

## ANÁLISE ESPACIAL DE SUPERFÍCIES

*Eduardo Celso Gerbi Camargo*

*Suzana Druck Fucks*

*Gilberto Câmara*

### 3.1 INTRODUÇÃO

Neste capítulo, apresentamos técnicas para tratamento e análise de dados de superfícies. De uma forma geral, estes dados estão disponíveis na forma de amostras pontuais, e para utilizá-los de forma efetiva em um ambiente de Geoprocessamento, necessitamos de um procedimento de interpolação, para gerar uma representação na forma de grade regular, como ilustrado na Figura 3-1.

As amostras são valores representativos do fenômeno estudado, usualmente obtidas a partir de levantamento de campo, e que apresentam consistência de metodologia e unidade. Conforme explicado no capítulo 1, essas amostras podem representar tanto variáveis naturais (como teor de argila no solo) como socioeconômicas (como taxa de homicídios).

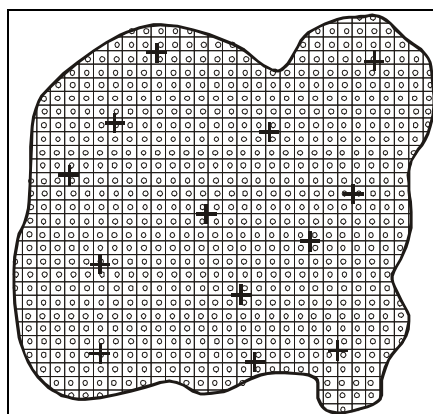


Figura 3-1 – Ilustração do processo de interpolação: amostras (cruzes) e aproximação da superfície por uma grade regular (círculos).

Para gerar superfícies que aproximem o fenômeno estudado de forma realista, é necessário modelar sua variabilidade espacial. Os modelos que objetivam gerar superfícies a partir de procedimentos de interpolação, de forma geral, representam a variável em estudo como uma combinação da variabilidade em larga e pequena

escala. Esse enfoque, entretanto, não é único. Assim, pode-se tomar três grandes abordagens: Para tanto, pode-se tomar três grandes abordagens:

- *Modelos determinísticos de efeitos locais*: cada ponto da superfície é estimado apenas a partir da interpolação das amostras mais próximas, utilizando funções como *inverso do quadrado da distância*. A suposição implícita é que predominam os efeitos puramente locais. Neste caso, não é feita qualquer hipótese estatística sobre a variabilidade espacial. Estes interpoladores serão apresentados na seção 3.2 deste capítulo.
- *Modelos determinísticos de efeitos globais*: a suposição implícita nesta classe de interpoladores é que, para a caracterização do fenômeno em estudo, predomina a variação em larga escala, e que a variabilidade local não é relevante. Este é caso do interpoladores por superfícies de tendência, apresentados na seção 3.3 deste capítulo.
- *Modelos estatísticos de efeitos locais e globais (krigagem)*: cada ponto da superfície é estimada apenas a partir da interpolação das amostras mais próximas, utilizando um estimador estatístico. Esses procedimentos requerem que a variabilidade local e global sejam modelada através de modelos apresentados como

$$Z(\mathbf{x}) = \sum_{j=1}^p \beta_j f_j + \varepsilon(\mathbf{x})$$

Nesse caso  $E\{Z(\mathbf{x})\} = \sum_{j=1}^p \beta_j f_j$  aonde  $\beta_j$  é um conjunto de parâmetros desconhecidos e  $f_j$  um conjunto de funções básicas, em geral, polinomiais. Esses estimadores apresentam propriedades de não ser tendenciosos e de procurar minimizar os erros inferenciais. Eles podem ser estimados através de procedimentos como a krigagem universal e as funções intrínsecas de ordem  $k$  não abordadas nesse capítulo.

Neste capítulo, iremos dar ênfase ao uso de técnicas de krigagem ordinária, ou seja a um caso particular desse modelo global em que  $p=1$  e  $k=0$ , aonde  $k$  representa a ordem da função  $f_j$ , e  $\beta_1$  igual a média local. A ênfase nesse procedimento é devido às suas propriedades, sua grande importância na modelagem de fenômenos naturais e também porque esse capítulo objetiva procedimentos que priorizam a interpolação espacial (predição). A modelagem de tendências ou variação em larga escala se faz necessária quando a etiologia de um fenômeno deve ser estudada e aonde a estimação da tendência é importante na compreensão do fenômeno. As técnicas da krigagem são discutidas a partir da seção 3.4. Para a comparação entre os interpoladores, foram utilizados dados da EMBRAPA – Solos,

obtidos na Fazenda Canchim, em São Carlos - SP. Trata-se de amostragem de 85 observações georreferenciadas coletadas no horizonte Bw (camada do solo com profundidade média de 1m), conforme ilustra a Figura 3-2. Dentre as variáveis disponíveis, selecionou-se para estudo o teor de argila, cujas estatísticas básicas amostrais são apresentadas na Tabela 3.1.

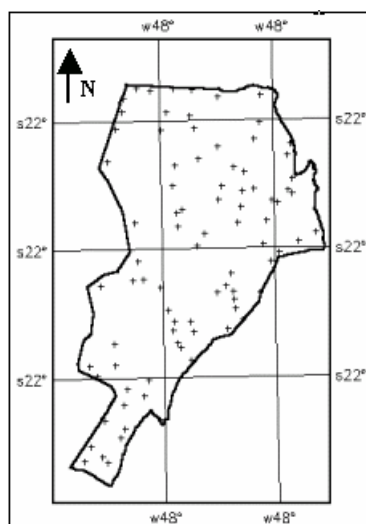
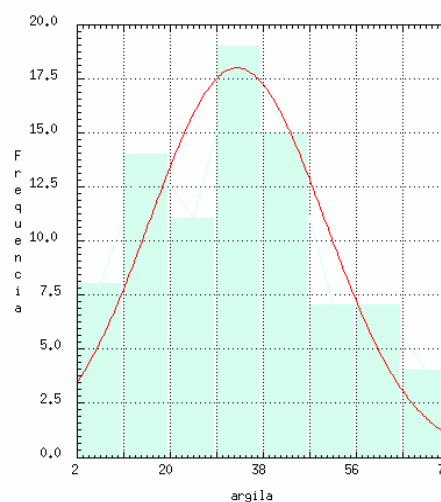


Figura 3-2- Disposição das amostras de teor de argila da Fazenda Canchim (EMBRAPA).

Tabela 3-1 - ESTATÍSTICAS DA AMOSTRA.

Número de observações	85
Média	33,035
Variância	288,034
Desvio Padrão	16,972
Coefficiente de variação	0,514
Coefficiente de assimetria	0,214
Coefficiente de curtose	2,344
Quartil Inferior	10
Mediana	33
Quartil superior	43



O histograma das amostras mostra que a distribuição do teor de argila é levemente alongada à direita. Neste caso, a distribuição é dita ser positivamente assimétrica, com coeficiente de assimetria de 0,214. Quanto ao grau de achatamento, o coeficiente de curtose (2,344) indica que a distribuição é

ligeiramente platicúrtica. Dentre outros valores apresentados na Tabela 3-1, nota-se que a média e a mediana, medidas que procuram caracterizar o centro da mesma distribuição de frequências, possuem valores próximos (33,035 e 33,0), respectivamente. Assim sendo, a distribuição da variável em estudo, pode ser considerada aproximadamente simétrica.

### 3.2 MODELOS DETERMINÍSTICOS LOCAIS

Uma alternativa simples para gerar uma superfície bidimensional a partir de amostras pontuais é ajustar uma função bidimensional sobre os amostras considerados, compondo uma superfície cujo valor será proporcional à local intensidade de amostras. A formulação geral para este tipo de interpolação é:

$$\hat{z}_i = \frac{\sum_{j=1}^n w_{ij} z_j}{\sum_{j=1}^n w_{ij}}, \quad (3.1)$$

onde:  $z_i$  é o valor de cota de um ponto  $i$  qualquer da grade,  $z_j$  é a cota de uma amostra  $j$  vizinha do ponto  $i$  da grade e  $w_{ij}$  é um fator de ponderação. A Figura 3-3 ilustra o procedimento de estimação.

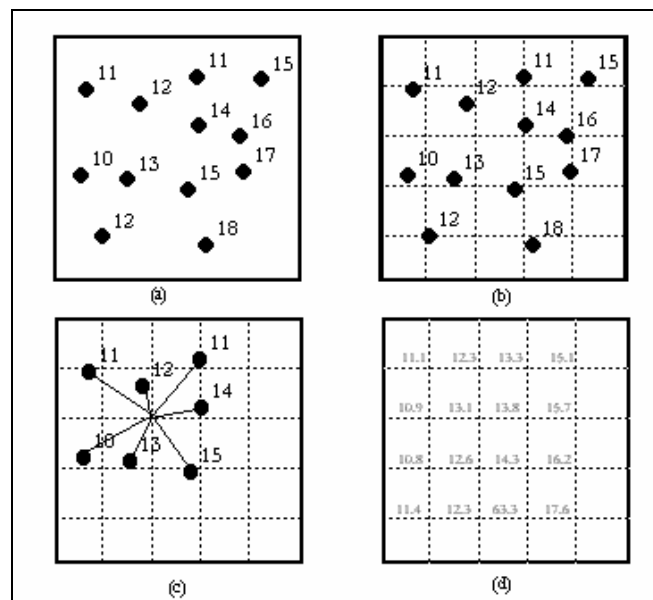


Figura 3-3 Ilustração do processo de interpolação por estimador local: (a) configuração original de amostras; (b) grade regular superposta às amostras; (c) interpolação de um valor a partir dos vizinhos; (d) grade regular resultante

Variações desse esquema básico são os interpoladores: (a) por vizinho mais próximo; (b) por média simples; (c) por média ponderada; Nos três primeiros casos, considera-se uma região em torno do ponto a ser interpolado como contendo os pontos que influenciam na interpolação. A *interpolação por vizinho mais próximo* é definida pela escolha de apenas uma amostra vizinha para cada ponto da grade. Este interpolador deve ser usado quando se deseja manter os valores de cotas das amostras na grade, sem gerar valores intermediários. A *interpolação por média simples* considera o valor de cota  $z$  do elemento da grade igual a média aritmética dos valores de cota das amostras vizinhas. Neste caso considera-se que o fator de ponderação  $w_{ij}$  é igual a  $1/n$  para qualquer amostra considerada. Na *interpolação por média ponderada* o valor de cota de cada elemento da grade é definido pela média ponderada dos valores de cota das amostras vizinhas. A ponderação mais usada na prática é o inverso da distância euclidiana do ponto da grade à amostra considerada ou seja:

$$w_{ij} = 1/d_{ij}^k, \quad (3.2)$$

onde:  $k$  é o expoente da distância, geralmente igual a 1 ou 2 e;  $d_{ij}$  é o valor de distância da amostra  $j$  ao ponto  $i$  da grade, expresso por:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (3.3)$$

Uma comparação visual entre os resultados desses interpoladores é mostrada na Figura 3-4 para os dados do teor de argila da Fazenda Canchim. Os mapas ilustram os defeitos típicos dessas funções simples: as funções de vizinho mais próximo e média simples tendem a produzir superfícies com variações abruptas; no caso do inverso do quadrado da distância, os máximos locais tendem a ser muito acentuados, formando “picos” artificiais.

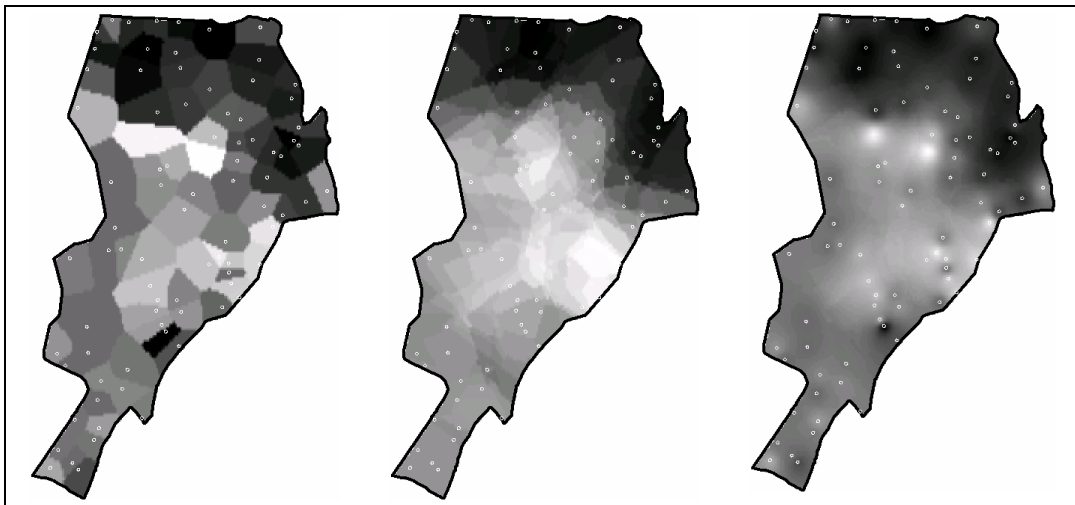


Figura 3-4 - Comparação entre interpoladores de média móvel, para o mesmo conjunto de amostras. À direita, inverso do quadrado da distância; no centro, média simples; à esquerda, vizinho mais próximo. Regiões mais claras representam alto valores e vice-versa.

Um refinamento desses estimadores é o uso de uma função de ponderação mais complexa que a *média simples* ou o *inverso do quadrado da distância*. Esta classe de estimadores é descrita na literatura como *kernel estimators*, ou *estimadores de densidade não-paramétricos*. Estes estimadores generalizam a idéia de média móvel local, ao supor que a densidade do fenômeno varia localmente de forma suave, sem “picos” nem “descontinuidades”. Seu objetivo é produzir superfícies mais suaves, que se espera mais representativas de fenômenos naturais e socioeconômicos. Estes estimadores são do mesmo tipo que os discutidos no capítulo 2 para o caso de eventos pontuais, agora generalizados para o caso de amostras.

Um *kernel estimator* é um estimador cujos parâmetros básicos são: (a) um raio de influência que define a vizinhança do ponto a ser interpolado; (b) uma função de estimação com propriedades “convenientemente” de suavização do fenômeno. Para toda posição  $z_i$  cujo valor queremos estimar, o estimador de intensidade será computado a partir dos valores das amostras  $\{z_1, \dots, z_n\}$  contidos num raio de tamanho  $\tau$ , e da distância euclidiana  $d_{ij}$  entre a  $i$ -ésima posição e a  $j$ -ésima amostra (como expresso na equação 3.3), a partir de funções do tipo

$$\hat{z}_i = \frac{\sum_{j=1}^n k\left(\frac{d_{ij}}{\tau}\right) z_j}{\sum_{j=1}^n k\left(\frac{d_{ij}}{\tau}\right)}, \quad d_{ij} \leq \tau \quad (3.4)$$

Esta fórmula é uma generalização da equação 3.1, na qual o cômputo dos pesos  $w_{ij}$  foi substituído por uma função generalizada dependente da distância. Exemplos destas funções incluem o *kernel* gaussiano

$$k(x, y, \tau) = \frac{1}{2\pi\tau} \exp\left(-\frac{d_{ij}^2}{2\tau^2}\right), \quad (3.5)$$

ou o *kernel* de quarta ordem

$$k(x, y, \tau) = \frac{3}{\pi\tau^2} \left(1 - \frac{d_{ij}^2}{\tau^2}\right)^2 \quad (3.6)$$

Para ilustrar esta classe de estimadores, foram geradas duas superfícies a partir das mesmas amostras usadas para produzir os mapas da Figura 3-4. A partir de um *kernel* de quarta ordem (equação 3.6), foram gerados dois mapas mostrados na Figura 3-5, com raios de busca de 500 e 1500 metros. A comparação entre os mapas mostra a grande importância de uma seleção apropriada do raio de busca no uso de *kernel estimators*. No primeiro mapa predominam os efeitos locais, pelo uso de um raio de busca reduzido; o segundo mapa evidencia melhor a distribuição do fenômeno, pelo uso de um raio mais apropriado aos dados.

Em resumo, os *kernel estimators* são uma alternativa viável a métodos mais sofisticados de interpolação, pois não requerem a parametrização da estrutura de correlação espacial (como no caso da geoestatística). As superfícies interpoladas são suaves e aproximam muitos fenômenos naturais e socioeconômicos. As desvantagens destes estimadores são a forte dependência no raio de busca e a excessiva suavização da superfície, que pode em alguns casos esconder variações locais importantes.

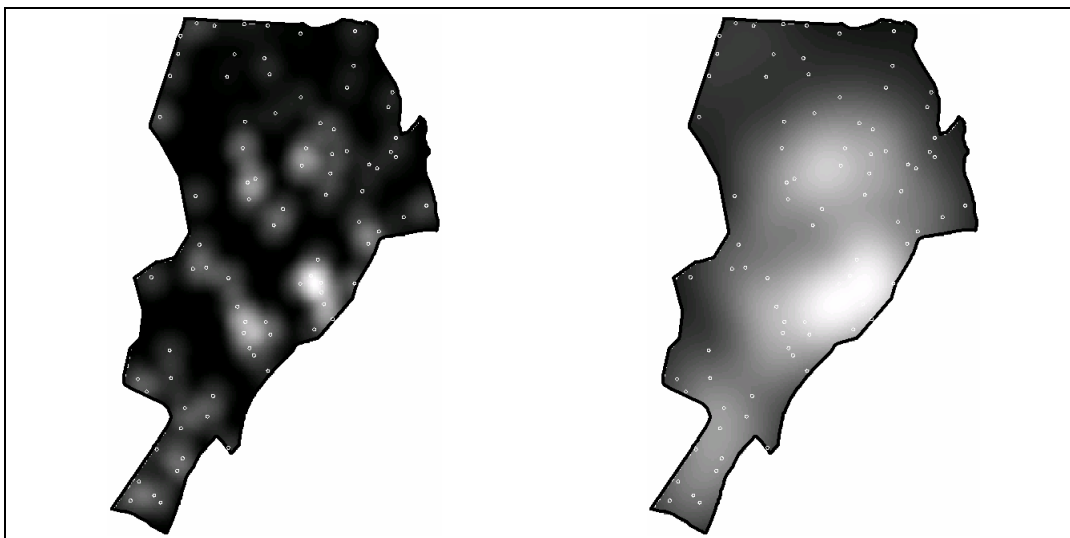


Figura 3-5- Superfícies de teor de argila interpoladas por kernel de quarta ordem. À esquerda, raio de busca de 500m; à direita, raio de busca de 1500m.

### 3.3 SUPERFÍCIES DE TENDÊNCIA

As superfícies de tendência são interpoladores determinísticos globais. A superfície é aproximada por um ajuste polinomial aos dados, através de um processo de regressão múltipla entre os valores do atributo e as localizações geográficas. Essa função polinomial é então utilizada para estimar os valores dos pontos em todas as localizações de uma grade regular que aproxima a superfície.

As superfícies de tendência buscam modelar a variação espacial em larga escala através de uma regressão múltipla entre os valores de atributo e as localizações geográficas. A saída é uma função polinomial na qual o valor do atributo é expresso em função das coordenadas da superfície, expressas em duas ou três dimensões. Exemplos incluem equações lineares do tipo:

$$z = \alpha_1 + \alpha_2 x + \alpha_3 y \quad (3.7)$$

e equações quadráticas como:

$$w = \alpha_1 + \alpha_2 x + \alpha_3 y + \alpha_4 xy + \alpha_5 x^2 + \alpha_6 y^2 \quad (3.8)$$

A suposição implícita nos interpoladores por superfícies de tendência é que, para a caracterização do fenômeno em estudo, predomina a variação em larga escala, e que a variabilidade local não é relevante. Neste modelo, a função de autocorrelação continua decaindo mesmo após ultrapassar a distância onde há influências locais; a covariância não se estabiliza com a distância e assim o fenômeno analisado é não-estacionário.



Para o caso dos dados de teor de argila da Fazenda Canchim (acima descritos), foi realizada uma análise de tendência usando uma regressão linear. Os ajustes indicaram um coeficiente de determinação ( $R^2$  ajustado) de apenas 17,3%, o que indica não haver efeitos espaciais significativos de larga escala. Deste modo, pode-se esperar que estes dados sejam modeláveis por interpoladores locais, sejam determinísticos (seção 3.2) ou estocásticos (seção 3.4 e seguintes).

Um exemplo típico de superfícies de tendência é o uso de dados de longitude, latitude e altitude para estimar a distribuição de temperatura. Neste caso, o objetivo foi estimar a distribuição de temperatura para o estado de Santa Catarina, para a época do plantio de soja, em intervalos de 10 dias (decêndios). Partindo da época recomendada para semeadura e do ciclo de diferentes cultivares de soja, determinou-se um período de análise compreendido entre 11/10 e 20/05 (22 decêndios), permitindo que cultivares com ciclos diferentes, semeadas dentro da época recomendada, tivessem todo o seu ciclo avaliado neste estudo. Foram coletados dados de temperatura média diária e precipitação diária de 27 estações meteorológicas monitoradas pela Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina S. A. – Epagri, com uma série histórica de aproximadamente cinco anos, mostrados na Figura 3-6.

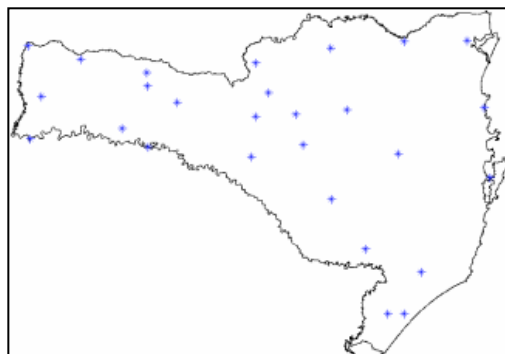


Figura 3-6– Distribuição espacial das estações monitoradas pela Epagri.

A partir dos dados diários, foi calculada a média decendial. Esta média das 27 estações foi utilizada no cálculo de superfícies de tendência a partir de uma equação do tipo:

$$z(x, y, h) = \alpha_1 x + \alpha_2 y + \alpha_3 h + \alpha_4 \quad (3.9)$$

onde  $z$  é a temperatura calculada a partir da longitude ( $x$ ), latitude ( $y$ ) e altitude ( $h$ ). Para o primeiro decêndio (11/10 a 20/10), os resultados estão mostrados na Tabela 3.1. Na análise dos coeficientes da regressão, mostrada na Tabela 3.2, a relação entre as variáveis independentes com a variável dependente (temperatura média decendial) foi verificada, inicialmente, pelo teste “F” e, depois,

pelo teste “t” de Student. Esta análise indicou todos os coeficientes como significativos. A normalidade dos resíduos foi avaliada pelo teste de Keifer-Salmon, e aceita a hipótese.

Tabela 3-2 - Coeficientes para Estimativa de Temperatura em Santa Catarina (Decêndio de 11/10 a 20/10).

	Valor	Teste F	Teste T	p-valor	Comentários
Intercepto	9,475	7,169			Significativo
Latitude	-0,447	0,169	-2,637		(idem)
Longitude	0,466	0,085	5,488		(idem)
Altitude	-0,005	0,000	-16,162		(idem)
R <sup>2</sup> ajustado	0,909				

A grande vantagem das superfícies de tendência é sua simplicidade e facilidade de cálculo. No entanto, a suposição implícita do modelo, em negligenciar a variabilidade local, não é realista para a maior parte dos dados naturais. Adicionalmente, os parâmetros estimados são muito sensíveis a valores extremos (*outliers*). Apesar destes problemas, as superfícies de tendência são úteis para remover efeitos de primeira ordem, quando a média varia de forma consistente no espaço. Outros usos importantes são a análise dos resíduos de estimação; tais resíduos também são bastante informativos, pois mostram a existência de sub-regiões que apresentam diferenças significativas na tendência geral.

No exemplo apresentado, trata-se de uma situação favorável, em que, em função do comportamento da temperatura, da época do ano e das características do estado de Santa Catarina, apenas a variação em larga escala foi capaz de produzir estimativas acuradas. Esta situação não é a mais usual. Na prática, na maior parte das vezes as variações locais não podem ser ignoradas. Neste caso, será preciso modelar o comportamento da variável e para isto, utiliza-se a abordagem geostatística, descrita a seguir.

### 3.4 MODELOS ESTATÍSTICOS DE EFEITOS LOCAIS E GLOBAIS: KRIGAGEM

#### 3.1.1 FUNDAMENTAÇÃO TEÓRICA

A krigagem compreende um conjunto de técnicas de estimação e predição de superfícies baseada na modelagem da estrutura de correlação espacial. A hipótese implícita no procedimento geoestatístico é que o processo estudado é estacionário (veja-se a definição de estacionariedade no capítulo 1 do livro). Os passos num estudo empregando técnicas de krigagem incluem:

- (a) análise exploratória dos dados;
- (b) análise estrutural (modelagem da estrutura de correlação espacial);
- (c) interpolação estatística da superfície.

O procedimento de interpolação é chamado de *krigagem* em honra a Daniel Krige, o pioneiro em introduzir o uso de médias móveis para evitar a superestimação sistemática de reservas em mineração. O que diferencia a krigagem de outros métodos de interpolação é a estimação de uma matriz de covariância espacial que determina os pesos atribuídos às diferentes amostras, o tratamento da redundância dos dados, a vizinhança a ser considerada no procedimento inferencial e o erro associado ao valor estimado. Além disso, a krigagem também fornece estimadores com propriedades de não tendenciosidade e eficiência.

A estrutura teórica da krigagem está baseada no conceito de *variável regionalizada*, desenvolvida por Georges Matheron. Uma variável regionalizada é uma variável distribuída no espaço (ou tempo) cujos valores são considerados como realizações de uma função aleatória (ou processo aleatório, ou campo aleatório, ou processo estocástico). Esta teoria permite incluir hipóteses estatísticas em processos espaciais locais. A variação espacial de uma variável regionalizada pode ser expressa pela soma de três componentes: a) uma componente estrutural, associada a um valor médio constante ou a uma tendência constante; b) uma componente aleatória, espacialmente correlacionada; e c) um ruído aleatório ou erro residual. Se o vetor  $\mathbf{x}$  representa uma posição em uma, duas ou três dimensões, então o valor da função aleatória  $Z$ , em  $\mathbf{x}$ , é dada por:

$$Z(\mathbf{x}) = \mu(\mathbf{x}) + \varepsilon'(\mathbf{x}) + \varepsilon'' \quad (3.10)$$

onde:

- $\mu(\mathbf{x})$  é uma função determinística que descreve a componente estrutural de  $Z$  em  $\mathbf{x}$ ;
- $\varepsilon'(\mathbf{x})$  é um termo estocástico correlacionado, que varia localmente;

- $\varepsilon''$  é um ruído aleatório não correlacionado, com distribuição normal com média zero e variância  $\sigma^2$ .

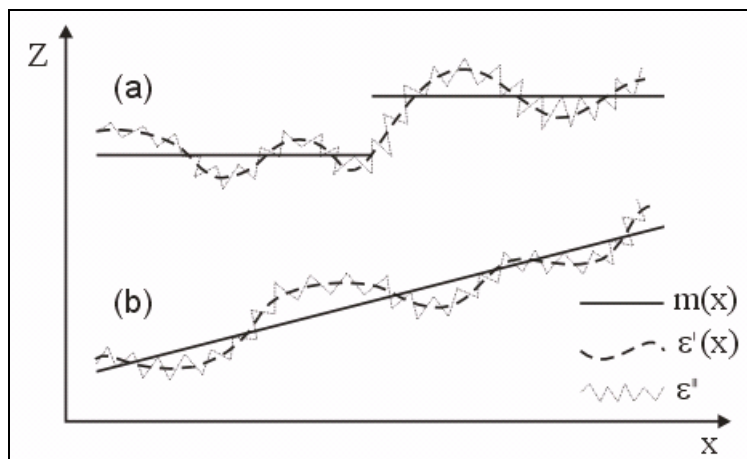


Figura 3-7- Componentes de uma variável regionalizada.

As Figura 3-7(a) e (b) ilustram as três componentes principais da variação espacial. A Figura 3.8(a) apresenta uma componente determinística que possui um comportamento regular (diferença entre os níveis médios), enquanto a componente determinística na Figura 3.8(b) apresenta uma tendência constante.

A hipótese mais simples sobre o comportamento da variável regionalizada é que a média do fenômeno,  $\mu(x)$ , seja constante na região de estudo, o que implica em não haver variação significativa na larga escala. Esta hipótese dá origem aos interpoladores de *Krigagem ordinária*, discutida a seguir. No caso de se querer modelar uma tendência, há vários métodos disponíveis: Krigagem Universal, Funções Aleatórias Intrínsecas de Ordem  $k$ , não discutidos neste capítulo.

Na hipótese da Krigagem ordinária,  $\mu(x)$  é constante e denotada por  $m$ . Deste modo, o valor esperado da função aleatória  $Z$  nas posições  $x$  e  $x + h$  são iguais a  $m$ . Isto implica que o valor esperado da diferença entre os valores observados em  $x$  e  $x + h$ , separados por um vetor de distância  $h$ , é nulo:

$$E [Z(x) - Z(x+h)] = 0 \quad (3.11)$$

Admite-se também que o fenômeno considerado seja *estacionário de segunda ordem*, isto é, a covariância entre dois pares quaisquer  $Z(x)$  e  $Z(x + h)$ , separados por um vetor distância  $h$ , existe e depende somente de  $h$ . Então:

$$C(h) = COV [ Z(x), Z(x+h) ] = E[Z(x).Z(x+h)] - m^2 \quad (3.12)$$

Adicionalmente, a estacionariedade da covariância implica na estacionariedade da variância:

$$\text{Var}(Z(\mathbf{x})) = E [Z(\mathbf{x}) - m]^2 = E[Z^2(\mathbf{x})] - 2E[Z(\mathbf{x})].m + m^2 \quad (3.13)$$

ou ainda

$$\text{Var}(Z(\mathbf{x})) = E[Z^2(\mathbf{x})] - 2m.m + m^2 = E[Z^2(\mathbf{x})] - m^2 = C(0) \quad (3.14)$$

Deste modo, verifica-se que as hipóteses de média constante e estacionariedade da covariância implicam que a determinação da função  $C(\mathbf{h})$  é suficiente para caracterizar a *variável regionalizada*. Isto quer dizer que, com base na Equação 3.10, a função  $C(\mathbf{h})$  permite caracterizar o termo estocástico  $\varepsilon'(\mathbf{x})$ . Para determinar  $C(\mathbf{h})$ , utiliza-se uma função auxiliar, chamada de função *variograma*  $2\gamma(\mathbf{h})$ , definida por:

$$2\gamma(\mathbf{h}) = E[Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})]^2 \quad (3.15)$$

que pode ser desenvolvida em:

$$2\gamma(\mathbf{h}) = E[Z^2(\mathbf{x}) - 2Z(\mathbf{x}).Z(\mathbf{x}+\mathbf{h}) - Z^2(\mathbf{x}+\mathbf{h})] \quad (3.16)$$

ou ainda

$$2\gamma(\mathbf{h}) = E[Z^2(\mathbf{x})] - 2E[Z(\mathbf{x}).Z(\mathbf{x}+\mathbf{h})] - E[Z^2(\mathbf{x}+\mathbf{h})] \quad (3.17)$$

Da equação (3.14), obtém-se

$$E[Z^2(\mathbf{x})] = E[Z^2(\mathbf{x}+\mathbf{h})] = C(0) + m^2 \quad (3.18)$$

e da equação (3.13) obtém-se

$$E[Z(\mathbf{x}).Z(\mathbf{x}+\mathbf{h})] = C(\mathbf{h}) + m^2 \quad (3.19)$$

Substituindo as equações (3.18) e (3.19) na equação (3.17), obtém-se:

$$2\gamma(\mathbf{h}) = 2C(0) - 2C(\mathbf{h}) \text{ ou } \gamma(\mathbf{h}) = C(0) - C(\mathbf{h}) \quad (3.20)$$

onde:

$\gamma(\mathbf{h})$  representa o *semivariograma*, que é metade do variograma. A relação em (3.20) indica que sob a hipótese de estacionariedade de 2ª ordem, que a covariância e o semivariograma são formas alternativas de caracterizar a autocorrelação dos pares  $Z(\mathbf{x})$  e  $Z(\mathbf{x}+\mathbf{h})$  separados pelo vetor  $\mathbf{h}$ .

### 3.1.2 DETERMINAÇÃO EXPERIMENTAL DO SEMIVARIOGRAMA

O semivariograma é uma ferramenta básica de suporte às técnicas de Krigagem, pois permite representar quantitativamente a variação de um fenômeno regionalizado no espaço. O semivariograma pode ser calculado experimentalmente, considerando o esquema de amostragem em duas dimensões mostrado na Figura 3-8, onde  $z(\mathbf{x})$  denota o valor de uma posição cujos componentes são  $(x_1, y_1)$ , e  $z(\mathbf{x}+\mathbf{h})$  o valor da amostra numa posição cujos componentes são  $(x_2, y_2)$ , sendo  $\mathbf{h}$  um vetor distância (módulo e direção) que separa os pontos.

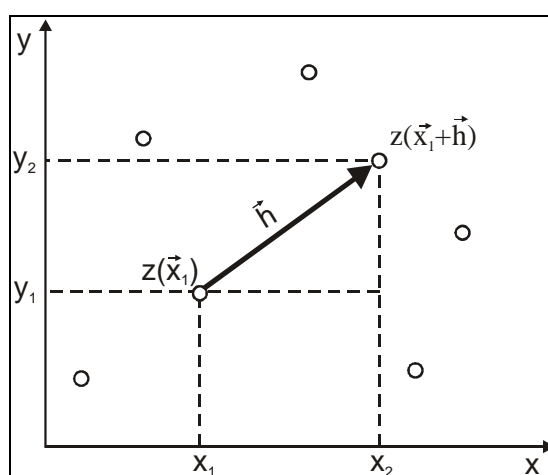


Figura 3-8 – Amostragem em duas dimensões.

A determinação experimental do semivariograma, para cada valor de  $\mathbf{h}$ , considera todos os pares de amostras  $z(\mathbf{x})$  e  $z(\mathbf{x}+\mathbf{h})$ , separadas pelo vetor distância  $\mathbf{h}$ , a partir da equação:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i+h)]^2 \quad (3.21)$$

onde:

$\hat{\gamma}(h)$  é o semivariograma estimado e  $N(\mathbf{h})$  é o número de pares de valores medidos,  $z(\mathbf{x})$  e  $z(\mathbf{x}+\mathbf{h})$ , separados pelo vetor  $\mathbf{h}$ . Esta fórmula, entretanto, não é robusta. Podem existir situações em que variabilidade local não é constante e se modifica ao longo da área de estudo (heteroscedasticidade). Um caso particular desse fato (denominado *efeito proporcional*) ocorre quando as distribuições são assimétricas e a média se correlaciona com a variância. O estimador de semivariograma apresentado em (3.22) não é resistente a esse efeito e apresenta tendências que impedem a estimação correta de seus parâmetros. Para expressões alternativas, deve-se consultar Cressie (1993).

Na prática, pode-se fazer a hipótese adicional de que o fenômeno é *isotrópico* (com comportamento igual em todas as direções). Neste caso, a determinação experimental do semivariograma depende apenas da distância entre as amostras e não da direção relativa entre elas. O tratamento da anisotropia (caso em que a estrutura espacial do fenômeno varia conforme a direção) é discutido no Apêndice deste capítulo.

As hipóteses de estacionariedade e média constante levam a postular um comportamento idealizado para o semivariograma experimental, mostrado na Figura 3-9. Espera-se que observações mais próximas geograficamente tenham um comportamento mais semelhante entre si do que aquelas separadas por maiores distâncias. Assim, o valor absoluto da diferença entre duas amostras  $z(x)$  e  $z(x+h)$  deveria crescer à medida que aumenta a distância entre elas, até um valor na qual os efeitos locais não teriam mais influência.

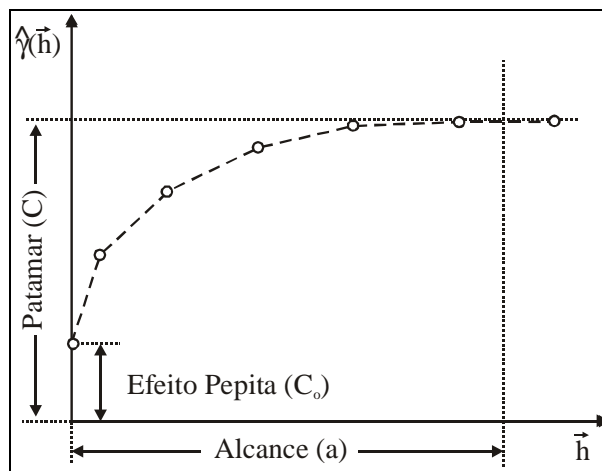


Figura 3-9 – Parâmetros do variograma.

Os parâmetros do semivariograma podem ser observados na Figura 3-9:

- *Alcance (a)*: distância dentro da qual as amostras apresentam-se correlacionadas espacialmente.
- *Patamar (C)*: é o valor do semivariograma correspondente a seu *alcance (a)*. Deste ponto em diante, considera-se que não existe mais dependência espacial entre as amostras, porque a variância da diferença entre pares de amostras ( $Var[Z(x) - Z(x+h)]$ ) torna-se aproximadamente constante.
- *Efeito Pepita (C₀)*: idealmente,  $\gamma(0)=0$ . Entretanto, na prática, à medida que  $h$  tende para zero,  $\gamma(h)$  se aproxima de um valor positivo chamado *Efeito Pepita (C₀)*, que revela a descontinuidade do semivariograma para distâncias menores do que a menor distância entre as amostras. O efeito pepita é o valor da

semivariância para a distância zero e representa a componente da variabilidade espacial que não pode ser relacionado com uma causa específica (variabilidade ao acaso). Parte desta descontinuidade pode ser também devida a erros de medição, sendo impossível quantificar se a maior contribuição provém dos erros de medição ou da variabilidade de pequena escala não captada pela amostragem.

### **3.1.3 MODELOS TEÓRICOS**

O gráfico do semivariograma experimental,  $\hat{\gamma}(h)$ , calculado através da Equação (3.22), é formado por uma série de valores, conforme ilustra a Figura 3-9, sobre os quais se objetiva ajustar uma função. É importante que o modelo ajustado represente a tendência de  $\hat{\gamma}(h)$  em relação a  $h$ . Deste modo, as estimativas obtidas a partir da krigagem serão mais exatas e, portanto mais confiáveis.

O procedimento de ajuste não é direto e automático, como no caso de uma regressão, por exemplo, mas sim interativo, pois nesse processo o intérprete faz um primeiro ajuste e verifica a adequação do modelo teórico. Dependendo do ajuste obtido, pode ou não redefinir o modelo, até obter um que seja considerado satisfatório.

Os modelos aqui apresentados são considerados modelos básicos, denominados modelos isotrópicos. Estão divididos em dois tipos: modelos com patamar e modelos sem patamar. Modelos do primeiro tipo são referenciados na geoestatística como modelos transitivos. Alguns dos modelos transitivos atingem o patamar ( $C$ ) assintoticamente. Para tais modelos, o alcance ( $a$ ) é arbitrariamente definido como a distância correspondente a 95% do patamar. Modelos do segundo tipo não atingem o patamar, e continuam aumentando enquanto a distância aumenta. Tais modelos são utilizados para modelar fenômenos que possuem capacidade infinita de dispersão. Os modelos transitivos mais utilizados são: modelo esférico (Sph), modelo exponencial (Exp) e modelo gaussiano (Gau). Estes modelos estão apresentados na Figura 3-10 com o mesmo alcance ( $a$ ).



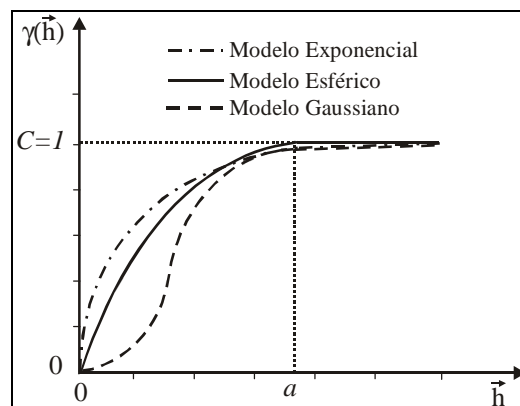


Figura 3-10 – Representação gráfica de modelos transitivos normalizados.

*Modelo Esférico*

O modelo esférico é um dos modelos mais utilizados e está representado na Figura 3-10. A equação normalizada deste modelo é:

$$\text{Sph}(\mathbf{h}) = \begin{cases} 0 & , \|\mathbf{h}\|=0 \\ 1,5\left(\frac{\|\mathbf{h}\|}{a}\right) - 0,5\left(\frac{\|\mathbf{h}\|}{a}\right)^3 & , 0 < \|\mathbf{h}\| \leq a \\ 1 & , \|\mathbf{h}\| > a \end{cases} \quad (3.22)$$

*Modelo Exponencial*

Um outro modelo bastante utilizado é o modelo exponencial, o qual é apresentado na Figura 3-10. A equação normalizada deste modelo é:

$$\text{Exp}(\mathbf{h}) = \begin{cases} 0 & , \|\mathbf{h}\|=0 \\ 1 - \exp\left(-\frac{\|\mathbf{h}\|}{a}\right) & , \|\mathbf{h}\| \neq 0 \end{cases} \quad (3.23)$$

Este modelo atinge o patamar assintoticamente, com o alcance prático definido como a distância na qual o valor do modelo é 95% do patamar.

*Modelo Gaussiano*

O modelo gaussiano é um modelo transitivo, muitas vezes usado para modelar fenômenos extremamente contínuos. Sua formulação é dada por:

$$\text{Gau}(\mathbf{h}) = \begin{cases} 0 & , \|\mathbf{h}\|=0 \\ 1 - \exp\left(-\frac{\|\mathbf{h}\|}{a}\right)^2 & , \|\mathbf{h}\| \neq 0 \end{cases} \quad (3.24)$$

Semelhante no modelo exponencial, o modelo gaussiano atinge o patamar assintoticamente e o parâmetro  $a$  é definido como o alcance prático ou distância na qual o valor do modelo é 95% do patamar. O que caracteriza este modelo é seu comportamento parabólico próximo à origem, conforme a Figura 3-10 .

Até este ponto foram apresentados os principais modelos básicos normalizados, os quais são utilizados para ajustar o semivariograma experimental. Na prática, os semivariogramas experimentais possuem valores de efeito pepita ( $C_o$ ) maior que zero e valores de patamar ( $C$ ) maiores que a unidade, conforme ilustrado na Figura 3-11.

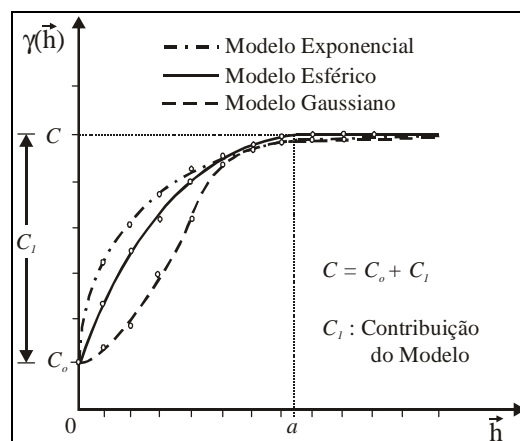


Figura 3-11 - Representação gráfica de semivariogramas experimentais e modelos teóricos.

Em resumo, os semivariogramas dos modelos transitivos básicos são assim definidos:

- *Modelo Esférico de Semivariograma:*

$$\gamma(h) = \begin{cases} 0 & , |h| = 0 \\ C_o + C_I \left[ \frac{3}{2} \left( \frac{|h|}{a} \right) - \frac{1}{2} \left( \frac{|h|}{a} \right)^3 \right] = C_o + C_I [Sph(|h|)] & , 0 < |h| \leq a \\ C_o + C_I & , |h| > a \end{cases} \quad (3.25)$$

- *Modelo Exponencial de Semivariograma:*

$$\gamma(h) = \begin{cases} 0 & , |h| = 0 \\ C_o + C_I \left[ 1 - \exp\left(-\frac{|h|}{a}\right) \right] = C_o + C_I [Exp(|h|)] & , |h| \neq 0 \end{cases} \quad (3.26)$$

- *Modelo Gaussiano de Semivariograma:*

$$\gamma(h) = \begin{cases} 0 & |h| = 0 \\ C_0 + C_1 \left[ 1 - \exp\left(-\frac{|h|}{a}\right)^2 \right] & |h| \neq 0 \end{cases} = C_0 + C_1 [\text{Gau}(|h|)], |h| \neq 0 \quad (3.27)$$

### Modelos Aninhados

Existem determinados fenômenos em que são necessários modelos mais complexos de semivariograma para explicar suas variações espaciais. Estes modelos são combinações de modelos simples, denominados *aninhados*; em muitos casos, os modelos aninhados são necessários para explicar a variação de fenômenos decorrentes da combinação de fatores independentes de formação. Por exemplo, um modelo aninhado útil em estudos de mineração e pesquisa de solo é o duplo esférico, definido como:

$$\gamma(h) = \begin{cases} C_0 + C_1 \left[ \frac{3}{2} \left( \frac{|h|}{a_1} \right) - \frac{1}{2} \left( \frac{|h|}{a_1} \right)^3 \right] = \gamma_1(h), & 0 < |h| \leq a_1 \\ C_0 + C_2 \left[ \frac{3}{2} \left( \frac{|h|}{a_2} \right) - \frac{1}{2} \left( \frac{|h|}{a_2} \right)^3 \right] = \gamma_2(h), & a_1 < |h| \leq a_2 \\ C_0 + C_1 + C_2 & |h| > a_2 \\ 0 & |h| = 0 \end{cases} \quad (3.28)$$

onde,

- $a_1$  e  $C_1$  correspondem aos parâmetros de alcance e contribuição, respectivamente, do primeiro modelo esférico ( $\gamma_1(h)$ ).
- $a_2$  e  $C_2$  correspondem aos parâmetros de alcance e contribuição, respectivamente, do segundo modelo esférico ( $\gamma_2(h)$ ).

Este modelo é mostrado na Figura 3-12, onde as linhas sólida e pontilhada representam os modelos de ajuste teórico ao semivariograma experimental.

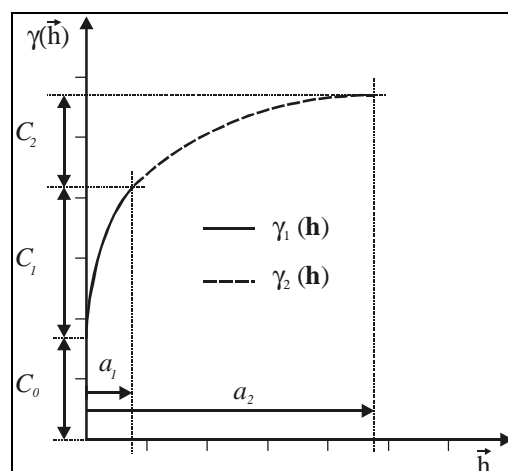


Figura 3-12 - Representação gráfica de um modelo duplo esférico.

Dependendo do fenômeno em estudo, outros modelos aninhados são necessários para caracterizar a variabilidade espacial.

### 3.5 KRIGAGEM

O termo *krigagem* é derivado do nome Daniel G. Krige, que foi o pioneiro a introduzir o uso de médias móveis para evitar a superestimação sistemática de reservas de mineração. Inicialmente, o método de krigagem foi desenvolvido para solucionar problemas de mapeamentos geológicos, mas seu uso expandiu-se com sucesso no mapeamento de solos, mapeamento hidrológico, mapeamento atmosférico e outros campos correlatos. A diferença entre a krigagem e outros métodos de interpolação é a maneira como os pesos são atribuídos às diferentes amostras. No caso de interpolação linear simples, por exemplo, os pesos são todos iguais a  $1/N$  ( $N$  = número de amostras); na interpolação baseada no inverso do quadrado das distâncias, os pesos são definidos como o inverso do quadrado da distância que separa o valor interpolado dos valores observados. Na Krigagem, o procedimento é semelhante ao de interpolação por média móvel ponderada, exceto que aqui os pesos são determinados a partir de uma análise espacial, baseada no semivariograma experimental. Além disso, a krigagem fornece, em média, *estimativas não tendenciosas e com variância mínima*<sup>1</sup>.

<sup>1</sup>Estimativas não tendenciosas significam que, em média, a diferença entre valores estimados e observados para o mesmo ponto deve ser nula; e variância mínima significa que estes estimadores possuem a menor variância dentre todos os estimadores não tendenciosos.

A krigagem engloba um conjunto de métodos de estimação, incluindo procedimentos estacionários (krigagem simples e ordinária), não estacionários (krigagem universal, funções intrínsecas de ordem  $k$ ), univariados e multivariados (co-krigagem etc). Este capítulo limita-se à apresentação da krigagem ordinária, descrita a seguir.

### 3.5.1. KRIGEAGEM ORDINÁRIA

Considere uma superfície sobre a qual se observe alguma propriedade do solo,  $Z$ , em  $n$  pontos distintos, com coordenadas representadas pelo vetor  $\mathbf{x}$ . Assim, tem-se um conjunto de valores  $\{z(\mathbf{x}_i), i=1, \dots, n\}$ , onde  $\mathbf{x}_i$  identifica uma posição em duas dimensões representada pelos pares de coordenadas  $(x_i, y_i)$ . Suponha que se objetive estimar o valor de  $Z$  no ponto  $\mathbf{c}$ . O valor desconhecido de  $Z(\mathbf{x}_0)$  pode ser estimado a partir de uma combinação linear dos  $n$  valores observados, adicionado a um parâmetro  $\lambda_0$ :

$$Z^*(\mathbf{x}_0) = \lambda_0 + \sum_{i=1}^n \lambda_i Z(\mathbf{x}_i) \quad (3.29)$$

Deseja-se um estimador não tendencioso, isto é,

$$E[Z(\mathbf{x}_0) - Z^*(\mathbf{x}_0)] = 0 \quad (3.30)$$

A relação acima impõe que as duas médias sejam iguais; assim aplicando-se a Equação 3.34 em 3.35, obtém-se:

$$E[Z(\mathbf{x}_0)] = E\left[\lambda_0 + \sum_{i=1}^n \lambda_i Z(\mathbf{x}_i)\right] \Rightarrow m = \lambda_0 + \sum_{i=1}^n \lambda_i m \quad (3.31)$$

A krigagem ordinária não requer o prévio conhecimento da média  $m$ . Neste caso, para que a igualdade da Equação 3.36 seja satisfeita é necessário que  $\lambda_0 = 0$  e  $\sum_{i=1}^n \lambda_i = 1$ . Portanto, o estimador de Krigagem ordinária é:

$$Z^*(\mathbf{x}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{x}_i), \text{ com } \sum_{i=1}^n \lambda_i = 1 \quad (3.32)$$

Minimizando a variância do erro ( $\text{Var}[Z(\mathbf{x}_0) - Z^*(\mathbf{x}_0)]$ ) na condição de  $\sum_{i=1}^n \lambda_i = 1$ , os pesos  $\lambda_i$  são obtidos a partir do seguinte sistema de equações, denominado sistema de krigagem ordinária:

$$\begin{cases} \sum_{j=1}^n \lambda_j C(\mathbf{x}_i, \mathbf{x}_j) - \alpha = C(\mathbf{x}_i, \mathbf{x}_0) & \text{para } i=1, \dots, n \\ \sum_{j=1}^n \lambda_j = 1 \end{cases} \quad (3.33)$$

onde,

- $C(\mathbf{x}_i, \mathbf{x}_j)$  e  $C(\mathbf{x}_i, \mathbf{x}_0)$  são, respectivamente, a semivariância entre os pontos  $\mathbf{x}_i$  e  $\mathbf{x}_j$  e entre os pontos  $\mathbf{x}_i$  e  $\mathbf{x}_0$ .
- $\alpha$  é o multiplicador de Lagrange necessário para a minimização da variância do erro.

A correspondente variância minimizada do erro, denominada variância de krigagem ordinária ( $\sigma_{ko}^2$ ), é dada pela expressão

$\sigma_{ko}^2 = \text{Var}[Z(\mathbf{x}) - Z^*(\mathbf{x}_0)] = C(\mathbf{0}) - \sum_{i=1}^n \lambda_i C(\mathbf{x}_i, \mathbf{x}_0) - \alpha$	(3.34)
---	--------

A krigagem ordinária é um interpolador exato no sentido de que, quando as equações acima forem usadas, os valores interpolados irão coincidir com os valores dos pontos amostrais. Além disso, a variância da krigagem ordinária, indicada na equação (3.35), fornece informação importante sobre a confiabilidade dos valores interpolados.

### 3.6 ESTUDO DE CASO

Tomemos como exemplo a distribuição amostral apresentada na Figura 3-2, cuja as estatísticas descritivas estão sumarizadas na Tabela 3-1. A análise da variabilidade espacial, do teor de argila, é realizada com o auxílio do semivariograma. Esta é uma das etapas mais importantes, pois o modelo de semivariograma escolhido representa a estrutura de correlação espacial a ser utilizada nos procedimentos inferenciais de krigagem. O resultado apresentado na Figura 3-13, mostra o semivariograma omnidirecional (caso isotrópico) e seu modelo de ajuste.

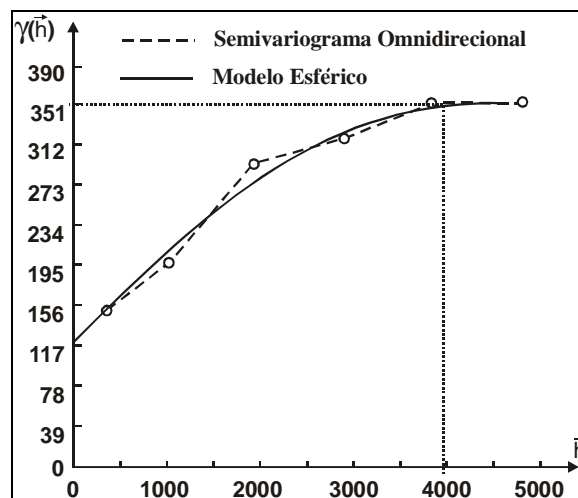


Figura 3-13 – Semivariograma omnidirecional e modelo esférico

O modelo de ajuste, mostrado na Figura 3-13, têm os seguintes parâmetros: Estrutura tipo Esférica, Efeito Pepita ( $C_0$ ) = 118,85; Contribuição ( $C_1$ ) = 230,89 e Alcance ( $a$ ) = 3989,20. O modelo teórico, normalizado em relação ao alcance, leva a seguinte notação:

$$\gamma(h) = C_0 + C_1 \left[ \text{Sph} \left( \frac{h}{a} \right) \right] = 118,85 + 230,89 \left[ \text{Sph} \left( \frac{h}{3989,20} \right) \right] \quad (3.35)$$

Uma vez definido o modelo e validado o mesmo, a etapa seguinte refere-se à estimação de krigagem ordinária. Como resultado têm-se uma grade de valores estimados e uma outra que refere-se à variância de krigagem. Ambas são convertidas em superfícies e apresentadas na Figura 3-14. Na Figura 3-14 à esquerda, regiões mais claras representam altos valores de teor de argila e vice-versa. Diferente dos métodos determinísticos (ver Figura 3-4), o uso da krigagem ordinária como método de interpolação espacial permitiu capturar e, portanto, representar com mais qualidade, a variabilidade espacial inerente à propriedade em estudo. Além disso, conforme ilustra a Figura 3-14 à direita, a krigagem ordinária fornece a variância da estimativa (denominada variância de krigagem). Tal informação pode ser útil para identificar regiões onde a amostragem pode ser melhorada.

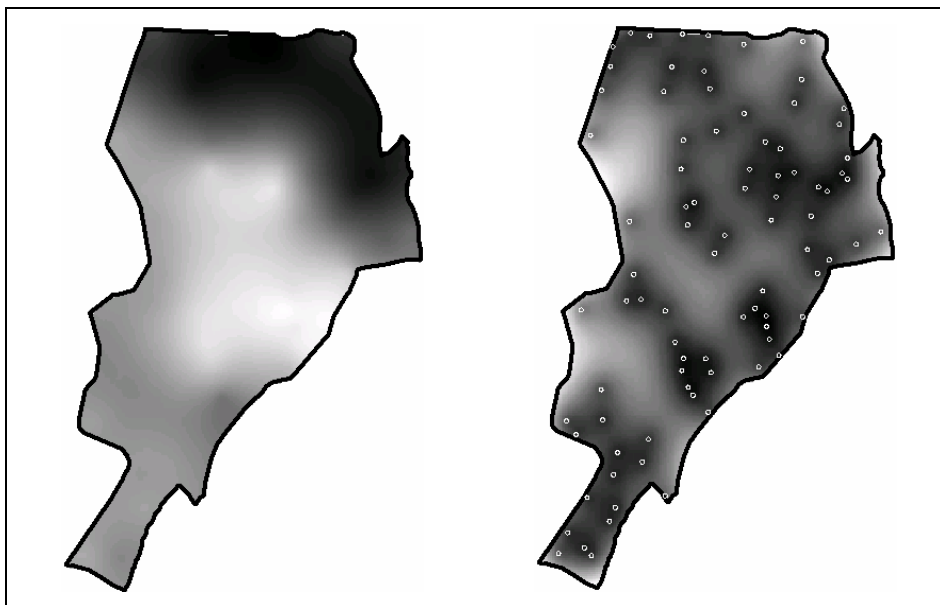


Figura 3-14 – À esquerda a superfície do teor de argila e à direita a variância de krigagem.

Com algumas ressalvas, o método da *média ponderada pelo inverso do quadrado da distância*, produz resultado que se assemelha ao resultado da krigagem ordinária. O ponto crítico, porém, ocorre em regiões onde há agrupamento (“clusters”) de amostras. A krigagem ordinária, por utilizar intrinsecamente uma estrutura de covariância, consegue tratar redundâncias (“clusters”), isto é, atribuir pesos adequados para os agrupamentos de amostras. Fato este não considerado nos procedimentos determinísticos. Além disso, na krigagem ordinária, a área de influência na interpolação é indicada pelo alcance; já nos procedimentos determinísticos, como o método da *média ponderada pelo inverso do quadrado da distância*, o raio de busca é arbitrário.

Os resultados produzidos pelos métodos *média simples* e *vizinho mais próximo*, são menos expressivos com relação aos demais. O método da *média simples* produz resultado que apresenta imbricação, principalmente na região central da área de estudo. Já o método de inferência relativo ao *vizinho mais próximo*, embora sendo o que pior expressa a variabilidade espacial do fenômeno estudado, revela a área de influência de cada ponto de observação. Tal informação é de grande valia, como, por exemplo, numa análise preliminar para detecção de valores amostrais suspeitos.

Um outro fato que merece atenção, é que os resultados apresentados na Figura 3-14 são oriundos de um modelo isotrópico. A suposição de isotropia, que é rara em fenômenos naturais, simplifica a modelagem por procedimentos geoestatísticos. Se a anisotropia existe, deve ser detectada e modelada, afim de representar com mais qualidade, a variabilidade espacial inerente à propriedade em estudo. No



apêndice ao Capítulo, são apresentados alguns tópicos sobre anisotropia e uma técnica para a modelagem da mesma.

### 3.7 CONCLUSÕES

Conclui-se que é possível melhorar a distribuição espacial das variáveis ambientais significativamente quando procedimentos geoestatísticos são aplicados. Ficou constatado que o teor de argila varia mais intensamente numa direção do que em outra. Tal fato refere-se à anisotropia da variável em estudo. Muitos aspectos particulares dos dados ficariam ocultos sem o uso de semivariogramas e da modelagem da anisotropia, mostrando, por exemplo, a tendência da distribuição espacial nos dados de teor de argila. Informações como estas não são apresentadas quando se usam apenas parâmetros estatísticos clássicos como médias e variâncias ou então, procedimentos determinísticos.

### 3.8 REFERÊNCIAS BIBLIOGRÁFICAS

A estrutura teórica da geoestatística está apresentadas na *Teoria das Variáveis Regionalizadas*, desenvolvida por Matheron (1971) e um artigo detalhado e teórico sobre geoestatística é escrito por Journel (1988). A referência básica sobre geoestatística, com um conjunto extensivo de exemplos é o livro de Issaks e Srivastava (1989). A descrição da GSLIB, uma das bibliotecas mais utilizadas para o desenvolvimento de programas em geoestatística, pode ser encontrada no livro de Deutsch e Journel (1992). Com relação à integração entre geoestatística e SIGs, o leitor deve referir-se a Camargo (1997), que descreve o desenvolvimento de um módulo geoestatístico no ambiente SPRING. Referências básicas sobre métodos de interpolação são descritas por Burrough (1987). O exemplo de superfícies de tendência está baseado no trabalho de Bönisch (2001).

Bönisch, S. (2001) *Geoprocessamento Ambiental com Tratamento de Incerteza: O Caso do Zoneamento Pedoclimático para a Soja no Estado de Santa Catarina*. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos.

Burrough, P. (1987). *Principles of geographical information systems for land resources assessment*. Oxford, Clarendon Press.

Camargo, E. (1997). *Desenvolvimento, Implementação e Teste de Procedimentos Geoestatísticos (Krigagem) no Sistema de Processamento de Informações Georreferenciadas (SPRING)*. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos.

- Deutsch, C. e A. Journel (1992). *GSLIB: Geostatistical Software Library and user's guide*. New York, Oxford University Press.
- Issaks, M. e E. Srivastava (1989). *An Introduction to Applied Geostatistics*. New York, Oxford University Press, 1989.
- Journel, A. (1988). *Fundamentals of geostatistics in five lessons*. California, Stanford Center for Reservoir Forecasting Applied Earth Sciences Department.
- Matheron (1963, 1971). *The theory of regionalized variables and its applications*. Paris, Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, 1971. 211p.