões, pelo método da leitura direta da distância.

Figura 22 - Autocorrelograma para os dados de umidade do solo nas tensões 9,8kPa (a) e 1,5MPa (b).

Utilizando o critério da escala integral o alcance do autocorrelograma é menor. Para a variável umidade do solo nas tensões 9,8kPa e 1,5 MPa os valores de κ foram iguais a 7,62m e 16,49m, respectivamente. Isto significa que utilizando o critério de E.I. a umidade do solo na tensão de 1,5MPa apresentou maior dependência espacial das amostras.

Ambos os autocorrelogramas apresentados na Figura 22 podem ser interpretados como espacialmente dependentes. O comportamento do autocorrelograma (A) da Figura 5, é o mais típico e pode estar associado ao comportamento da variável umidade do solo na tensão 9,8kPa, em que o valor de $r(h)$ declina gradualmente até nivelar-se em torno de zero, mostrando a parti daí a independência entre as amostras.

O caso (C) da Figura 5 descreve melhor o comportamento da variável umidade do solo na tensão 1,5MPa, em que o coeficiente de autocorrelação diminui lentamente indicando independência entre as amostras a longo alcance.

Observa-se que o alcance visualmente determinado teve relação com os semivariogramas e que o autocorrelograma também demonstrou que a variável umidade do solo apresentou maior dependência espacial das amostras quando comparada a variável densidade do solo.

4.2.4 Interpolação por krigagem

Modelos isotrópicos e anisotrópicos foram considerados, contudo para aplicação da técnica de interpolação por *krigagem* proposta inicialmente optou-se pelo modelo isótrópico, pois apesar de detalhar menos a estrutura espacial de correlação sua aplicação é mais simples, uma vez que possui menos parâmetros.

- Densidade do Solo

A *krigagem* para densidade do solo na área em estudo considerou o modelo ajustado para o semivariograma. Os resultados apontaram que a maior concentração dos valores ocorre para valores de densidade menores que $1530 kg.m^{-3}$, mostrando uma área de maior densidade do solo na sub-região à direita inferior (Figura 23).

A superfície dos valores de densidade do solo estimados pelo método de *krigagem* apresentou a mesma sub-região descrita na análise da superfície de valores, no entanto, foi apresentado um efeito suavizador das diferenças, visto que o

interpolador de *krigagem* superestima para baixas medidas e subestima para elevados valores, ou seja, tende à média dos dados (Figura 24).

Figura 23 - Área interpolada para densidade do solo ($kg.m^{-3}$) pelo método de *krigagem* ordinária pontual considerando o modelo gaussiano.

Figura 24 - Superfície de valores de densidade do solo ($kg.m^{-3}$), interpolada pelo método de *krigagem*.

- Umidade do Solo

Quanto aos dados de umidade do solo na área em estudo foram considerados os modelos esférico e gaussiano, ajustados para as tensões 9,8kPa e 1,5MPa,

respectivamente.

O padrão de distribuição da umidade do solo não seguiu o mesmo comportamento quando aplicamos diferentes tensões de retenção de água no solo, visto observarmos algumas diferenças nos contornos das sub-regiões (Figura 25).

Figura 25 - Áreas interpoladas para umidade do solo em ambas tensões 9,8kPa (a) e 1,5MPa (b) pelo método de *krigagem* ordinária pontual.

Existem regiões que perdem água mais facilmente quando comparadas com outras (Figura 26a), pois a medida que aplicamos pressão sobre o solo para retirar água, a umidade restante nas partículas permanecem presas às tensões muito maiores, chegando ao ponto em que não é mais possível a retirada de água.

A superfície dos valores de umidade do solo estimados pela *krigagem* apresentou a mesma sub-região apontada na análise da superfície de valores, observando um predomínio de área onde os valores de umidade do solo estão entre as tensões de 19 e 26Pa (Figura 26) e existe um aumento da região onde a umidade do solo foi retida às pressões maiores, $\geq 20,6$ Pa (Figura 26b).

Diferentes tensões apresentaram comportamentos particulares, visto que áreas com maior quantidade de água no solo na tensão da capacidade de campo (9,8kPa) nem sempre corresponderam a mesma região da superfície de valores na

tensão do ponto de murcha permanente (1,5MPa).

Figura 26 - Superfícies de valores de umidade do solo nas tensões 9,8kPa (a) e 1,5MPa (b), interpoladas pelo método de *krigagem*.

Do ponto de vista prático significou que um desconhecimento da regionalização do fenômeno poderiam levar a superestimativas (pequeno valor de umidade no Pmp e grande valor na Cc) ou subestimativas (o inverso) provocando excesso de água em algumas sub-áreas e falta em outras, portanto ineficiência da irrigação.

Ressalta-se ainda, que a partir do conhecimento do semivariograma o planejamento de futuras amostragem para determinações de água disponível pode ser melhor dimensionado.

4.2.5 Validação cruzada

- Densidade do Solo

Um modelo com ajuste adequado foi aquele que apresentou, na validação cruzada, seus erros padronizados com distribuição próxima da normal padrão. Três modelos foram testados e a suas estatísticas estão na Tabela 4.

Observou-se que o modelo mais adequado para descrever o fenômeno de dependência espacial da variável densidade do solo foi o modelo gaussiano, apesar do modelo exponencial ter apresentado menor valor do efeito pepita, $C_0 = 2650$.

Tabela 4 - Resumo estatístico da validação cruzada para os erros padronizados de três modelos ajustados ao semivariograma experimental da variável densidade do solo.

	Exponencial	Esférico	Gaussiano
Média ($kg.m^{-3}$)	-0,004	-0,004	-0,003
Variância ($kg.m^{-3}$) ²	0,994	1,014	1,00

Lembrando que nos mapas dos erros de *krigagem* os pontos marcados com sinal “+” correspondem aos valores superestimados e “x” aos valores subestimados, quanto maior o tamanho do “x” maior a descontinuidade espacial dos dados. A Figura 27 apresenta o gráfico de dispersão dos erros de *krigagem* com aparente uniformidade e sem tendência nos valores calculados. O desvio padrão e média dos erros padronizados da validação cruzada para o modelo gaussiano foram de 1 e -0,003, respectivamente.

Figura 27 - Mapa de erros dos valores observados em relação aos estimados ajustado ao modelo gaussiano com 16 vizinhos.

A dispersão dos erros positivos e negativos foram semelhantes (Figura 28), observando maior afastamento da reta 1:1 para os erros positivos nas menores

medidas de densidade do solo. Quanto aos erros negativos as subestimativas da variável foram observadas nos maiores valores de densidade do solo e os demais equilibraram-se entre os valores superestimados e os subestimados, fato que concorda com a característica de não viés do estimador de *krigagem* (Figura 28b).

Figura 28 - Gráficos de dispersão das estimativas de *krigagem* em relação aos valores observados (a) e as diferenças padronizadas (b) na validação cruzada para a variável densidade do solo ($kg.m^{-3}$) no modelo gaussiano.

Outras estatísticas da validação cruzada para densidade do solo podem ser vistas na Tabela 5.

Tabela 5. Resumo estatístico da validação cruzada para o modelo gaussiano com 16 vizinhos.

	$Z(s)$	$\hat{Z}(s)$	$Z(s) - \hat{Z}(s)$	Desvio padrão da estimação	Diferenças padronizadas
Média	1490,4	1490,20	-0,22	62,51	-0,003
Desvio padrão	71,14	34,62	62,50	0,61	1,00
Mínimo	1270	1430,48	-151,59	62,15	-2,430
1º quartil	1450	1465,47	-39,62	62,16	-0,633
Mediana	1490	1486,16	-3,52	62,16	-0,057
3º quartil	1530	1503,76	37,07	63,29	0,596
Máximo	1730	1622,56	201,18	65,19	3,235

Comparando os valores amostrados, $Z(s)$ com os valores estimados, $\hat{Z}(s)$, observou-se menor amplitude de variação para os valores estimados, com menor desvio padrão ($34,62kg.m^{-3}$), isto ocorreu porque na validação cruzada os valores estimados são médias dos vizinhos e, portanto, as variações foram amenizadas.

As diferenças padronizadas devem ter, em uma situação ideal, média nula, variância unitária e ser normalmente distribuídas. As estatísticas assinalaram valores próximos do ideal (Figura 29). A padronização foi realizada a partir do desvio padrão de *krigagem* obtido para cada um dos pontos da malha.

Figura 29 - Histograma de frequência dos erros padronizados da densidade do solo ($kg.m^{-3}$), ajustado ao modelo gaussiano com 16 vizinhos.

O histograma de frequências dos erros padronizados apresentou uma pequena assimetria negativa concentrando os valores negativos dos erros, ou seja, a proporção de subestimação é maior que a de superestimação. A distribuição dos erros não se afastou da normal padrão indicando um ajuste muito bom do modelo gaussiano, ou seja, o modelo ajustado descreve adequadamente a dependência espacial dos dados de densidade do solo.

- Umidade do Solo

Três modelos foram testados e as suas estatísticas estão apresentadas na Tabela 6, sendo os modelos mais adequados para descrever o fenômeno de dependência espacial da variável umidade do solo nas tensões 9,8kPa e 1,5MPa os modelos esférico e gaussiano, respectivamente.

Tabela 6. Resumo estatístico da validação cruzada para os erros padronizados de três modelos ajustados ao semivariograma experimental da variável umidade do solo em diferentes tensões 9,8kPa (Cc) e 1,5MPa (Pmp).

	Umidade na Cc		Umidade no Pmp	
	Média	Variância	Média	Variância
Exponencial	0,002	0,914	0,004	0,863
Esférico	0,003	1,040	0,002	0,947
Gaussiano	0,003	1,212	0,001	0,947

A Figura 30 apresenta uma aparente uniformidade e sem tendência nos valores calculados para ambas as tensões. A média e os desvios padrão dos erros padronizados da validação cruzada para os modelos esférico e gaussiano foram 0,003 e 1,020; 0,001 e 0,973, respectivamente, indicando esses resultados uma boa aproximação para a distribuição normal padrão.

Figura 30 - Gráficos de erros dos valores observados em relação aos estimados ajustados aos modelos esférico referente a tensão 9,8kPa (a) e gaussiano referente a tensão 1,5MPa (b).

O gráfico de erros dos valores observados de umidade do solo na tensão 9,8kPa em relação aos valores esperados apresentou maior quantidade de erros nega-

tivos (“x”), ou seja, o estimador de *krigagem* para o modelo esférico tende a subestimar mais os pontos com valores de umidade mais elevados favorecendo uma simetria positiva dos dados (Figura 30).

Os valores estimados dispersaram menos que as medidas observadas e os erros padronizados (Figura 31), isto porque o estimador de *krigagem* suaviza o efeito de valores extremos.

Figura 31 - Gráficos de dispersão das estimativas de *krigagem* em relação aos valores observados e as diferenças padronizadas na validação cruzada para a umidade do solo (Pa) nas tensões 9,8kPa (a) e 1,5MPa (b) nos modelos esférico e gaussiano, respectivamente.

Para a variável umidade do solo na tensão de 1,5MPa (Pmp) o comportamento dos gráficos de dispersão (Figura 31b) apresentou-se mais regular, com pouca dispersão e maior aproximação à reta 1:1, isto significa que o modelo esférico erra pouco quando estima valores próximos da média amostral e as diferenças

padronizadas revelaram um comportamento bastante simétrico.

As Tabelas 7 e 8 apresentam as estatísticas descritivas da validação cruzada, das quais destacaram-se a coincidência dos valores da média e mediana em ambos os modelos avaliados, esférico e gaussiano.

Tabela 7. Resumo estatístico da validação cruzada para o modelo esférico com 16 vizinhos.

	$Z(s)$	$\hat{Z}(s)$	$Z(s) - \hat{Z}(s)$	Desvio padrão da estimação	Diferenças padronizadas
Média	26,24	26,24	0,006	1,580	0,003
Desvio padrão	1,99	1,14	1,606	0,027	1,020
Mínimo	20,6	22,54	-4,59	1,56	-2,93
1º quartil	24,53	25,59	-0,935	1,56	-0,579
Mediana	26,49	26,33	-0,029	1,56	0,018
3º quartil	27,47	26,97	1,04	1,61	0,659
Máximo	31,39	28,86	3,69	1,71	2,36

Tabela 8. Resumo estatístico da validação cruzada para o modelo gaussiano com 16 vizinhos.

	$Z(s)$	$\hat{Z}(s)$	$Z(s) - \hat{Z}(s)$	Desvio padrão da estimação	Diferenças padronizadas
Média	19,51	19,51	0,002	1,660	0,001
Desvio padrão	1,98	1,06	1,62	0,039	0,973
Mínimo	14,72	16,83	-4,53	1,64	-2,73
1º quartil	17,66	18,74	-0,986	1,64	-0,598
Mediana	19,62	19,59	0,039	1,64	0,023
3º quartil	20,60	20,22	1,12	1,71	0,672
Máximo	23,54	22,12	4,87	1,84	2,97

Sendo a *krigagem* uma estimativa do valor esperado de $Z(s_0)$, este resultado garante que a hipótese inicial é satisfeita, ou seja, dado um ponto qualquer s_i a esperança da variável regionalizada $Z(s_i)$ é um constante m , qualquer que seja o ponto i .

O histograma de frequência dos erros padronizados para a variável umidade do solo na tensão 9,8kPa (Cc) apontou uma distribuição bem próxima da normal padrão e com boa simetria. Quanto aos dados da tensão 1,5MPa (Pmp) observou-se uma variabilidade menor que o gráfico da tensão 9,8kPa, destacando uma pequena assimetria positiva com caudas mais leves (Figura 32).

Figura 32 - Histogramas de frequência dos erros padronizados da umidade do solo nas tensões 9,8kPa (a) e 1,5MPa (b).

5 CONCLUSÕES

De acordo com a metodologia desenvolvida e com as análises dos dados das variáveis consideradas, pode-se concluir que:

- os métodos geoestatísticos utilizados são adequados para descrever a estrutura de dependência espacial das variáveis densidade do solo e umidade do solo nas diferentes tensões;
- as análises exploratórias descritiva e espacial identificam claramente o comportamento inicial das variáveis em estudo e auxiliam na decisão do tipo de hipótese de estacionaridade que deve ser assumida;
- tanto para a variável densidade do solo quanto para umidade do solo na capacidade de campo e ponto de murcha permanente existe estrutura de dependência espacial na escala utilizada;
- A variável densidade do solo apresenta magnitude de dependência espacial à escala menor quando comparada com a variável umidade do solo nas diferentes tensões;
- a análise anisotrópica revela a anisotropia dos fenômenos em estudo;
- existe uma estrutura de correlação entre as amostras adjacentes, em que a dependência de um ponto a outro é diminuída à medida que as amostras estão mais distantes.

Como sugestão para a continuidade deste estudo, utilizando métodos estatísticos de estimação mais precisos, estimativas e modelagens semivariográficas

podem ser comparadas adotando-se como parâmetros iniciais resultados obtidos neste trabalho sem esquecer do aspecto anisotrópico do fenômeno.

Outro ponto a ser considerado é o estudo conjunto dos dados das variáveis densidade e umidade do solo, de modo a estimar parâmetros de uma variável utilizando dados de outra através de semivariogramas e autocorrelogramas cruzados além da *co-krigagem*.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALMEIDA, C.F.P.; RIBEIRO JÚNIOR, P.J. **Estimação da distribuição espacial de retenção de água em um solo utilizando krigagem indicatriz**. Curitiba: UFPR, Departamento de Estatística, 1996. 38p. (Relatório Técnico do Laboratório de Estatística)
- BURGESS, T.M.; WEBSTER, R. Optimal interpolation and isarithmic mapping of soil properties. I. The semivariogram and punctual kriging. **Journal Soil of Science**, v.31, p.315-331, 1980.
- CALVETE, F.J.S.; RAMIREZ, J.C. **Geoestatística**: aplicaciones e la hidrogeologia subterránea. Barcelona: Gráfica Torres, 1990. 484p.
- CRESSIE, N. Fitting variogram model by weighted least squares. **Mathematical Geology**, v.17, n.5, p.563-586, 1988.
- CRESSIE, N. **Statistics for spatial data**. New York: John Wiley, 1991. 900p.
- DEUTSCH, C.V.; JOURNEL, A.G. **GSLIB: geostatistical software library and user's guide**. New York: Oxford University Press, 1992. 336p.
- DOURADO NETO, D. Variabilidade espacial das alturas de chuva e irrigação e de potenciais da solução do solo. Piracicaba, 1989. 180p. Dissertação (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Sistema brasileiro de classificação do solo**. Brasília: EMBRAPA, CNPS, 1999. 412p.

- ENGLUND, E.; SPARKS, A. **Geo-EAS geostatistical environmental assessment software**. Las Vegas: U.S. Environmental Protection Agency, 1988. 1v.
- GONÇALVES, A.C.A. Variabilidade espacial de propriedades físicas do solo para fins de manejo da irrigação. Piracicaba, 1997. 118p. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- GONÇALVES, A.C.A ; FOLEGATTI, M.V.; VIEIRA, S.R. Padrões de amostragem e intensidade de krigagem na caracterização do armazenamento de água no solo, em área irrigada por pivô central. **Revista Brasileira de Ciência do Solo**, v.23, n.3, p.1485-495, 1999.
- GUERRA, P. A. G. **Geoestatística operacional**. Brasília: Departamento Nacional da Produção Mineral, 1988. 145p.
- HAMLETT, J.M.; HORTON, R.; CRESSIE, N.A.C. Resistant and exploratory techniques for use in semivariogram analysis. **Soil Science Society of America Journal**, v.50, n.3, p.868-875, 1986.
- HOAGLIN, D.C.; MOSTELLER, F.; TUKEY, J.W. **Análise exploratória de dados: técnicas robustas**. Lisboa: Salamandra, 1992. 446p.
- ISAAKS, E.H.; SRIVASTAVA, R.M. **Applied geostatistics**. New York: Oxford University Press, 1989. 561p.
- JOURNEL, A.G.; HUIJBREGTS, C.J. **Mining geostatistics**. London: Academic Press, 1978. 600p.
- KITANIDIS, P.K. **Introduction to geostatistics: applications to hydrogeology**. Cambridge: Cambridge University Press, 1997. 249p.
- KRIGE, D.G. A statistical approach to some basic mine evaluation problems on the Witwatersrand. **Journal of Chemical Metall Mining Society South African**, v.52, p.119-139, 1951.

- LANDIM, P.M.B. **Análise estatística de dados geológicos**. São Paulo: UNESP, 1998. 226p.
- LIBARDI, P.L.; MANFRON, P.A.; MORAES, S.O.; TUON, R.L. Variabilidade da umidade gravimétrica de um solo hidromórfico. **Revista Brasileira de Ciência do Solo**, v.20, n.1, p.1-12, 1996.
- MCBRATNEY, A.B.; WEBSTER, R. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. **Journal of Soil Science**, v.37, p.617-639, 1986.
- MORAES, S.O. Heterogeneidade hidráulica de uma terra roxa estruturada. Piracicaba, 1991. 141p. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- PANNATIER, Y. **VARIOWIN**: software for spatial data analysis in 2D. New York: Springer-Verlag, 1996. 90p.
- REICHARDT, K. **Processos de transferência no sistema solo-planta-atmosfera**. 4. ed. Campinas: Fundação Cargill, 1984, 445p.
- REICHARDT, K.; VIEIRA, S.R.; LIBARDI, P.L. Variabilidade espacial de solos e experimentação de campo. **Revista Brasileira de Ciência do Solo**. v.10, n.1, p.1-6, 1986.
- RIBEIRO JÚNIOR, P.J. Abordagem geoestatística em um estudo de variabilidade espacial. Piracicaba, 1995. 99p. Dissertação (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- SILVA, A.P. Variabilidade espacial de atributos físicos do solo. Piracicaba, 1988. 98p. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- SMITH, L.H. Plot arrangement for variety experiment with corn. **Proceedings of the American Society of Agronomy**, v.107, n.09, p.84-89, 1910.

- TRANGMAR, B.B.; YOST, R.S.; UEHARA, G. Application of geostatistics to spacial studies of soil properties. **Advances in Agronomy**, v.38, p.45-94, 1987.
- VALENTE, J. M. G. P. **Geomatemática**: lições de geoestatística. 2. ed. Ouro Preto : Fundação Gorceix, 1989. v.3, p.511-713.
- VAUCLIN, M.; VIEIRA, S.R.; BERNARD, R.; HATFIELD, J.L. Spatial variability of surface temperature long two transects of bare soil. **Water Resources Research** v.18, p.1677-1686, 1982.
- VIEIRA, S.R.; HATFIELD, J.L.; NIELSEN, D.R.; BIGGAR, J. W. Geostatistical theory and application to variability of some agronomical properties. **Hilgardia**, v.51, n.3, p.1-75, June 1983.
- VIEIRA, S.R. **Geoestatística**. Botucatu: UNESP, 1996. 55p. (Apostila do curso de extensão universitária)
- WARRICK, A.W.; NIELSEN, D.R. Spatial variability of soil physical properties in the field. In: HILLEL, D. **Applications of soil physics**. San Diego: Academic Press, 1980. p.319-344.
- WARRICK, A.W.; MYERS, D.E; NIELSEN, D.R. Geostatistical methods applied to soil science. In: BLACH, C.A. **Methods of soil analysis**. Madison: American Society of Agronomy; Soil Science Society of American, 1986. p.53-82.
- WAYNICK, D.D. Variability in soils an its significance to past and future soil investigations. I.Statistical study of nitrification in soil. **University of California Publications Agronomy Science**, v.3, n.9, p.243-270, 1918.
- WEBSTER, R.; BURGESS, T.M. Sampling and bulding strategis for estimating soil properties in small regions. **Journal of Soil Science**, v.35, n.1, p.127-140, 1984.

WILDING, L.P.; DREES, L.R. Spatial variability and pedology. In:
WILDING,L.P., SMECHAND, N.E.; HALL, G.F. **Pedogenesis and soil ta-
xonomy**. New York: Elsevier, v.11a, p.83-113, 1983.