

## A GEOESTATÍSTICA E OS SISTEMAS DE INFORMAÇÕES GEOGRÁFICAS

**Suzana Druck**

Embrapa/CPAC-Centro de Pesquisa Agropecuária dos Cerrados  
Caixa Postal 08223, 73301-970, Planaltina, DF, Brasil.  
[suzana@cpac.embrapa.br](mailto:suzana@cpac.embrapa.br)

Recebido 10 de novembro de 2001; revisado 19 de dezembro; aceito 5 de janeiro de 2002

**RESUMO** - Esse trabalho objetiva enfatizar a importância da associação dos procedimentos geoestatísticos às técnicas de análise espacial. Tradicionalmente essas técnicas estão implementadas nos Sistemas de Informações Geográficas. A associação da geoestatística aos SIG's vem se tornando relevante nos procedimentos de modelagem, pois acrescentam aos resultados obtidos medidas de incerteza decorrente dos procedimentos inferenciais (ou de inferências). O trabalho apresenta os principais fundamentos da geoestatística e enfatiza os procedimentos de *krigeagem* mais utilizados como a *krigeagem* simples e ordinária.

**Palavras-Chave** – geoestatística, *krigeagem*, análise espacial.

**ABSTRACT** - This work aims to emphasize the association of geostatistics procedures to spatial analysis, which are generally used in Geographical Information Systems (GIS). The associations between geostatistics procedures and GIS have become relevant in modeling procedures because they aggregate uncertainty measures derived from inferential procedures. This work summarizes geostatistical structure and presents the kriging procedures that are mostly applied as simple and ordinary kriging.

**Keywords** – geostatistics, kriging, spatial analysis

## INTRODUÇÃO

Os Sistemas de Informações Geográficas (SIGs), devido à sua crescente utilização, vem incorporando novos princípios e procedimentos de análise espacial. Esse esforço de desenvolvimento deve-se ao consenso existente de que o futuro sucesso da tecnologia dos SIGs depende da potência de seus procedimentos na transformação dos dados observados principalmente no que se refere a capacidade dos SIGs desenvolver ferramentas de modelagem espacial

A literatura recente mostra que há uma crescente integração de modelos, sejam eles determinísticos ou estocásticos, aos Sistemas de Informações Geográficas principalmente quando estes são aplicados a fenômenos ambientais. Hoje é possível implementar funções complexas aos SIGs e tem havido um significativo esforço neste sentido (De Roo *et al*, 1992). Entretanto, essas inovações, que são relevantes na análise de dados espaciais, não motivaram a implementação de procedimentos que associados a elas, pudessem medir a qualidade da informação gerada (Burrough 1992). Verificar a performance de um modelo escolhido na representação de um fenômeno espacial, comparar comportamentos de diferentes modelos utilizados na explicação de um mesmo processo ou otimizar procedimentos na obtenção de um dado objetivo torna-se difícil se os SIGs não processam medidas indicadoras da qualidade do que foi produzido. Assim, se a associação de modelos ambientais aos SIGs representa um crescente avanço na automatização e na análise da informação espacial, simultaneamente, há um desconhecimento sobre qualidade da informação fornecida aos usuários. Uma das razões para esta aparente contradição é a atenção insuficiente dos modelos utilizados para a representação da variabilidade espacial. A maior parte dos trabalhos que utilizam SIG armazenam

representações de um fenômeno contínuo, como são os processos ambientais, de forma discreta, não favorecendo a uma análise da continuidade espacial, nem a verificação de seus efeitos na imprecisão dos fenômenos modelados.

A representação convencional da variabilidade espacial nos SIGs reflete as hipóteses da cartografia convencional em que a representação da variação contínua de múltiplas propriedades que compõem determinado fenômeno é feita em termos de limites precisos e na homogeneidade das suas variações espaciais em uma determinada área. Desta forma, as modificações de fenômenos contínuos são representadas como abruptas, só ocorrendo a partir de um limite estabelecido sem etapas intermediárias ou mudanças graduais.

Esta representação é uma excessiva simplificação da verdadeira natureza da variação de alguns fenômenos espaciais e, uma vez que essa informação armazenada de forma resumida começa a ser desagregada, para ser combinada de diferentes formas uma significativa propagação de erros no produto gerado pode ser evidenciada (Goodchild *et al*, 1992).

A importância de associar aos resultados obtidos nos SIGs, a distribuição da propagação de erros efetuados é verificada por Arbia (1993), que a esse respeito argumenta:.. *”o que é nos SIGs sua maior força, isto é, a possibilidade de lidar e combinar informações de diversos tipos e fontes, pode ser também sua maior fraqueza, pois em algumas circunstâncias os procedimentos realizados podem enfatizar de tal forma os erros gerados que se corre o risco de se ter resultados sem informações substanciais”*.

A moderna tecnologia dos SIG's vem evoluindo no sentido de melhor representar o contínuo dos dados ambientais. O modelo de dados *raster* nos SIG's particiona o espaço em pequenos *pixels* que compõem uma grade regular, assim, quanto menor o *pixel* mais ele é homogêneo e maiores são os esforços necessários de armazenamento. Nesse modelo, a definição do número de *pixels* de uma grade depende do equilíbrio entre a capacidade de armazenamento dos SIG's e a homogeneidade interna requerida pela escala da variabilidade espacial. Cada uma dos atributos considerados tem um plano de representação possibilitando uma interpretação específica da distribuição espacial de cada uma dessas propriedades. O problema que se apresenta é que geralmente somente uma amostra de locais observados compõe a informação disponível e é preciso definir um modelo de representação espacial que possa associar valores do atributo a todos os pontos da grade definida na estrutura raster. Assim pontos não amostrados devem também ter os valores associados. Definida dessa forma a representação *raster* está intimamente ligada à interpolação espacial uma vez que é necessário estimar, a partir de uma amostra observada, valores de atributos em pontos não amostrados.

Os procedimentos geoestatísticos possibilitam a representação da variabilidade espacial através de estimadores que consideram a estrutura de dependência espacial, seus algoritmos utilizam inferências não somente variáveis quantitativas como também outras formas de informação não-quantitativa, que representam também uma grande parte da informação utilizada nos SIG's.

Assim, uma alternativa para a representação espacial convencional em fenômenos contínuos baseia-se na potência da computação moderna, na tecnologia de banco de dados, que possibilita o armazenamento de informações

pontuais georreferenciadas e na utilização dos procedimentos geoestatísticos. Dessa forma, a associação de procedimentos de geoestatística aos SIG's é uma decorrência natural da evolução dos SIG's na implementação de representações espaciais com hipóteses mais próximas do contínuo dos fenômenos ambientais. E o crescente direcionamento dos SIG's em processos de modelagem, nos quais a geoestatística e os SIG's são parceiros naturais, alicerçam as bases dessa associação.

### FUNDAMENTOS DE GEOESTATÍSTICA

A geoestatística envolve um conjunto de procedimentos de análise e inferência dos fenômenos espaciais, que apresentem uma dependência espacial expressa, por exemplo, numa função de autocorrelação no espaço. Esses procedimentos utilizam um modelo de autocorrelação espacial definido a priori e objetivam representar a variabilidade espacial de um atributo considerado em uma superfície contínua. São, portanto, procedimentos que podem ser aplicados a vários tipos de fenômenos, tais como concentração de poluentes, variação do teor de zinco no solo, preço do petróleo no tempo, etc. apesar de ter sua origem na área de mineração (Krige, 1951).

A estrutura das técnicas geoestatísticas está fundamentada na *Teoria das Variáveis Regionalizadas*, desenvolvida por Matheron (1971). A variável regionalizada é uma variável distribuída no espaço (ou tempo) e a teoria geoestatística associa o conceito de variável regionalizada aos procedimentos estocásticos através do conceito de Função Aleatória (ou processo aleatório, ou campo aleatório, ou processo estocástico). Nesse caso os valores observados ou não a cada ponto do espaço  $x$ , em uma determinada região A, são compre-

endidos como realizações de uma variável aleatória e o conjunto de variáveis aleatórias em  $A$ ,  $\{Z(x); x \in A\}$  como uma Função Aleatória ou seja uma F.A. caracterizada por sua função de autocorrelação.

Portanto, na utilização de procedimentos geoestatísticos a análise da dependência estocástica, que infere o modelo da estrutura de dependência espacial é relevante. Essa análise é denominada de análise estrutural ou modelagem do variograma e antecede as inferências em localizações não amostradas realizadas pela “*Krigeagem*”.

Entretanto a análise estrutural não é o único procedimento que antecede a *krigeagem*, resumidamente, quando se utilizam técnicas geoestatísticas os seguintes passos são necessários:

- análise exploratória dos dados que evidencia as principais características da amostra de dados observados. Alguma delas, se existentes nos dados, podem dificultar ou interferir no procedimento inferencial da estrutura de autocorrelação espacial. Esses aspectos são: distribuições assimétricas com a presença de dados discrepantes, a formação de “clusters espaciais” espaciais, ou seja, agrupamentos de dados no espaço etc.
- análise estrutural (cálculo e modelagem do variograma) que infere um variograma experimental e ajusta a um modelo permissível;
- validação que analisa a performance na *krigeagem* quando determinado modelo de variograma é utilizado. A validação consiste em inferir, utilizando um modelo de autocorrelação, nos locais já observados e por

meio de uma análise exploratória nos resíduos para verificar o bom ajuste; e

- realização de inferências (*Krigeagem* ou Simulação) que se processa através da escolha do procedimento inferencial de *krigeagem*.

A variação espacial de uma variável aleatória espacial pode ser expressa pela soma de dois componentes:

- uma componente determinística, que representa um valor médio ou uma tendência observada em larga escala; e
- uma componente aleatória, que representa a variabilidade em pequena escala, espacialmente correlacionada;

Se  $\mathbf{x}$  representa uma posição em uma, duas ou três dimensões, então o valor da variável  $Z$ , em  $\mathbf{x}$ , é dado por:

$$Z(x) = m(x) + e(x) \quad (1)$$

Onde  $m(x)$  é uma função determinística que descreve a componente estrutural de  $Z$  em  $\mathbf{x}$  e  $e(x)$  é um termo estocástico, que varia localmente e é dependente espacialmente.

## **ESTRUTURA DE DEPENDÊNCIA ESPACIAL**

Diferente de outras técnicas de inferência, a *Krigeagem* requer que hipóteses de estacionariedade sejam assumidas para que a estrutura de covariância espacial possa ser inferida, uma vez que em cada ponto observado existe uma única observação (Burrough, 1987, David, 1977).

A estacionariedade de segunda ordem considera que a média ou tendência é constante sobre determinada região e a covariância somente em função do vetor  $\mathbf{h}$ , ou seja:

$$E[Z(x)] = m(x) = \mathbf{m} \quad (2)$$

$$Cov[Z(x), Z(x+h)] = C(h) \quad (3)$$

Entretanto a estacionariedade de segunda ordem, para alguns fenômenos, é restritiva, principalmente quando a variância não converge, isto é,  $Var[Z(x)] = \infty$ . Assim para tornar os modelos de inferências mais abrangentes uma estacionariedade mais fraca que a anterior é definida que é a estacionariedade intrínseca.

A estacionariedade intrínseca estabelece que os incrementos  $z(x) - z(x+h)$  tem esperança zero e variância somente em função de  $\mathbf{h}$ .

$$E[z(x) - z(x+h)] = 0 \quad (4)$$

$$Var[z(x) - z(x+h)] = 2\mathbf{g}(h) \quad (5)$$

A função  $\mathbf{g}(h)$  é chamada de semivariograma e pode ser definida considerando (4) como:

$$\mathbf{g}(h) = \frac{1}{2} E[Z(x) - Z(x+h)]^2 \quad (6)$$

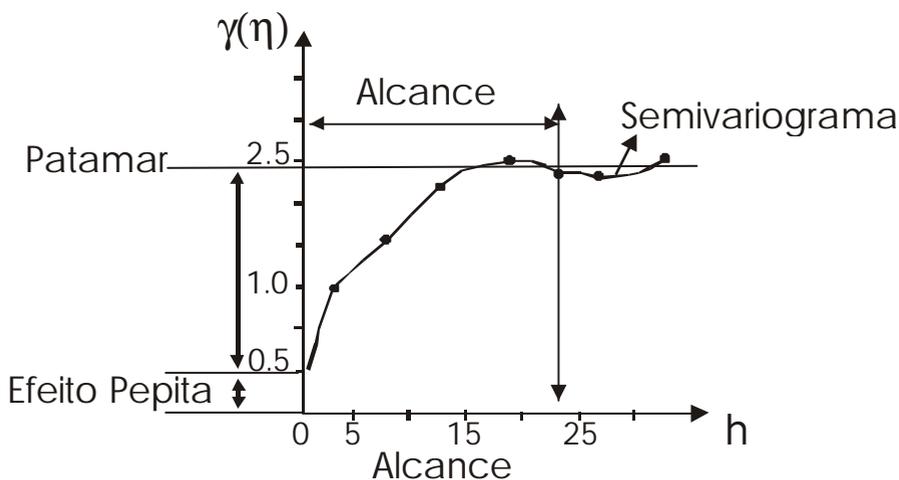
Se as condições especificadas pela hipótese de estacionariedade são contempladas, então o semivariograma pode ser estimado a partir do dado amostral:

$$\hat{g}(\vec{h}) = \frac{1}{2N(h)} \sum_{i=1}^n [z(x_i + h) - z(x_i)]^2 \quad (7)$$

onde  $N(\mathbf{h})$  é o número de pares de pontos amostrais separados pelo vetor distância  $\mathbf{h}$ .

O gráfico  $\gamma(\mathbf{h})$  versus  $\mathbf{h}$  é conhecido como semivariograma experimental. A **Figura 1** ilustra um semivariograma experimental com características muito próximas do ideal.

Seu padrão representa o que, intuitivamente, se espera de dados de campo, isto é, que as diferenças  $\{Z(\mathbf{x}_i) - Z(\mathbf{x}_i + \mathbf{h})\}$  decresçam à medida que  $\mathbf{h}$  diminui. É esperado que observações mais próximas geograficamente tenham um comportamento mais semelhante entre si do que aquelas separadas por maiores distâncias. Desta maneira, é esperado que  $\gamma(\mathbf{h})$  aumente com a distância  $\mathbf{h}$ .



**Figura 1** - Parâmetros do variograma.

Os parâmetros do semivariograma podem ser observados diretamente da

**Figura 1.** Esses parâmetros são:

- *Alcance (a)*: distância dentro da qual as amostras apresentam-se correlacionadas espacialmente. Na **Figura 1**, o alcance ocorre próximo de 25m.
- *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,
- *Efeito Pepita (C<sub>0</sub>)*: por definição,  $\gamma(\mathbf{0})=0$ . Entretanto, na prática, à medida que **h** tende para 0 (zero),  $\gamma(\mathbf{h})$  se aproxima de um valor positivo chamado *Efeito Pepita (C<sub>0</sub>)*. O valor de C<sub>0</sub> revela a descontinuidade do semivariograma para distâncias menores do que a menor distância entre as amostras. Parte desta descontinuidade pode ser também devida a erros de medição (Isaaks & Srivastava, 1989), mas é impossível quantificar se a maior contribuição provém dos erros de medição ou da variabilidade de pequena escala não captada pela amostragem.
- *Contribuição (C<sub>1</sub>)*: é a diferença entre o patamar (C) e o Efeito Pepita (C<sub>0</sub>).

O semivariograma experimental fornece valores experimentais do semivariograma para um número finito de *lags* em diferentes direções. Funções contínuas devem ser ajustadas a esses valores experimentais para que os valores de semivariogramas possam ser deduzidos para qualquer *lag* nos procedimentos inferências.

O procedimento de ajuste, do semivariograma experimental 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.

Resumidamente, os modelos permissíveis básicos de ajuste estão divididos em dois tipos (Isaaks e Srivastava, 1989): modelos com patamar e modelos sem patamar. Modelos do primeiro tipo são referenciados na geoestatística como modelos transitivos. Exemplos de modelos transitivos são: Esférico, Gaussiano e Exponencial.

*Esférico*

$$\mathbf{g}(h) = \begin{cases} 1.5 \frac{h}{a} - 0.5 \left( \frac{h}{a} \right)^3 & h \leq a \\ C & h. > a \end{cases}$$

*Gaussiano*

$$\mathbf{g}(h) = C \left( 1 - \exp \left( - \frac{h^2}{a^2} \right) \right)$$

Onde  $C > 0$  e  $a > 0$  são os dois parâmetros desse modelo. Esse modelo alcança o patamar  $C$  assintoticamente e na prática amplitude  $a$  é calculada como  $a \approx \frac{7a}{4}$ .

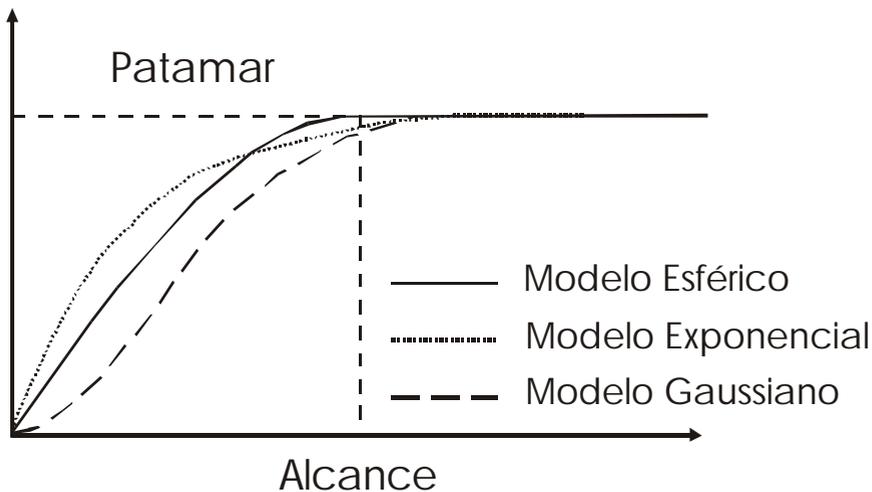
*Exponencial*

$$\mathbf{g}(h) = C \left( 1 - \exp \left( - \frac{h}{a} \right) \right)$$

Onde,  $C > 0$  e  $a > 0$ . O alcance desse modelo é assintótico a  $a$  e aproximadamente igual a  $a \approx 3a$ . A **Figura 2** apresenta esses modelos.

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. Como exemplo desse último pode-se citar o modelo Potência.

Uma vez inferido o modelo semivariograma melhor ajustado ao fenômeno em estudo o procedimento de krigagem deve ser escolhido e aplicado.



**Figura 2** - Modelos de variogramas transitivos.

## KRIGEAGEM

Os modelos de variabilidade espaciais podem ser definidos como soma de dois fatores

$$Z(x) = m(x) + \mathbf{e}(x) \quad (8)$$

Os procedimentos de krigagem linear se dividem conforme a hipótese considerada para a média ou tendência  $m(x)$ . Assim, na krigagem simples essa média ou tendência é constante e conhecida à priori. Na krigagem ordinária ela é constante desconhecida e na krigagem universal ou krigagem não estacionária a média flutua sobre toda região e pode ser expressa como:

$$m(x) = x^T \mathbf{b} \quad (9)$$

onde  $x$  é um vetor de funções conhecidas e  $\mathbf{b}$  um conjunto de parâmetros desconhecidos a serem estimados.

Este trabalho focaliza as técnicas de Krigagem Simples e Ordinária, uma vez que são as mais utilizadas.

Quanto aos métodos de inferência de Krigagem Simples e Ordinária, seus estimadores são obtidos conforme as equações (10) e (11) respectivamente (Journel, 1988):

$$\hat{Z}_{xo} = m + \sum_{i=1}^n I_i [Z(x_i) - m] \quad (10)$$

$$\hat{Z}_{xo} = \sum_{i=1}^n I_i Z(x_i) \quad \text{com} \quad \sum_{i=1}^n I_i = 1 \quad (11)$$

Nas equações (8) e (9) o que difere o estimador de Krigagem Simples do estimador de Krigagem Ordinária, é que este último não requer o prévio conhecimento da média amostral  $m$ .

A grande diferença entre a Krigagem e outros métodos de inferência são a maneira como os pesos são atribuídos às diferentes amostras. No caso da interpolação por média 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.

Journal (1988) mostra que, os sistemas de Krigagem simples e ordinários podem ser escritos em notações matriciais como:

$$\mathbf{K} \cdot \boldsymbol{\lambda} = \mathbf{k} \Rightarrow \boldsymbol{\lambda} = \mathbf{K}^{-1} \cdot \mathbf{k} \tag{12}$$

onde,  $\mathbf{K}$  e  $\mathbf{k}$  são matrizes das covariâncias (ou variogramas) e  $\mathbf{l}$  o vetor dos pesos. Por exemplo, no caso da Krigagem Ordinária as matrizes são assim definidas:

$$\mathbf{K} = \begin{vmatrix} c_{11} & c_{12} & \dots & c_{1n} & 1 \\ c_{21} & c_{22} & \dots & c_{2n} & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{vmatrix}, \quad \mathbf{l} = \begin{vmatrix} 1 \\ 2 \\ \vdots \\ n \end{vmatrix} \text{ e } \mathbf{k} = \begin{vmatrix} c_{10} \\ c_{20} \\ \vdots \\ c_{n0} \\ 1 \end{vmatrix}$$

Determinado os pesos, equação (10), aplica-se o estimador da equação (8) ou (9). Além da inferência de valores, esta técnica fornece também uma medida de imprecisão denominada de Variância de Krigagem (Simples ou Ordinária).

ria), dada por (Journal, 1988) como:

$$S_k^2 = \mathbf{C}(\mathbf{0}) - I^T \cdot \mathbf{k} \quad (13)$$

## REFERÊNCIAS BIBLIOGRÁFICAS

- Arbia, G. (1993). The use of GIS in Spatial Statistical Surveys. *International Statistics Review* 61, 339-359.
- Burrough, P. A. (1992). The tecnologic paradox in soil survey: new methods and techniques of data capture and handling. *Soil Survey: Perspectives and Strategies for the 21st century*. ITC publication,. Number 21, p. 4-20.
- David, M. (1997). *Geostatistical ore reserve estimation*. Elsevier Scientific, New York, 364p.
- Goodchild, M.F.; Guoqing, S. & Shiren, Y. (1992). Development and test of an error model for categorical data. *International Journal of Geographical Information Systems*. nº2 p. 87-104.
- Krige, D.G. (1951). A statistical approaches to some basic mine evaluation problems on the Witwatersrand. *Johanesburg Chemistry Metallurgy Mining Society South African*, 52 (6): 119-139.
- Matheron, G. (1971). The Theory of regionalizes variables and its applications. *Cahiers du Centre de Morphologie Mathématique* No 5, 211p
- Roo, A. de, Hazelhoff, L. & Heuvelink, G. B. M. (1992). Estimating de effects of spatial variability of infiltration on the output of a distributed runoff and soil erosion model using Monte Carlo methods. *Hydrological Process*. p. 127-143.