

*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Solos
Ministério da Agricultura, Pecuária e Abastecimento
Secretaria de Desenvolvimento Agropecuário e Cooperativismo
Departamento de Propriedade Intelectual e Tecnologia da Agropecuária*

GEOESTATÍSTICA APLICADA NA AGRICULTURA DE PRECISÃO UTILIZANDO O VESPER

*Ronaldo Pereira de Oliveira
Célia Regina Grego
Ziany Neiva Brandão
Editores Técnicos*

Embrapa
Brasília, DF
2015

Capítulo 3

Conceitos Básicos da Geoestatística

Célia Regina Grego

Ronaldo Pereira de Oliveira

3.1 Origem e Conceitos

Os primeiros relatos do uso da geoestatística datam de 1911, com W.B. Mercer e A. D. Hall que examinaram a variação da produção de culturas no espaço de pequenos lotes. Mas somente na década de 60 a geoestatística foi consolidada. Na área de mineração, em 1951, Daniel G. Krige, engenheiro de minas de ouro na África do Sul, observou ganhos na precisão das estimativas com dados de concentração do mineral quando considerou as amostras vizinhas (KRIGE, 1951). Desta forma, a autocorrelação espacial tornou-se prática nas minas de ouro, considerando que para encontrar sentido nas variações das concentrações de ouro, era preciso levar em conta as distâncias entre as amostras. Na evolução das técnicas, em 1963, Matheron, um matemático da escola francesa, teve a mesma preocupação de melhorar as estimativas de ocorrência de minérios a partir de dados autocorrelacionados, derivando posteriormente soluções para o problema de estimativa a partir da teoria das variáveis regionalizadas (MATHERON, 1965; MATHERON, 1971). Conjunto de técnicas este que caracteriza um ramo específico da estatística espacial, denominado de geoestatística (WEBSTER; OLIVER, 2007).

A geoestatística parte do princípio que quanto mais próximas as amostras, mais parecidas entre si elas se apresentaram. Segundo Soares (2006), abrange um conjunto de métodos, técnicas e instrumentos estatísticos que caracterizam os fenômenos espaciais naturais e tem

por objetivo a caracterização da dispersão espacial e espaço-temporal das grandezas que definem a quantidade e a qualidade de recursos naturais e outros fenômenos em que os atributos manifestem certa estrutura no espaço e ou no tempo. Neste contexto, destaca-se o método de interpolação por krigagem, no qual se pode estimar o valor de uma dada propriedade para um local onde não foi medida.

Com este método, pode-se estimar o valor de uma dada propriedade em local não observado, utilizando uma função de correlação espacial entre os dados, sem viés e com variância mínima (VIEIRA, 2000). A interpolação por krigagem, ou simplesmente krigagem, cujo nome foi dado em homenagem a Daniel G. Krige, consiste em ponderar os vizinhos mais próximos do ponto a ser estimado, obedecendo os critérios de não tendenciosidade, que significa que em média a diferença entre valores estimados e observados para o mesmo ponto dever ser nula e ter mínima variância, ou seja, que os estimadores possuam a menor variância dentre todos os estimadores não tendenciosos.

Fazendo uma comparação entre a estatística clássica e a geoestatística, podemos dizer necessita da normalidade e independência espacial entre os dados enquanto que a geoestatística requer a correlação e dependência espacial. A estatística clássica assume que os pontos de observação são independentes, o que, na maioria dos casos, não acontece nos estudos envolvendo as ciências da terra (SRIVASTAVA, 1996).

A intensificação do uso da geoestatística deveu-se ao fato de assumir-se que a distribuição espacial de pontos de observação apresenta correlação, ou seja, que exista dependência espacial. Esta presuposição não é verdadeira quando se assume que os pontos de observação são independentes, o que, na maioria dos casos, não acontece nos estudos envolvendo as ciências da terra (SRIVASTAVA, 1996; WEBSTER, 1985). Srivastava (1996) comenta que a geoestatística incorpora, além da análise da distribuição estatística dos dados coletados, também as relações espaciais entre estes, na forma de correlação entre os pontos amostrados, pois de acordo com Trangmar et al. (1985), amostras mais próximas, dentro de uma mesma mancha de solo, são mais parecidas do que as mais distantes. Por essas diferenças, segundo Goovaerts (1997), os problemas da ciência da terra são efetivamente analisados atualmente por técnicas da geoestatística,

quando a interpretação da distribuição espacial dos dados tem forte impacto sobre os resultados e sobre a tomada de decisão.

Com a análise geoestatística é possível extrair e organizar os dados disponíveis numa relação espacial dos mesmos de acordo com a semelhança entre vizinhos georreferenciados. Suas técnicas são bastante consolidadas em estudos de solos, caracterizando a dependência espacial de diversos atributos físicos e químicos (ORTIZ, 2002; VIEIRA et al., 1981; WALTER et al., 2001) independentemente do tamanho da área amostrada (GOOVAERTS, 1997; GREGO; VIEIRA, 2005; WARRICK; NIELSEN, 1980). Estas análises apresentam um potencial de aplicação em diversas outras questões envolvendo ciências da terra e do ambiente (SOARES, 2006). Portanto, segundo Molin (2012), traz grande contribuição para a AP, principalmente na definição de unidades de manejo a partir de mapas de produtividade. Desta forma, a AP que propõe o manejo aplicado de acordo com a variabilidade local, seja com foco na produtividade da cultura ou da fertilidade do solo, pode em muito se beneficiar destas técnicas para obter vantagens econômicas e ambientais na sua aplicação.

A geoestatística é um modelo probabilístico que baseia-se na aleatoriedade dos dados para avaliar a correlação espacial; entre o valor de uma variável em local conhecido e o valor da mesma variável localizada em sua vizinhança (HUIJBREGTS, 1975). Na análise da autocorrelação espacial, a variável que apresenta um certo grau de correlação na distribuição espacial de seus valores pode ser considerada regionalizada (LIMA, 2006). Característica esta comum a diversas variáveis do meio físico. Este comportamento dos dados está fundamentado em modelos probabilísticos que definem a chamada teoria das variáveis regionalizadas (MATHERON, 1965; STURARO, 1994).

A função aleatória é considerada estacionária quando se assume que o seu valor médio esperado é constante em qualquer direção (STURARO, 1994). A hipótese de estacionaridade considera que a variável deve ser estatisticamente homogênea e isotrópica, permitindo que se façam inferências estatísticas (VIEIRA et al., 1983). Como na prática esta hipótese básica é difícil de ser verificada, trabalha-se alternativamente com as hipóteses de segunda ordem e a hipótese intrínseca conforme apresentadas em Pannatier (1996) e Vieira et al. (1983).

Os métodos são geralmente divididos em duas etapas metodológicas: a Análise Variográfica e Modelos de Estimção. A análise variográfica está estruturada no conceito de variáveis regionalizadas. Uma variável regionalizada é uma variável distribuída no espaço (ou no tempo) cujos valores são considerados como realizações de uma função aleatória (i.e.: processo aleatório, campo aleatório ou processo estocástico), no sentido de que os valores das medições feitas podem variar consideravelmente entre si dentro de uma dada distância. Entretanto, espera-se que, pares de pontos separados por distâncias menores do que uma determinada amplitude variográfica sejam espacialmente correlacionadas (i.e.: tenham valores mais próximos). Por outro lado, pares mais distantes do que esta amplitude não apresentam uma dependência espacial (LIMA, 2006).

Após a análise variográfica e verificada a possibilidade de estimção por técnicas geoestatísticas, pode-se proceder a uma estimção de valores em locais não amostrados. Constitui-se essa, numa tarefa importantíssima dos estudos ambientais, principalmente no que diz respeito a espacialização e representação cartográfica de diversos fenômenos de interesse. A krigagem constitui-se em um método de estimção por médias móveis e tem como característica particular, que o diferencia e o torna superior aos demais métodos de estimção, o fato de permitir o cálculo do erro associado às estimativas, chamado de variância de estimção.

3.2 Análise do Variograma

O variograma é a representação gráfica da dependência espacial obtido pela variância versus a distância. Segundo Vieira et al. (2008), é a grandeza mais aproximada para decidir se a dependência espacial existe ou não. Se a dependência espacial existir, haverá um crescimento na semivariância até uma determinada distância, a partir da qual o variograma se estabiliza. Caso contrário, se o variograma não apresentar nenhum crescimento com a distância, as amostras são independentes e ocorre o efeito de aleatoriedade, comumente chamado de efeito pepita puro.

A ocorrência da dependência espacial expressa no variograma segue a hipótese básica na qual dados da vizinhança são mais pareci-

dos do que dados distantes. Em termos simples, podemos dizer que o variograma é um medidor do grau de semelhança entre vizinhos. Como brevemente discutido acima, os variogramas são construídos partindo das pressuposições de estacionaridade da hipótese intrínseca e cálculo da semivariância $\gamma(h)$ dada pela Equação 2:

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

Onde $N(h)$ é o número de pares dos valores medidos $Z(x_i)$, $Z(x_i + h)$, separados por um vetor h . Segundo Vieira (2000), a caracterização da dependência espacial no variograma (Figura 3.1) dá-se a partir da semivariância na distância zero (i.e. no valor do efeito pepita - C_0), de forma que ocorra um aumento gradual de $\gamma(h)$, proporcional ao aumento da distância h , até um valor máximo (i.e.: intervalo que representa uma variância estruturada - C_1), ponto a partir do qual o variograma se estabiliza em um patamar ($C_0 + C_1$) que correspondente diretamente à distância limite de dependência espacial (i.e.: o valor de alcance - a).

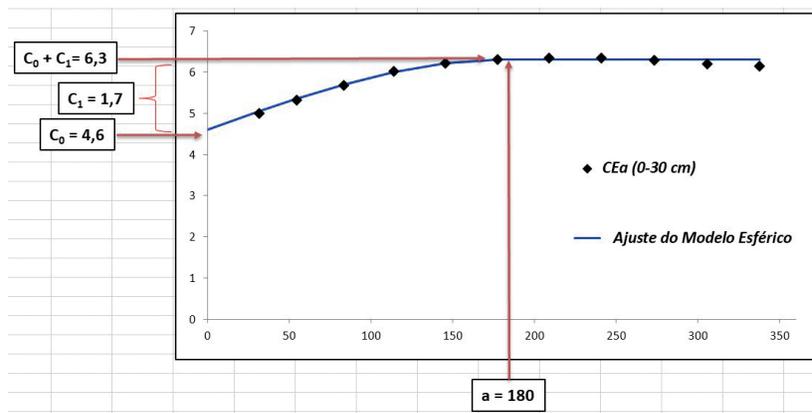


Figura 3.1. Identificação dos parâmetros de ajuste do variograma de CE_a, indicando uma autocorrelação espacial entre valores da variável distantes até 180 m.

O gráfico do variograma (i.e.: variância $\gamma(h)$ versus distância h) representa uma série de pontos discretizados por um espaçamento específico e recursivo no eixo X, usualmente denominado de *lag* (termo em analogia a passo ou pulo), plotando os valores obtidos no cálculo

da semivariância média de todos os pares de observações separados pela mesma distância. Na Figura 3.1 podemos observar um *lag* de aproximadamente 30 m. Este, representa o modelo do variograma empírico, também conhecido como experimental, que é calculado para os valores obtidos nas observações de campo e corresponde ao valor da distância na qual uma função contínua, denominada de modelo do variograma teórico, deverá ser ajustada. Para o cálculo da dependência espacial, os modelos dos variogramas empíricos devem ser ajustados pelo modelo matemático de melhor correspondência (i.e. modelos teóricos), para que se possa obter os valores dos parâmetros de ajuste do variograma. Estes parâmetros são:

- Amplitude variográfica, “alcance” ou “*range*” (notação: “*a*”): Distância na qual a máxima variabilidade é atingida e que corresponde ao aumento da distância entre as amostras;
- Patamar ou “*sill*” (notação: “ C_1 ”): Representa o nível de variabilidade onde o variograma se estabiliza. Corresponde à diferença entre o ponto de maior correlação ou a origem do variograma e o ponto que teoricamente representa a variância populacional e a variabilidade se estabiliza;
- Efeito pepita ou “*nugget effect*” (notação: “ C_0 ”): Descontinuidade na origem do variograma, correspondendo à diferença entre as amostras de maior proximidade e gerada por microrregionalizações, erros de amostragens ou erros de medidas.

Várias formulações e algoritmos utilizando parâmetros do variograma visam classificar e quantificar a estrutura espacial da variação (DIGGLE; RIBEIRO JUNIOR, 2007; OLIVEIRA et al., 2007; PRINGLE et al., 2003; TISSEYRE; MCBRATNEY, 2007). Garcia (1988) considera que uma relação entre os parâmetros “ C_0 ” e “ C_1 ” expressa o grau de aleatoriedade do fenômeno regionalizado, e pode ser avaliada por “*E*”, que representa o efeito pepita relativo. Garcia (1988) categoriza o efeito de pepita relativo da seguinte forma: a) $E < 0,15$ - componente aleatória pequena; b) $0,15 < E < 0,30$ - componente aleatória significativa; e c) $E > 0,30$ - componente aleatória muito significativa. Outros modelos também considerando razões entre estes parâmetros são usados para classificar a estrutura espacial da variação em: forte, média e fraca (CAMBARDELLA et al., 1994; ZIMBACK, 2001).

Deve-se salientar que o ajuste do variograma é um dos aspectos mais importantes da análise geoestatística. O método de tentativa e erro, aliado ao exame de indicadores de melhor ajuste. Alguns coeficientes, como o coeficiente de correção, a raiz quadrada do erro médio (RMSE) e o índice com o de Akaike (IAC) podem auxiliar na validação do modelo escolhido e dos parâmetros para o ajuste do variograma. O ajuste por métodos automáticos, embora possa ser usado, não é o mais adequado para acervos com baixa densidade amostral. Segundo Vieira et al. (2010), uma poderosa ferramenta de validação é o método “*Jack-knifing*” que calcula os parâmetros do erro absoluto e reduzido da estimativa, tornando-a valiosa. Além disso indica qual a vizinhança ideal para a estimativa.

McBratney e Webster (1986) indicam que os modelos mais adequados para os mais variados tipos de situações, na maioria dos casos, serão esférico, exponencial ou gaussiano. A Figura 3.2 mostra o comportamento destes três modelos. Vieira et al. (2010) sugerem que o usuário escolha um destes três modelos segundo o comportamento de seus variogramas para pequenas distâncias (menor do que o alcance), faça o ajuste usando algum método de otimização dos parâmetros C_0 , C_1 e a , e submeta este modelo ao processo de validação cruzada pelo método “*Jack-knifing*”. Esta técnica pode ser utilizada para: avaliar a qualidade do método de estimação; definir o melhor número de vizinhos do ponto sendo estimado; ou avaliar o ajuste variograma teórico aos dados do variograma empírico (SOUZA, 1992). Embora trabalhoso, o método irá eliminar qualquer possibilidade de ajuste inadequado, porque seus resultados indicarão se o ajuste está dentro dos padrões estatísticos requisitados.

Os modelos teóricos para ajuste do variograma mais consagrados em aplicações envolvendo as ciências da terra são apresentados a seguir:

Modelo Exponencial:

O modelo exponencial é o que representa processos que tem a maior perda de semelhança com a distância. Dados de precipitação pluviométrica normalmente são ajustados por este modelo. Atinge o patamar exponencialmente, por isso possui apenas uma estrutura onde d é a máxima distância na qual o variograma é definido. O parâmetro a

é determinado visualmente como a distância após a qual o variograma se estabiliza. Este modelo é determinado pela Equação 3 como segue:

$$\gamma(h) = C_0 + C_1 [1 - \exp(-3 \frac{h}{a})] \quad 0 < h < d \quad (3)$$

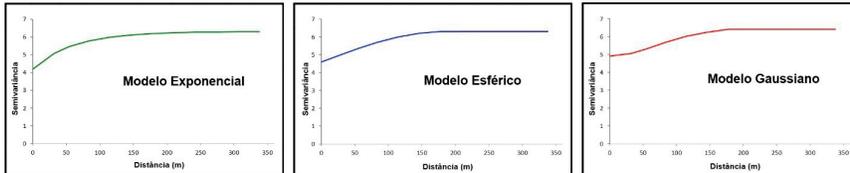


Figura 3.2. Forma gráfica dos três modelos teóricos mais utilizados no ajuste de variogramas típicos de variáveis de processos geomorfológicos.

Modelo Esférico:

O modelo esférico possui duas estruturas, uma com a distância entre zero e o alcance, e outra maior do que o alcance, onde atinge o patamar. O modelo esférico é linear até aproximadamente 1/3 do alcance e pode ser calculado pela Equação 4.

$$\gamma(h) = C_0 + C_1 \left[\frac{3}{2} \left(\frac{h}{a} \right) - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right] \quad 0 < h < a \quad (4)$$

$$\gamma(h) = C_0 + C_1 \quad h > a$$

Onde C_0 , C_1 e a são os parâmetros de ajuste, h a distância e $\gamma(h)$ o valor da semivariância para o modelo de ajuste.

Modelo Gaussiano:

O modelo gaussiano inicia com uma queda nos primeiros valores de semivariância para as menores distâncias, e após este início há um crescimento da semivariância até o alcance, onde é atingido o patamar. Tem um crescimento bastante lento no início, até atingir um ponto de inflexão antes do alcance. Os dados mais comuns de ocorrência de modelo gaussiano são cotas topográficas. Ele é definido pela Equação 5, representando os processos naturais mais contínuos que se tem conhecimento.

$$\gamma(h) = C_0 + C_1 [1 - \exp(-3 (\frac{h}{a})^2)] \quad 0 < h < d \tag{5}$$

3.3 Exemplo de Ajuste Manual

A título de auxiliar na compreensão do ajuste do variograma, serão detalhados a seguir os passos do processo de ajuste do modelo numa planilha eletrônica de cálculo (Figura 3.3):

1. Calcular a semivariância (γ) do conjunto de dados da variável no Vesper;
2. Levar arquivo “*Semivar.txt*” para planilha de cálculo, por exemplo no Excel;
3. Dispor as 3 colunas na seguinte ordem: Número de pares – Distância - Semivariância da variável em estudo;
4. Para que a linha do ajuste inicie na distância 0 (zero), inserir uma linha em branco antes da primeira distância e adicionar o valor 0 (zero) para a primeira linha da coluna;
5. Plotar um gráfico de dispersão da distância versus a semivariância;
6. Em campos vazios escolher adicionando manualmente os valores correspondentes aos parâmetros efeito pepita (C_0), variância estrutural (C_1) e alcance (a);

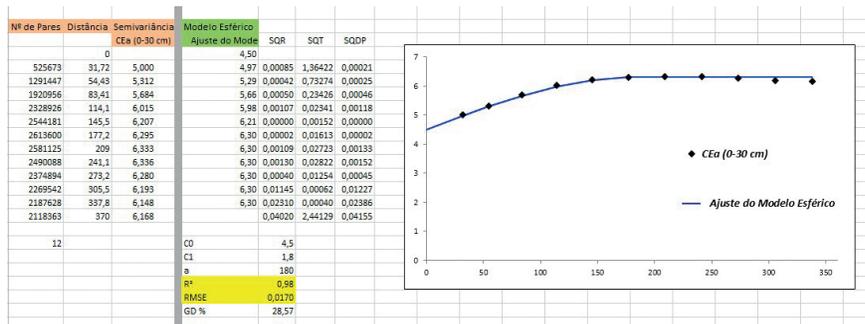


Figura 3.3. Resultado do procedimento para geração do variograma empírico e ajuste do modelo teórico em uma planilha eletrônica.

7. Inserir numa coluna a fórmula do ajuste e calcular para todas as distâncias o modelo que melhor represente os pontos do gráfico

do variograma conforme descritos anteriormente (i.e.: Item 3.1) os modelos para esférico, exponencial ou gaussiano;

8. Inserir a coluna de resultados do cálculo do modelo de ajuste no gráfico do variograma e formatar para linha. O variograma será ajustado ao modelo escolhido;
9. O Grau de Dependência (*GD*) espacial (se fraco <25%, moderado >25% até 75%, forte > 75%) pode ser calculado pela Equação 6 definida por Zimback (2001):

$$GD(\%) = C1 * 100 / (C0 + C1) \quad (6)$$

Para validar o ajuste do variograma, ou seja, para verificar se os parâmetros de ajuste e o modelo de ajuste escolhidos foram adequados, podem ser calculados alguns coeficientes como coeficiente de correlação (r^2) e raiz quadrada do erro médio (RMSE), sendo que, quanto mais próximo do 1 o r^2 for, e quanto menor o valor de RMSE, melhor o ajuste. Para isso é necessário definir as respectivas equações como segue:

1. Adicionar colunas para calcular a Soma de Quadrado do Resíduo (SQR). O SQR mede a soma dos quadrados dos resíduos, exclusivamente relacionados ao erro aleatório medida dentro dos grupos nas diferentes distâncias (Equação 7);

$$SQR = (\gamma(\text{modelo}) - \gamma(\text{variável}))^2 \quad (7)$$

2. Adicionar colunas para calcular a Soma de Quadrado Total (SQT). O SQT mede a variação geral de todas as observações (Equação 8);

$$SQT = (\gamma(\text{variável}) - \text{média } \gamma(\text{variável}))^2 \quad (8)$$

3. Adicionar colunas para calcular a Soma de Quadrado dos Desvios Ponderados (SQDP). O SQDP é a soma dos quadrados dos desvios de cada dado em relação à média do conjunto (Equação 9);

$$SQDP = \left(\frac{nPares}{\sum nPares} \right) * SQR \quad (9)$$

4. Adicionar colunas para calcular o Coeficiente de Determinação (r^2). O r^2 é definido como a relação que mede a proporção da variação total da variável dependente que é explicada pela variação da variável independente (Equação 10);

$$r^2 = 1 - \left(\frac{SQR}{SQT} \right) \tag{10}$$

5. Adicionar colunas para calcular a Raiz Quadrada do Erro Quadrático Médio (*RMSE*). O *RMSE* é a medida da magnitude média dos erros estimados (Equação 11);

$$RMSE = \left(\frac{1}{Contagem\ nPares} \right) * \sqrt{\sum SQDP} \tag{11}$$

Na Figura 3.4 são apresentados os ajustes dos modelos teóricos (i.e.: gaussiano, esférico e exponencial) ao modelo empírico dos dados de CE_a , juntamente com os respectivos cálculos de r^2 , *RMSE* e *GD* (%).

Nr de Pares	Distância (m)	Modelo Empírico (CEa 30 cm)				Modelo Gaussiano (4,9; 1,5; 170)				Modelo Esférico (4,6; 1,7; 180)				Modelo Exponencial (4,2; 2,1; 180)				
		SQR	SQT	SQDP	GD %	SQR	SQT	SQDP	GD %	SQR	SQT	SQDP	GD %					
0																		
525673	31,72	5,147	5,048764175	0,0097	1,7662	0,0024	5,044715082	0,0105	1,7662	0,0026	5,062280371	0,0072	24,1488	0,0018				
1291447	54,43	5,417	5,29711982	0,0144	1,1215	0,0088	5,347589039	0,0048	1,1215	0,0029	5,452300714	0,0012	26,6514	0,0008				
1920956	83,41	5,086	5,874478762	0,0002	0,6241	0,0002	5,897063986	0,0001	0,6241	0,0001	5,777029029	0,0083	30,6568	0,0075				
2328926	114,1	5,900	6,011666666	0,0125	0,3318	0,0137	5,999916446	0,0100	0,3318	0,0110	5,986428954	0,0075	34,5400	0,0082				
2544181	145,5	6,063	6,233396526	0,0290	0,1706	0,0349	6,212307841	0,0223	0,1706	0,0268	6,11419595	0,0026	37,1951	0,0031				
2613000	177,2	6,185	6,4	0,0462	0,0847	0,0570	6,299386162	0,0131	0,0847	0,0161	6,190452385	0,0000	39,2550	0,0000				
2581125	209	6,276	6,4	0,0154	0,0400	0,0187	6,3	0,0006	0,0400	0,0007	6,23551973	0,0016	39,2550	0,0020				
2490088	241,1	6,343	6,4	0,0032	0,0177	0,0038	6,3	0,0018	0,0177	0,0022	6,262235886	0,0065	39,2550	0,0077				
2374894	273,2	6,393	6,4	0,0000	0,0069	0,0001	6,3	0,0086	0,0069	0,0097	6,277882718	0,0133	39,2550	0,0149				
2269542	305,5	6,429	6,4	0,0008	0,0022	0,0009	6,3	0,0156	0,0022	0,0118	6,287080953	0,0201	39,2550	0,0216				
2187628	337,8	6,456	6,4	0,0031	0,0004	0,0032	6,3	0,0243	0,0004	0,0251	6,292463992	0,0267	39,2550	0,0276				
2118363	370	6,476	6,4	0,1346	4,1660	0,1437	6,3	0,1128	4,1660	0,1151	6,2951	388,7221	0,0952					
			CO	4,9							CO	4,6						
			C1	1,5							C1	1,7						
			a	170							a	180						
			R ²	0,97							R ²	0,97						
			RMSE	0,0316							RMSE	0,0281						
			GD %	23,4375							GD %	26,9841						
			CO	4,2							CO	4,2						
			C1	2,1							C1	2,1						
			a	180							a	180						
			R ²	0,95							R ²	0,95						
			RMSE	0,0022							RMSE	0,0022						
			GD %	59,3750							GD %	59,3750						

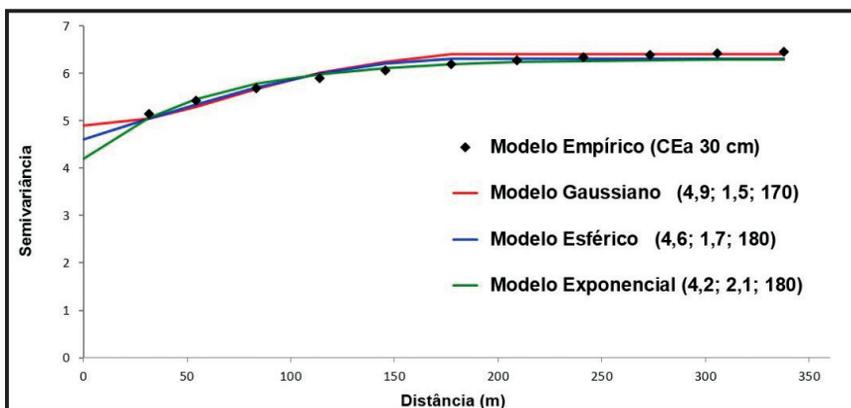


Figura 3.4. Ajuste de três modelos teóricos ao variograma empírico de CE_a , destacando os valores de melhor ajuste de cada modelo entre parênteses, sendo: (C_0 ; C_1 ; a).

3.4 Características da Dependência Espacial

Características específicas da dependência espacial pressupõem o uso de aplicações mais avançadas como processamento em lote, retirada de tendências, análise de anisotropia e cokrigagem não estão no escopo deste livro, bem como não são diretamente disponíveis na interface do Vesper. Estas análises exigem um nível mais avançado de conceitos da geoestatística e são melhor considerados em outros aplicativos disponíveis. Entretanto, é possível considerar estes tipos de análise no Vesper através do uso de linhas de comando no arquivo de Controle (i.e.: "**Control.txt**") como ilustrado na figura abaixo. Este arquivo pode ser utilizado tanto para krigagens repetitivas e sequenciais, por exemplo, para um acervo de dados temporais de um talhão ou para conjuntos de dados espacialmente contíguos de uma mesma variável.

Para avaliar o comportamento da variabilidade espacial de uma variável, os variogramas são elaborados experimentalmente e submetidos à análise de suas características estruturais. Huijbregts (1975) ilustra as propriedades estruturais do variograma como sendo: suporte, zona de influência, estruturas superpostas, anisotropia, continuidade espacial ou comportamento da variável próxima a origem, correção regionalização. A análise destas propriedades exige funções avançadas que não estão disponíveis no Vesper e, portanto, não terão uma apresentação aprofundada. Estas funções envolvem conceitos fundamentais para uma visão mais abrangente do uso das técnicas, e por isso uma breve introdução é feita a seguir sem entrar no mérito das respectivas formulações matemáticas.

Anisotropia

A anisotropia é uma característica comum nos processos de dispersão de elementos dos recursos naturais. O termo indica que a variabilidade ou distribuição espacial de tais elementos ocorre mais intensamente numa direção e menos intensamente em outra direção, não apresentando uma dependência espacial radialmente homogênea. Considerando o mapeamento de teores de potássio, é improvável que esta propriedade se disperse igualmente em todas as direções. A estatística clássica é limitada para tratar apropria-

damente os efeitos anisotrópicos, pois não considera a estrutura de autocorrelação espacial da propriedade. Modelos mais adequados para este objetivo são propostos na geoestatística (CAMARGO et al., 2004).

Aplicações da anisotropia tratam a dependência espacial de acordo com variação numa determinada orientação. Ou seja, são modelos ajustados a um variograma direcional. No caso da variação isotrópica (i.e.: omnidirecional), o variograma amostral depende somente da distância de separação, considerando o mesmo comportamento da variabilidade em todas as direções.

A representação isotrópica caracteriza uma distribuição espacial que pode ser modelada por um único modelo teórico, ou seja, o ajuste do variograma será o mesmo para qualquer direção analisada (Figura 3.5.a). Em contraposto, se os variogramas não são iguais para todas as direções, a distribuição é denominada anisotrópica. Se a anisotropia é observada, e ainda reflete um mesmo patamar com diferentes alcances para o mesmo modelo, então ela pode ainda ser denominada de anisotropia geométrica (Figura 3.5.b). Os modelos de anisotropia buscam dependências espaciais que variam conforme a orientação através da observação dos variogramas obtidos para pares de pontos alinhados em diferentes direções. As convenções direcionais usadas na geoestatística são ilustradas na Figuras 3.6.

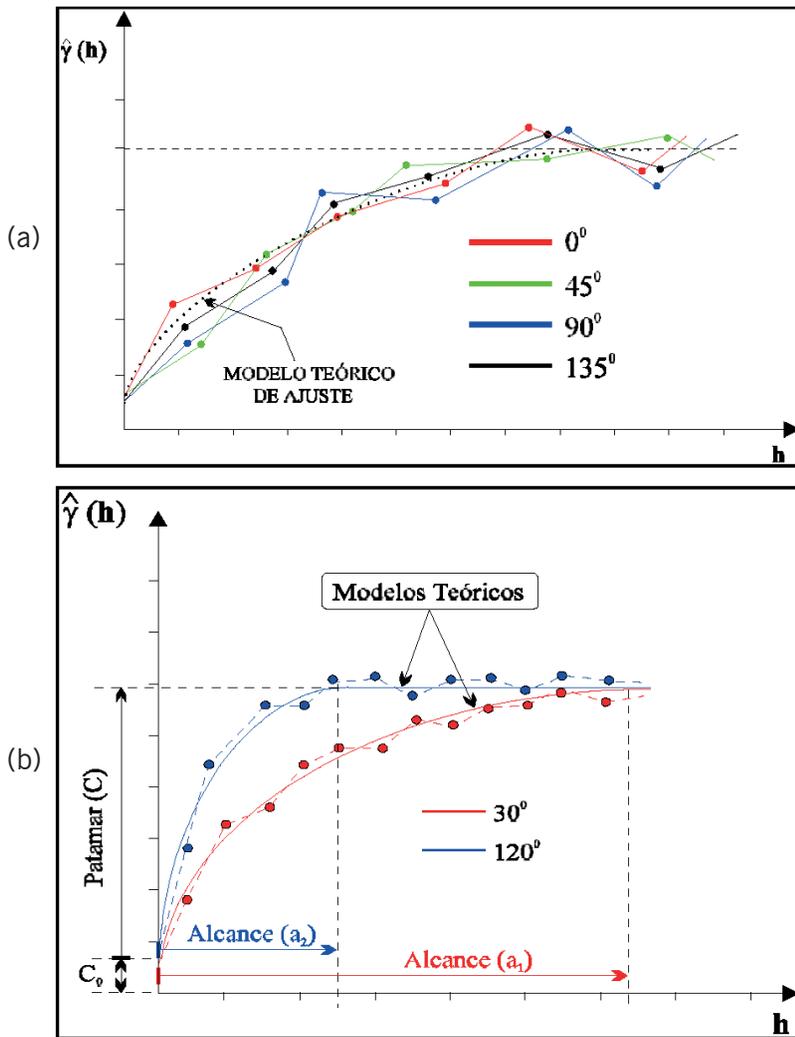


Figura 3.5. Comportamentos da autocorrelação espacial considerando diferentes direções: a) isotrópico, com ajuste único para todas as direções; e b) anisotrópico geométrico, com ajustes em diferentes alcances para direções distintas.

Fonte: adaptado de Pereira et al. (2010).

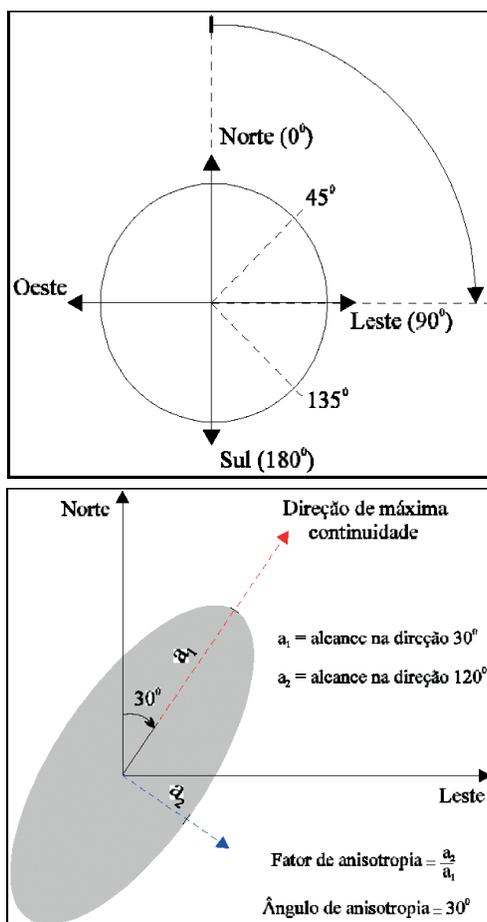


Figura 3.6. Convenções direcionais de anisotropia usadas na geoestatística.
Fonte: adaptado de Pereira et al. (2010).

Tendência

Para a obtenção de um ajuste preciso do variograma, é suposto que a variável regionalizada tenha um comportamento fracamente estacionário, onde os valores esperados, assim como sua covariância espacial, sejam os mesmos dentro de uma determinada área. Nem sempre o comportamento espacial de uma variável tem a característica estacionária, onde a média permanece constante. Quando isto não ocorre, a variável apresenta uma deriva, uma tendência regional (i.e.: "trend" ou "drift"). O que significa que o valor médio da variação

esperada dentro de uma área de vizinhança varia sistematicamente. A Figura 3.7 demonstra que o conceito de estacionaridade está diretamente relacionado com a escala das observações. Se considerarmos o traço como um todo, então as hipóteses estacionárias de segunda ordem serão observadas. Entretanto, o monitoramento intensivo entre os pontos C e D estará refletindo uma tendência não estacionária bem definida. Já no caso do intervalo entre A e B existe a estacionaridade de segunda ordem na mesma escala considerada na seção CD.

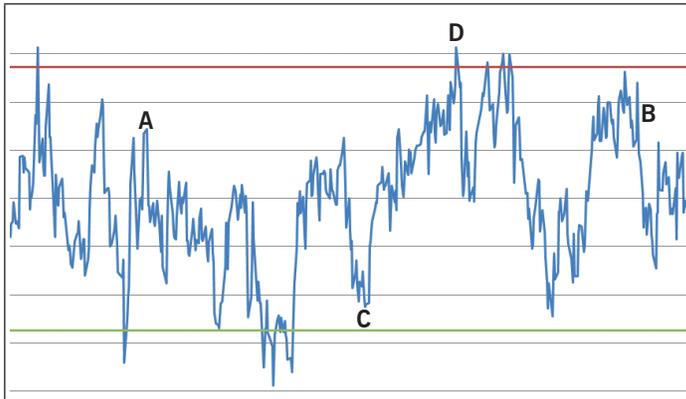


Figura 3.7. Gráfico da relação entre os conceitos de estacionaridade e escala amostral das observações, sendo um processo estacionário de A a B e não estacionário de C a D.

Fonte: adaptado de Pereira et al. (2010).

Em aplicações com ocorrências de não estacionaridade, deve-se considerar a remoção de tendências nos valores das variáveis (CARMARGO et al., 2004), como no caso de distribuições não normais que são comuns para atributos do solo. Nestes casos, formulações numéricas aplicadas aos valores das variáveis possibilitam transformar distribuições não normais em distribuições aproximadamente normais (e.g.: funções logarítmicas e de raiz quadrada). Outro tipo de correção de tendências pode ser aplicado quando a não estacionaridade da média é observada. Nestes casos, a tendência pode ser removida em função das coordenadas (i.e.: X e Y), ajustando-se uma equação polinomial de ordem “n” pelo método dos mínimos quadrados, onde, posteriormente, o resíduo servirá para o ajuste do variograma da variável em estudo (LANDIM; CORSI, 2001).

Camargo et al. (2004) sugerem que a variação espacial de uma variável regionalizada pode ser expressa pela soma de três componentes: a) um ruído aleatório ou erro residual; b) um componente aleatório espacialmente correlacionado; e c) um componente estrutural correlacionado e associado a um valor médio ou a uma tendência constantes. Estes comportamentos são observados na forma geral de variogramas típicos (Figura 3.8) que indicam estas estruturas de variação espacial, respectivamente sendo: a) variogramas aleatórios (i.e.: efeito pepita puro), onde não se observa uma covariância espacial entre os valores e a geoestatística não se aplica (Figura 3.8.a); b) variogramas ideais, onde se observa um patamar bem definido demonstrando um processo estacionário de segunda ordem (Figura 3.8.b); e c) variogramas com tendência, sem um patamar claro que pode indicar um processo não estacionário (Figura 3.8.c).

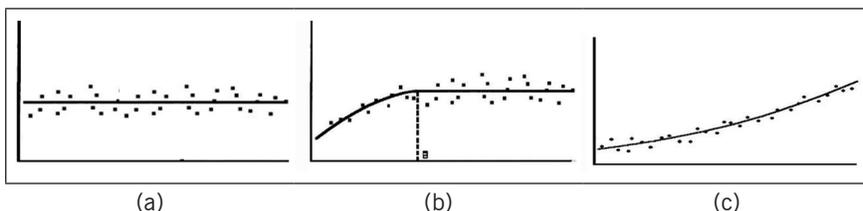


Figura 3.8. Variogramas típicos dos três principais componentes na estrutura de variação de uma variável regionalizada: a) efeito pepita puro; b) componente aleatório espacialmente correlacionado; e c) componente correlacionado e associado a uma tendência.

3.5 Interpolação por Krigagem

A krigagem é um método de estimação geoestatística com denominação sugerida por Matheron em homenagem aos trabalhos pioneiros de Daniel Krige, ao introduzir o uso de médias móveis para evitar a superestimação sistemática de reservas em mineração. (SOARES, 2006). O objetivo da krigagem é estimar valores para qualquer local não observado. Considera uma estimativa sem tendência, com variância mínima e onde a média não deve superestimar ou subestimar valores, minimizando assim a incerteza da estimativa. O método foi concebido sob a hipótese de que a variável regionalizada resultava de um

processo estocástico estacionário de segunda ordem, dando origem as krigagens denominadas simples e ordinária. A krigagem simples considera uma média conhecida, sendo esta constante em todo o domínio, enquanto a krigagem ordinária assume uma média constante, ainda que desconhecida. A krigagem calcula um valor no ponto central de cada célula da grade de interpolação. Este valor estimado está condicionado a estrutura espacial da variável observada (variograma empírico) e aos parâmetros da função de correlação espacial (variograma teórico) ajustada para estes dados.

Segundo Vieira (2000), existem interpoladores que não consideram a dependência espacial, como o inverso do quadrado da distância, média móvel ponderada e outros que interpolam por meio de ajustes de polinômios bidimensionais. Entretanto, estes interpoladores não consideram a estrutura da variação espacial dos valores medidos revelado pelo variograma e, portanto, não determinam com exatidão a variação dos dados interpolados. Na krigagem, o procedimento é semelhante ao de interpolação por média móvel ponderada, exceto que os pesos associados aos vizinhos mais próximos são determinados pelo ajuste do variograma empírico.

3.6 Tipos de Krigagem

A interpolação por krigagem é uma técnica que deriva diferentes métodos de estimação, incluindo: krigagem simples, krigagem ordinária, krigagem universal, krigagem disjuntiva e cokrigagem; entre outras de caráter não lineares como: krigagem indicativa e krigagem fractal (CAMARGO, 1998; SANTOS, 2010). Estas variações do método estendem ou adaptam os algoritmos básicos que não se aplicavam a todas as estruturas de variação espacial. Aplicam-se em função de processos específicos como: estacionários (e.g.: krigagem simples e ordinária), não estacionários (e.g.: krigagem universal e funções intrínsecas) e multivariados (e.g.: cokrigagem). Ainda que o Vesper considere apenas a krigagem ordinária, um resumo dos métodos mais comumente utilizados são apresentados.

Krigagem Ordinária

A krigagem ordinária é um método de estimação linear para uma variável regionalizada que satisfaça a hipótese intrínseca. O método é definido por um algoritmo que considera que os valores de uma variável regionalizada apresentam uma média constante, porém desconhecida. Destina-se a calcular ponderadores ótimos, que minimizem a variância do erro de estimação (DEUTSCH, 1996). A krigagem ordinária é a técnica de interpolação geoestatística mais difundida (LANDIM et al., 2002; SOARES, 2006), já implementada em módulos de análise espacial de vários ambientes SIG. Considerando a dificuldade em quantificar o erro e a variância para os pontos estimados, dado o desconhecimento dos valores reais, a krigagem ordinária faz uso de uma função aleatória que permite atribuir pesos às amostras usadas nas estimativas. Por isso, no algoritmo de krigagem ordinária não há a necessidade de se conhecer o valor da média a priori; isto é, o valor da média é constante, mas como não entra no cálculo, pode continuar desconhecido.

A krigagem ordinária é considerada por Sturaro (1988) como o melhor estimador linear sem viés, em função das seguintes características: a) Linear - as estimativas são feitas através de uma combinação linear dos dados; b) Sem viés - o método objetiva que o erro residual médio seja igual a zero; e c) Melhor estimador - o método objetiva minimizar a variância dos erros. Como em outras formas de krigagem, ela pode ser executada com estimativas por ponto ou por bloco, onde os resultados por bloco apresentam uma maior suavização na superfície estimada.

Krigagem por Ponto ou por Blocos

Os mesmos cálculos da krigagem ordinária podem ser aplicados para estimativas em um ponto, ou em uma área (i.e.: estimativas por blocos). A krigagem por pontos é utilizada para estimar uma variável de interesse em um ponto não observado, a partir dos valores conhecidos de observações vizinhas. A krigagem pontual é uma técnica muito aplicada na representação de dados geofísicos (e.g.: mapas de isovalores e superfícies tridimensionais). Em geral, modelando variações mais abruptas de valores absolutos, sendo adequada para alguns processos da AP.

A krigagem por blocos é uma técnica que tem origem na estimativa de teores médios em painéis, ou blocos, de mineração. Difere da estimativa pontual, à medida que áreas e volumes maiores são representados por pontos amostrais coletados em função da variabilidade natural do depósito sendo avaliado. Onde a estimativa de apenas um ponto central esta unidade (i.e.: área ou volume) não será suficiente para representá-la. A técnica considera uma região com uma área conhecida "A" e um ponto central X_0 ; para a qual, as variâncias entre os pontos amostrados (i.e.: $X_1, X_2, X_3, \dots, X_n$) e o ponto X_0 são substituídas pela média das variâncias entre os pontos amostrados e os pontos que ocorrem dentro da área "A". Obviamente, uma krigagem com blocos de 0 x 0 m, equivale a uma krigagem por ponto. Em relação ao uso da técnica em AP, esta estimativa é uma opção a ser avaliada em dados de sensores de produtividade, considerando-se a variação da produtividade dentro de uma área definida pela largura da plataforma no implemento agrícola e a distância percorrida durante o intervalo de leitura (i.e.: 1 segundo). O erro de estimativa associado à krigagem por blocos será menor do que aquele associado à krigagem pontual. Em contrapartida, a estimativa por blocos resulta em mapas mais suavizados (i.e.: valores mais homogêneos). Onde, quanto maior o bloco, mais suavizada fica a representação.

Krigagem Simples

A krigagem simples é utilizada quando estatisticamente se assume que a média é conhecida e constante para toda a área (i.e.: ajuste do variograma global). Difere da krigagem ordinária que considera uma média flutuante ou móvel por toda a área (i.e.: ajuste do variograma local). Estas presunções são muito restritivas para modelar, com realismo, processos geoambientais e, por isso, é uma técnica pouco utilizada. Em geral, transformações de normalização dos dados ou procedimentos de extração de tendências são necessários para viabilizar esta técnica.

Krigagem Universal

É um método similar ao da krigagem ordinária, aplicado quando a variável regionalizada não é estacionária mas apresenta uma tendência e seus resíduos contém a hipótese intrínseca. Esta ocorrên-

cia indica uma média que não é constante e, conseqüentemente, um variograma que não viabiliza a modelagem precisa da estrutura de correlação espacial dos dados considerados. Uma variável regionalizada não estacionária pode ser considerada como constituída por dois componentes: 1) a tendência, que consiste no valor médio esperado dessa variável, dentro de uma certa vizinhança, que varia sistematicamente; e 2) o resíduo que é a diferença entre os valores reais e o valor do deslocamento causado pela tendência. Nestes casos, a krigagem é executada sobre os resíduos. O que impõem o ajuste do variograma sobre os resíduos, para que se possa descrever a estrutura da variação sem tendência. De forma mais direta, se a variável regionalizada representar um processo não estacionário, trabalha-se sobre a estacionariedade residual da variável. Na krigagem universal, o ajuste local é feito mediante um polinômio de ordem “ n ” em função das coordenadas dos pontos estimados (i.e.: X e Y).

Cokrigagem

A cokrigagem é espécie de extensão da krigagem, a qual permite estimar uma variável a partir de suas próprias informações em conjunto com as informações disponíveis sobre outras variáveis espacialmente correlacionadas. É um procedimento de interpolação no qual diversas variáveis regionalizadas podem ser estimadas em conjunto, com base numa correlação espacial multivariada entre si. Aplica-se quando duas ou mais variáveis apresentam uma alta correlação em seus valores numéricos e tenham sido observadas em locais próximos dentro de uma mesma área. Este método é comumente utilizado quando a variável de interesse (i.e.: “primária”) apresenta uma menor densidade amostral, ou correlação espacial, em relação às demais variáveis de apoio (i.e.: “secundárias”). Onde, para cada local amostrado, obtém-se um vetor de valores em lugar de um único valor. A cokrigagem busca melhorar a estimativa da variável primária utilizando outras mais densamente amostradas; ainda que, em geral, apenas uma variável secundária seja usada.

Krigagem Indicativa ou Indicadora

Este tipo de interpolação utiliza procedimentos não lineares para modelar a variabilidade dos atributos. Possibilita a inferência de uma

aproximação discretizada do modelo da distribuição de probabilidade para a modelagem da incerteza sobre seus valores (LIMA, 2006). Uma vantagem importante destes métodos é a possibilidade de se modelar dados temáticos qualitativos, além dos dados de natureza numérica. Assim, pode-se trabalhar com propagação de incertezas para modelos computacionais que envolvam atributos numéricos e temáticos. A krigagem indicativa é aplicada de forma discretizada para encontrar a função de distribuição acumulada de cada ponto a ser estimado. É um método não paramétrico que não considera nenhum tipo de distribuição de probabilidade, a priori, para a variável aleatória. Em contrapartida, ela possibilita a construção de uma aproximação discretizada da função de distribuição que pode ser usada diretamente na estimativa de valores característicos da distribuição, tais como: quartis, valores esperados e variâncias. Portanto, não fica restrita à modelagem de distribuições simétricas, como, por exemplo, a gaussiana (LIMA, 2006). Estes procedimentos possibilitam modelar atributos com alta variabilidade espacial sem o rigor da análise exploratória que busca filtrar amostras com valores discrepantes (i.e.: “*outliers*”) de uma tendência. Entretanto, a krigagem indicadora requisita um alto grau de embasamento teórico na interatividade com suas funções, exigindo o ajuste de um variograma para cada valor de corte considerado.