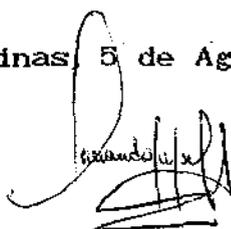


PLANOS AMOSTRAIS PARA VARIÁVEIS ESPACIAIS UTILIZANDO
GEOESTATÍSTICA

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pelo Sr. *Marcelo Silva de Oliveira* e aprovada pela Comissão Julgadora.

Campinas, 5 de Agosto de 1991



Prof. Dr. Armando Mário Infante
Orientador

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de *Mestre em Estatística*.

DEDICATÓRIA

Ao Fundamento e Razão do Universo, meu Padrão, meu Exemplo, minha Referência.

A Pessoa que tem uma honestidade linda, uma firmeza tranquilizadora, humildade de quem conhece à Deus, sabedoria pujante, fé vencedora, amor cativante.

Aquele que Deus escolheu para a restauração de todas as coisas.

Ao Alfa e o Ômega, o Princípio e o Fim.

A Cristo Jesus,

Dedico.

AGRADECIMENTOS

A Deus.

A minha esposa Nilma, que tem cumprido com amor e dedicação a ordem: "Serás uma ajudadora que lhe seja idônea" (Gênesis 2.18).

Aos colegas do Departamento de Ciências Exatas da Escola Superior de Agricultura de Lavras (ESAL) pelo apoio, amizade e incentivo constantes.

Ao povo do Brasil e a meus pais, Sr. Sebastião Melo de Oliveira e D. Marília Silva de Oliveira, que apoiaram e financiaram meus estudos.

Aos professores, funcionários e colegas do Instituto de Matemática, Estatística e Ciência da Computação da Universidade Estadual de Campinas (UNICAMP) pela transferência de conhecimentos e experiências. Em especial aos professores Dr. Ademir José Petenate, Dr. Mauro Sérgio de Freitas Marques e Dra. Regina Célia de Carvalho Pinto Moran pelo apoio decisivo, e ao meu orientador, professor Dr. Armando Mário Infante, pela disposição de me treinar e com o qual aprendi para a vida.

Ao professor Dr. Armando Zaupa Remacre do Instituto de Geociências da UNICAMP pelo auxílio amigo e ao Dr. Sidney Rosa Vieira do Instituto Agronômico de Campinas, pela consideração fraternal.

Agradeço.

RESUMO

O planejamento da amostragem de uma variável espacial é tratado nesta dissertação utilizando a teoria geoestatística.

Planos amostrais para mapeamento de uma variável espacial com semivariograma não-decrescente e isotrópico são definidos em primeiro lugar. Eles levam à um acréscimo relativamente pequeno na variância máxima do erro de predição, se a melhor configuração para a amostra (malha triangular) for trocada pela malha quadrada, cuja alocação no campo é menos onerosa. Caso o semivariograma seja anisotrópico geométrico com modelo linear, demonstra-se formalmente que para controlar o risco é suficiente usar uma malha retangular cuja razão entre os lados do retângulo seja a mesma que entre os comprimentos dos eixos da elipse de anisotropia.

Planos amostrais para a estimação da média espacial de uma variável com semivariograma isotrópico do tipo linear que se distribui sobre uma subregião quadrada, são satisfatórios se prescrevem uma malha também quadrada, interna à subregião.

Algoritmos para a determinação do tamanho da amostra e o espaçamento entre os pontos amostrados são fornecidos em todos os casos estudados. Estes planejamentos dependem de uma estimação prévia do semivariograma, possivelmente através de uma pré-amostragem.

Aplicações numéricas em ciência do solo são utilizadas como ilustração em toda a dissertação.

ABSTRACT

This dissertation deals with sampling of a spatial variable using the geostatistical theory.

Sample designs for mapping a variable with non decreasing, isotropic semivariogram are first discussed. They cause a slight increase in the maximum error variance of prediction if the best sample configuration (which is a triangular grid) is replaced by a square grid. A formal proof of the optimality of the rectangular grid is then given in the case of a linear semivariogram with geometrical anisotropy when the ratio of sides coincides with the ratio between the principal axes of the anisotropy ellipse.

Sample designs for the estimation of the spatial mean of a variable with linear isotropic semivariogram, distributed in a square subregion, are found adequate when based on an internal square grid.

Algorithms for the determination of sample size and spacing between sample points are given.

Numerical examples from soil science are used as illustrations.

LISTA DE NOTAÇÕES

Símbolo	Denota (página)
$\{ Z(x):x \in R \subset \mathbb{R}^2 \}$	Processo estocástico real (7)
R	Região (7)
\mathbb{R}	Conjunto dos números reais (7)
$\{ z(x):x \in R \subset \mathbb{R}^2 \}$	Realização do processo estocástico $\{ Z(x):x \in R \subset \mathbb{R}^2 \}$, ou população de valores da variável z, associada à região R (8)
Var	Variância (8)
$\gamma_z(\cdot)$	Função semivariograma do processo estocástico de 2 ^a ordem $\{ Z(x):x \in R \subset \mathbb{R}^2 \}$ que satisfaz a hipótese intrínseca (8)
\mathbb{N}	Conjunto dos números naturais (9)
Cov	Covariância (9)
$R(x)$	Subregião de R centrada em x (10)
Min	Mínimo (13)
$\sigma_K^2(x)$	Variância de krigagem (14)
$\bar{\gamma}_z(x_i, R(x))$	Semivariância média entre o ponto x_i e todos os pontos de $R(x)$ (15)
$\sigma_{KB}^2(x)$	Variância de krigagem de bloco (15)

(Continua)

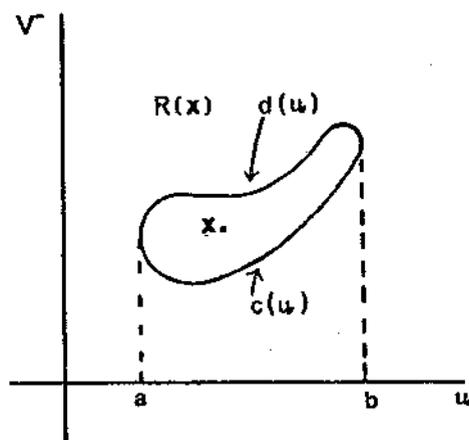
(continuação)

Símbolo	Denota (página)
$\bar{\gamma}_Z (R(x))$	Semivariância média entre todos os pares de pontos internos à $R(x)$ (15)
σ_{max}^2	Risco máximo associado à um plano amostral (27, 66)
σ_{KMTL}^2	Variância máxima de krigagem numa malha triangular sob semivariograma linear (36)
σ_{KMQL}^2	Variância máxima de krigagem numa malha quadrada sob semivariograma linear (39)
	Matriz simétrica (39)
σ_{KM}^2	Variância máxima de krigagem (48)
DIV e MOD	<p>Quociente e resto, respectivamente, da divisão inteira: sejam a e b números inteiros positivos; os números $c = a \text{ DIV } b$ e $d = a \text{ MOD } b$ são inteiros tais que:</p> <p>$a = c.b + d$ com $0 \leq d < b$.</p> <p>Por exemplo: $5 \text{ DIV } 3 = 1$, $5 \text{ MOD } 3 = 2 \Rightarrow 5 = 1.3 + 2$ (71)</p>
$\int_{R(x)} \int_{R(x)} \int_Z \kappa(s,t) ds dt$	<p>Integral no sentido de Riemann:</p> $\int_{a(x)}^{b(x)} \int_{c(u(x))}^{d(u(x))} \int_Z \kappa((u,v), (r,w)) dv du dw dr$

(Continua)

(conclusão)

Símbolo	Denota (página)
---------	-----------------

Onde $R(x)$ é:

$$\int_{R(x)} Z(t) dt =$$

$$\int_{R(x_1, x_2)} \int Z(u, v) dudv$$

Integral estocástica em média quadrática (m.q.):

$$\lim_{\substack{p \rightarrow \infty \\ q \rightarrow \infty}} \text{m.q.} \sum_{i=0}^{p-1} \sum_{j=0}^{q-1} Z(u_i, v_j) \cdot (u_{i+1} - u_i) \cdot (v_{j+1} - v_j) \quad (10)$$

INDICE

	Página
1. INTRODUÇÃO	01
2. GEOESTATÍSTICA	05
2.1. Teoria	07
2.1.1. Modelo	07
2.1.2. Inferências	13
2.2. Geoestatística e Ciência do Solo	17
2.2.1. Estudo de Solos	18
2.2.2. Breve Histórico	20
3. PLANO AMOSTRAL PARA MAPEAMENTOS	24
3.1. Semivariograma Isotrópico	33
3.1.1. Variância Máxima de Krigagem	34
3.1.2. Configuração e Tamanho da Amostra	45
3.2. Semivariograma Anisotrópico	52
3.2.1. Configuração	54
3.2.2. Tamanho da Amostra	57
3.3. Pré-amostragem para Semivariograma	60
3.4. Conclusão	64
4. PLANO AMOSTRAL PARA MÉDIAS ESPACIAIS	66
4.1. Variância de Krigagem de Bloco	71
4.2. Tamanho da Amostra	73
4.3. Conclusão	77
5. CONSIDERAÇÕES FINAIS	80
APÊNDICE	84
REFERÊNCIAS BIBLIOGRÁFICAS	97

Capítulo 1. INTRODUÇÃO

Em muitas áreas da ciência e da tecnologia é necessário conhecer a distribuição dos valores de variáveis no espaço. Por exemplo, o interesse do estudo pode ser a distribuição dos teores de um minério em uma jazida, ou a distribuição da quantidade de chuva sobre uma região da Terra, ou a variabilidade espacial da velocidade de infiltração da água no solo, ou ainda a distribuição da concentração de poluentes gasosos, líquidos ou sólidos em um meio ambiente especificado. Em regra, o conhecimento da distribuição de variáveis no espaço é obtido através de uma amostragem espacial. Um dos problemas que se apresenta ao pesquisador é a construção de um plano para realizar a amostragem de modo satisfatório. Num plano de amostragem espacial essencialmente dois elementos devem ser definidos: a *configuração da amostra* (isto é, a posição relativa dos pontos amostrados entre si) e o *espaçamento entre os pontos amostrados*. A configuração e o espaçamento definirão o tamanho da amostra espacial.

Esta dissertação tem o objetivo específico de estudar os problemas envolvidos no planejamento da amostragem espacial para dois fins: a determinação dos valores de uma variável em cada ponto de uma região (o problema do mapeamento) e a

determinação dos valores médios de uma variável em pequenas regiões (o problema de médias espaciais).

No planejamento da amostragem espacial para fins de mapeamento as seguintes questões estão envolvidas:

- (a) Definição de um plano amostral "aceitável", e das configurações que a amostra espacial possa ter para que o plano seja aceitável.
- (b) Possibilidade de estabelecer uma hierarquia, segundo algum critério, entre as configurações descritas em (a).
- (c) Cálculo do espaçamento entre os pontos amostrados e o tamanho da amostra, em função da qualidade desejada para o plano.

Questões paralelas às acima podem ser reconhecidas no estudo de planos amostrais para estimação de médias espaciais.

Todos estes pontos serão tratados nesta dissertação usando uma teoria estatística relativamente recente chamada geoestatística. Procurar-se-á estudar aspectos dessa teoria, propondo-se métodos para a solução de alguns problemas de amostragem encontrados na prática.

O capítulo 2, apresenta elementos de um caso particular da geoestatística, suas pressuposições e alguns conceitos, como os de *semivariograma* e *krigagem*. Estes são os instrumentos teóricos básicos para a solução dos problemas de planejamento da amostra espacial tratados nos capítulos 3 e 4. O capítulo 2, também apresenta subsídios para uma melhor compreensão dos

exemplos numéricos encontrados nos capítulos posteriores.

O capítulo 3 discute as questões (a), (b) e (c) envolvidas na construção de um plano para mapeamento. Mesmo sendo estas questões em parte respondidas pela literatura, aqui apresentam-se argumentações e demonstrações novas, levando à fórmulas e algoritmos aplicáveis na prática do planejamento.

Neste contexto é apresentada a demonstração de que a razão entre os comprimentos dos lados dos retângulos que compõe uma malha retangular deve ser igual a razão que há entre os comprimentos dos eixos da elipse de anisotropia, quando o semivariograma é linear.

Tanto para o caso isotrópico quanto para o caso anisotrópico, algoritmos com exemplos numéricos de aplicação em ciência do solo serão fornecidos. (Esta aplicação numérica foi executada usando o módulo IML do SAS (versão para PC), instalado no Laboratório de Estatística do Instituto de Matemática, Estatística e Ciência da Computação da Universidade Estadual de Campinas.) Estas instruções ordenadas poderão ser úteis ao pesquisador interessado em usar o computador como ferramenta no planejamento da amostragem espacial. Ainda neste capítulo uma seção é dedicada à discussão da pré-amostragem, necessária à estimação do semivariograma, que deve preceder às soluções dos problemas de planejamento amostral.

O capítulo 4 discute as questões (a), (b) e (c) apresentadas anteriormente, visando a estimação de uma média espacial. Uma sequência de instruções que permite o cálculo da

variância de krigagem de bloco em função do tamanho da amostra é construída. Nela é baseada um algoritmo computacional para a determinação do tamanho da amostra e do espaçamento entre pontos amostrados. Um exemplo numérico utilizando variáveis e valores típicos da ciência do solo ilustra uma possível aplicação prática do método. Os cálculos numéricos foram efetuados usando o software LOTUS 1-2-3 (microcomputador PC) no Laboratório de Computação do Departamento de Ciências Exatas da Escola Superior de Agricultura de Lavras.

O capítulo 5 examina os possíveis obstáculos ao uso dos planos propostos nesta dissertação, analisando mais detidamente o compromisso que deve existir entre o risco do plano e seu custo.

O apêndice contém algumas demonstrações usadas na fundamentação de resultados apresentados nos capítulos 3 e 4. Uma lista de notações (apresentada no início) e as referências bibliográficas utilizadas neste trabalho completam esta dissertação.

Capítulo 2. GEOESTATÍSTICA

Desde à muito tempo, cientistas de diferentes áreas do conhecimento estão estudando processos de variação espacial e/ou temporal (Cressie, 1990), por exemplo:

- (i) A distribuição espacial do teor de estanho, expresso em lb/ton, em uma jazida (Clark, 1979).
- (ii) A distribuição espacial da quantidade de chuva, em mm, que cai em uma certa região, num determinado período de tempo (Switzer, 1979; Cabannes, 1979).
- (iii) O estudo da distribuição espaço-temporal da poluição atmosférica, por exemplo chuva ácida, em uma determinada região geográfica (Cressie, 1989).
- (iv) A variabilidade espacial da taxa de infiltração da água no solo, medida no campo (Vieira, Nielsen e Biggar, 1981).

Nestes estudos, principalmente dois elementos podem ser discernidos:

- (i) A necessidade de considerar explicitamente *ab initio* as relações de dependência entre os valores de uma variável em pontos vizinhos.

- (ii) A construção de preditores (ou estimadores) utilizando a dependência acima e a informação fornecida por dados amostrais.

Dentre os muitos autores que contribuíram para este estudo, atribui-se à G. Matheron no início da década de 60 na França a apresentação/formalização da teoria que se passou a chamar *Geoestatística* (Cressie, 1990). Esta teoria incorpora os elementos (i) e (ii) acima. Matheron estava inicialmente voltado para a solução de problemas de variação espacial na mineração, mas hoje a teoria é aplicada em vários outros campos da ciência, como por exemplo às distribuições das variáveis espaciais e temporais mencionadas acima. Porém, nesta dissertação, serão consideradas somente variáveis espaciais. É também conveniente salientar que a geoestatística não é hoje (como seu nome poderia sugerir) o conjunto de todas as técnicas estatísticas aplicáveis em ciência da terra nem tampouco restringe sua aplicação somente à ciência geológica.

O objetivo deste capítulo é fornecer ao leitor os subsídios necessários para a compreensão dos capítulos 3 e 4, assim como também informações técnicas e históricas sobre a junção da geoestatística com a ciência do solo, pois os exemplos numéricos de aplicações dos capítulos 3 e 4 pertencem a essa área. A seção 2.1, trata da teoria geoestatística e a seção 2.2, relaciona a geoestatística com a ciência do solo.

2.1. Teoria

Esta seção apresenta a parte da teoria que é fundamental para a solução de alguns problemas em planejamento da amostragem espacial, assunto a ser tratado nos capítulos 3 e 4. Uma apresentação mais completa pode ser encontrada nos textos de Journel e Huijbregts (1978) ou Isaaks e Srivastava (1989).

Em 2.1.1, o modelo geoestatístico é apresentado e em 2.1.2, certos tipos de inferências são examinados com algum detalhe.

2.1.1. Modelo

Um *processo estocástico (real)* é uma coleção $\{Z(x):x \in R \subset \mathbb{R}^p\}$ de variáveis aleatórias reais, definidas sobre um mesmo espaço de probabilidade, indexadas em um subconjunto R do espaço vetorial p -dimensional \mathbb{R}^p . Este último é chamado *espaço de índices* do processo estocástico. Nesta dissertação, o subconjunto R será denominado *região* e prescindir-se-á de outras considerações sobre sua geometria. Em particular, um processo estocástico é dito ser de 2^{a} ordem se a esperança matemática $E[Z(x)]^2$ é finita, $\forall x \in R$ (veja p. ex., Cramér e Leadbetter, 1967). O espaço de índices \mathbb{R}^p é definido de tal maneira que seja possível representar variações aleatórias em espaços de qualquer dimensão, p. ex., $p = 1$ para variações no tempo (como as estudadas nas séries de tempo), $p = 2$ para variações em superfícies, $p = 3$ para variações no espaço

tridimensional e $p = 4$ para variação no espaço-tempo.

A geoestatística baseia-se no suposto de que a distribuição espacial (e/ou temporal) de uma variável z em uma região R é uma realização $\{z(x):x \in R \subset \mathbb{R}^p\}$ do processo estocástico $\{Z(x):x \in R \subset \mathbb{R}^p\}$. Esta realização será também chamada nesta dissertação de *população* de valores da variável z associada à região R . Nesta dissertação, dotar-se-á o processo estocástico básico com algumas características, explicitadas abaixo.

Seja uma população $\{z(x):x \in R \subset \mathbb{R}^2\}$, realização do processo estocástico de 2ª ordem $\{Z(x):x \in R \subset \mathbb{R}^2\}$, que satisfaz o seguinte suposto de estacionaridade, chamada tradicionalmente de *hipótese intrínseca* (Journel, 1978; Cressie, 1988, 1989):

(i) $E[Z(x)] = \mu_z, \forall x \in R$, onde μ_z é uma constante real desconhecida

(ii) $\text{Var}[Z(x) - Z(y)] = E[Z(x) - Z(y)]^2 = 2\gamma_z(x-y), \forall x,y \in R$, onde $\gamma_z(\cdot)$ é uma função real não-negativa.

A função $2\gamma_z(\cdot)$ é chamada *variograma* e $\gamma_z(\cdot)$ é chamada *semivariograma*. O valor $\gamma_z(h)$ é chamado *semivariância* entre dois pontos separados pelo vetor h . Se a função semivariograma $\gamma_z(\cdot)$ só depende da distância entre os pontos x e y (isto é, se $\gamma_z(x-y)$ só depende da norma $|x - y|$), o semivariograma é dito ser *isotrópico*. Se $\gamma_z(\cdot)$ depende não só da distância, mas também da direção da reta que passa pelos pontos x e y , o semivariograma é dito ser *anisotrópico*.

As seguintes propriedades para a função $\gamma_Z(\cdot)$ decorrem de sua própria definição: (i) $\gamma_Z(0) = 0$ (ii) $\gamma_Z(x-y) = \gamma_Z(y-x)$, $\forall x, y \in \mathbb{R}$, (iii) $\sum_{i=1}^m \sum_{j=1}^m a_i a_j \gamma_Z(x_i - x_j) \leq 0$, $\forall x_1, \dots, x_m \in \mathbb{R}$, $\forall a_1, \dots, a_m \in \mathbb{R}$ tal que $\sum_{i=1}^m a_i = 0$, $\forall m \in \mathbb{N}$ (veja Cressie, 1988).

O sentido físico do semivariograma $\gamma_Z(\cdot)$ pode ser percebido quando se analisa sua definição: ele mede a variabilidade das diferenças entre as realizações da variável aleatória de interesse, de tal maneira que quanto menor a semivariância, menor a variação dessas diferenças. A compreensão deste fato pode ser melhorada quando se considera um conjunto de processos estocásticos que satisfazem a hipótese intrínseca: o conjunto dos processos estacionários de 2^a ordem ou estacionários com respeito à covariância (covariance stationary, em inglês). Estes são processos de 2^a ordem que satisfazem (a) $E[Z(x)] = \mu_Z, \forall x \in \mathbb{R}$ (b) $Cov[Z(x), Z(y)] = \kappa_Z(x, y) = \kappa_Z(x-y)$, $\forall x, y \in \mathbb{R}$, onde $\kappa_Z(\cdot)$ é denominada função de covariância.

Nesse caso, pode-se mostrar que $\gamma_Z(x-y) = \kappa_Z(0) - \kappa_Z(x-y)$. Essa relação permite ver que, enquanto $\kappa_Z(x-y)$ cresce (os pontos ficam mais correlacionados), a semivariância $\gamma_Z(x-y)$ decresce.

Alguns modelos para a função $\gamma_Z(\cdot)$ tem sido propostos, todos eles satisfazendo as exigências (i), (ii) e (iii) acima (veja Journel e Huijbregts, 1978). Especialmente um deste é interessante nesta dissertação, devido a sua simplicidade: o modelo linear, representado por $\gamma(h) = a_h \cdot |h|$, onde a_h é um

número real que depende da direção do vetor h . Esta função pode modelar o semivariograma de processos que satisfazem a hipótese intrínseca mas que não são estacionários com respeito à covariância. Um modelo para este último poderia ser o denominado *esférico*. Alguns fenômenos exigem uma descontinuidade na origem do semivariograma, chamada *efeito pepita*, que expressa limitação na estimação do semivariograma em pequenos espaçamentos. Sua ocorrência pode estar associada ao aumento acentuado na dependência espacial entre pares de pontos próximos, como quando há grãos ou palhetas de metal nativo, em mineração, particularmente de ouro ("pepitas") (Cressie, 1988). De fato, o semivariograma pode ser modelado completamente pelo efeito pepita, isto é, $\gamma(h) = \gamma_0$ para $h \neq 0$, $\gamma_0 \geq 0$. Isto é um caso extremo e estaria associado a um processo onde toda a variação seria aleatória, isto é, há uma correlação zero entre os pontos da região. O mais comum é o efeito pepita estar combinado aos modelos usuais, por exemplo, com o linear: $\gamma_z(h) = \gamma_0 + a_h \cdot |h|$.

Nesta dissertação, além das propriedades naturais, mais uma propriedade será exigida para a função $\gamma_z(\cdot)$: ela deve ser não-decrescente. A razão para isto pode ser percebida quando consideramos um processo estacionário com respeito à sua covariância: $\kappa_z(h) = \kappa_z(0) - \gamma_z(h)$. A medida que $|h|$ cresce (os pontos ficam mais distantes), a covariância (e a correlação) diminuem. Assim, pontos mais distantes fornecem menos informação sobre o valor de z em um determinado ponto em estudo. Isto cria uma associação direta entre distância e incerteza,

facilitando a correspondência entre resultados geométricos e estatísticos.

Suponha que para $x \in R$ é considerada a subregião $R(x)$ de forma e dimensão fixas, contendo o ponto x (veja quadro 2.1). Assuma ainda que a subregião $R(x) \subset R$ e que a função $\mu_z(\cdot, \cdot)$ sejam tais que a integral

$$\int_{R(x)} \int_{R(x)} \mu_z(s, t) ds dt$$

é finita, e que $R(x)$ não impossibilite a existência da variável aleatória

$$S(x) = |R(x)|^{-1} \int_{R(x)} Z(t) dt$$

definida como a *média espacial* da variável z na subregião $R(x)$. (Aqui $|R(x)|$ é o volume de $R(x)$.) Uma realização $s(x)$ de $S(x)$ é a média aritmética da *subpopulação* $\{z(t): t \in R(x) \subset R\}$. Nessas condições, pode ser demonstrado (David, 1978) que o processo $\{S(x): R(x) \subset R \subset \mathbb{R}^2\}$ é de 2ª ordem e que satisfaz a hipótese intrínseca, isto é:

(i) $E[S(x)]$ é constante e coincide com μ_z , $\forall R(x) \subset R$.

R.

(ii) $\text{Var}[S(x) - S(y)] = E[S(x) - S(y)]^2 =$

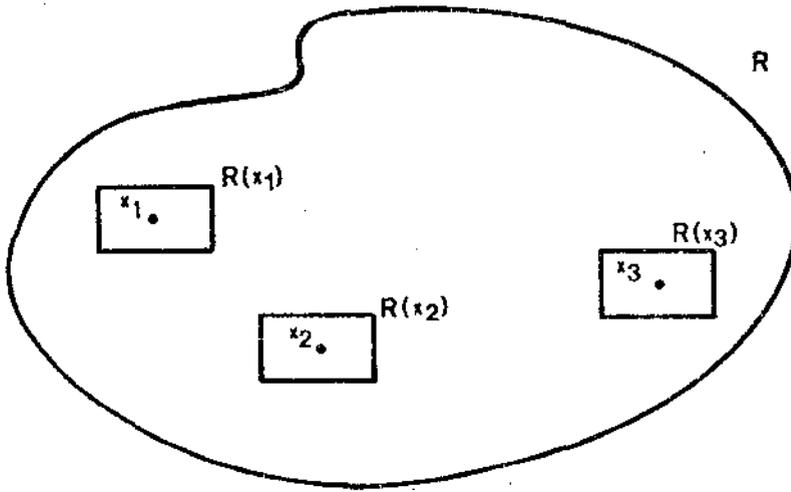
$$2 \cdot \gamma_z(x-y) =$$

$$= 2 \cdot |R(x)|^{-2} \cdot \left\{ \int_{R(x)} \int_{R(y)} \gamma_z(s-t) ds dt - \int_{R(x)} \int_{R(x)} \gamma_z(s-t) ds dt \right\}$$

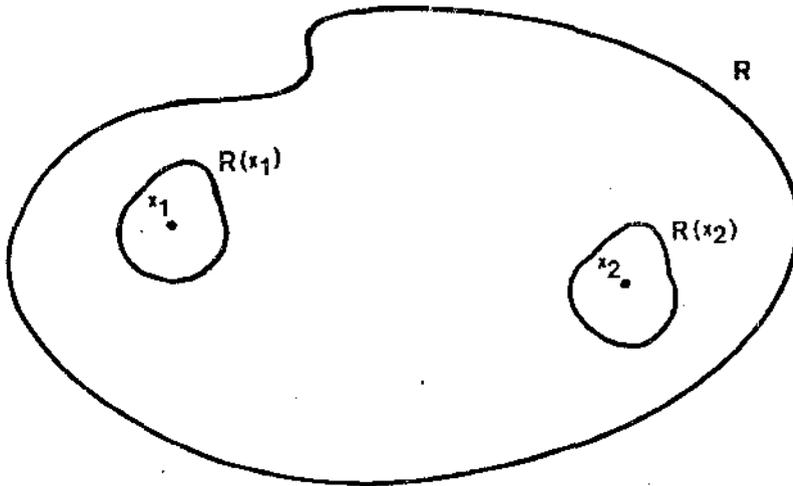
Uma *amostra* de tamanho n da população $\{z(x): x \in R \subset \mathbb{R}^2\}$ é um subconjunto finito $\{z(x_1), \dots, z(x_n)\}$ da mesma.

Definidos os supostos teóricos adotados, prosseguir-se-á com os tipos usuais de inferências.

(a)



(b)



QUADRO 2.1. Duas formas diferentes para as subregiões $R(x)$ distribuídas na região R : (a) regular (b) forma arbitrária.

2.1.2. Inferências

As inferências no modelo estatístico definido na subseção anterior podem agora ser discutidas. Para desenvolvê-las é necessário realizar determinações de três tipos diferentes:

(i) Estimação da função semivariograma $\gamma_z(\cdot)$, usada para descrever a dependência espacial no processo sob estudo.

(ii) Predição da variável $z(x)$, $x \in R$. Predizer $z(x)$ baseado numa amostra $\{z(x_1), \dots, z(x_n)\}$, é construir uma função real mensurável $g_0(\cdot, \dots, \cdot)$ tal que o número $\hat{z}(x) = g_0(z(x_1), \dots, z(x_n))$ opere como substituto do valor desconhecido de $z(x)$; o valor $\hat{z}(x)$ é chamado *predição* de $z(x)$. A variável aleatória $\hat{Z}(x) = g_0(Z(x_1), \dots, Z(x_n))$ é dita ser um *preditor* de $Z(x)$. Como a predição é feita para a realização da variável aleatória em um ponto $x \in R$, tal predição é dita *pontual*.

(iii) Predição da variável $s(x)$ para $R(x) \subset R$ usando a amostra $\{z(x_1), \dots, z(x_n)\}$. A variável aleatória $S(x)$ é predita pelo preditor $\hat{S}(x)$.

O conhecimento do semivariograma $\gamma_z(\cdot)$ é fundamental para desenvolver as predições. A função $\gamma_z(\cdot)$ é estimada usando uma amostra, não necessariamente a mesma que será usada nas predições: sua estimação poderá se basear numa pré-amostragem. Estimadores e procedimentos para estimação de semivariogramas

não são tratados aqui; eles podem ser encontrados nos textos de Isaaks e Srivastava (1989), Zimmerman e Zimmerman (1991) ou Journal e Huijbregts (1978).

Na predição pontual, uma atitude usual é considerar preditores que são uma combinação linear $\hat{Z}(x) = \sum_{i=1}^n \lambda_i \cdot Z(x_i)$ dos elementos da amostra. Dentre estes, procura-se um *preditor ótimo*, no sentido de que seja *não-tendencioso* e que possua o menor *erro quadrático médio de predição* dentre todos os preditores lineares não-tendenciosos. Assim, o preditor procurado $\hat{Z}(x)$ deve satisfazer as condições:

- (i) $E[\hat{Z}(x)] = E[Z(x)] = \mu_Z, \forall x \in R.$
 (ii) $E[\hat{Z}(x) - Z(x)]^2 = \text{Var}[\hat{Z}(x) - Z(x)] =$
 $\min \text{Var}[Z^*(x) - Z(x)], \forall x \in R,$ onde o mínimo é obtido entre todos os preditores $Z^*(x)$ que satisfazem (i).

Em outros termos, procura-se um BLUP (Best Linear Unbiased Predictor) no contexto do modelo definido em 2.1.1 (Cressie, 1990). A condição de não-tendenciosidade para todo $\mu_Z \in \mathcal{R}$ implica em $\sum_{i=1}^n \lambda_i = 1$. Para obter os valores de $(\lambda_1, \dots, \lambda_n) = \lambda$, que minimizam $E[Z(x) - \sum_{i=1}^n \lambda_i Z(x_i)]^2$, basta resolver a equação matricial

$$\begin{bmatrix} \lambda \\ \alpha \end{bmatrix} = \begin{bmatrix} \Gamma & | & 1 \\ \hline 1' & | & 0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} r \\ \hline 1 \end{bmatrix}$$

onde α é um multiplicador de Lagrange associado à restrição $\sum_{i=1}^n \lambda_i = 1$, Γ é uma matriz $n \times n$ cujo elemento na posição (i,j)

é $\gamma_z(x_i - x_j)$, $\mathbf{1}$ é o vetor $n \times 1$ com todos os elementos iguais à 1 e $\boldsymbol{\gamma}$ é um vetor $n \times 1$ cujo i -ésimo elemento é $\gamma_z(x - x_i)$. O preditor ótimo assim obtido recebe o nome de *preditor de krigagem linear ordinária* ou resumidamente *preditor de krigagem* (Cressie, 1990). Definindo $Z' = (Z(x_1), \dots, Z(x_n))$ pode-se anotá-lo como

$$\hat{Z}_k(x) = \boldsymbol{\lambda}'Z$$

Seu erro quadrático médio de predição é chamado *variância de krigagem*. Ela é dada pela fórmula

$$\sigma_k^2(x) = \boldsymbol{\lambda}'\boldsymbol{\gamma} + \frac{1}{2} [k_z(x,x) - \boldsymbol{\lambda}' \text{diag}(K)] + \alpha$$

onde $\text{diag}(K)$ é a diagonal principal da matriz com o elemento (i,j) dado por $k_z(x_i, x_j)$. Por causa do desconhecimento da função de covariância, usa-se outra expressão equivalente

$$\sigma_k^2(x) = 2\boldsymbol{\lambda}'\boldsymbol{\gamma} - \boldsymbol{\lambda}'T\boldsymbol{\lambda}$$

Os números $\lambda_1, \dots, \lambda_n$ são chamados *pesos de krigagem*. Note que a variância de krigagem não depende da realização $\{z(x_1), \dots, z(x_n)\}$, pois apenas as semivariâncias e os pesos de krigagem (que por si também só dependem das semivariâncias) são requeridos para calculá-la. Porém, ela é função da configuração da amostra, pois depende das distâncias e da posição relativa dos pontos amostrados entre si. Demonstrações e maiores informações sobre a krigagem podem ser encontradas nos textos de Journel e Huijbregts (1978), Clark (1979), e Cressie (1988, 1990).

Um preditor para $S(x)$ (com $R(x) \subset R$), pode ser construído de modo análogo. A combinação linear $\hat{S}(x) = \sum_{i=1}^n v_i \cdot Z(x_i)$ fornece o preditor não-tendencioso de mínimo erro quadrático médio dentre todos os preditores lineares não-tendenciosos, se os pesos de krigagem $(v_1, \dots, v_n) = \mathbf{v}'$ forem a solução da equação matricial

$$\begin{bmatrix} \mathbf{v} \\ \beta \end{bmatrix} = \begin{bmatrix} \Gamma & | & \mathbf{1} \\ \hline \mathbf{1}' & | & 0 \end{bmatrix}^{-1} \begin{bmatrix} \bar{\boldsymbol{\gamma}} \\ 1 \end{bmatrix}$$

onde β é um multiplicador de Lagrange associado à restrição $\sum_{i=1}^n v_i = 1$ e $\bar{\boldsymbol{\gamma}}$ é um vetor $n \times 1$ cujo i -ésimo elemento é $\bar{\gamma}_Z(x_i, R(x)) = |R(x)|^{-1} \cdot \int_{R(x)} \gamma_Z(x_i - s) ds$. Preditor de krigagem de bloco é sua denominação usual. Seu erro quadrático médio de predição é chamado *variância de krigagem de bloco*. Em linguagem matricial:

$$\hat{S}_{KB}(x) = \mathbf{v}' \cdot \mathbf{Z}$$

$$\sigma_{KB}^2(x) = \kappa_S(x, x) - |R(x)|^{-1} \int_{R(x)} \kappa_Z(t, t) dt - \mathbf{v}' \cdot \Gamma \cdot \mathbf{v} + 2\mathbf{v}' \cdot \bar{\boldsymbol{\gamma}}$$

onde $\kappa_S(x, x) = \text{Cov}[S(x), S(x)]$. Pelo fato das covariâncias serem desconhecidas, usa-se a expressão equivalente

$$\sigma_{KB}^2(x) = \mathbf{v}' \cdot \bar{\boldsymbol{\gamma}} + \beta - \bar{\gamma}_Z(R(x))$$

onde

$$\begin{aligned} \bar{\gamma}_Z(R(x)) &= |R(x)|^{-2} \cdot \int_{R(x)} \int_{R(x)} \gamma_Z(s-t) ds dt = \\ &= |R(x)|^{-1} \int_{R(x)} \bar{\gamma}_Z(t - R(x)) dt \end{aligned}$$

Outra vez pode-se notar que a variância de krigagem é função apenas da estrutura de dependência espacial do processo sob estudo, da forma e dimensão da subregião $R(x)$ e da posição relativa dos pontos amostrados entre si e entre pontos amostrados e a subregião. Assim, os valores da amostra não influem no risco, sendo este função apenas da configuração e tamanho amostrais. Informações suplementares podem ser encontradas nos textos de Clark (1979), David (1978) e Cressie (1990).

Observe que a geoestatística incorpora na sua teoria a estrutura de dependência espacial, através do semivariograma, e fornece um preditor espacial ótimo, a krigagem, com uma medida calculável para sua qualidade, a variância de krigagem. Isto é uma vantagem definida sobre outros métodos de interpolação/predição propostos na literatura e, em problemas onde é razoável supor a validade da hipótese intrínseca, pode fornecer um argumento conclusivo para a escolha do preditor (veja Cressie, 1989).

2.2. Geoestatística e Ciência do Solo

Esta seção visa definir convenientemente alguns estudos realizados em ciência do solo, como os exemplos numéricos dos capítulos 3 e 4.

Em 2.2.1 define-se objetivos e exemplos do estudo de solos. Em 2.2.2 é apresentado um histórico resumido.

2.2.1. Estudo de Solos

É um fato estabelecido em ciências agrárias que um solo mal utilizado coloca em risco a sua produção agrícola. Para que um solo seja adequadamente usado é necessário que seja bem conhecido. Seu conhecimento envolve a descrição e estudo de propriedades, levando à propostas adequadas de manejo. Uma classificação das propriedades detectáveis de um solo está apresentada no quadro 2.2. Seu estudo pode ser realizado qualitativa ou quantitativamente.

O estudo dos valores de variáveis do solo de uma região poderia ser feito sem incertezas se uma amostragem completa fosse realizada, isto é, se todo o solo fosse avaliado, exaustivamente. Porém isto é impraticável, pois o trabalho de coleta de dados, as medições no laboratório e mesmo a computação dos dados são sempre feitos com recursos limitados. A alternativa é então amostrar só em determinados pontos, fazendo um levantamento incompleto, chamado *amostragem* e preenchendo as lacunas de informação sobre os pontos não amostrados através de predições. A descrição de um solo pode também ser realizada mediante valores da média espacial de uma variável em determinadas sub-regiões $R(x)$ contidos em R - isto leva à estudar a população $(s(x):R(x) \subset R \subset \mathbb{R}^2)$. Outra possibilidade é o *mapeamento* de uma variável espacial distribuída sobre R . Esta tarefa pode ser tratada estatisticamente estabelecendo-se *isolinhas* (curvas de

Descrição	Variável	Determinação	Referência
Física	Velocidade de infiltração básica da água no solo.	No campo, usando um infiltrômetro de anel. Unidade de medida: mm/h.	Vieira et alii (1981).
	Retenção de água no solo.	No laboratório, usando o princípio de equilíbrio do potencial via placa porosa. Unidade de medida: g água/g solo.	Klute, A. (1986).
Química	pH	Métodos colorimétricos ou eletrométricos. Unidade de medida: Logaritmo na base 10 do inverso da concentração do íon H^+ .	McLean, E. O. (1982).
Biológica	População de nematóides.	Métodos de separação entre o ser vivo e o solo (centrifugação, etc). Unidade de medida: número de indivíduos/ m^2 .	Van Gundy, S.D. (1982).

QUADRO 2.2. Algumas propriedades descritivas do solo.

nível) de predições pontuais $\hat{z}(x)$ dentro de R . A isolinha associada ao valor z_i é o lugar geométrico de todos os pontos $x \in R$ tal que $\hat{z}(x) = z_i$. O conjunto das isolinhas associadas aos valores z_i , $i=1, \dots, k$ poderia ser denominado o *mapeamento estatístico* da população $\{z(x): x \in R \subset \mathbb{R}^2\}$ para as *cotas* (z_1, \dots, z_k) .

2.2.2. Breve Histórico

Um estudo quantitativo de solos pode ser definido como uma descrição conveniente de populações de variáveis espaciais, tendo como base uma amostra de solo. Estudos deste tipo já são feitos pelo menos desde o início deste século. Alguns marcos históricos são apontados abaixo. Veja Oliver (1989), para mais detalhes.

Registros históricos antigos sobre estudos da variabilidade espacial em solos para uso agrícola são os ensaios de uniformidade de Mercer e Hall em 1911, realizados na estação experimental de Rothamsted, na Inglaterra. Eles dividiram campos de 2 culturas (trigo e nabo) em centenas de pequenas parcelas e mediram suas produções. Após aglomeração destas parcelas pequenas em parcelas maiores com tamanhos crescentes, eles mediram a alteração produzida na variação da produção. Sua descoberta foi que, quando as parcelas eram de aproximadamente 0,01 hectare (100 m²), a variação entre parcelas era quase completamente anulada. Diante desse resultado, recomendaram

parcelas desse tamanho para futuros experimentos comparativos entre tratamentos agronômicos.

Fairfield Smith em 1938 estudando também a variação das produções de parcelas em ensaios de uniformidade descobriu que a variância da produção tende a aumentar indefinidamente com o crescimento do tamanho da parcela, formulando uma importante lei empírica. Aqui tem-se um dos primeiros estudos da variação espacial com um modelo para descrevê-la.

Ainda nos anos 30, R.A. Fisher em Rothamsted tinha influenciado decisivamente a estatística agrícola, mostrando como separar a variação nos experimentos de campo e em levantamentos amostrais de acordo com as diferentes causas. Youden e Mehlich em 1937 adaptaram as técnicas de Fisher para analisar dados de solos vindos de levantamentos espaciais com estrutura hierárquica. Nesta situação estimaram os componentes da variância do pH do solo para espaçamentos de 3 m, 30 m, 300 m e aproximadamente 1600 m. Acumularam os componentes começando nos menores espaçamentos, obtendo assim os primeiros protótipos do variograma. Seu trabalho não foi explorado adequadamente e seus resultados foram redescobertos sucessivamente por L.C. Hammond e outros em 1958 e por Webster e Butler em 1976.

F. Yates em 1948 trabalhou no planejamento de levantamentos amostrais em agricultura à procura de planos eficientes para estimação da produção de uma cultura e de uma propriedade do solo. Para isso usou semivariâncias calculadas sobre medições ao longo de transeções ("transects") no terreno.

Webster e Beckett em 1968 trabalharam na Inglaterra estudando o solo usando o modelo que em ciência do solo é denominado usualmente de "estatística clássica" (veja Burgess, Webster e McBratney, 1981; McBratney, Webster e Burgess, 1981). De acordo com este modelo o solo é dividido em estratos (subregiões), chamadas *classes de solo*, a variação dentro dessas classes é espacialmente não-correlacionada e a mudança das características do solo de uma classe para outra é abrupta. Este modelo mostrou-se inadequado quando exposto ao conhecimento já adquirido em ciência do solo.

P. Beckett voltou às idéias de Fairfield Smith e estudou o aumento na variância "dentro das classes" como uma função do tamanho das mesmas. Técnicas de série de tempo foram introduzidas no trabalho de Webster e Cuanalo em 1975.

Mais recentemente tem sido feitos esforços no sentido de aplicar os instrumentos conceituais da geoestatística no estudo de solos. Burgess e Webster (1980a, 1980b), Webster e Burgess (1980), Burgess, Webster e McBratney (1981) apresentaram parte da teoria geoestatística que pode ser usada no estudo de solos. Simultaneamente Vieira, Nielsen e Biggar (1981) e Gajem, Warrick e Myers (1981) utilizavam geoestatística para o estudo de solos. Desde então esta teoria tem sido explorada na descrição científica do solo.

Esse histórico mostra diferentes opções para modelar a distribuição de variáveis espaciais em um solo. Há essencialmente duas propostas estatísticas para modelar populações do

tipo $\{ z(x):x \in R \subset \mathbb{R}^2 \}$ (de Gruijter e ter Braak, 1990):

- A primeira delas consiste em considerar $\{ z(x):x \in R \subset \mathbb{R}^2 \}$ como uma população no sentido da teoria clássica da amostragem (Cochran, 1977). A mesma considera a população como um conjunto de números, sem qualquer relação quantitativa explícita de dependência entre eles. Fazendo assim, não levanta nenhum suposto questionável para viabilizar e/ou validar suas análises, porém não toma a vantagem de considerar a estrutura de dependência que possivelmente exista em dados espaciais.

- A segunda proposta considera $\{ z(x):x \in R \subset \mathbb{R}^2 \}$ como uma amostra retirada de uma superpopulação, que é uma "população" de populações do tipo de $\{ z(x):x \in R \subset \mathbb{R}^2 \}$. Esta proposta expressa-se na teoria de processos estocásticos, da qual tanto a geoestatística como a "estatística clássica" são casos particulares. A geoestatística necessita de supostos para ser válida (a hipótese intrínseca), mas considera a correlação espacial dos dados. A "estatística clássica" apresenta desvantagens pois exige supostos (não-correlação, normalidade, etc.), não considerando a estrutura de dependência espacial (Burgess, Webster e McBratney, 1981).

Capítulo 3. PLANO AMOSTRAL PARA MAPEAMENTOS

Em 2.1.2 ficou claro que as inferências no modelo geoestatístico são baseadas na amostra $\{z(x_1), \dots, z(x_n)\}$ da população $\{z(x): x \in R \subset \mathbb{R}^2\}$. Um plano para amostragem espacial indica de que maneira essa amostra será obtida. Tal plano inclui dois elementos essenciais:

- (i) A *configuração da amostra*, isto é, a posição relativa dos pontos amostrados entre si.
- (ii) O *espaçamento entre os pontos amostrados*, isto é, a distância entre pares consecutivos de pontos.

Observe que a configuração e o espaçamento definem o *tamanho da amostra*, isto é, o número de pontos amostrados na região, e a *densidade de amostragem*, isto é, a razão entre o tamanho da amostra e a área da região amostrada.

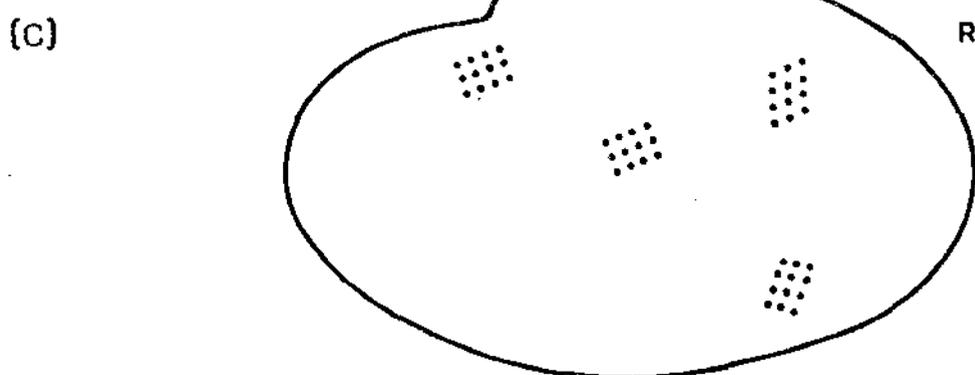
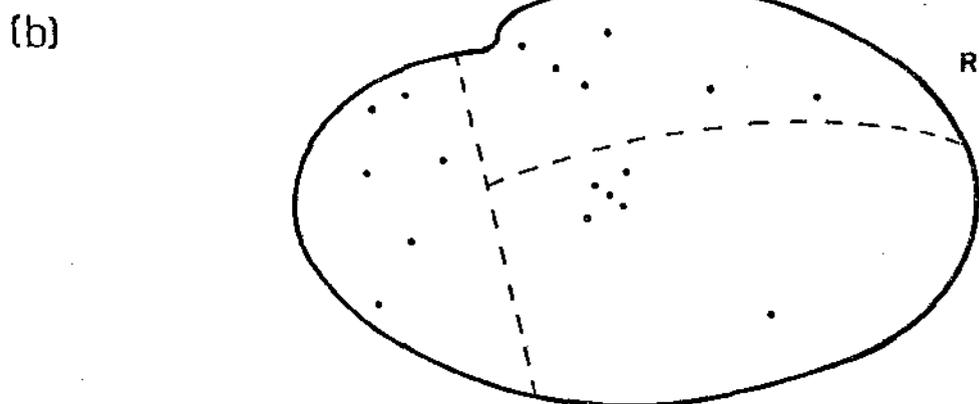
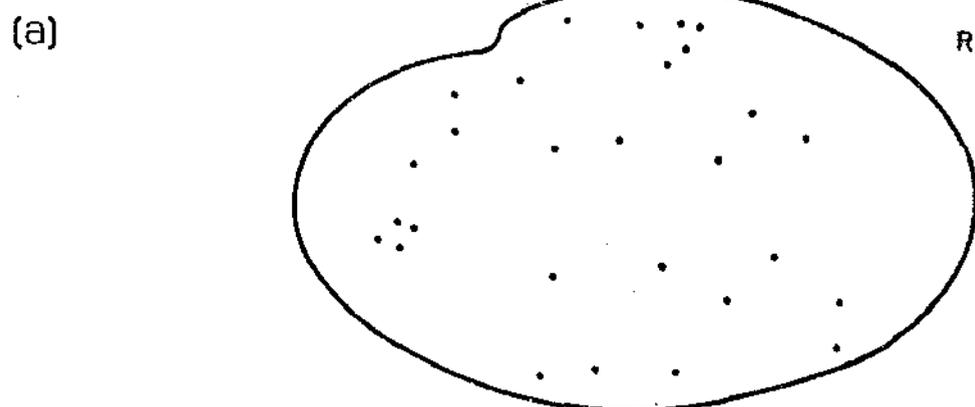
O espaçamento é dado numa escala de comprimento adequada ao plano amostral, p. ex. em metros. A configuração deverá ser escolhida entre *configurações aleatórias* e *configurações sistemáticas*. Em amostragem espacial algumas configurações sistemáticas são preferidas, especificamente as chamadas *malhas* (*reticulados, tesselações*) (veja quadro 3.2.), que são partições da região usando polígonos. Os pontos amostrados são os vértices de cada polígono e coincidem com os nós da malha. Se

os polígonos são regulares (isto é, polígonos com todos os ângulos internos iguais e todos os lados iguais), a malha recebe o nome de *malha regular*. Os quadros 3.1 e 3.2 mostram exemplos de configurações espaciais. Maiores informações podem ser encontradas em Petersen e Calvin (1986), Wilding e Drees (1983), Cochran (1977) e Ripley (1981). O planejamento da amostragem espacial será proposto, nesta dissertação, num contexto onde vale a teoria geostatística definida na seção 2.1.

Uma utilidade da amostra espacial é fornecer uma base para krigagem num mapeamento. O plano amostral é construído levando-se em consideração duas quantidades.

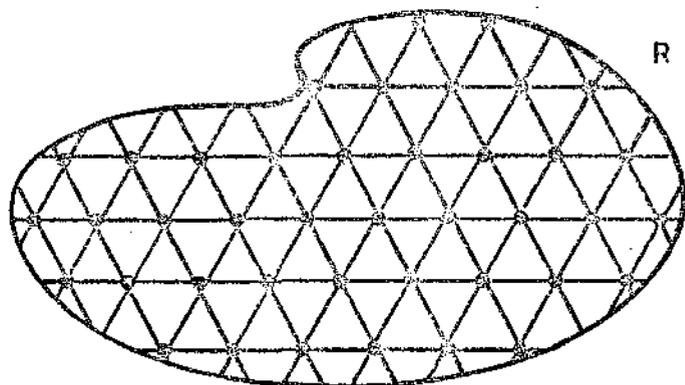
- (i) O *risco do plano para mapeamento* da população $\{z(x):x \in R \subset \mathbb{R}^2\}$, definido aqui como a máxima variância de krigagem produzida pela amostra.
- (ii) O *custo do plano para mapeamento* da população $\{z(x):x \in R \subset \mathbb{R}^2\}$, definido aqui como a quantidade de recursos financeiros necessária para executar o plano.

Configuração e espaçamento num plano para mapeamento serão determinados em função do risco e do custo. Porém, para chegar à propostas factíveis para o planejamento devemos inverter a relação entre risco (e custo) e características do plano. Por isso, neste capítulo será apresentado um estudo do risco do plano para mapeamento em função da configuração e espaçamento dos pontos amostrados. O custo também será considerado, incorporando-o na tarefa do planejamento.

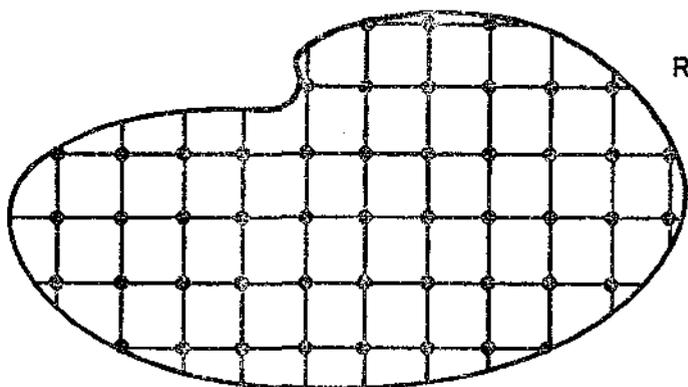


QUADRO 3.1. Alguns exemplos de configurações aleatórias possíveis para uma amostra espacial: (a) Amostra aleatória simples (b) Amostra aleatória estratificada (c) Amostra aleatória em conglomerados.

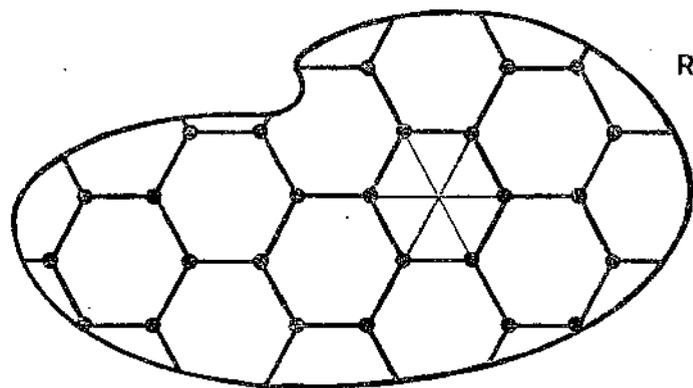
(a)
(a.1)



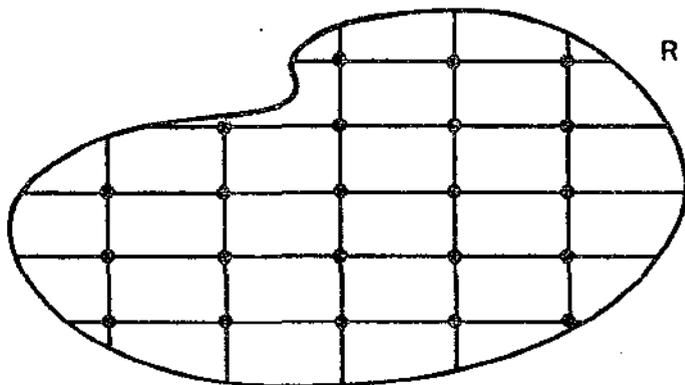
(a.2)



(a.3)



(b)



QUADRO 3.2. Malhas para amostragem. (a) Malhas regulares: (a.1) Triangular (a.2) Quadrada (a.3) Hexagonal. (b) Malha retangular.

A principal necessidade que leva ao planejamento da amostragem espacial é o controle do risco : estabelece-se um risco máximo σ_{\max}^2 para as inferências e planeja-se a amostragem para que o risco não ultrapasse esse máximo (Cabannes, 1979; Burgess, Webster e McBratney, 1981). Um plano tal que o risco seja menor ou igual ao máximo estabelecido σ_{\max}^2 será denominado aqui *plano σ_{\max}^2 -aceitável*.

A seção 3.1 estuda o caso no qual o semivariograma é isotrópico. Nessa situação uma série de problemas devem ser resolvidos, entre os quais se destacam os seguintes. (A identificação dos problemas corresponde às questões propostas no capítulo 1; p.ex., à questão (a) foi dividida nos dois problemas, A1 e A2.)

Problema A1. Determinar a configuração da amostra para mapeamento.

Se um risco máximo σ_{\max}^2 estabelecido para um plano amostral for tão alto que não poderá ser ultrapassado nem por uma amostra de tamanho $n = 1$, então não há razão para planejar esta amostragem, pois basta amostrar um ponto x qualquer pertencente à região R para que o plano seja σ_{\max}^2 -aceitável. Isto é uma situação extrema, quase nunca encontrada na prática. Uma outra situação, bem mais comum, apresenta-se quando é necessário escolher uma configuração e um tamanho de amostra que atribuam σ_{\max}^2 -aceitabilidade à um plano que pode não ser σ_{\max}^2 -aceitável. Em casos assim, a única maneira de garantir σ_{\max}^2 -aceita-

bilidade ao plano é fazê-lo prescrever uma amostragem sistemática. Este resultado é apresentado em Burgess, Webster e McBratney (1981) mediante argumentos de plausibilidade, considerando a minimização da variância máxima de krigagem. A necessidade de definir qual tipo de amostra sistemática será adotada leva à outro problema:

Problema A2. Escolher uma configuração sistemática para mapeamento.

Burgess, Webster e McBratney (1981), McBratney, Webster e Burgess (1981), Yfantis, Flatman e Behar (1987) consideram este problema resolvido (sem demonstrar) quando a amostragem é feita em malhas regulares. O critério usado para se chegar à esta conclusão é ainda a minimização da variância máxima de krigagem. Como os únicos polígonos regulares capazes de formar uma malha no plano são o triângulo equilátero, o quadrado e o hexágono regular (Matérn, 1986, p. 74), uma malha regular para mapeamento será uma dentre três tipos: malha triangular, malha quadrada ou malha hexagonal. A comparação entre elas leva ao estudo dos problemas que se seguem.

Problema B1. Localizar a variância máxima de krigagem

Também sem prova, Burgess, Webster e Mcbratney (1981) estabelecem o centro de cada polígono constituinte da malha como o ponto onde ocorre a máxima variância de krigagem. Para que essa conclusão seja válida, é necessário considerar a

amostragem executada sobre todo o conjunto \mathbb{R}^2 . Esta pressuposição é necessária para evitar o problema das bordas (o contorno) de uma região finita.

Problema B2. Minimizar a variância máxima de krigagem

Burgess, Webster e Mcbratney (1981) concluem a partir de estudos numéricos que na malha triangular a variância máxima de krigagem é menor do que na quadrada e na hexagonal. Neste capítulo, uma confirmação também numérica é obtida e um argumento de plausibilidade é apresentado para a solução deste problema. Ainda é necessário pressupor malhas infinitas. A malha triangular produz a menor variância, a quadrada é ligeiramente menos precisa e a malha hexagonal é ainda menos precisa do que as anteriores.

A malha hexagonal não será considerada nesta dissertação como opção viável na amostragem espacial, por causa de sua inferioridade na escala acima e porque não apresenta vantagens em termos de custo em comparação com a triangular e a quadrada. A escolha fica então entre as malhas triangular e quadrada.

Problema B3. Comparar o risco e o custo dos planos com malhas triangular e quadrada.

Ainda um estudo numérico levou Burgess, Webster e McBratney (1981) à concluir que o aumento no risco do plano ao trocar de malha triangular para quadrada é pequeno e é compensado pela redução no custo do plano, argumentando que a

alocação no campo de uma amostra em malha quadrada é menos custosa do que uma triangular. Semelhantemente, neste capítulo, um estudo numérico confirmará a pequena diferença entre triangular e quadrada. Para o desenvolvimento desse estudo numérico e também para propôr um algoritmo que determine o espaçamento entre pontos amostrados, é necessário ainda resolver os problemas que se seguem:

Problema C1. Determinar a vizinhança de pontos amostrados que deve ser usada na krigagem.

A literatura parece não conter critérios objetivos para resolver este problema. Porém, Vieira, Hatfield, Nielsen e Biggar (1983) propõem a utilização de uma validação cruzada, enquanto Journel (1988) sugere um número mínimo de pontos nos octantes em torno do ponto sob krigagem, para enfrentar o problema. Neste capítulo, uma discussão deste problema e uma proposta de solução são apresentadas.

Problema C2. Calcular a variância máxima de krigagem

Em 2.1.2 mostrou-se uma fórmula que permite calcular a variância de krigagem em qualquer caso. Uma particularização será apresentada para o semivariograma linear, pressupondo malhas infinitas. Estas fórmulas são explícitas e permitem o cálculo da máxima variância tanto em malhas triangulares como quadradas.

A necessidade de adequar os resultados para malha infinita à realidade prática, que trata com malhas internas a uma

região finita, leva à formulação de mais um problema:

Problema C3. Calcular a variância máxima de krigagem em função do semivariograma, da área da região R (finita) e do tamanho da amostra.

Fórmulas e um algoritmo não encontrados na literatura são apresentados para solucionar este problema. Em malhas que amostram uma região finita não se pode garantir que a variância calculada para o centro do polígono seja a máxima, pois não há controle das bordas. Portanto, a solução aqui apresentada ao problema C3 será somente aproximada.

Todos os problemas formulados acima tem por objetivo fornecer informações para planejar amostras para mapeamento quando o semivariograma é isotrópico. Para os problemas A1, A2 e B1 serão adotadas as soluções sugeridas na literatura. Os problemas B2, B3, C1, C2, C3 serão tratados na seção 3.1, subsidiando uma proposta para a configuração e tamanho da amostra espacial, incluindo um algoritmo e um exemplo de aplicação na ciência do solo.

Um estudo para o caso anisotrópico será apresentado na seção 3.2, utilizando uma proposta de Burgess, Webster e McBratney (1981), os quais sugerem (sem demonstração) uma maneira de planejar a amostragem no caso anisotrópico. Uma proposição é demonstrada, que justifica a proposta dos autores citados. Ela é rerepresentada na forma de um algoritmo, com um

exemplo numérico de aplicação.

Como ficará claro no desenvolver deste capítulo, uma estimação prévia do semivariograma se fará necessária. Na seção 3.3 apresentam-se considerações sobre a pré-amostragem visando estimá-lo.

A seção 3.4 conclui este capítulo com comentários sobre os resultados encontrados; alguns problemas que ainda aguardam solução também são apresentados.

3.1. Semivariograma Isotrópico

Nesta seção receberão atenção alguns problemas contidos no estudo do risco de planos para mapeamento, no caso isotrópico.

As sugestões da literatura para solução dos problemas A1, A2 (relativos às configurações) e B1 (localização da variância máxima de krigagem) são adotadas nesta dissertação, que passa a discutir a sequência de problemas à partir do problema B2 (minimização da variância máxima de krigagem). Em 3.1.1 trata-se de B2, B3, (hierarquia das configurações) e C1, C2, C3 (cálculo do espaçamento e tamanho da amostra). Configuração e tamanho da amostra espacial para mapeamento são conclusivamente determinadas em 3.1.2.

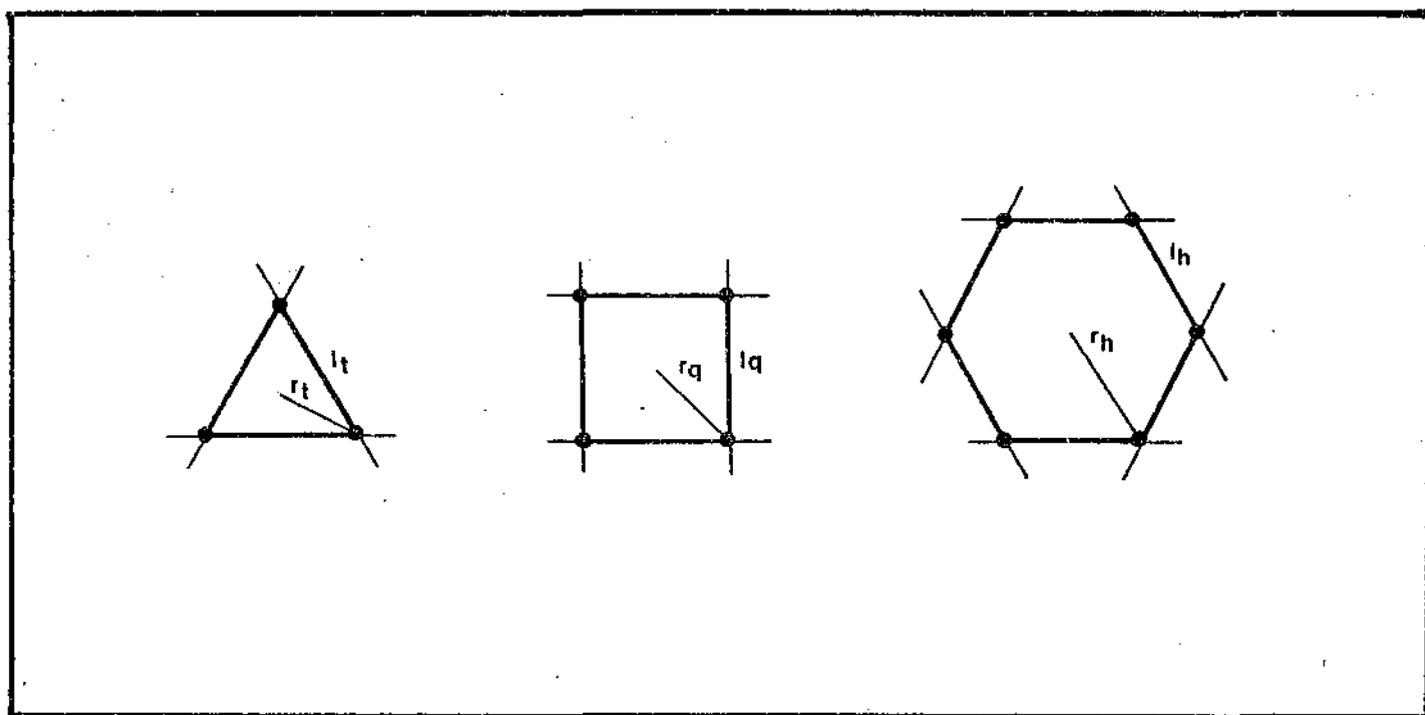
3.1.1. Variância Máxima de Krigagem

Problema B2. Minimizar a variância máxima de krigagem.

Burgess, Webster e McBratney (1981) e McBratney, Webster e Burgess (1981) afirmam que a malha triangular produz a menor variância máxima de krigagem dentre os três tipos possíveis. A formalização dessa afirmação desenvolve-se nos resultados abaixo:

Lema 3.1. (Relações entre polígonos das malhas). Seja uma malha regular que cobre todo o \mathbb{R}^2 . Seja δ uma densidade de amostragem fixa. Se l_t é o lado de um dos triângulos equiláteros de uma malha triangular com densidade δ , então os lados correspondentes dos quadrados e hexágonos regulares em malhas quadrada e hexagonal com densidade δ são $l_q = (3^{1/2}/2)^{1/2} \cdot l_t \cong 0,9306 \cdot l_t$ e $l_h = (2/3)^{1/2} \cdot l_t \cong 0,8165 \cdot l_t$. Se r_t é o raio da circunferência que passa pelos vértices de um dos triângulos equiláteros de uma malha triangular com densidade δ , então os raios correspondentes dos quadrados e hexágonos regulares em malhas quadrada e hexagonal com densidade δ são $r_q = \left[3^{1/2} \cdot (3/4) \right]^{1/2} \cdot r_t \cong 1,1398 \cdot r_t$ e $r_h = 2^{1/2} \cdot r_t \cong 1,4142 \cdot r_t$ (veja quadro 3.3).

A demonstração deste lema encontra-se no apêndice.



QUADRO 3.3. Raio r da circunferência que passa pelos vértices e lado l dos polígonos que formam malhas triangular, quadrada e hexagonal com mesma densidade de amostragem.

Lema 3.2. (Otimidade da malha triangular). Seja a população $\{z(x):x \in \mathbb{R}^2\}$ realização do processo estocástico $\{Z(x):x \in \mathbb{R}^2\}$ que satisfaz a hipótese intrínseca com semivariograma não-decrescente e isotrópico. Seja um plano-amostral para mapeamento dessa população que estabelece uma malha regular com densidade de amostragem fixa δ sobre todo \mathbb{R}^2 . Então a malha triangular minimiza o risco desse plano.

Este lema não terá uma demonstração formal nesta dissertação, restringindo-se tão somente à um argumento de plausibilidade, apresentado no apêndice. Este argumento permite também estabelecer a hierarquia das malhas quanto ao valor da variância máxima de krigagem: triangular é superior à quadrada, que por sua vez é superior à hexagonal.

Problema B3. Comparar o risco e o custo dos planos para malhas triangular e quadrada.

A necessidade de efetuar cálculos do risco impedirá o tratamento desse problema neste ponto, pois ainda não há um algoritmo para execução dos cálculos necessários. Ele será discutido na consideração do problema C3.

Problema C1. Determinar a vizinhança de pontos amostrados que deve ser usada na krigagem.

Sugestões de critérios para o estabelecimento dessa vizinhança podem ser encontrados em Journel (1988), Burgess, Webster e McBratney (1981) e Vieira, Hatfield, Nielsen e Biggar

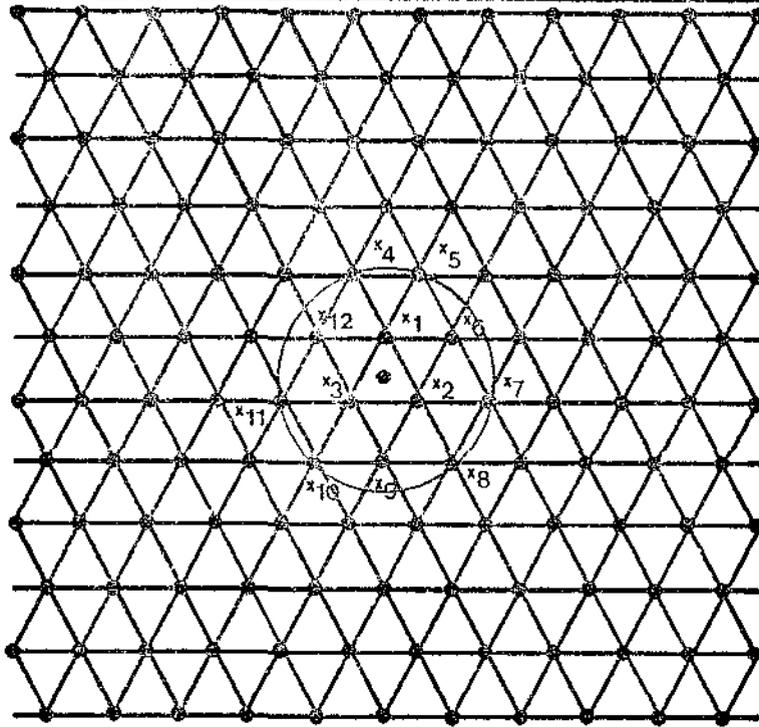
(1983). Todos estes são baseados no critério que consiste em "não deixar de fora nenhum ponto que contenha informação relevante para a krigagem, mas não incluir nenhum ponto que forneça informação duplicada". Esta propriedade da vizinhança quanto à economia de informação é algumas vezes referida como *efeito de tela (screen effect)* da vizinhança. Tanto em malha triangular quanto em quadrada, diferentes vizinhanças podem ser consideradas, em função do número de pontos amostrados que elas encerram, veja quadro 3.4. Neste trabalho será analisada uma vizinhança de 12 pontos, estabelecida após uma análise do efeito de tela proporcionado por ela: todo ponto fora desta vizinhança guarda informação que já foi considerada pelo menos em parte por no mínimo um ponto incluso na vizinhança.

Problema C2. Calcular a variância máxima de krigagem

Depois de estabelecer uma vizinhança de 12 pontos para as krigagens, pode-se realizar o cálculo do risco de um plano.

Lema 3.3. (Cálculo do risco). Seja o processo $\{Z(x):x \in \mathbb{R}^2\}$ que satisfaz a hipótese intrínseca, com o semivariograma isotrópico linear (efeito pepita γ_0 e inclinação a). Se uma malha triangular com triângulos de lado l_t é usada para amostrar uma realização desse processo, então o risco do plano amostral é dado por $\sigma_{KMTL}^2 = 2\lambda_T' \gamma_T - \lambda_T' \Gamma_T \lambda_T$ onde $\Gamma_T = \Gamma_{RT} \cdot a \cdot l_t + (1 - D) \cdot \gamma_0$, $\gamma_T = \gamma_{RT} \cdot a \cdot l_t + 1 \cdot \gamma_0$ e λ_T é obtido conforme a equação matricial apresentada em 2.1.2. A matriz simétrica

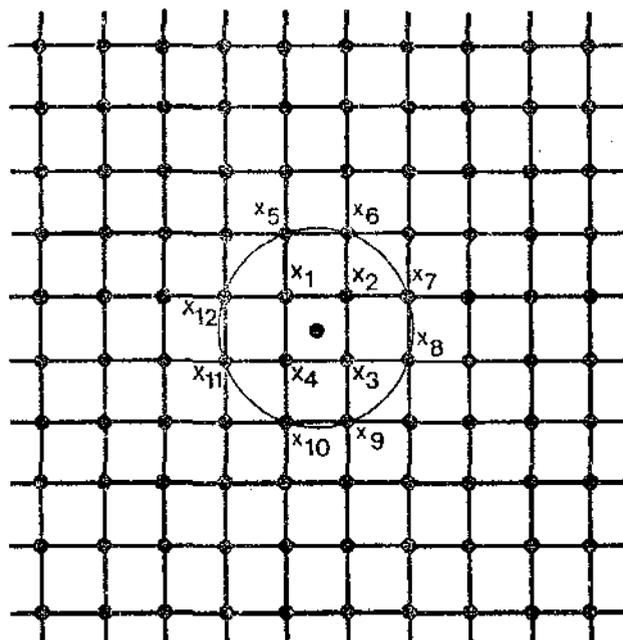
(a)



Número n de pontos da vizinhança	Valor do coeficiente T_n	
	Exato	Aproximado
3	$\sqrt{3}/3$	0,5774
6	$\sqrt{12}/3$	1,1547
12	$\sqrt{21}/3$	1,5275
18	$\sqrt{39}/3$	2,0817
21	$\sqrt{48}/3$	2,3094
27	$\sqrt{57}/3$	2,5166
36	$\sqrt{84}/3$	3,0551
42	$\sqrt{93}/3$	3,2146
48	$\sqrt{111}/3$	3,5119
54	$\sqrt{129}/3$	3,7859
63	$\sqrt{147}/3$	4,0415

QUADRO 3.4. O raio da circunferência que passa por pontos amostrados delimitando uma vizinhança para a krigagem no centro do polígono tem comprimento que depende do número n de pontos encerrados pela circunferência. Seu valor é: (a) T_n na malha triangular com triângulos equiláteros de lado 1.

(b)



Número n de pontos da vizinhança	Valor do coeficiente Q_n	
	Exato	Aproximado
4	$\sqrt{2}/2$	0,6580
12	$\sqrt{10}/2$	1,4714
16	$\sqrt{18}/2$	1,9741
24	$\sqrt{26}/2$	2,3726
32	$\sqrt{34}/2$	2,7132
44	$\sqrt{50}/2$	3,2902

QUADRO 3.4. (b) $Q_n \cdot \frac{1}{a}$ na malha quadrada com quadrados de lado $\frac{1}{a}$

Γ_{RT} e o vetor γ_{RT} são constantes para toda densidade de amostragem, todo valor do efeito pepita e da inclinação do semivariograma e são dados por

$$\Gamma_{RT} = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & \sqrt{3} & 2 & \sqrt{3} & 2 & \sqrt{3} & 1 \\ 0 & 1 & 2 & \sqrt{3} & 1 & 1 & 1 & 1 & \sqrt{3} & 2 & \sqrt{3} & \\ 0 & \sqrt{3} & 2 & \sqrt{3} & 2 & \sqrt{3} & 1 & 1 & 1 & 1 & 1 & \\ 0 & 1 & \sqrt{3} & \sqrt{7} & 3 & \sqrt{7} & \sqrt{7} & \sqrt{7} & 2 & 1 & & \\ & 0 & 1 & 2 & \sqrt{7} & \sqrt{7} & 3 & \sqrt{7} & \sqrt{3} & & & \\ & & 0 & 1 & \sqrt{3} & 2 & \sqrt{7} & \sqrt{7} & 2 & & & \\ & & & 0 & 1 & \sqrt{3} & \sqrt{7} & 3 & \sqrt{7} & & & \\ & & & & 0 & 1 & 2 & \sqrt{7} & \sqrt{7} & & & \\ & & & & & 0 & 1 & \sqrt{3} & 2 & & & \\ & & & & & & 0 & 1 & \sqrt{3} & & & \\ & & & & & & & 0 & 1 & & & \\ & & & & & & & & 0 & 1 & & \\ & & & & & & & & & 0 & 1 & \\ & & & & & & & & & & 0 & 1 \\ & & & & & & & & & & & 0 \end{bmatrix}, \gamma_{RT} = \frac{1}{3} \begin{bmatrix} \sqrt{3} \\ \sqrt{3} \\ \sqrt{3} \\ \sqrt{21} \\ \sqrt{21} \\ \sqrt{12} \\ \sqrt{21} \\ \sqrt{21} \\ \sqrt{12} \\ \sqrt{21} \\ \sqrt{21} \\ \sqrt{12} \end{bmatrix}$$

Se a malha é quadrada com quadrados de lado l_q a compoendo, então o risco torna-se $\sigma_{KMQL}^2 = 2 \cdot \lambda'_\alpha \gamma_\alpha - \lambda'_\alpha \Gamma_\alpha \lambda_\alpha$ onde $\Gamma_\alpha = \Gamma_{RQ} \cdot a \cdot l_q + (1.1' - D) \cdot \gamma_0$ e $\gamma_\alpha = \gamma_{RQ} \cdot a \cdot l_q + 1 \gamma_0$. A matriz e vetor constantes são dadas por

$$\Gamma_{RQ} = \begin{bmatrix} 0 & 1 & \sqrt{2} & 1 & 1 & \sqrt{2} & 2 & \sqrt{5} & \sqrt{5} & 2 & \sqrt{2} & 1 \\ 0 & 1 & \sqrt{2} & \sqrt{2} & 1 & 1 & \sqrt{2} & 2 & \sqrt{5} & \sqrt{5} & 2 & \\ 0 & 1 & \sqrt{5} & 2 & \sqrt{2} & 1 & 1 & \sqrt{2} & 2 & \sqrt{5} & & \\ & 0 & 2 & \sqrt{5} & \sqrt{5} & 2 & \sqrt{2} & 1 & 1 & \sqrt{2} & & \\ & & 0 & 1 & \sqrt{5} & \sqrt{8} & \sqrt{10} & 3 & \sqrt{5} & \sqrt{2} & & \\ & & & 0 & \sqrt{2} & \sqrt{5} & 3 & \sqrt{10} & \sqrt{8} & \sqrt{5} & & \\ & & & & 0 & 1 & \sqrt{5} & \sqrt{8} & \sqrt{10} & 3 & & \\ & & & & & 0 & \sqrt{2} & \sqrt{5} & 3 & \sqrt{10} & & \\ & & & & & & 0 & 1 & \sqrt{5} & \sqrt{8} & & \\ & & & & & & & 0 & \sqrt{2} & \sqrt{5} & & \\ & & & & & & & & 0 & 1 & & \\ & & & & & & & & & 0 & 1 & \\ & & & & & & & & & & 0 & 1 \\ & & & & & & & & & & & 0 \end{bmatrix}, \gamma_{RQ} = \frac{1}{2} \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \\ \sqrt{2} \\ \sqrt{2} \\ \sqrt{10} \end{bmatrix}$$

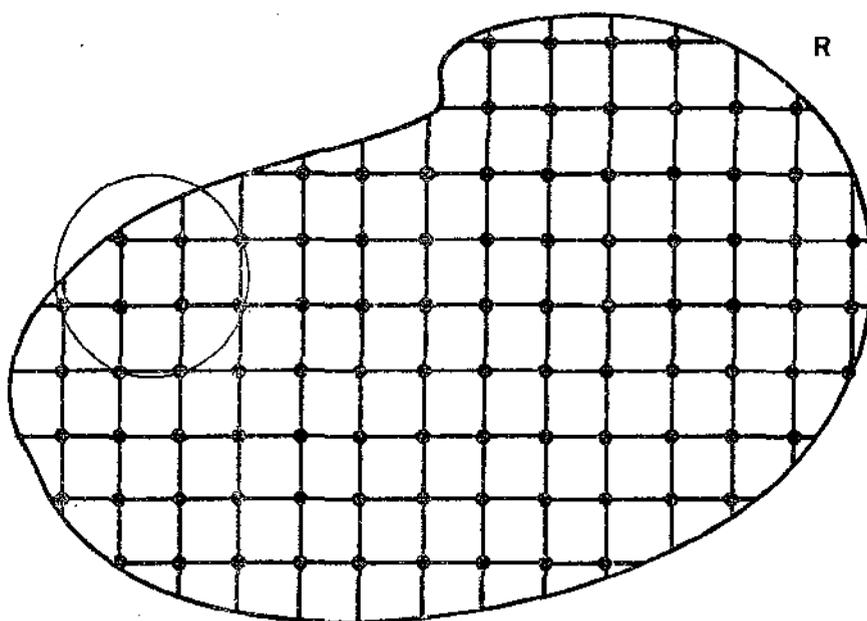
Demonstração. Veja apêndice.

O lema 3.3 é uma solução para o problema C2. Porém, a prática do planejamento amostral não lida com malhas infinitas. O problema que surge então é calcular o risco do plano para malhas que amostram uma região finita.

Problema C3. Calcular a variância máxima de krigagem em função do semivariograma, da área da região R (finita) e do tamanho da amostra.

Na amostragem de populações $\{z(x):x \in R \subset \mathbb{R}^2\}$ onde a região R é limitada não se pode garantir que a variância máxima de krigagem ocorre nos centros dos polígonos por causa do efeito das bordas (veja quadro 3.5). Por isso cria-se uma faixa fronteira onde a variância de krigagem pode ultrapassar os valores obtidos nos centros dos polígonos mais internos. Esta faixa é função da forma do contorno da região. Nesta dissertação desprezar-se-á este efeito e considerar-se-á o risco do plano como a variância de krigagem que ocorre no centro de um polígono com vizinhança completa (12 pontos). Para chegar-se ao cálculo do risco em função do semivariograma, da área da região e do tamanho da amostra necessita-se de um resultado auxiliar:

Lema 3.4. (Relação entre malha e região). Seja a região $R \subset \mathbb{R}^2$, com área s . Se uma malha regular estende-se sobre R com densidade de amostragem δ , então: (i) o tamanho n da amostra em R é $\delta \cdot s$ (ii) o lado l dos triângulos equiláteros



QUADRO 3.5. Em malhas que amostram uma região limitada, a ocorrência da variância máxima de krigagem desloca-se do centro do polígono para algum outro ponto porque a vizinhança de pontos amostrados torna-se assimétrica.

constituintes de uma malha triangular é $\left[(2.3^{1/2} \cdot s) / (3 \cdot n) \right]^{1/2}$

(iii) o lado l_q dos quadrados constituintes de uma malha quadrada é $(s/n)^{1/2}$.

Demonstração. Veja apêndice.

O lema 3.5 resolve aproximadamente o problema C3.

Lema 3.5. (Cálculo do risco em regiões limitadas). Seja um processo $\{Z(x): x \in R \subset \mathbb{R}^2\}$ satisfazendo a hipótese intrínseca, com semivariograma isotrópico linear, com efeito γ_0 e inclinação a . Seja uma malha regular para amostrar uma realização desse processo, estendida sobre R . Se a área de R é s e a densidade de amostragem da malha é tal que aloque n pontos em R , então a variância de krigagem obtida no centro dos polígonos que tem uma vizinhança de pontos amostrados completa (12 pontos) é dada por (i) malha triangular $\sigma_{KMTL}^2 = 2\lambda'_T \gamma_T - \lambda'_T \Gamma_T \lambda_T$ onde $\Gamma_T = \left[2.3^{1/2}/3 \right]^{1/2} \cdot a \cdot s^{1/2} \Gamma_{RT} / n^{1/2} + (1.1' - I) \cdot \gamma_0$, $\gamma_T = \left[2.3^{1/2}/3 \right]^{1/2} \cdot a \cdot s^{1/2} \gamma_{RT} / n^{1/2} + 1 \cdot \gamma_0$ e λ_T é obtido conforme a equação matricial apresentada em 2.1.2. A matriz Γ_{RT} e o vetor γ_{RT} são os do lema 3.3. (ii) malha quadrada: $\sigma_{KMQL}^2 = 2\lambda'_Q \gamma_Q - \lambda'_Q \Gamma_Q \lambda_Q$ onde $\Gamma_Q = a \cdot s^{1/2} \Gamma_{RQ} / n^{1/2} + (1.1' - I) \cdot \gamma_0$, $\gamma_Q = a \cdot s^{1/2} \gamma_{RQ} / n^{1/2} + 1 \cdot \gamma_0$ e λ_Q é obtido conforme a equação matricial de 2.1.2. A matriz Γ_{RQ} e o vetor γ_{RQ} são os do lema 3.3.

Demonstração. Veja apêndice.

Aplicação numérica:

Uma aplicação numérica do lema 3.5 pode surgir no estudo da capacidade de retenção de água do solo na camada 0 - 20 cm de profundidade sob pressão de 0,33 atm (g água/g solo). Este estudo foi realizado no Campus da Escola Superior de Agricultura de Lavras, juntamente com o estudo de algumas outras variáveis espaciais. Uma malha quadrada com 100 pontos no espaçamento igual à 10 metros amostrou um latossolo vermelho-amarelo com respeito à várias características físicas, entre elas a citada acima, que apresentou média de 28,27 g água/g solo e desvio-padrão de 2,78 g água/g solo. O semivariograma da capacidade de retenção foi achado anisotrópico geométrico, com o modelo linear (Carvalho, 1991) com a inclinação na direção do eixo menor da elipse de anisotropia igual à $0,16 \text{ (g água/g solo)}^2/\text{m}$ e na direção do eixo maior $0,07 \text{ (g água/g solo)}^2/\text{m}$, sendo o efeito pepita igual à $1,5 \text{ (g água/g solo)}^2$. Nenhuma das variáveis foi achada sendo de variação isotrópica. Um exemplo para o caso isotrópico pode ser encontrado modificando-se ligeiramente os resultados acima, criando uma situação fictícia para a distribuição espacial da variável em questão. Assim, considerar-se-á para efeito de ilustração um semivariograma isotrópico linear, com inclinação $a = 0,14 \text{ (g água/g solo)}^2/\text{m}$ e efeito pepita γ_0 variável no conjunto $\{0; 1,5; 7,5; 15,0; 1500000,0 \text{ (g água/g solo)}^2\}$. A razão para variar o efeito pepita está baseada no fato de que este parâmetro mede a quantidade de variação espacial que não é explicada pela continuidade

espacial na população sob estudo. Com esses parâmetros procedeu-se o cálculo do risco de um plano amostral para mapeamento da variável espacial 'capacidade de retenção de água do solo na camada 0-20 cm de profundidade sob pressão de 0,33 atmosfera' numa região R com área $s = 10.000 \text{ metros}^2$, usando malha triangular e quadrada, veja quadro 3.6. Nesse quadro resume-se o cálculo do risco para cada tipo de malha em função do tamanho n da amostra alocada na região e do efeito pepita γ_0 , p. ex., se o efeito pepita é $\gamma_0 = 1,5 \text{ (g água/ g solo)}^2$ e uma amostra de 35 pontos for alocada na região em malha triangular, o risco do plano é aproximadamente $3,12 \text{ (g água/ g solo)}^2$; se a alocação for feita em malha quadrada o risco sobe para $3,17 \text{ (g água/ g solo)}^2$. Os gráficos do quadro 3.7 resumem toda a informação numérica do quadro 3.6.

Tanto pelos resultados numéricos como pelos gráficos percebe-se a pequena diferença entre os riscos do plano usando malha triangular e malha quadrada, diminuindo esta diferença à medida que o efeito pepita aumenta. Este resultado é coerente com os encontrados por Burgess, Webster e Mcbratney (1981). Estes resultados serão usados para propôr uma configuração de amostra, também determinando seu tamanho.

3.1.2. Configuração e Tamanho da Amostra

Os resultados de 3.1.1 mostram o pequeno decréscimo no risco do plano ao trocar uma malha quadrada por uma triangular.

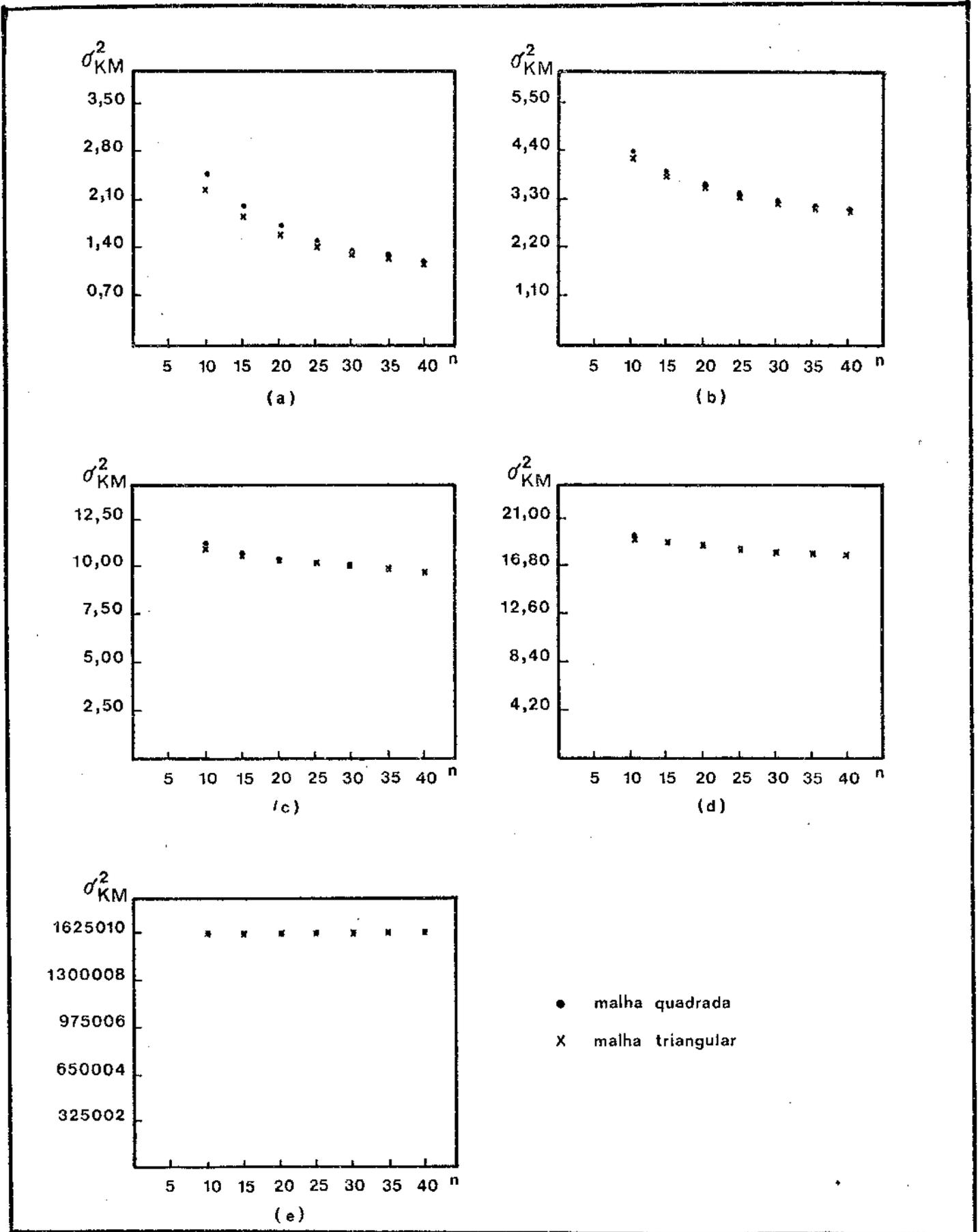
Risco do plano para mapeamento				
γ_0	n	Triangular	Quadrada	
		σ^2_{KMTL}	σ^2_{KMQL}	
0	10	2,30	2,44	
	15	1,88	2,00	
	20	1,63	1,73	
	25	1,45	1,55	
	30	1,33	1,41	
	35	1,23	1,31	
	40	1,15	1,22	
1,5	10	4,25	4,35	
	15	3,81	3,89	
	20	3,55	3,62	
	25	3,36	3,43	
	30	3,23	3,28	
	35	3,12	3,17	
	40	3,04	3,08	

(Continua)

QUADRO 3.6. Risco do plano amostral para mapeamento da variável espacial "capacidade de retenção de água do solo na camada 0 - 20 cm de profundidade sob pressão de 0,33 atm", numa região com área de 10.000 m². Neste exemplo (fictício), o semivariograma isotrópico é linear, com inclinação 0,14 ((g água/g solo)²/m). O risco está em função do efeito pepita γ_0 , do tamanho n da amostra e do tipo de malha.

Risco do plano para mapeamento			
γ_0	n	<u>Triangular</u>	<u>Quadrada</u>
		σ^2_{KMTL}	σ^2_{KMQL}
7,5	10	11,29	11,35
	15	10,77	10,81
	20	10,45	10,49
	25	10,23	10,26
	30	10,06	10,09
	35	9,93	9,96
	40	9,82	9,85
	15,0	10	19,65
15		19,07	19,10
20		18,72	18,75
25		18,48	18,50
30		18,29	18,31
35		18,15	18,17
40		18,03	18,05
1 500 000,0		10	1 625 003,81
	15	1 625 003,11	1 625 003,11
	20	1 625 002,69	1 625 002,69
	25	1 625 002,41	1 625 002,41
	30	1 625 002,20	1 625 002,20
	35	1 625 002,03	1 625 002,03
	40	1 625 001,90	1 625 001,90

QUADRO 3.6. (Conclusão)



QUADRO 3.7. Risco de um plano para mapeamento em função da malha de amostragem, tamanho n da amostra e efeito pepita. (a) efeito pepita $\gamma_0=0,0$ (b) $\gamma_0=1,5$ (c) $\gamma_0=7,5$ (d) $\gamma_0=15,0$ (e) $\gamma_0=1500000,0$.

Burgess, Webster e McBratney (1981) sustentam que esse decréscimo não é compensado pelo maior custo de alocação da malha triangular, concluindo que uma malha quadrada deve ser preferida quando deseja-se uma amostra para subsidiar o mapeamento de uma variável com semivariograma isotrópico numa região de solo limitada. Esta dissertação aceita esta proposta e propõe um algoritmo para a determinação do tamanho da amostra quando o semivariograma (isotrópico) é linear, veja quadro 3.8. Esta sequência de instruções permite a determinação do tamanho da amostra em malha quadrada que será alocada numa região de solo limitada e também o espaçamento entre os pontos amostrados. Paralelamente à este algoritmo prático apresenta-se no quadro 3.9 um exemplo numérico de aplicação, associado à situação prática envolvendo a variável 'capacidade de retenção de água do solo na camada 0 - 20 cm de profundidade sob pressão de 0,33 atmosfera', apresentada em 3.1.1. Nesse exemplo considerou-se a necessidade de um plano amostral com um risco menor ou igual à 4,00 (g água/g solo)². A determinação do número de pontos foi feita por unidade de área, cada unidade com 1 ha (10000 m²), contida na região R. Para que o plano seja aceitável, bastaria amostrar 14 pontos em cada unidade de área de 10000 m², em malha quadrada, espaçados de 26,7 m.

A teoria, algoritmos e exemplos tratados até aqui estão baseados na isotropia do semivariograma. O caso anisotrópico será tratado na próxima seção.

 Determinação do Tamanho da Amostra

Passo	Procedimento
1	Especificar o valor do efeito pepita γ_0 , da inclinação a , da área s da região à ser amostrada e do risco máximo tolerável σ_{max}^2 .
2	Usar os resultados do lema 3.5 para construir uma tabela como no quadro 3.6 e/ou um gráfico como no quadro 3.7 obtendo risco em função de n (o tamanho da amostra), n variando até um número de pontos amostrados máximo, de acordo com os recursos disponíveis.
3	Usar a tabela e/ou gráfico do passo 2 para escolher o valor de n tal que o risco associado seja menor ou igual à σ_{max}^2 . O espaçamento entre pontos amostrados é $\sqrt{\frac{s}{n}}$.

QUADRO 3.8. Algoritmo para a determinação do tamanho da amostra e do espaçamento entre pontos amostrados para fins de mapeamento. Esta amostra será alocada numa região R com área s , em malha quadrada. O semivariograma da variável de interesse deve ser isotrópico e linear, com efeito pepita γ_0 e inclinação a . Esta malha faz o plano amostral ser σ_{max}^2 - aceitável.

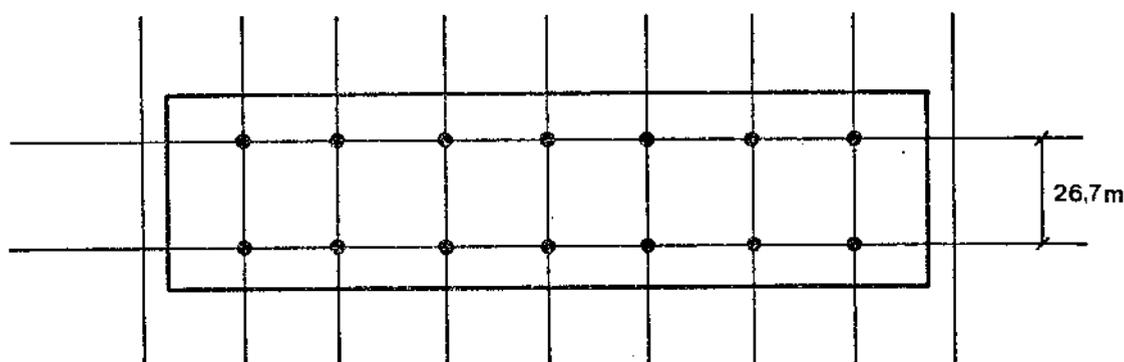
Determinação do Tamanho da Amostra

1^o Passo

Região		Plano de Amostragem	
Semivariograma	Unidade de área	Risco	
γ_0	a	s	σ_{max}^2
1,50	0,14	10000	4,00

2^o Passo. Veja quadros 3.6 e 3.7.

3^o Passo. $n = 14$ pontos amostrados na região, em malha quadrada. O espaçamento entre pontos amostrados será de $l = 100/\sqrt{14} = 26,7$.



Escala 1:2000
(1cm = 20m)

QUADRO 3.9. Exemplo de aplicação do algoritmo do quadro 3.8. Unidades das medidas: área (m^2), efeito pepita e risco ($(g \text{ água}/g \text{ solo})^2$), inclinação ($(g \text{ água}/g \text{ solo})^2/m$) e espaçamento (m).

3.2. Semivariograma Anisotrópico

Nesta seção estudar-se-á o risco de um plano amostral para mapeamento quando o semivariograma é anisotrópico. Um semivariograma de modelo linear é dito ser *anisotrópico geométrico* se ele pode ser escrito na forma seguinte (Journal, 1988; Isaaks e Srivastava, 1989):

$$\gamma(h) = \begin{cases} 0 & , |h| = 0 \\ \gamma_0 + a_h \cdot |h| & , |h| \neq 0 \end{cases}$$

onde a_h (a inclinação na direção do vetor h) depende da direção do vetor h , de tal maneira que os inversos de a_h para cada direção formam uma elipse, veja quadro 3.10. Isto pode ser expresso de outra maneira:

$$\gamma(h) = \begin{cases} 0 & , |h| = 0 \\ \gamma_0 + |A \cdot R \cdot h| & , |h| \neq 0 \end{cases}$$

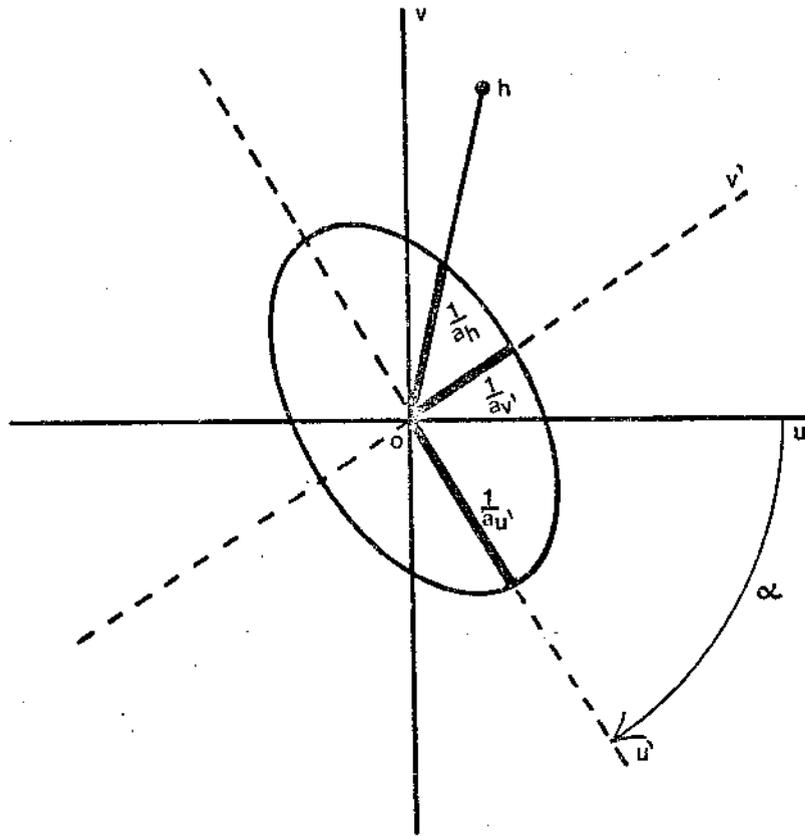
onde:

(i) a matriz $R = \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix}$ rotaciona o sistema

original de coordenadas (O_u, O_v) para o sistema $(O_{u'}, O_{v'})$ definido pelos eixos da elipse (α é o ângulo entre O_u e $O_{u'}$).

(ii) a matriz $A = \begin{bmatrix} a_{u'} & 0 \\ 0 & a_{v'} \end{bmatrix}$ transforma a elipse numa

circunferência.



O vetor h é expressado no sistema de coordenadas (O_u, O_v) . O sistema $(O_{u'}, O_{v'})$ é o sistema definido pelos eixos da ellipse de anisotropia.

$a_{u'}$ é o coeficiente angular na direção do eixo $O_{u'}$.

$a_{v'}$ é o coeficiente angular na direção do eixo $O_{v'}$.

a_h é o coeficiente angular na direção do vetor h .

QUADRO 3.10. Anisotropia geométrica num semivariograma linear.

Burgess, Webster e McBratney (1991) sugerem, sem demonstração, uma configuração de amostra e um método para determinar o espaçamento entre seus pontos, que confere σ_{\max}^2 - aceitabilidade à um plano amostral quando o semivariograma linear é anisotrópico geométrico. Em 3.2.1 demonstra-se um lema que fundamenta e dá validade a essa proposta. Em 3.2.2 apresenta-se um algoritmo para determinação do tamanho dessa amostra e o espaçamento entre os pontos, com um exemplo de aplicação em ciência do solo.

3.2.1. Configuração

Devido ao fato de que pode-se levar uma elipse de anisotropia à uma circunferência de isotropia através da matriz A , é possível utilizar a solução apresentada para o caso isotrópico também no caso anisotrópico, à menos de uma pequena modificação na malha. Burgess, Webster e McBratney (1981) propõem uma amostragem com malha retangular quando há anisotropia geométrica. O lema seguinte garante o acerto dessa proposta.

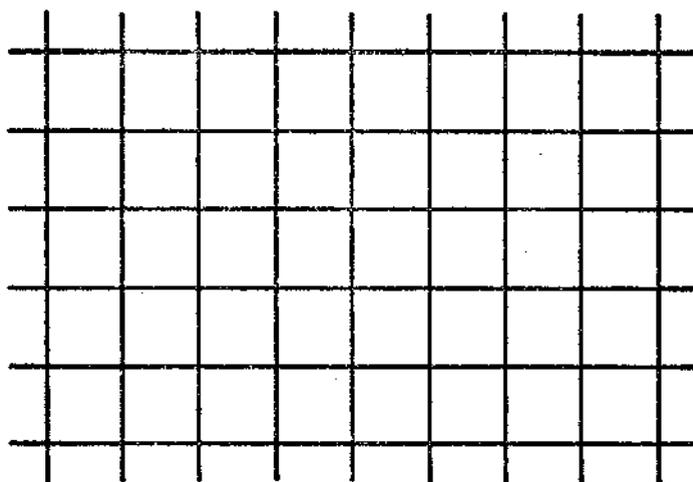
Lema 3.6. (Malha retangular σ_{\max}^2 - aceitável). Seja σ_{\max}^2 o risco máximo tolerável para um mapeamento. Seja o processo $\{W(x): x \in \mathbb{R}^2\}$ que satisfaz a hipótese intrínseca, com semivariograma isotrópico linear com efeito pepita γ_0 e inclinação a . Seja uma malha quadrada que amostra uma realização

$\langle w(x): x \in \mathbb{R}^2 \rangle$ desse processo. Sua densidade de amostragem é δ_i , o espaçamento entre pontos amostrados é l_i e a variância máxima de krigagem é $\sigma_{KM_i}^2$. Seja o processo $\langle Z(x): x \in \mathbb{R}^2 \rangle$ que também satisfaz a hipótese intrínseca, com semivariograma anisotrópico geométrico linear com efeito pepita γ_0 , inclinação na direção do eixo $O_{u'}$, igual à $a_{u'}$, e na direção ortogonal, inclinação $a_{v'} = a$. Seja uma malha retangular que amostra uma realização $\langle z(x): x \in \mathbb{R}^2 \rangle$ desse processo. Sua densidade de amostragem é δ_a , os lados dos retângulos estão descansando sobre as direções dos eixos da elipse de anisotropia (sendo que o lado sobre o eixo $O_{v'}$ tem comprimento l_i e o lado na direção ortogonal tem comprimento $f.l_i$, $f = a_{v'} / a_{u'}$, veja quadro 3.11, e sua variância máxima de krigagem é $\sigma_{KM_a}^2$. Então: (i) a densidade de amostragem na malha retangular é $\delta_a = \delta_i / f$ (ii) dada uma região $R \subset \mathbb{R}^2$ com área s finita, o número de pontos amostrados dentro de R em malha quadrada é $n_i = s / l_i^2$ e o número n_a de pontos amostrados dentro de R em malha retangular é $n_a = n_i / f$ (iii) a variância máxima de krigagem $\sigma_{KM_a}^2$ da malha retangular é igual a variância máxima de krigagem $\sigma_{KM_i}^2$ da malha quadrada.

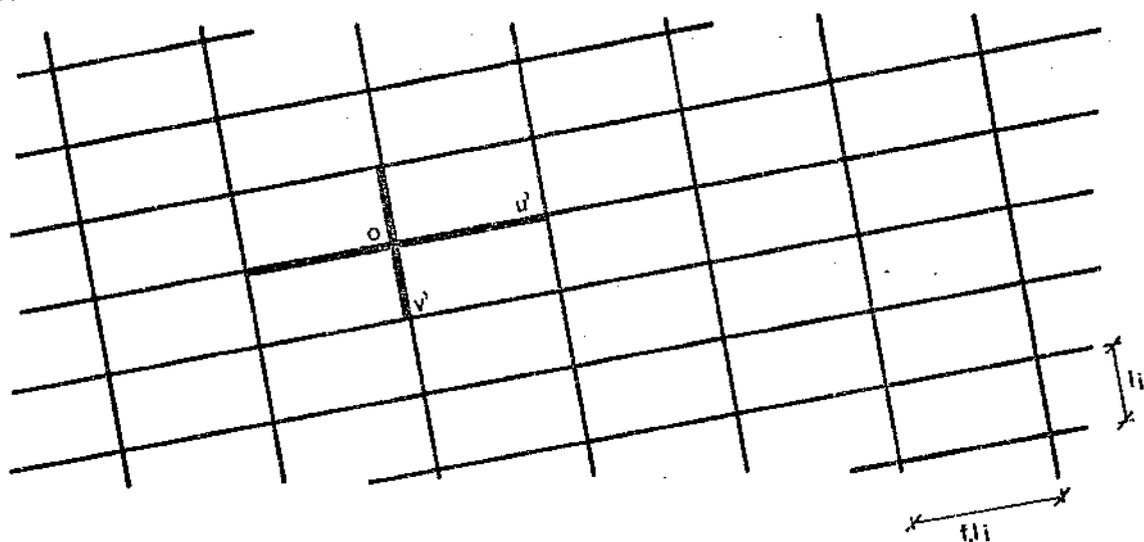
Demonstração. Veja apêndice.

Escolhida uma malha retangular para amostrar sob anisotropia resta a tarefa de determinar o número de pontos à serem amostrados e seus espaçamentos.

(a)



(b)



QUADRO 3.11. Disposição da malha: (a) sob isotropia, a orientação espacial da malha quadrada é irrelevante (b) sob anisotropia geométrica, a orientação espacial da malha retangular deve estar de acordo com os eixos de anisotropia $0_{u'}$ e $0_{v'}$.

3.2.2. Tamanho da Amostra

A determinação do número de pontos à serem amostrados no estudo de uma variável espacial que possui semivariograma anisotrópico geométrico com modelo linear é feita em função dos espaçamentos na malha retangular. No quadro 3.12 apresenta-se uma seqüência de instruções para dimensionar uma malha retangular numa situação prática. O quadro 3.13 é um paralelo aplicativo em ciência do solo, no estudo da variável espacial 'capacidade de retenção de água do solo na camada 0 - 20 cm de profundidade sob pressão de 0,33 atm', medida em (g água/g solo). Os valores dados às inclinações são fictícios. Nesse exemplo considerou-se um plano amostral com um risco menor ou igual à $3,50 \text{ (g água/g solo)}^2$, amostrando uma região dividida em unidades de área com 10000 m^2 . O valor do efeito pepita é de $1,50 \text{ (g água/g solo)}^2$. Um croqui mostra a região e a malha que a amostra, uma malha retangular com 10 pontos amostrados em cada 10000 m^2 de área.

A determinação da configuração, do espaçamento e do número de pontos amostrados para o mapeamento de uma variável de interesse foi tratada até aqui, tanto para o caso isotrópico quanto para um caso anisotrópico. Em ambos fica patente a necessidade de se conhecer o semivariograma antes de planejar a amostragem. Para a estimação do semivariograma uma pré-amostragem é sugerida na próxima seção.

 Determinação do Tamanho da Amostra

Passo	Procedimento
1	Especificar o valor do efeito pepita γ_0 , das inclinações $a_{u'}$ e $a_{v'}$, da área s da região a ser amostrada e do risco máximo tolerável σ_{\max}^2 . Obs. As direções dos eixos da elipse de anisotropia devem ser conhecidas.
2	Escolhe-se a direção de um dos eixos da elipse: p. ex., $0_{v'}$. Usando o coeficiente angular do semivariograma nesta direção ($a_{v'}$) dimensiona-se, pelo algoritmo do quadro 3.8, o espaçamento $l_{v'}$ entre pontos amostrados na direção $0_{v'}$.
3	O espaçamento $l_{u'}$ na direção $0_{u'}$ é obtido fazendo-se $l_{u'} = f.l_{v'}$, onde $f = a_{v'}/a_{u'}$.
4	O número de pontos amostrados nessa malha retangular construída é $n_a = s/(l_{u'} \cdot l_{v'}) = s/f.l_{v'}^2$
5	O risco do plano que prevê esta malha é o risco associado ao passo 2 deste algoritmo.

QUADRO 3.12. Algoritmo para a determinação do espaçamento entre pontos amostrados (e do tamanho da amostra) numa malha retangular estendida sobre uma região R com área s . O semivariograma da variável de interesse deve ser anisotrópico geométrico com modelo linear, com efeito pepita γ_0 e inclinações $a_{u'}$ e $a_{v'}$ nas direções dos eixos da elipse. Esta malha faz o plano amostral σ_{\max}^2 - aceitável.

Determinação do Tamanho da Amostra

1^o Passo.

Região		Plano de Amostragem		
Semivariograma		Unidade de área		Risco
γ_0	$a_{u'}$	$a_{v'}$	s	σ_{max}^2
1,50	0,06	0,14	10000	3,50

2^o Passo. Com $\gamma_0 = 1,5$, $a = a_{v'} = 0,14$ entra-se no algoritmo do quadro 3.8. O resultado obtido é $l_{v'} =$

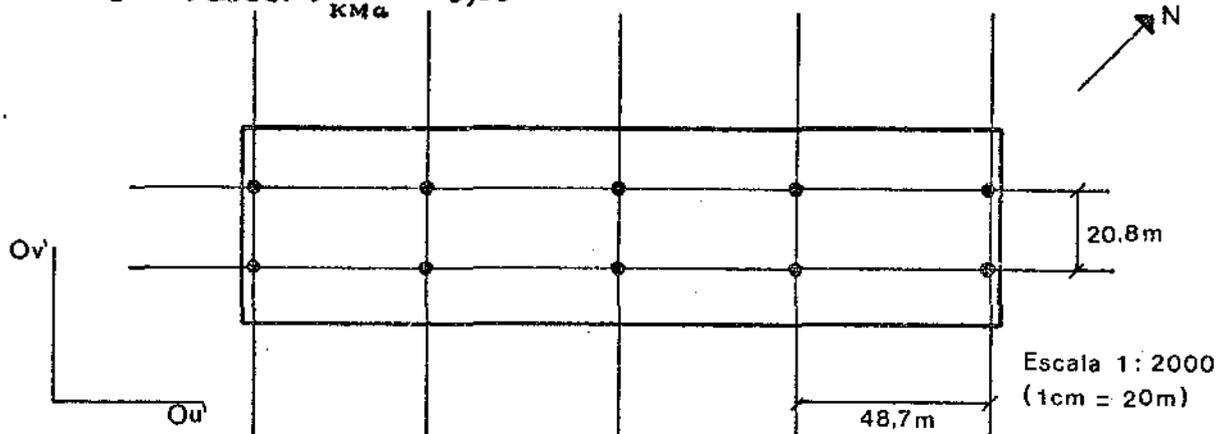
$$\frac{100}{\sqrt{23}} = 20,85.$$

3^o Passo. $f = \frac{0,14}{0,06} = 2,33$

$$l_{u'} = f \cdot l_{v'} = \frac{0,14}{0,06} \cdot \frac{100}{\sqrt{23}} = 48,7$$

4^o Passo. $n_a = \frac{10000}{(20,85)^2 \cdot 2,33} \cong 10$

5^o Passo. $\sigma_{km_a}^2 = 3,50$



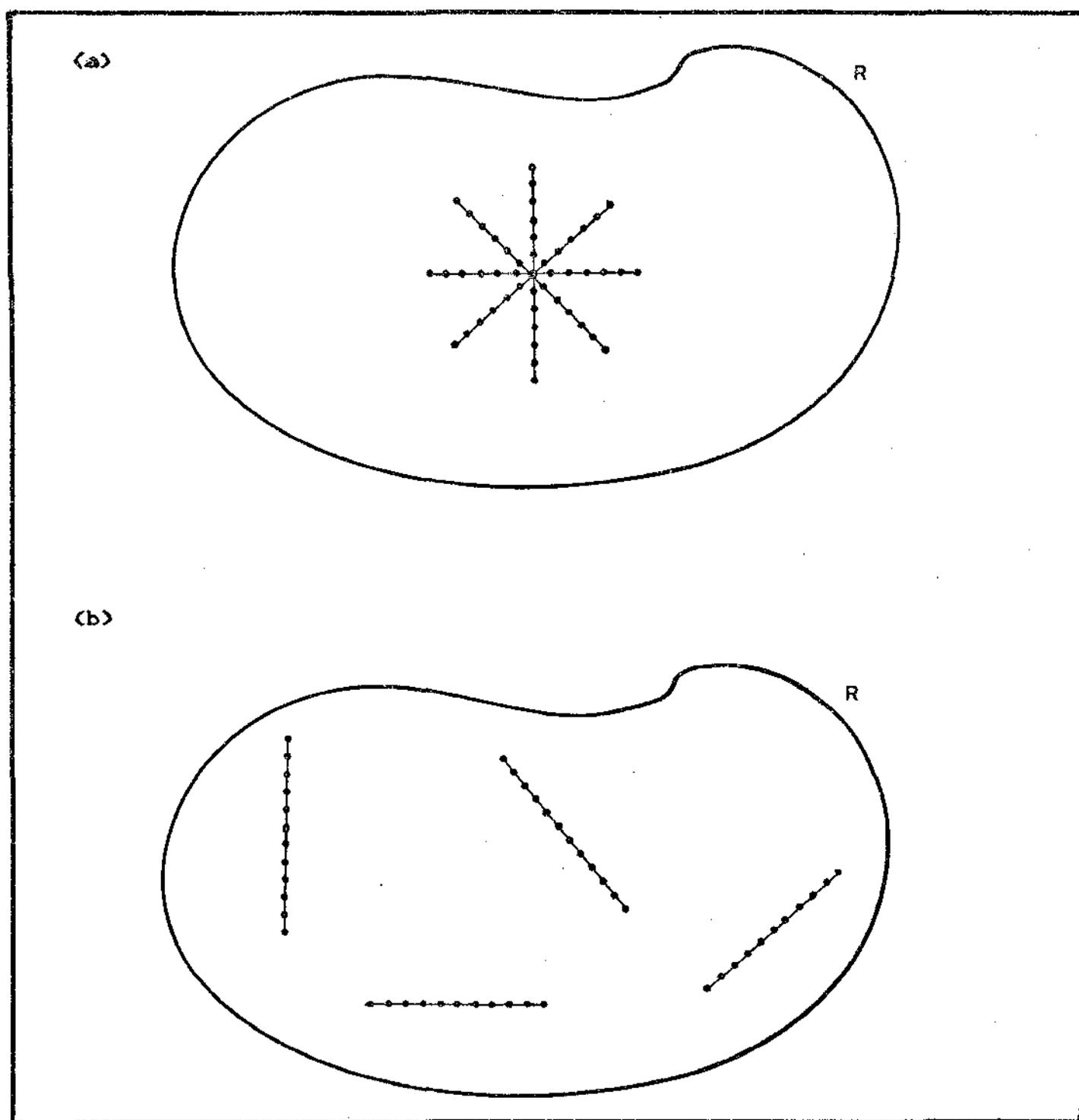
QUADRO 3.13. Exemplo de aplicação do algoritmo do quadro 3.12. As unidades dos dados são: área (m^2), efeito pepita e risco ($(g \text{ água}/g \text{ solo})^2$), inclinação ($(g \text{ água}/g \text{ solo})^2/m$) e espaçamento (m).

3.3. Pré-Amostragem para Semivariograma

As propostas para o planejamento da amostragem espacial apresentadas nesta dissertação pressupõem que um modelo para o semivariograma já foi determinado, possivelmente utilizando uma pré-amostragem. Nesta seção apresentar-se-á sugestões para conduzi-la.

Seja um processo $\{ Z(x): x \in R \subset \mathbb{R}^2 \}$ que satisfaz a hipótese intrínseca, com semivariograma desconhecido. Então uma realização desse processo pode ser amostrada para estimar o semivariograma segundo o plano abaixo:

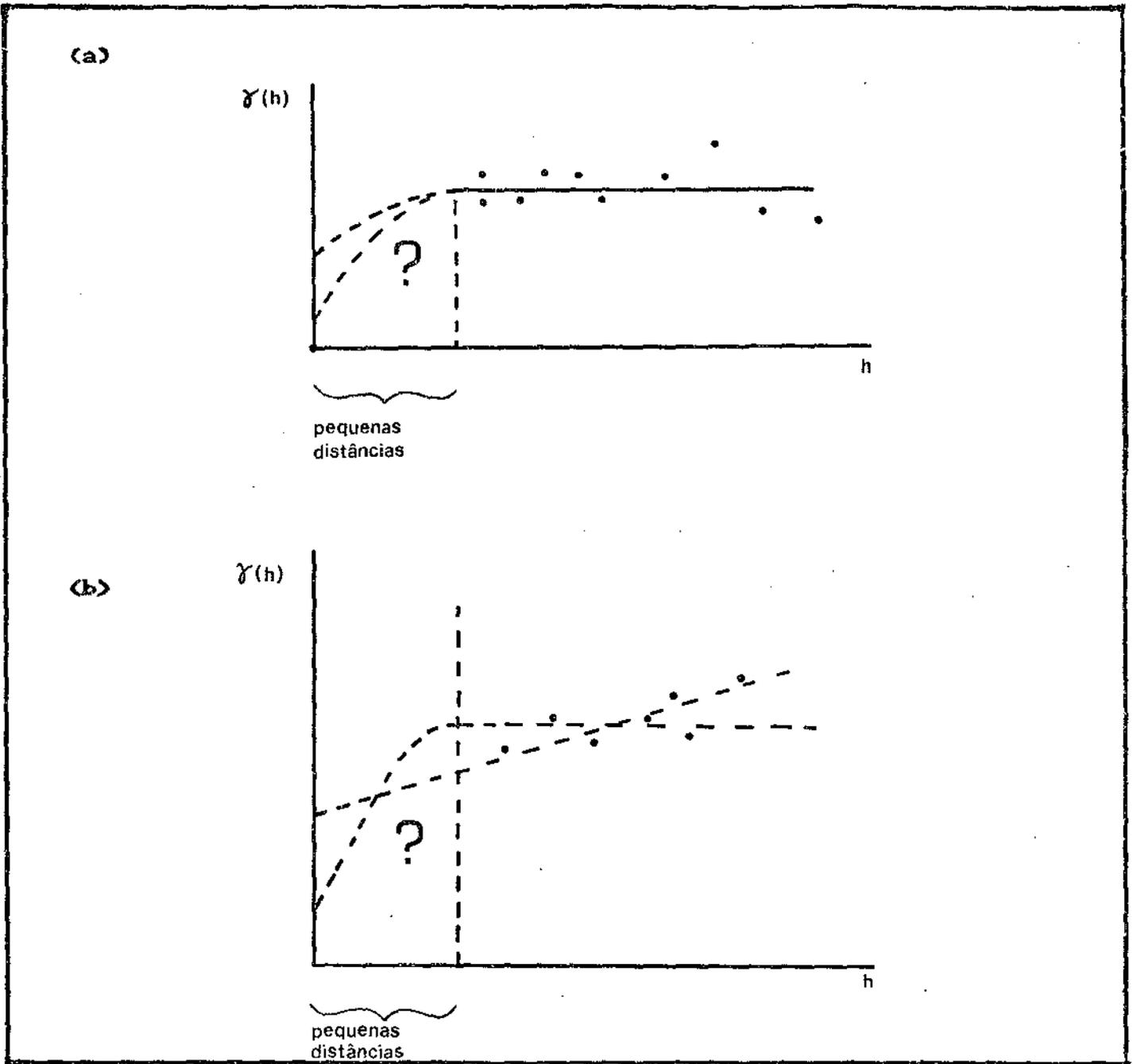
- (i) São realizadas transeções sobre R , procurando-se amostrar pontos sobre estas linhas separados por distâncias menores do que possivelmente serão as distâncias no plano de amostragem para as krigagens (Burgess, Webster e McBratney, 1981). Isto permitirá determinar com detalhe a estrutura de dependência espacial à pequenas distâncias.
- (ii) Estas linhas de (pré) amostragem deverão ter várias direções dentro de R para poder-se detectar a existência de possível anisotropia. Esta alocação poderá ser feita pelo menos de duas maneiras, veja quadro 3.14.
- (iii) O número de pontos amostrados poderá consumir até um terço do esforço total do levantamento (Young, 1976; citado em Burgess, Webster e McBratney,



QUADRO 3.14. Duas possíveis alocações de transeções para amostragem de uma região R : (a) centrado ("em cruz")
(b) não-centrado.

1981). Isto permitirá uma melhor estimação do semivariograma, principalmente à pequenas distâncias.

A necessidade de estimar bem o semivariograma à pequenas distâncias reside no fato de que é de fundamental importância conhecer a dependência espacial entre um ponto e sua vizinhança. Por exemplo, a qualidade da estimação do efeito pepita depende da qualidade da estimação do semivariograma à pequenas distâncias. Como pode ser visto nos gráficos do quadro 3.7, o tamanho da amostra necessário para fazer um plano σ_{\max}^2 -aceitável, depende sensivelmente do valor do efeito pepita. Por exemplo, no quadro 3.7, se um plano deve ser 2,00 (g água/g solo)²-aceitável, o tamanho da amostra muda de aproximadamente 15 para mais de 40, quando o efeito pepita muda de 0,00 para 1,5 (g água/g solo)². Além disso, e conclusivamente, o próprio modelo à ser adotado depende da qualidade dessa estimação, veja quadro 3.15. Com o objetivo de melhorar a (pré) estimação do semivariograma à pequenas distâncias, sugere-se que nas transeções propostas nos itens (i) e (ii) acima contenham segmentos onde os espaçamentos entre os pontos seja tão pequenos quanto as condições de medição o permitam. Naturalmente, a qualidade de todas as inferências futuras (krigagens pontuais em mapeamentos e krigagens de bloco na estimação de médias espaciais) e do próprio planejamento da amostragem dependerá da qualidade da estimação do semivariograma.



QUADRO 3.15. O modelo do semivariograma e em particular o efeito pepita dependem criticamente da informação da (pré)amostragem à pequenas distâncias. Exemplos: (a) semivariograma esférico com diferentes efeitos pepita (b) Dois modelos diferentes ajustados para a mesma nuvem de pontos.

3.4. Conclusão

Um plano de amostragem espacial para conhecer uma população $\{ z(x): x \in R \subset \mathbb{R}^2 \}$ realizada sob semivariograma não-decrescente terá um risco mínimo se uma malha triangular for usada. Apesar disso, uma malha quadrada é recomendada, porque seu risco não é muito maior do que a triangular (principalmente se o efeito pepita for grande) e seu custo de implantação é menor. Para o exemplo apresentado no quadro 3.9, o que se perde em usar uma malha quadrada em vez de uma triangular é no máximo da ordem aproximada de 6%, em termos do aumento no risco.

O estudo de uma variável espacial que apresenta um semivariograma anisotrópico geométrico, com modelo linear, pode ser feito levando-se em consideração os resultados para o caso isotrópico, modificando-se a malha de amostragem de quadrada para retangular. Esta alteração consiste em alongar (ou contrair) o lado do quadrado na direção de um dos eixos da elipse segundo a razão entre os comprimentos desses eixos.

O investimento numa pré-amostragem para a estimação do semivariograma pode ser compensador, visto que permitirá um planejamento calcado em bases mais seguras.

Finalmente, estudos que demonstrem rigorosamente as soluções apresentadas aos problemas A1, A2, B1 e B2, que permitam uma medição melhor da eficiência relativa entre as malhas triangular e quadrada (problema B3) e forneçam um critério objetivo para solucionar o problema C1, estão ainda sendo

necessários. Os problemas C2 e C3 podem ser resolvidos para outros tipos de semivariograma (por exemplo o esférico) e o problema das bordas da região R, contido no problema C3, merecem maior atenção. Um estudo para um semivariograma anisotrópico geométrico com o modelo esférico poderia ser também feito, bem como para outros tipos de anisotropia (zonal, por exemplo). Por último, sem dúvida um estudo mais objetivo estudo sobre a pré-amostragem para estimação do semivariograma seria bem recebido.

Capítulo 4. PLANO AMOSTRAL PARA MÉDIAS ESPACIAIS

No capítulo 3, propôs-se um plano amostral para a amostragem espacial para mapeamento de uma população $\{z(x):x \in R \subset \mathbb{R}^2\}$. Outra utilidade da amostra espacial é fornecer uma base para predição de uma média espacial. Em 2.2.2 define-se essa média e sua predição geoestatística, que é a krigagem de bloco. O objetivo deste capítulo é propôr um plano amostral para estimação da média espacial de uma subpopulação $\{z(t):t \in R(x) \subset R \subset \mathbb{R}^2\}$, realizada sob um semivariograma isotrópico com modelo linear.

Um plano amostral para estimação de uma média espacial é construído considerando-se duas quantidades:

- (i) O risco do plano para estimação da média espacial $s(x)$ da subpopulação $\{z(t):t \in R(x) \subset R \subset \mathbb{R}^2\}$, definido aqui como a variância de krigagem de bloco produzida pela amostra.
- (ii) O custo do plano para estimação da média espacial $s(x)$ da subpopulação $\{z(t):t \in R(x) \subset R \subset \mathbb{R}^2\}$, definido aqui como a quantidade de recursos financeiros necessária para executar o plano.

A principal necessidade que leva ao planejamento da amostragem espacial para estimação de uma média espacial é o

controle do risco: um risco máximo σ_{\max}^2 é estabelecido e a amostra é planejada para que o risco do plano não ultrapasse esse limite (Webster e Burgess, 1984). Um plano tal que seu risco seja menor ou igual ao máximo estabelecido σ_{\max}^2 é um plano σ_{\max}^2 - aceitável. Configuração e espaçamento num plano de amostragem para a estimação de uma média espacial serão determinados em função do risco do plano. Como no planejamento para mapeamento, o custo do plano influenciará decisivamente a escolha da configuração da amostra.

Sob semivariograma isotrópico, linear, uma série de problemas devem ser resolvidos até se chegar à uma proposta para o plano visando médias espaciais. Esta sequência é análoga à apresentada no capítulo 3 e está apresentada como segue. (A identificação dos problemas guarda correspondência com as questões levantadas no capítulo 1).

Problema A. Determinar a configuração da amostra para estimação de uma média espacial.

Considerações análogas aquelas desenvolvidas para concluir a superioridade da amostragem sistemática para mapeamentos levam à conclusão de que também para médias espaciais é melhor a amostragem sistemática. Como antes, essa superioridade traduz-se na qualidade de garantir σ_{\max}^2 - aceitabilidade numa situação onde o plano puder não ser σ_{\max}^2 - aceitável (numa situação onde qualquer amostra produz um plano σ_{\max}^2 - aceitável não há necessidade de planejar no sentido definido nesta

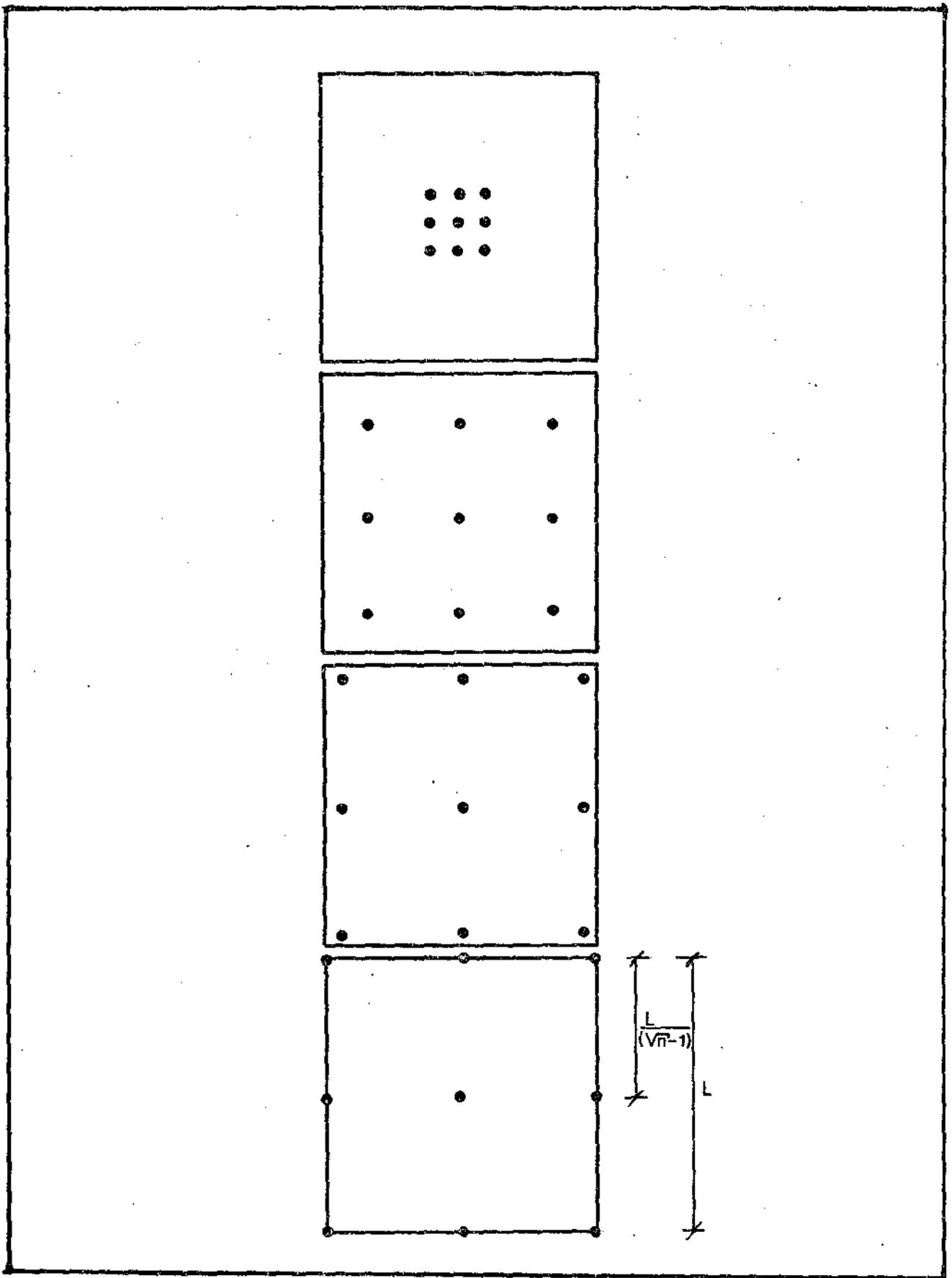
dissertação). Webster e Burgess (1984) aceitam implicitamente tal superioridade e neste capítulo ela também será aceita sem demonstração formal.

Problema B. Escolher uma configuração sistemática para estimação de uma média espacial.

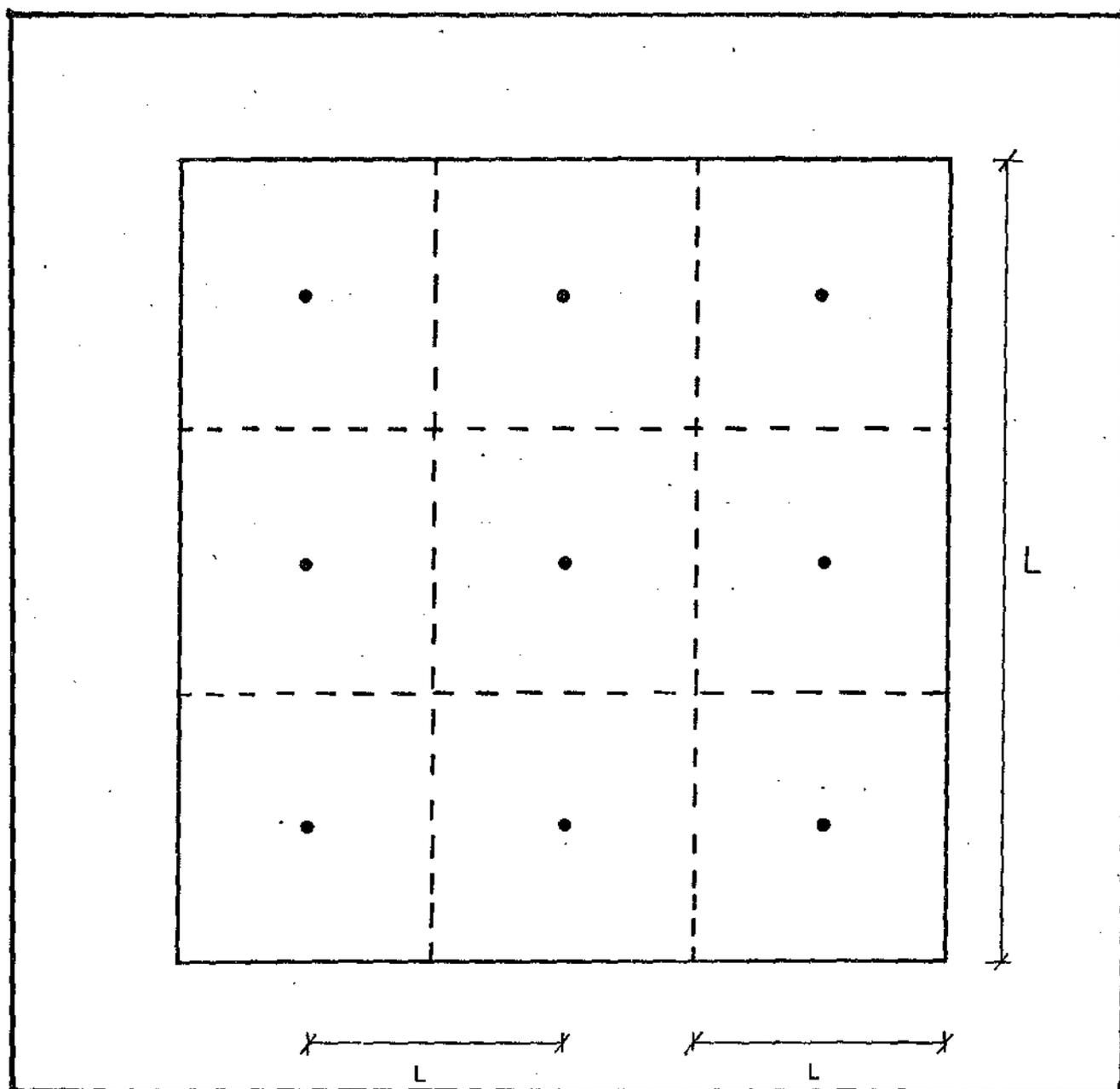
Se a subregião $R(x) \subset R \subset \mathbb{R}^2$ tem forma quadrada, Webster e Burgess (1984) afirmam que uma malha quadrada interna à subregião deve ser usada. Esta afirmação é adotada aqui como solução do problema B quando a subregião é quadrada. Necessita-se agora determinar o tamanho da amostra e o espaçamento entre pontos amostrados. Para que este objetivo seja atingido é necessário solucionar ainda dois problemas:

Problema C1. Minimizar a variância de krigagem de bloco em função do espaçamento, para um tamanho de amostra fixo.

Dada uma subregião quadrada, várias são as maneiras de alocar-lhe internamente uma malha quadrada com n nós: o espaçamento entre pontos amostrados varia de 0 à $L/(n^{1/2} - 1)$ (L é o comprimento do lado da subregião $R(x)$ quadrada), veja quadro 4.1. Webster e Burgess (1984) apresentam a solução à este problema (sem demonstrá-la), dando como $L/(n^{1/2})$ o espaçamento que minimiza a variância de krigagem de bloco, veja quadro 4.2. Neste capítulo este resultado é usado sem demonstração.



QUADRO 4.1. Possíveis malhas quadradas (como exemplo $n = 9$) internas à uma subregião $R(x)$ quadrada. Os espaçamentos entre os pontos amostrados variam de 0 à $L/(\sqrt{n} - 1)$.



QUADRO 4.2. Espaçamento entre pontos amostrados numa malha quadrada que minimiza a variância de krigagem de bloco para esta região $R(x)$ quadrada (um exemplo para $n = 9$). Os pontos amostrados estão no centro das tesselas quadradas que compõem $R(x)$.

Problema C2. Calcular a variância de krigagem de bloco.

Definido o espaçamento ótimo, resta a tarefa de calcular o risco do plano que prevê uma amostra de tamanho n . Neste capítulo, um algoritmo para esse cálculo é proposto.

Os problemas A até C2 são tratados na seção 4.1. Eles conduzem à uma proposta para a determinação do tamanho da amostra suficiente para tornar um plano σ_{\max}^2 -aceitável. Ela é apresentada como um algoritmo na seção 4.2, onde também um exemplo de aplicação é dado, num estudo com dados fictícios, modificados de dados de um estudo de solo. A seção 4.3 conclui este capítulo salientando resultados importantes e expondo alguns problemas ainda em aberto no planejamento da amostragem espacial para estimação de uma média espacial.

4.1. Variância de Krigagem de Bloco

Nesta dissertação os problemas A, B e C1 tem suas soluções aceitas conforme expostas na literatura. O problema C2 também é enfrentado na literatura (Webster e Burgess, 1984) porém sem uma sequência de instruções sobre o cálculo da variância de krigagem de bloco. Esta seção procura suprir esta falta apresentando um algoritmo adequado. O lema seguinte é apresentado como subsídio para a construção do algoritmo.

Lema 4.1 (Cálculo da variância de krigagem de bloco).
Seja um processo $\{Z(x); x \in R \subset \mathbb{R}^2\}$ que satisfaz a hipótese intrínseca, com semivariograma isotrópico, linear, com inclinação

a e efeito pepita γ_0 . Seja uma subpopulação $\{z(t): t \in R(x) \subset R \subset \mathbb{R}^2\}$, realização do processo acima dentro da subregião $R(x)$ quadrada, de lado L , contida em R . Se uma amostra de n pontos em malha quadrada é alocada dentro de $R(x)$ no espaçamento $L/n^{1/2}$, então a variância de krigagem de bloco é dada por $\sigma_{KB}^2(x) = \nu' \bar{\gamma} + \beta - \bar{\gamma}_Z(R(x))$ onde:

(i) ν e β são dados por

$$\begin{bmatrix} \nu \\ \beta \end{bmatrix} = \begin{bmatrix} \Gamma & | & 1 \\ \hline 1' & | & 0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \bar{\gamma} \\ 1 \end{bmatrix}$$

Γ é igual à $a.L.\Gamma_R/n^{1/2} + \gamma_0.(1.1' - I)$ onde a matriz Γ_R é uma matriz simétrica $n \times n$, com o elemento na posição (i,j) dado por $f(i,j) = [(u_i - u_j)^2 + (v_i - v_j)^2]^{1/2}$, $u_k = [(k-1) \text{ DIV } n^{1/2}] + 1$ e $v_k = [(k-1) \text{ MOD } n^{1/2}] + 1$.

(ii) $\bar{\gamma}$ é igual à $a.L.\bar{\gamma}_R/n^{1/2} + \gamma_0.1$ onde $\bar{\gamma}_R$ é um vetor $n \times 1$ com o i -ésimo elemento dado por $l_{1i}.l_{2i}.g(l_{1i}, l_{2i}) + l_{2i}.l_{3i}.g(l_{2i}, l_{3i}) + l_{3i}.l_{4i}.g(l_{3i}, l_{4i}) + l_{4i}.l_{1i}.g(l_{4i}, l_{1i})/n$, $l_{1i} = (2.u_i - 1)/2$, $l_{2i} = (2.v_i - 1)/2$, $l_{3i} = n^{1/2} - l_{1i}$ e $l_{4i} = n^{1/2} - l_{2i}$, $g(p,q) = \{2.(p^2+q^2)^{1/2} + (p^2/q). \ln [(q+(p^2+q^2)^{1/2})/p] + ((q^2/p). \ln [(p+(p^2+q^2)^{1/2})/q])\}/6$.

$$\begin{aligned} \text{(iii)} \quad \bar{\gamma}_Z(R(x)) &= |R(x)|^{-2} \cdot \int_{R(x)} \int_{R(x)} \gamma_Z(s-t) ds dt = \\ &= \gamma_0 + a.L \cdot \{(2+\sqrt{2}) + 15 \cdot \ln(1+\sqrt{2})\}/15. \end{aligned}$$

Demonstração. Veja apêndice.

O algoritmo do quadro 4.3 está baseado no lema 4.1. Ele conduz ao cálculo da variância de krigagem de bloco numa subregião $R(x)$ quadrada sob semivariograma isotrópico, linear. Um exemplo de sua aplicação prática se dá com dados fictícios, modificados de um estudo da variável espacial 'capacidade de retenção de água do solo na camada 0-20cm de profundidade sob pressão de 0,33 atm '(g água/g solo), sobre uma subregião $R(x)$ quadrada com lado $L = 100m$. Seu semivariograma apresenta um efeito pepita de $1,5 (g \text{ água/g solo})^2$ e uma inclinação de $0,14 (g \text{ água/g solo})^2/m$. A variância de krigagem de bloco foi calculada para vários tamanhos de amostra sob diferentes efeitos pepita e os resultados estão resumidos nos quadros 4.4 e 4.5.

Os resultados desta seção fundamentam a proposta para determinação do tamanho da amostra espacial para estimação de uma média espacial, tema da próxima seção.

4.2. Tamanho da Amostra

A configuração de uma amostra espacial que fornecerá uma base para a estimação da média espacial de uma variável que apresenta semivariograma isotrópico com modelo linear, distribuída sobre uma subregião quadrada de lado L está escolhida como uma malha quadrada interna à subregião, com o espaçamento entre os n pontos amostrados igual à $L/n^{1/2}$. Resta pois determinar n para que o plano seja σ_{\max}^2 - aceitável.

 CALCULO DA VARIANCIA DE KRIGAGEM DE BLOCO

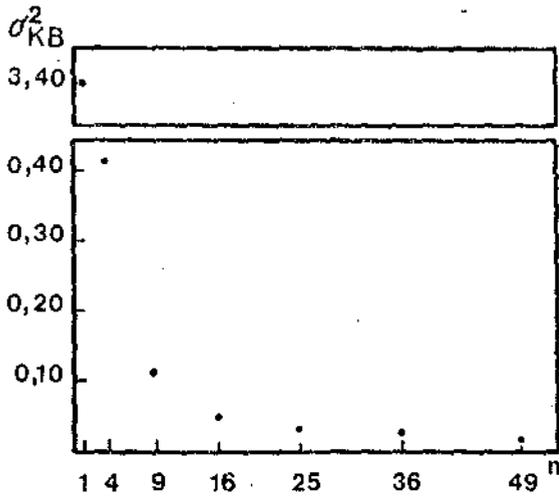
Passo	Procedimento
1	Especificar o valor do efeito pepita γ_0 , da inclinação a e o comprimento L do lado da subregião $R(x)$.
2	Determinar o tamanho n da amostra em malha quadrada à ser alocada em $R(x)$, $n = \{1^2, 2^2, 3^2, 4^2, 5^2, \dots\}$.
3	<p>Calcular as quantidades $u_i = [(i-1) \text{DIV } n^{1/2}] + 1$, $v_i = [(i-1) \text{MOD } n^{1/2}] + 1$, $l_{1i} = (2u_i - 1)/2$, $l_{2i} = (2v_i - 1)/2$, $l_{3i} = n^{1/2} - l_{1i}$, $l_{4i} = n^{1/2} - l_{2i}$, $g(l_{1i}, l_{2i})$, $g(l_{2i}, l_{3i})$, $g(l_{3i}, l_{4i})$, $g(l_{4i}, l_{1i})$ onde $g(p, q) = \{2 \cdot (p^2 + q^2)^{1/2} + ((p^2/q) \cdot \ln(q + (p^2 + q^2)^{1/2})/p) + ((q^2/p) \cdot \ln(p + (p^2 + q^2)^{1/2})/q)\} / 6$, para cada ponto x_i da amostra.</p>
4	<p>Determinar o vetor ν de pesos de krigagem, o número β, o vetor de semivariâncias médias $\bar{\gamma}$ e a quantidade $\bar{\gamma}_Z(R(x))$ através das fórmulas seguintes:</p> <p>$\bar{\gamma} = a \cdot (L/n^{1/2}) \cdot \gamma_R + \gamma_0 \cdot \mathbf{1}$ onde γ_R é o vetor $n \times 1$ cujo i-ésimo elemento é $[l_{1i} \cdot l_{2i} \cdot g(l_{1i}, l_{2i}) + l_{2i} \cdot l_{3i} \cdot g(l_{2i}, l_{3i}) + l_{3i} \cdot l_{4i} \cdot g(l_{3i}, l_{4i}) + l_{4i} \cdot l_{1i} \cdot g(l_{4i}, l_{1i})] / n$. ν e β são solução de</p> $\begin{bmatrix} -\nu \\ \beta \end{bmatrix} = \begin{bmatrix} -\Gamma & \mathbf{1} \\ \mathbf{1}' & 0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} -\bar{\gamma} \\ 1 \end{bmatrix}$ <p>onde a matriz $n \times n$ Γ é igual a $a \cdot (L/n^{1/2}) \cdot \Gamma_R + \gamma_0 \cdot (\mathbf{1}\mathbf{1}' - \mathbf{I})$, sendo Γ_R a matriz $n \times n$ cujo elemento (i, j) é dado por $f(i, j) = [(u_i - u_j)^2 + (v_i - v_j)^2]^{1/2}$. A quantidade $\bar{\gamma}_Z(R(x))$ é dado por $\gamma_0 + 0,5214054 \cdot a \cdot L$.</p>
5	<p>Calcular a variância de krigagem de bloco:</p> $\sigma_{KB}^2(x) = \nu' \bar{\gamma} + \beta - \bar{\gamma}_Z(R(x)).$

QUADRO 4.3. Algoritmo para cálculo da variância de krigagem de bloco em uma subregião quadrada, amostrada em malha quadrada.

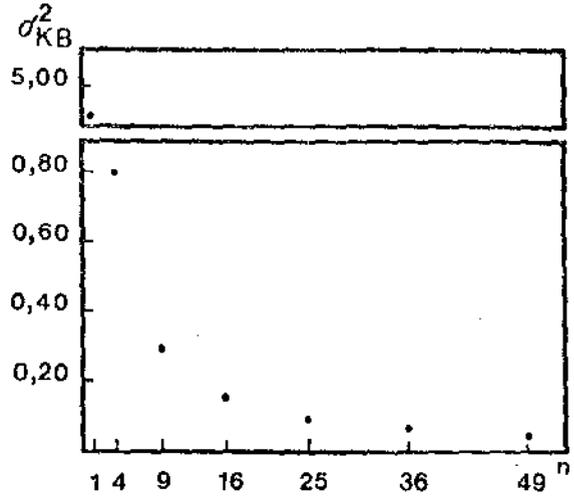
RISCO DO PLANO PARA ESTIMAÇÃO DE MEDIA ESPACIAL

γ_0	n	σ_{KB}^2	γ_0	n	σ_{KB}^2
0	1	3,41	15,0	1	18,41
	4	0,42		4	4,17
	9	0,12		9	1,79
	16	0,05		16	0,99
	25	0,03		25	0,63
	36	0,02		36	0,43
	49	0,01		49	0,32
	1,5	1		4,91	1 500 000,0
4		0,80	4	375 000,40	
9		0,29	9	166 666,70	
16		0,15	16	93 750,05	
25		0,09	25	60 000,02	
36		0,06	36	41 666,68	
49		0,04	49	30 612,25	
7,5		1	10,91		
	4	2,30	4		
	9	0,96	9		
	16	0,52	16		
	25	0,33	25		
	36	0,22	36		
	49	0,16	49		

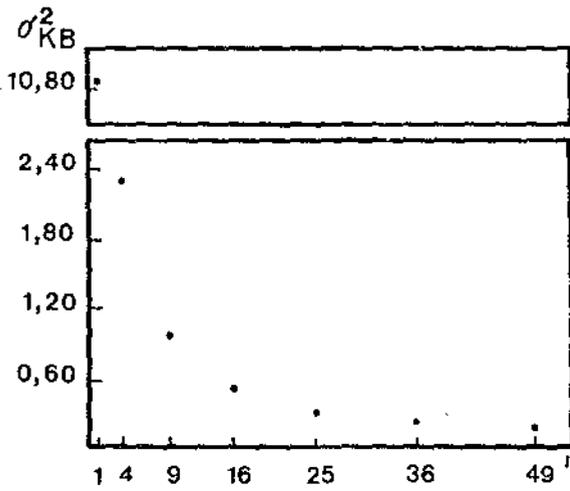
QUADRO 4.4. Risco do plano amostral para estimação da média espacial da variável "capacidade de retenção de água do solo na camada 0-20 cm sob pressão de 0,33 atm", numa subregião quadrada de lado 100 m. Neste exemplo (fictício), o semivariograma isotrópico é linear, com inclinação de 0,14 (g água/g solo)²/m.



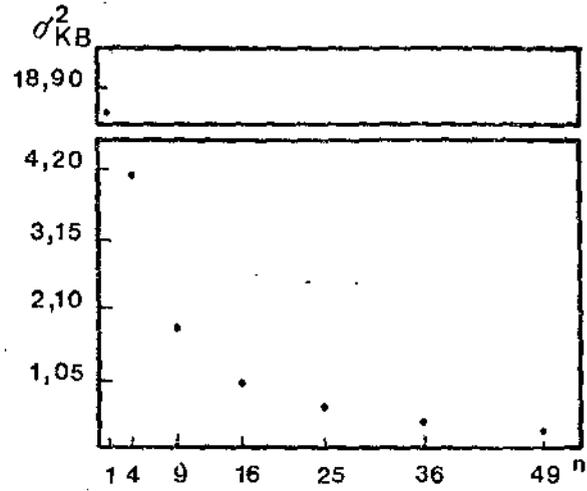
(a)



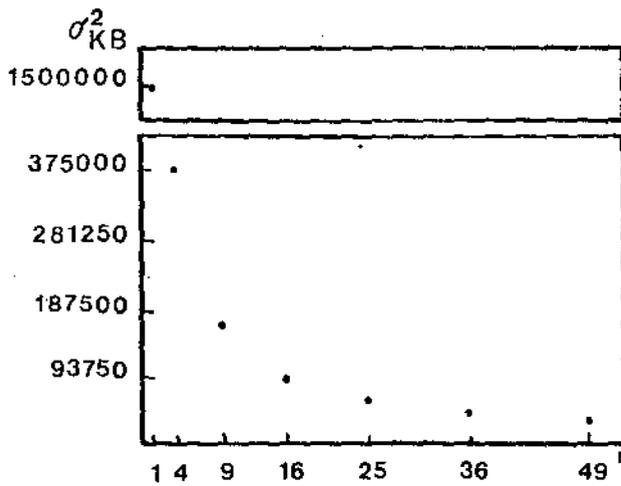
(b)



(c)



(d)



(e)

QUADRO 4.5. Risco de um plano para estimação de uma média espacial. Valores do efeito pepita: (a) 0,00 (b) 1,5 (c) 7,5 (d) 15,0 (e) 1 500 000,0

A proposta para a determinação de n é apresentada em modo algoritmo no quadro 4.6. Um exemplo de aplicação no planejamento de uma amostra espacial para estimação da média espacial da 'capacidade de retenção de água do solo na camada 0-20cm de profundidade à pressão de 0,33 atm' (g água/g solo) sobre uma subregião quadrada de 10000 m² de área, que apresenta um semivariograma isotrópico, com efeito pepita de 1,5 (g água/g solo)² e inclinação de 0,14 (g água/g solo)²/m, está apresentado no quadro 4.7. Uma amostra de 4 pontos espaçados de 50 m, dentro de $R(x)$, confere ao plano 4,00 (g água/g solo)²-aceitabilidade.

4.3. Conclusão

Um plano amostral para se estimar a média espacial da subpopulação $\{ z(t): t \in R(x) \subset R \subset \mathbb{R}^2 \}$ de uma variável que apresenta um semivariograma isotrópico, modelo linear, pode ser feito em malha quadrada se a subregião $R(x)$ é quadrada. Este estudo pode ainda ser generalizado em pelo menos três direções: (i) considerar subregiões com outros formatos, por exemplo, o retangular (ii) considerar outros modelos para o semivariograma, p.ex. o esférico (iii) considerar o planejamento sob anisotropia.

 DETERMINAÇÃO DO TAMANHO DA AMOSTRA

Passo	Procedimento
1	Especificar o valor do efeito pepita γ_0 , da inclinação a , do comprimento L do lado da subregião quadrada e o risco máximo tolerável σ_{max}^2 para este plano.
2	Usar o algoritmo proposto no quadro 4.3 para construir uma tabela como no quadro 4.4 e/ou gráfico como no quadro 4.5 para a variância de krigagem de bloco em função do tamanho n da amostra, n variando de 1 até o maior número inteiro quadrado possível, dentro dos recursos disponíveis.
3	Usar a tabela e/ou gráfico para escolher o valor de n que faz a variância de krigagem de bloco ser menor ou igual à σ_{max}^2 . O espaçamento entre os pontos amostrados é $L/n^{1/2}$.

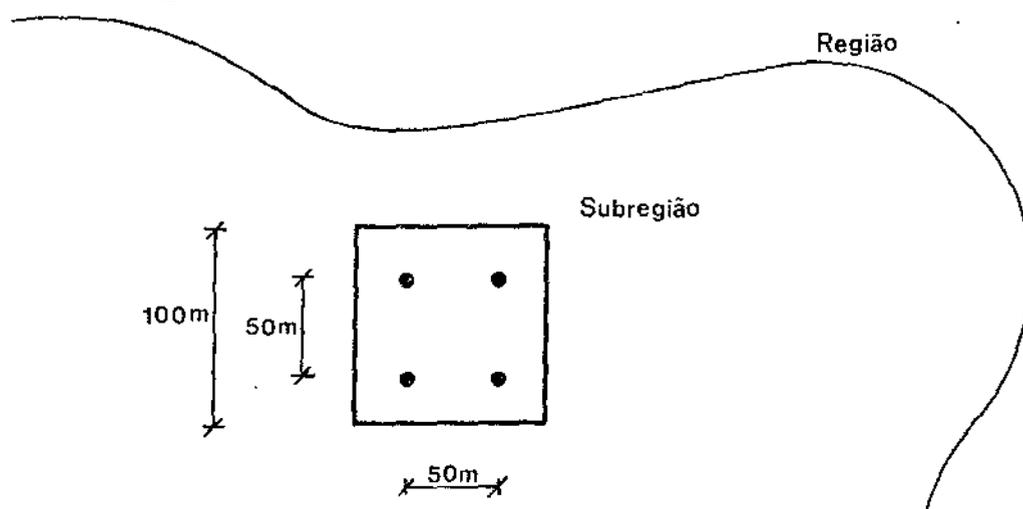
QUADRO 4.6. Algoritmo para determinação do tamanho da amostra (e do espaçamento entre pontos amostrados) para estimação da média espacial de uma variável espacial que apresenta um semivariograma isotrópico, modelo linear com efeito pepita γ_0 e inclinação a , distribuída sobre uma subregião quadrada com lado L . Esta amostra disposta em malha quadrada no interior da subregião confere σ_{max}^2 - aceitabilidade ao plano amostral.

1^o Passo.

Região		Plano de Amostragem	
Semivariograma	Subregião	Risco	
γ_0	a	σ_{max}^2	
1,50	0,14	100	4,00

2^o Passo. Veja quadros 4.4 e 4.5.

3^o Passo. $n = 4$ pontos amostrados na subregião, em malha quadrada. O espaçamento entre pontos amostrados será de $\frac{L}{\sqrt{n}} = \frac{100}{2} = 50$.



Escala 1:4000
(1cm = 40m)

QUADRO 4.7. Exemplo de aplicação do algoritmo do quadro 4.6. As unidades dos dados são: comprimento e espaçamento (m), efeito pepita e risco $((g \text{ água}/g \text{ solo})^2)$ e inclinação $((g \text{ água}/g \text{ solo})^2)/m$.

Capítulo 5. CONSIDERAÇÕES FINAIS

Nesta dissertação foi exibida a utilidade da teoria geoestatística no estudo da variação de uma variável espacial, quando a hipótese intrínseca é válida. Nesse contexto mostrou-se a conveniência de se usar uma malha quadrada na amostragem para fins de mapeamento quando o semivariograma é não-decrescente e isotrópico. Numa aplicação numérica, verificou-se a pequena diminuição da variância máxima de krigagem (no máximo apenas aproximadamente de 6%), ao se trocar uma malha quadrada por uma triangular. Esta vantagem não compensa a desvantagem na alocação de campo, pois é mais custoso demarcar em malha triangular do que em quadrada.

Conhecendo-se uma estimativa prévia para o semivariograma da variável espacial e a área da região sob estudo, pode-se chegar, através de algoritmo proposto no quadro 3.8, à determinação do tamanho da amostra e do espaçamento entre pontos amostrados, para que o plano seja σ_{\max}^2 -aceitável. No exemplo com aplicação numérica do quadro 3.9, o mapeamento da variável 'capacidade de retenção de água no solo, na camada 0-20 cm sob pressão de 0,33 atm' (g água/g solo) (semivariograma isotrópico com modelo linear, efeito pepita 1,50 (g água/g solo)² e inclinação 0,14 ((g água/g solo)²/m)) numa região de solo, poderia

ser feito com 14 pontos amostrados para cada 10000 m², em malha quadrada, espaçados de 26,7 m, sob um risco máximo de 4,00 (g água/g solo)².

Caso o semivariograma seja anisotrópico geométrico, com modelo linear, basta usar uma malha retangular cuja razão entre os lados do retângulo é a mesma que entre os comprimentos dos eixos da elipse de anisotropia. O quadro 3.12 apresenta um algoritmo para a determinação do tamanho da amostra e do espaçamento entre pontos amostrados, nessa situação. A aplicação numérica do quadro 3.13 mostra como 10 pontos amostrados em 10000 m² pode conferir 3,50 (g água/g solo)² - aceitabilidade à um plano, para mapear a 'capacidade de retenção de água no solo na camada 0-20 cm sob pressão de 0,33 atm' se seu semivariograma é anisotrópico geométrico, com efeito pepita igual à 1,50 (g água/g solo)², inclinação de 0,06 ((g água/g solo)²/m) na direção NE - SO e 0,14 ((g água/g solo)²/m) na direção ortogonal.

A estimação da média de uma variável espacial com semivariograma isotrópico do tipo linear, que se distribui sobre uma subregião quadrada, pode ser efetuada por uma malha também quadrada interna à subregião. O número de pontos amostrados e o espaçamento entre eles, calculados pelo algoritmo do quadro 4.6, conferirá ao plano σ_{max}^2 - aceitabilidade. Para que a média espacial da 'capacidade de retenção de água no solo na camada 0-20 cm sob pressão de 0,33 atm' (semivariograma isotrópico, do tipo linear, com efeito pepita 1,50 (g água/g solo)² e inclina-

ção $0,14 ((\text{g água/g solo})^2/\text{m})$ numa subregião quadrada com área 10000 m^2 , seja estimada com um risco máximo de $4,00 (\text{g água/g solo})^2$, basta que o plano prescreva uma amostra de 4 pontos espaçados de 50 m, dentro da subregião (veja quadro 4.7).

Para que os planos de amostragem espacial sejam aplicáveis em várias situações práticas diferentes, é necessário ainda explorar outros problemas. Por exemplo, estudar o planejamento sob diferentes tipos de anisotropia, planejar estimação de médias espaciais em subregiões com formas diferentes da quadrada, tornar mais segura e confiável a pré-estimação do semi-variograma.

Pode-se distinguir três possíveis obstáculos à utilização dos planos amostrais propostos neste trabalho:

- (i) Não se pode admitir a validade da hipótese intrínseca na variabilidade espacial da característica sob estudo.
- (ii) A insuficiência dos recursos computacionais para o planejamento (máquina e/ou programa).
- (iii) O custo do planejamento e execução. (Este custo está diretamente relacionado com o número de pontos que devem ser amostrados para se alcançar o controle desejado no risco).

Somente (i) é intransponível quando verificado. Com o avanço da informatização na pesquisa científica e na empresa, o obstáculo (ii) tende a desaparecer mais e mais. Nesta

dissertação tal desafio foi enfrentado utilizando um microcomputador compatível com o IBM-PC e softwares como o módulo IML do SAS e o LOTUS 1-2-3. Transposto (i) e (ii) resta o terceiro obstáculo: vencê-lo implica em encontrar - como é frequente em amostragem - um ponto de equilíbrio entre o risco do plano e seu custo. Quando o custo é limitante, o número total de pontos à serem amostrados na região é função do custo máximo, e o risco é então dependente do custo. Se o risco é limitante, então o tamanho da amostra é função do risco, e conseqüentemente também o custo.

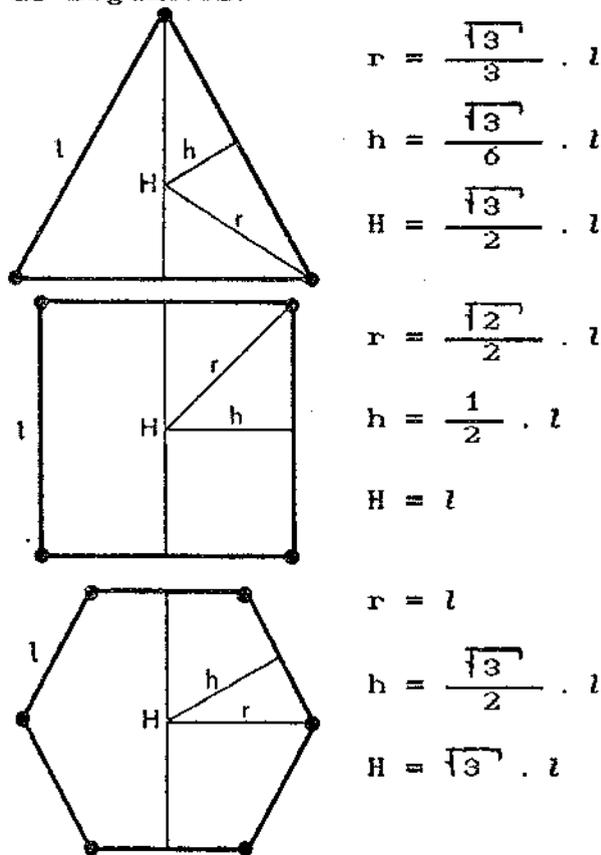
Transpostas estas barreiras, ter-se-á em mãos uma ferramenta para o estudo da variabilidade espacial. Sua utilidade deverá ser avaliada após consideração de diversos estudos substantivos e poderá influenciar decisivamente na continuação da exploração desse campo da estatística.

APÊNDICE

A.1. Demonstração do Lema 3.1.

Fixada uma densidade de amostragem δ , os lados dos polígonos regulares que formam malhas com densidades δ são l para triângulo equilátero, $(3^{1/2}/2)^{1/2} \cdot l$ para o quadrado e $(2/3)^{1/2} \cdot l$ para o hexágono regular, veja Yfantis, Flatman e Behar (1987).

Seja H a altura do polígono regular, h seu apótema e r o raio da circunferência que passa pelos seus vértices. Utilizando o teorema de Pitágoras, as relações entre essas três medidas para o triângulo equilátero, o quadrado e o hexágono regular são as seguintes:



$$r = \frac{\sqrt{3}}{3} \cdot l$$

$$h = \frac{\sqrt{3}}{6} \cdot l$$

$$H = \frac{\sqrt{3}}{2} \cdot l$$

$$r = \frac{\sqrt{2}}{2} \cdot l$$

$$h = \frac{1}{2} \cdot l$$

$$H = l$$

$$r = l$$

$$h = \frac{\sqrt{3}}{2} \cdot l$$

$$H = \sqrt{3} \cdot l$$

Posto isso, é imediato que os raios da circunferência que passam pelos vértices dos polígonos regulares que formam malhas com densidade δ são dados por r para o triângulo equilátero, $[3^{1/2}, (3/4)]^{1/2} \cdot r$ para o quadrado e $2^{1/2} \cdot r$ para o hexágono regular.

A.2. Argumento de Plausibilidade para o Lema 3.2.

Sejam os pontos amostrados dispostos em malha regular. Para x centro de um polígono, a variância (máxima) de krigagem é dada por

$$\sigma_K^2(x) = 2 \cdot \sum_i \lambda_i \cdot \gamma_Z(x_i - x) - \sum_i \sum_j \lambda_i \cdot \lambda_j \cdot \gamma_Z(x_i - x_j)$$

(veja 2.1.2). Como o semivariograma é não-decrescente, pelos resultados do lema 3.1. conclui-se que o primeiro termo é minimizado quando a malha é triangular enquanto que o segundo termo é maximizado nesta mesma malha. Como o primeiro termo é positivo e o segundo é negativo, a variância máxima de krigagem é minimizada numa malha triangular.

A.3. Demonstração do Lema 3.3.

Num semivariograma $\gamma(\cdot)$ isotrópico, as semivariâncias só dependem da norma $|h|$ do argumento h . As distâncias $|h|$ entre os pares de pontos amostrados e entre os pontos amostrados e o ponto à ser predito tanto em malha triangular como em quadrada são do tipo

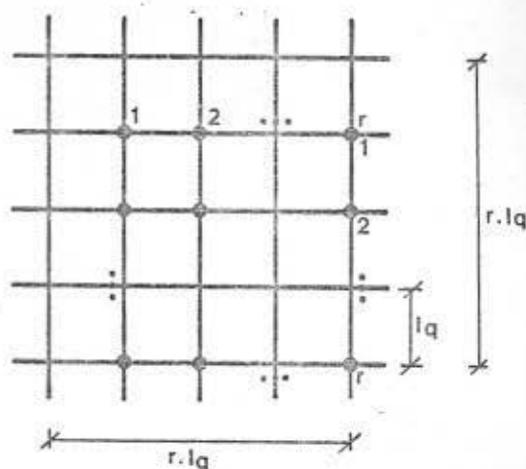
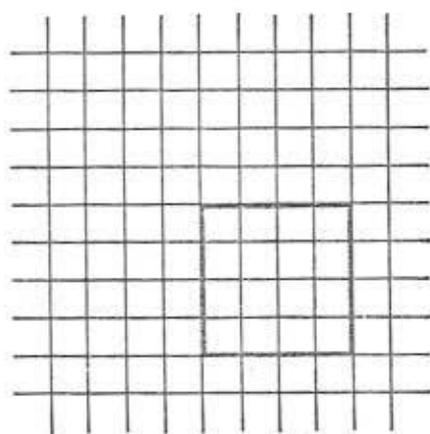
$$|h| = \sqrt{(c.l)^2 + (d.l)^2}$$

onde l é o lado do polígono (l_t ou l_q) e c e d são coeficientes que dependem dos pares de pontos considerados (por exemplo, na malha triangular do quadro 3.4(a), $c = 1/2$ e $d = 3^{1/2}/2$ para o par (x_1, x_4) pois $|x_1 - x_4| = [(l_t/2)^2 + (3^{1/2}.l_t/2)^2]^{1/2}$). Aliando os resultados apresentados em 2.1.2 para Γ , γ e λ , chegam-se aos resultados do lema 3.3.

A.4. Demonstração do Lema 3.4.

(i) Decorre da própria definição de δ .

(iii) Seja uma malha quadrada de lado l_q . Nesta malha, r^2 pontos amostrados ocupam uma área de $(r.l_q)^2$:



A densidade δ de amostragem é então $n^2 / (n^2 l_q^2) = 1/l_q^2$. Mas δ é também n/s , logo

$$\frac{n}{s} = \frac{1}{l_q^2} \Rightarrow l_q = \sqrt{s} \cdot \frac{1}{\sqrt{n}}$$

(i) Pelo lema 3.1, $l_q/l_t = (3^{1/2}/2)^{1/2}$. Logo,

$$\frac{n}{s} = \frac{1}{(3^{1/2}/2)l_t} \Rightarrow l_t = \sqrt{\frac{2\sqrt{3}}{3}} \cdot \sqrt{s} \cdot \frac{1}{\sqrt{n}}$$

A.5. Demonstração do Lema 3.5.

(i) O lema 3.4. mostrou que $l_t = \sqrt{\frac{2\sqrt{3}}{3}} \cdot \sqrt{s} \cdot \frac{1}{\sqrt{n}}$

O lema 3.3. apresentou σ_{KMTL}^2 como $2\lambda'_T \cdot \gamma_T - \lambda'_T \cdot \Gamma_T \cdot \lambda_T$ onde $\Gamma_T = \Gamma_{RT} \cdot a \cdot l_t + (1 - 1' - D) \cdot \gamma_0$ e $\gamma_T = \gamma_{RT} \cdot a \cdot l_t + 1\gamma_0$.

A conjugação dos dois resultados leva à

$$\Gamma_T = \Gamma_{RT} \cdot a \cdot (2 \cdot 3^{1/2}/3)^{1/2} \cdot s^{1/2}/n^{1/2} + (1 - 1' - D)\gamma_0$$

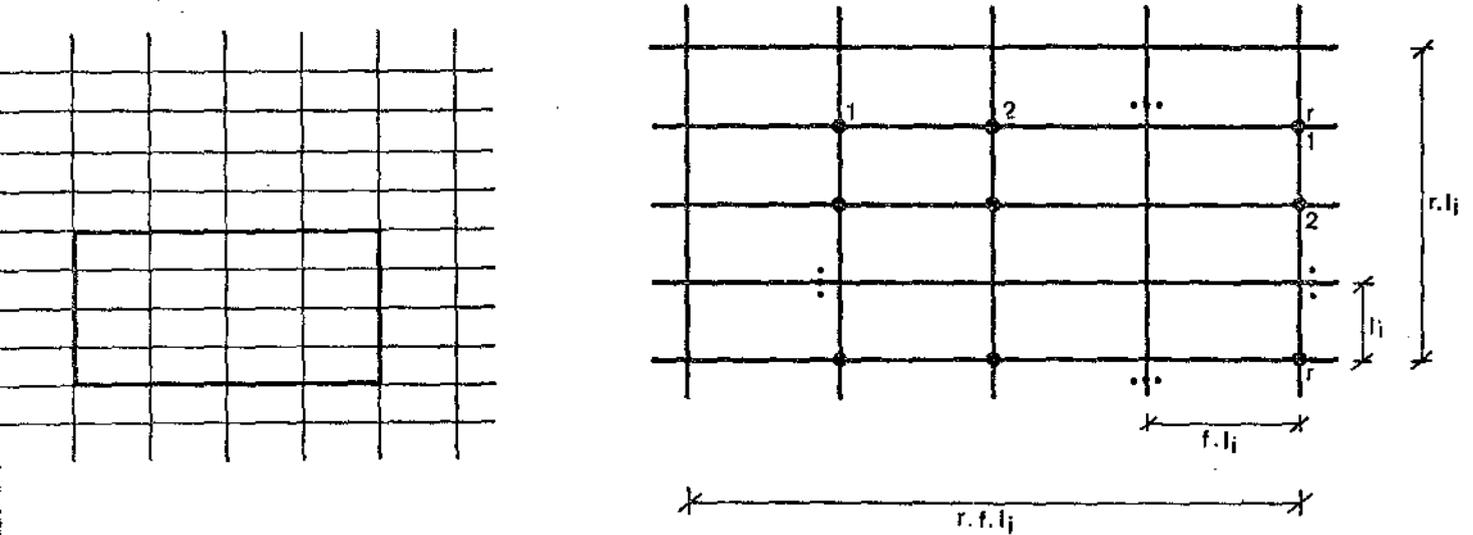
$$\text{e } \gamma_T = \gamma_{RT} \cdot a \cdot (2 \cdot 3^{1/2}/3)^{1/2} \cdot s^{1/2}/n^{1/2} + 1\gamma_0,$$

como se quer demonstrar.

(ii) É demonstrado de maneira análoga.

A.6. Demonstração do Lema 3.5.

- (i) Seja uma malha retangular de lados l_i e $f.l_i$.
 Nesta malha, r^2 pontos amostrados ocupam uma área de $(r^2 \cdot f.l_i^2)$.



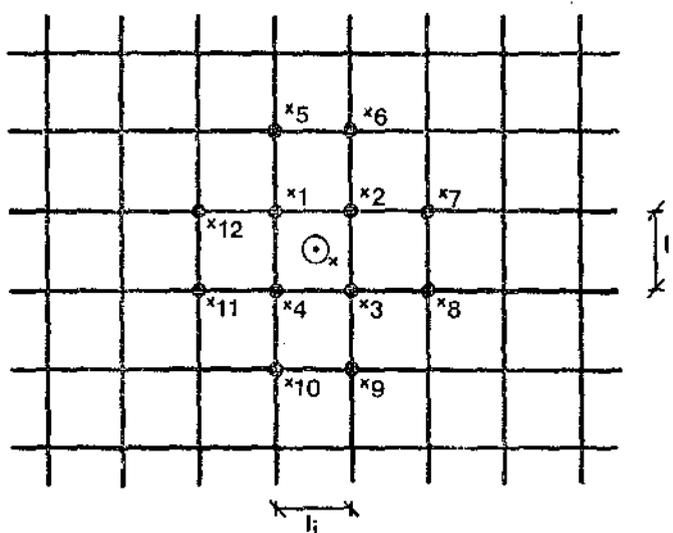
A densidade δ_a de amostragem é então $r^2 / (r^2 \cdot f.l_i^2) = 1 / (f.l_i^2)$. Como $\delta_i = 1/l_i^2$ tem-se que $\delta_a = \delta_i / f$.

- (ii) Como $\delta_a = n_a / s$ e $\delta_i = n_i / s$ tem-se que $n_a = n_i / f$.

- (iii) Seja uma malha quadrada amostrando uma população gerada por um processo com semivariograma isotrópico. A variância máxima de krigagem é função das distâncias dos pontos amostrados entre si e entre os pontos amostrados e o centro do quadrado. Sua fórmula é:

$$\sigma_{KM_i}^2 = 2 \cdot \sum_i \lambda_i \cdot \gamma_w(|x_i - x|) - \sum_i \sum_j \lambda_i \lambda_j \gamma_w(|x_i - x_j|).$$

Todas as distâncias entre os pontos $(x, x_1, x_2, \dots, x_{12})$ serão do tipo $[(c.l_i)^2 + (d.l_i)^2]^{1/2}$, onde c e d dependem do par de pontos considerados:



Sendo assim, todas as semivariâncias serão do tipo

$$\gamma_w = \sqrt{(c.l_i)^2 + (d.l_i)^2} =$$

$$= \begin{cases} 0 & , \sqrt{(c.l_i)^2 + (d.l_i)^2} = 0 \\ \gamma_0 + a \cdot \sqrt{(c.l_i)^2 + (d.l_i)^2} & , \sqrt{(c.l_i)^2 + (d.l_i)^2} \neq 0 \end{cases}$$

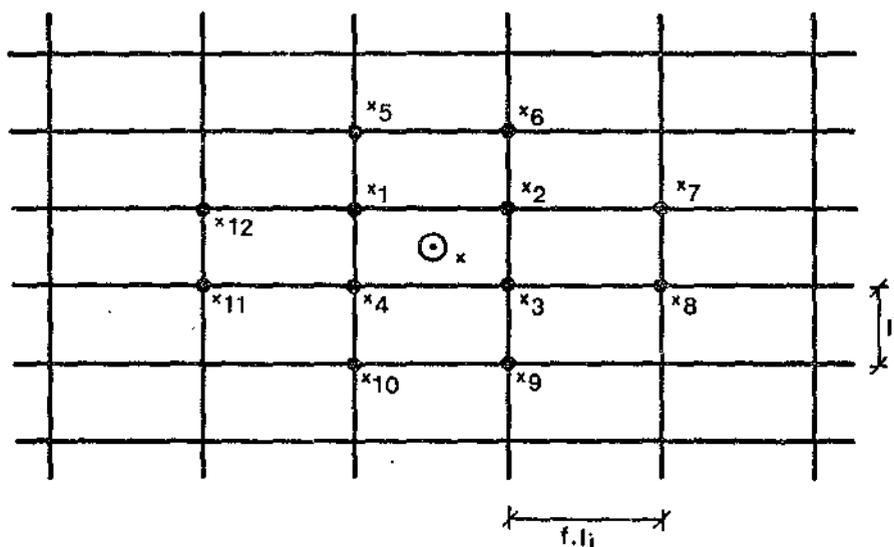
onde a inclinação 'a' independente da direção .

Seja uma malha retangular amostrando sob um semivariograma anisotrópico geométrico, linear. A variância máxima de krigagem é função das distâncias dos pontos amostrados entre si e entre os pontos amostrados e o centro do retângulo e também é função da direção da reta que passa pelo par de pontos

considerado, através do semivariograma:

$$\sigma_{KM\alpha}^2 = 2 \cdot \sum_i \lambda_i \cdot \gamma_Z(x - x_i) - \sum_i \sum_j \lambda_i \cdot \lambda_j \cdot \gamma_Z(x_i - x_j)$$

Semelhantemente à malha quadrada (caso isotrópico), todas as distâncias entre os pontos $\{x, x_1, x_2, \dots, x_{12}\}$ são do tipo $[(c.f.l_i)^2 + (d.l_i)^2]^{1/2}$, onde c e d são os mesmos coeficientes do correspondente par de pontos na malha quadrada:

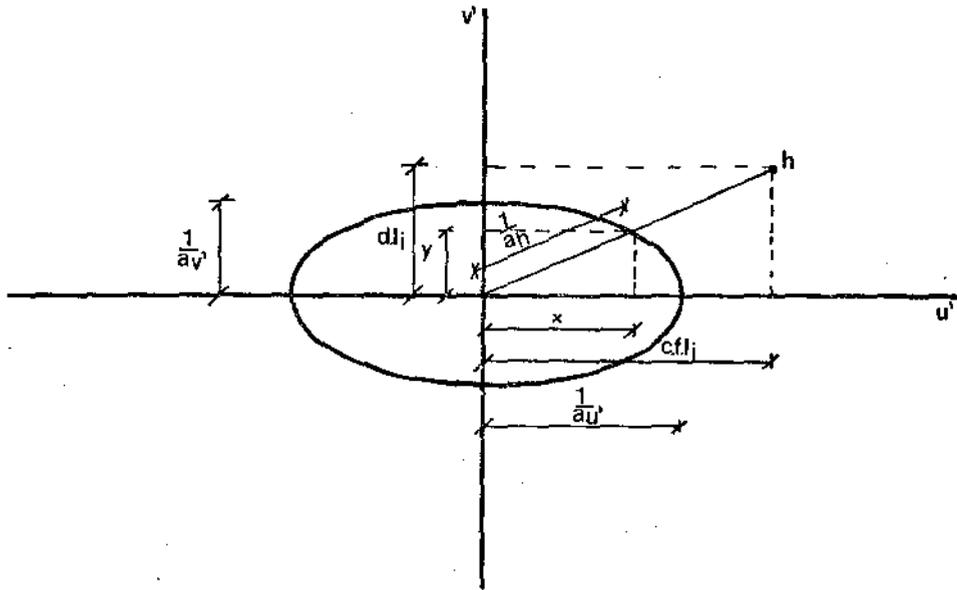


Assim, todas as semivariâncias são do tipo

$$\gamma_Z(h) = \begin{cases} 0 & , |h| = 0 \\ \gamma_0 + a_h \cdot \sqrt{(c.f.l_i)^2 + (d.l_i)^2} & , |h| \neq 0 \end{cases}$$

onde $|h| = \sqrt{(c.f.l_i)^2 + (d.l_i)^2}$. A inclinação a_h depende da direção do vetor h . Apesar dessa dependência que a_h tem da direção de h , $a_h \cdot [(c.f.l_i)^2 + (d.l_i)^2]^{1/2}$ resulta ser igual à $a_v \cdot [(c.l_i)^2 + (d.l_i)^2]^{1/2}$. A demonstração desse fato depende de um resultado intermediário, que é provado abaixo:

Seja a elipse que descreve a anisotropia geométrica para o semivariograma linear:



Nessa elipse, por semelhança de triângulos, conclui-se que:

$$\frac{\frac{x}{1}}{\frac{1}{a_h}} = \frac{c.f.l_i}{(\sqrt{c^2 f^2 + d^2}) l_i} = \frac{c.f.}{\sqrt{c^2 f^2 + d^2}}$$

$$\frac{\frac{y}{1}}{\frac{1}{a_h}} = \frac{d.l_i}{(\sqrt{c^2 f^2 + d^2}) l_i} = \frac{d}{\sqrt{c^2 f^2 + d^2}}$$

Pela equação da elipse:

$$\left[\frac{x}{\frac{1}{a_{u'}}} \right]^2 + \left[\frac{y}{\frac{1}{a_{v'}}} \right]^2 = 1 \Rightarrow \left[\frac{1}{\frac{a_h}{\sqrt{c^2 f^2 + d^2}}} \right]^2 + \left[\frac{1}{\frac{a_h}{\sqrt{c^2 f^2 + d^2}}} \right]^2 = 1 \Rightarrow$$

$$\Rightarrow \frac{a_{u'}^2 \cdot c^2 \cdot f^2}{c^2 \cdot f^2 + d^2} + \frac{a_{v'}^2 \cdot d^2}{c^2 \cdot f^2 + d^2} = a_h^2 \Rightarrow a_h = \sqrt{\frac{a_{u'}^2 \cdot c^2 \cdot f^2 + a_{v'}^2 \cdot d^2}{c^2 \cdot f^2 + d^2}}$$

Este resultado é aplicado em $a_h \cdot [(c \cdot f \cdot l_i)^2 + (d \cdot l_i)^2]^{1/2}$ da seguinte maneira:

$$a_h \cdot \sqrt{(c \cdot f \cdot l_i)^2 + (d \cdot l_i)^2} = a_h \cdot \sqrt{(c^2 \cdot f^2 \cdot d^2) \cdot l_i^2} =$$

$$= \frac{(a_{u'}^2 \cdot c^2 \cdot f^2 + a_{v'}^2 \cdot d^2) \cdot (c^2 \cdot f^2 + d^2) \cdot l_i^2}{c^2 \cdot f^2 + d^2} =$$

$$= \sqrt{(a_{u'}^2 \cdot c^2 \cdot f^2 + a_{v'}^2 \cdot d^2) l_i^2} = \sqrt{(a_{v'}^2 \cdot c^2 \cdot f^2 + a_{u'}^2 \cdot d^2) l_i^2} =$$

$$= a_{v'} \cdot \sqrt{(c \cdot l_i)^2 + (d \cdot l_i)^2} \text{ desde que } f = a_{v'} / a_{u'}$$

Este resultado permite então escrever para o caso anisotrópico:

$$\gamma_z(h) = \begin{cases} 0 & , |h| = 0 \\ \gamma_0 + a_{v'} \cdot \sqrt{(c \cdot l_i)^2 + (d \cdot l_i)^2} & , |h| \neq 0 \end{cases}$$

onde $|h| = \sqrt{(c.l_i)^2 + (d.l_i)^2}$. Ora, então diante disso conclui-se que

$$\gamma_z(h) = \gamma_w \left(\sqrt{(c.l_i)^2 + (d.l_i)^2} \right)$$

para h tal que $|h| = \sqrt{(c.l_i)^2 + (d.l_i)^2}$, isto é, a semivariância entre pares de pontos que são correspondentes nas malhas quadrada (caso isotrópico) e retangular (caso anisotrópico) são iguais. Como os pesos de krigagem só dependem das semivariâncias, os pesos na malha retangular sob um semivariograma anisotrópico serão os mesmos que os da malha quadrada, sob isotropia. Como a variância máxima de krigagem só depende dos pesos de krigagem e das semivariâncias, conclui-se que a variância máxima na malha retangular é igual à variância máxima na malha quadrada:

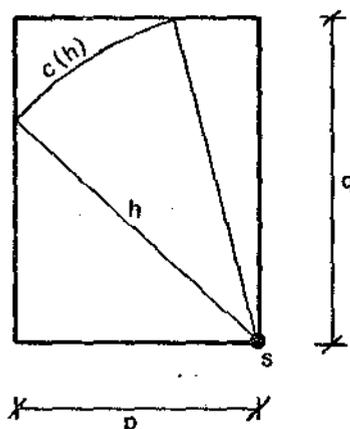
$$\sigma_{KM\alpha}^2 = \sigma_{KM\beta}^2$$

A.7. Lema auxiliar à demonstração do Lema 4.1. (Semivariograma médio entre uma região e um ponto). Seja um processo $\{Z(x): x \in R \subset \mathbb{R}^2\}$ que satisfaz a hipótese intrínseca, com semivariograma isotrópico, linear, com efeito pepita γ_0 e inclinação a . Se uma subregião r retangular, de lados com comprimento p e q é definida dentro da região R , então a média das semivariâncias entre os pontos de r e um de seus vértices s é dado por $\bar{\gamma}(s,r) = \gamma_0 + a.g(p,q)$ onde $g(p,q) = \{2.(p^2+q^2)^{1/2}\} + \{(p^2/q).\ln [q + (p^2 +$

$$q^2)^{1/2}/p)] + ((q^2/p) \ln [(p+(p^2+q^2)^{1/2})/q])/6.$$

Demonstração:

Por ser uma média sobre em contínuo de valores, $\bar{\gamma}(s,r)$ pode ser escrita como $|r|^{-1} \int_r \gamma(s,t) dt$, que pode ser escrita como $(pq)^{-1} \int_0^{(p^2+q^2)^{1/2}} c(h) \cdot \gamma(h) dh$, veja figura abaixo:



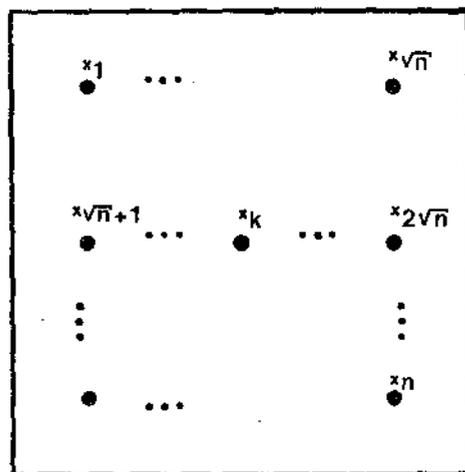
O comprimento de arco $c(h)$ depende do valor de h da seguinte forma: se $h \leq p$ então $c(h) = \pi h/2$; para $p < h \leq q$ $c(h) = h[\pi/2 - \arccos(p/h)]$ e se $q < h \leq (p^2+q^2)^{1/2}$ então $c(h) = h[\pi/2 - \arccos(p/h) - \arccos(q/h)]$. Portanto, $\bar{\gamma}(s,r) = (p \cdot q)^{-1} \left(\int_0^p (\pi h/2) \cdot \gamma(h) dh + \int_p^q h[\pi/2 - \arccos(p/h)] \cdot \gamma(h) dh + \int_q^{(p^2+q^2)^{1/2}} h[\pi/2 - \arccos(p/h) - \arccos(q/h)] \cdot \gamma(h) dh \right)$ que se particulariza para semivariograma linear como $\bar{\gamma}(s,r) = \gamma_0 + a_g(p,q)$.

A.8. Demonstração do Lema 4.1.

(i) De 2.1.2 tem-se que a matriz simétrica $n \times n$ Γ tem seu elemento (i,j) igual à $\gamma_z(x_i - x_j)$. Sendo o semivariograma isotrópico, ele é função apenas da norma de $x_i - x_j$. Para o modelo linear sua expressão fica

$$\gamma_z(x_i - x_j) = \begin{cases} 0 & , |x_i - x_j| = 0 \\ \gamma_0 + a \cdot |x_i - x_j| & , |x_i - x_j| \neq 0 \end{cases}$$

Para um conjunto de n pontos (x_1, \dots, x_n) arranjados em malha quadrada interna à uma subregião quadrada de lado L , qualquer distância $|x_i - x_j|$ é dado por $\{(u_i - u_j) \cdot L/n^{1/2}\}^2 + \{(v_i - v_j) \cdot L/n^{1/2}\}^2 = (L/n^{1/2}) \cdot [(u_i - u_j)^2 + (v_i - v_j)^2]^{1/2} = (L/n^{1/2}) \cdot f(i, j)$, onde u_k e v_k dados por $u_k = \lfloor (k - 1) \text{DIV } n^{1/2} \rfloor + 1$ e $v_k = \lfloor (k - 1) \text{MOD } n^{1/2} \rfloor + 1$ são respectivamente a fila horizontal e a fila vertical do ponto x_k , veja figura abaixo:



Logo $\Gamma = \gamma_0 \cdot (11' - D) + a \cdot L \cdot \Gamma_R / n^{1/2}$.

(ii) $\bar{\gamma}$ é o vetor $n \times 1$ que tem seu i -ésimo elemento igual à $\bar{\gamma}(x_i, R(x))$, veja 2.1.2. Neste caso especial onde $R(x)$ é quadrada tem-se que $\bar{\gamma}(x_i, R(x)) =$

$$\begin{aligned}
 &= |R(x)|^{-1} \cdot \int_{R(x)} \gamma_Z(x_i - s) ds = \\
 &= |R(x)|^{-1} \cdot \left[\sum_{k=1}^4 \int_{r_k} \gamma_Z(x_i - s) ds \right] = \\
 &= |R(x)|^{-1} \cdot \left[\sum_{k=1}^4 |r_k| \cdot |r_k|^{-1} \cdot \int_{r_k} \gamma_Z(x_i - s) ds \right] = \\
 &= |R(x)|^{-1} \cdot \left[\sum_{k=1}^4 |r_k| \cdot \bar{\gamma}(x_i, r_k) \right]. \text{ Pelo lema au-} \\
 &\text{xiliar anunciado antes desta demonstração con-} \\
 &\text{clui-se que } \bar{\gamma}(x_i, R(x)) = \\
 &= \gamma_0 + a(L/n^{1/2}) \cdot (l_{1i} \cdot l_{2i} \cdot g(l_{1i}, l_{2i}) + \\
 &l_{2i} \cdot l_{3i} \cdot g(l_{2i}, l_{3i}) + l_{3i} \cdot l_{4i} \cdot g(l_{3i}, l_{4i}) \\
 &+ l_{4i} \cdot l_{1i} \cdot g(l_{4i}, l_{1i})) / n.
 \end{aligned}$$

(iii) Este resultado está apresentado sem prova em Webster e Burgess (1984). Sua demonstração pode ser feita analogamente à do lema auxiliar anunciado antes dessa demonstração.

REFERÊNCIAS BIBLIOGRÁFICAS

- BURGESS, T.M. & WEBSTER, R. (1980a). Optimal interpolation and isarithmic mapping of soil properties I. The semivariogram and punctual kriging. *Journal of Soil Science* 31, 315-331.
- BURGESS, T.M. & WEBSTER, R. (1980b). Optimal interpolation and isarithmic mapping of soil properties II. Block kriging. *Journal of Soil Science* 31, 333-341.
- BURGESS, T.M.; WEBSTER, R. & McBRATNEY, A.B. (1981). Optimal interpolation and isarithmic mapping of soil properties. IV Sampling strategy. *Journal of Soil Science* 32, 643-659.
- CABANNES, A.F. (1979). Estimation of random fields from network observations. *Technical Report 26*. Stanford: Department of Statistics, Stanford University.
- CARVALHO, T.M. de (1991). Comunicação pessoal sobre levantamentos amostrais. Dissertação de Mestrado em Irrigação e Drenagem, ESAL.
- CLARK, I. (1979). *Practical Geostatistics*. London: Applied Science Publishers.
- COCHRAN, W.G. (1977). *Sampling Techniques*. New York: John Wiley and Sons.
- GRAMER, H & LEADBETTER, M.R. (1967). *Stationary and Related Stochastic Processes*. John Wiley and Sons.
- GRESSIE, N. (1988). Spatial prediction and ordinary kriging. *Mathematical Geology* 20(4), 405-421.
- GRESSIE, N. (1989). Geostatistics. *The American Statistician*. 43(4), 197-202.
- GRESSIE, N. (1990) The origins of kriging. *Mathematical Geology*, 22(3), 239-252.

- DAVID, M. (1978). Sampling and estimation problems for three dimensional spatial stationary and non stationary stochastic processes as encountered in the mineral industry. *Journal of Statistical Planning and Inference* 2, 211-244.
- DE GRUIJTER, J.J. & TER BRAAK, C.J.F. (1990). Model-free estimation from spatial samples: a reappraisal of classical sampling theory. *Mathematical Geology*, 22(4), 407-415.
- GAJEM, Y.M.; WARRICK, A.W. & MYERS, D.E. (1981). Spatial dependence of physical properties of a typic torrifluent soil. *Soil Science Society of America Journal* 45, 709-715.
- ISAAKS, E.H. & SRIVASTAVA, R.M. (1989). *Applied Geostatistics*. Oxford: University Press.
- JOURNEL, A.G. (1988). *Geostatistics For The Environmental Sciences*. Las Vegas: Report of United States Environmental Protection Agency.
- JOURNEL, A.G. & HUIJBREGTS, C.J. (1978). *Mining Geostatistics*. London: Academic Press.
- KLUTE, A. (1986). Water retention: laboratory methods, p. 635-661 em: KLUTE, A. (ed.) *Methods of Soil Analysis, Part 1*, 2nd edition. Madison: American Society of Agronomy & Soil Science Society of America.
- MATERN, B. (1986). *Spatial Variation*, 2nd edition. Lecture Notes in Statistics, 36. Berlin: Springer-Verlag.
- McBRATNEY, A.B.; WEBSTER, R. & BURGESS, T.M. (1981). The design of optimal sampling schemes for local estimation and mapping of regionalized variables-I. Theory and method. *Computers and Geosciences* 7(4), 331-334.
- McLEAN, E.O. (1982). Soil pH and lime requirement, p. 199-223 em: PAGE, A.L. (ed.) *Methods of Soil Analysis, Part 2*, 2nd edition. Madison: American Society of Agronomy & Soil Science Society of America.

- OLIVER, M. (1989). Geostatistics in soil science: how did it happen? *De Geostatisticis* 7, 7-9.
- PETERSEN, R.G. & CALVIN, L.D. (1986). Sampling, p. 33-51 em: KLUTE, A. (ed.) *Methods of Soil Analysis, Part 1*, 2nd edition. Madison: American Society of Agronomy & Soil Science Society of America.
- RIPLEY, B.D. (1981). *Spatial Statistics*. New York: John Wiley and Sons.
- SWITZER, P. (1979). Statistical considerations in network design. *Water Resources Research* 15(6), 1712-1716.
- VAN GUNDY, S.D. (1982). Nematodes, p. 1121-1129 em: PAGE, A.L. (ed.) *Methods of Soil Analysis, Part 2*, 2nd edition. Madison: American Society of Agronomy & Soil Science Society of America.
- VIEIRA, S.R.; NIELSEN, D.R. & BIGGAR, J.W. (1981). Spatial variability of field-measured infiltration rate. *Soil Science Society of America Journal* 45, 1040-1048.
- VIEIRA, S.R.; HATFIELD, J.L.; NIELSEN, D.R. & BIGGAR, J.W. (1983). Geostatistical Theory and Application to Variability of Some Agronomical Properties. *Hilgardia* 51(3), 1-75.
- WEBSTER, R. & BURGESS, T.M. (1980). Optimal interpolation and isarithmic mapping of soil properties III. Changing drift and universal kriging. *Journal of Soil Science* 31, 505-524.
- WEBSTER, R. & BURGESS, T.M. (1984). Sampling and bulking strategies for estimating soil properties in small regions. *Journal of Soil Science*, 35, 127-140.
- WILDING, L.P. & DRESS, L.R. (1983). Spatial variability and pedology, p. 83-116 em: WILDING, L.P.; SMECK, N.E. & HALL, G.F. (ed.) *Pedogenesis and Soil Taxonomy 1. Concepts and Interactions*. Developments in Soil Science 11A. Amsterdam: Elsevier Science Publishers.

YFANTIS, E.A.; FLATMAN, G.T. & BEHAR, J.V. (1987). Efficiency of kriging estimation for square, triangular, and hexagonal grids. *Mathematical Geology*, 19(3), 183-205.

ZIMMERMAN, D.L. & ZIMMERMAN, M.B. (1991). A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors. *Technometrics* 33(1), 77-91.

Composicao: AROCOM INFORMATICA

(095) 821.3122 - Lavras - MG