

# UTILIZAÇÃO DE GEOESTATÍSTICA PARA O TRATAMENTO DE DADOS DE PROSPECÇÃO GEOQUÍMICA

ADRIANO H. GRIPP\*

**ABSTRACT** GEOSTATISTICS APPLICATIONS TO GEOCHEMICAL EXPLORATION DATA TREATMENT. The present work describes with minimum emphasis to mathematical formalism, the basic ideas of the Geostatistics theory (regionalized variables), and its possible applications to geochemical exploration data treatment. Some applications already used are cited and others are proposed, in response to the problems of interpreting geochemical exploration data. The main objective is to spread and to motivate the use of Geostatistics by professionals involved in geochemical exploration. The available experimental data and the geostatistical model are in perfect agreement in this case.

*Keywords:* Geostatistics, geochemical exploration data processing.

**RESUMO** Este trabalho apresenta de forma descritiva, com o mínimo possível de formalismo matemático, as idéias de base da teoria geoestatística (variáveis regionalizadas), e algumas aplicações possíveis no tratamento de dados de prospecção geoquímica. São citadas algumas aplicações já realizadas, e propostas outras, que respondem diretamente aos problemas de interpretação dos dados de prospecção geoquímica. O objetivo principal é divulgar e incentivar a utilização da geoestatística pelos profissionais envolvidos em prospecção geoquímica, considerando a perfeita concordância entre o tipo de dados experimentais disponíveis e o modelo geoestatístico.

*Palavras-chaves:* Geoestatística, tratamento de dados de prospecção geoquímica.

**INTRODUÇÃO** Durante o tratamento de dados de prospecção geoquímica, aparece o problema crítico da definição das anomalias; nem sempre é evidente ou fácil identificar áreas anômalas, ou mesmo, definir o limite que separa valores anômalos dos normais.

Ao longo do tempo, vêm sendo utilizadas várias técnicas para tentar resolver este problema, desde uma simples análise visual de mapas de isolinhas (cartografia das concentrações dos elementos analisados), passando pela utilização de técnicas de estatística clássica univariável, até a utilização de técnicas de estatística multivariável.

Os métodos estatísticos uni ou multivariáveis, classicamente utilizados, não consideram de forma alguma a repartição espacial das concentrações dos elementos estudados, a saber, a sua posição relativa (coordenadas geográficas), a variabilidade ou as correlações que porventura existam por causa da distância de um ponto amostrado a outro. Da mesma forma, a maior parte dos métodos de interpelação utilizados em cartografia geoquímica, também não consideram a variabilidade espacial da variável estudada.

A geoestatística, desenvolvida inicialmente por G. Matheron (Matheron 1965), visando a solução de problemas de estimação de reservas minerais, é um método topo-probabilístico, isto é, que utiliza um modelo probabilístico e a posição espacial (coordenadas  $x, y, z$ ) das observações, para estudar a variabilidade dos valores observados.

Com o posterior desenvolvimento da teoria, verificou-se ser possível a aplicação da geoestatística a diversas outras áreas, tais como hidrogeologia, oceanografia, engenharia florestal e, inclusive, prospecção geoquímica (David & Dagbert 1974, Croissant 1977). No entanto, encontramos uma quantidade relativamente pequena de aplicações em prospecção geoquímica, já publicadas, provavelmente devido à pequena divulgação da teoria geoestatística junto aos profissionais que trabalham nesta área.

A intenção deste trabalho é, então, apresentar resumidamente alguns aspectos da geoestatística, passíveis de serem utilizados com êxito (alguns já o foram) no tratamento de dados de prospecção geoquímica.

**APRESENTAÇÃO RESUMIDA DA GEOESTATÍSTICA O modelo geoestatístico** Uma série de variáveis geológicas (teores, densidade, potência de camadas etc.), quando observadas, apresentam características interessantes:

- cada observação está associada a um ponto bem definido no espaço (por suas coordenadas  $x, y, z$ ) e existem variações espaciais mais ou menos evidentes no domínio considerado;
- quando observadas a pequena escala, estas variações apresentam-se aparentemente caóticas, como se fossem totalmente aleatórias;
- no entanto, pode-se notar um comportamento médio, mais ou menos bem definido para estas variações, ou seja, existe uma certa estrutura espacial que caracteriza a variável considerada.

A estas variáveis damos o nome de variáveis regionalizadas (VR). A geoestatística utiliza o conceito de funções aleatórias (FA) a fim de construir um modelo para a realidade física, fazendo aparecer estas duas características aparentemente contraditórias: aleatório e estruturado.

Assim, o valor de uma variável regionalizada, observado no ponto  $x$  (por exemplo, a concentração medida de um elemento traço em uma amostra),  $z(x)$ , é interpretado como uma realização particular de uma função aleatória  $Z(x)$ .  $Z(x)$  é aleatória, pois é uma variável aleatória (VA), mas é função porque  $Z(x)$  e  $Z(x+h)$  - dois valores da VR  $Z(x)$  nos pontos  $x$  e  $x+h$  - não são independentes: existe uma correlação entre eles, que caracteriza a estrutura espacial do fenômeno estudado.

\* Departamento de Engenharia de Minas, Universidade Federal de Minas Gerais, Rua Espírito Santo, 35, CEP 30160-000, Belo Horizonte, MG, Brasil

**Hipóteses de trabalho** Na realidade, tem-se acesso a uma única realização  $z(x)$  da FA  $Z(x)$ ; por não ser possível inferir a lei de distribuição de probabilidades da VA considerada; faz-se necessário, então, algumas hipóteses restritivas:

- Hipótese de estacionaridade: a FA é estacionária se a lei de probabilidades de  $Z(x)$  for a mesma em qualquer ponto  $x$  considerado. Poderíamos tentar inferir a lei de  $Z(x)$  a partir das observações experimentais  $z(x)$ . No entanto, em geoestatística linear, só nos interessamos pelos momentos de ordem um e dois da FA e podemos trabalhar com uma hipótese mais fraca (menos restritiva);
- Hipótese de estacionaridade de ordem 2: a FA  $Z(x)$  é estacionária de ordem 2 se:

$E [Z(x)] = m$ , a esperança matemática de  $Z(x)$  é constante e não depende do ponto  $x$ ,

Para todo par de pontos  $x_1$  e  $x_2$  existe a covariância de  $Z(x)$ , dependendo exclusivamente da distância  $h = x_1 - x_2$ :

$$C(h) = E [Z(x) Z(x + h)] - m^2 \quad (1)$$

Estas duas propriedades implicam na existência da variância:

$$\text{Var} [Z(x)] = \text{Var} [Z(x + h)] = C(0) \quad (2)$$

e do variograma, função definida como:

$$2\gamma(h) = \text{Var} [Z(x) - Z(x + h)] \quad (3)$$

No entanto, existem situações para as quais a variância não é finita ( $C(0)$  é indefinido) e deve-se trabalhar com uma hipótese ainda mais fraca, suficiente à maior parte das aplicações de geoestatística linear:

- Hipótese intrínseca: a FA  $Z(x)$  é intrínseca se:
- A esperança matemática de  $Z(x)$  existe e não depende do ponto de implantação  $x$ :  $E [Z(x)] = m$ ;
- Para todo vetor  $h$ , a diferença  $(Z(x) - Z(x + h))$  tem uma variância finita, que não depende de  $x$ , mas unicamente do vetor  $h$ . Esta variância permite definir a função intrínseca chamada variograma:

$$2\gamma(h) = \text{Var} [Z(x) - Z(x + h)] \quad (4)$$

Se a FA considerada for estacionária de ordem 2, então:

$$\gamma(h) = C(0) - C(h) \quad (5)$$

**O variograma** O variograma é a ferramenta fundamental da geoestatística; caracteriza as estruturas e as correlações espaciais da variável regionalizada em estudo. Sob a hipótese intrínseca, é possível sua inferência estatística a partir dos dados experimentais  $z(x_i)$ , sendo o variograma experimental  $\gamma^*(h)$  calculado como:

$$\gamma^*(h) = \frac{1}{2N} \sum_{i=1}^N [Z(x_i) - Z(x_i + h)]^2 \quad (6)$$

Deve-se notar que, originalmente, o nome variograma foi dado à função  $2\gamma(h)$ , sendo  $\gamma(h)$  o semi-variograma, no entanto, tradicionalmente confunde-se o nome variograma com a função semi-variograma por uma questão de simplicidade de nomenclatura e de utilização.

O variograma experimental deve ser ajustado por um modelo teórico autorizado: a função  $\gamma(h)$  deve ser de tipo positivo condicional. Na prática, alguns modelos já bem definidos e estudados, e/ou combinações lineares dos

mesmos, são suficientes para ajustar a maioria dos variogramas experimentais.

A função variograma é, por construção, simétrica ( $\gamma(h) = \gamma(-h)$ ), positiva ( $\gamma(h) > 0$ ) e nula para  $h=0$  ( $\gamma(0) = 0$ ). Em geral,  $\gamma(h)$  aumenta quando  $h$  aumenta, ou seja, quanto mais os pontos  $x$  e  $x+h$  estão distantes, menos eles são correlacionados.

Na maioria dos casos, o variograma atinge um patamar, caracterizando o que chamamos um "fenômeno de transição". A distância a partir da qual o patamar é atingido chama-se "amplitude" e corresponde à noção física de zona de influência de uma amostra.

A existência de eventuais erros de medida (devidos ao processo de coleta de amostras, ao método analítico etc.) e/ou de correlações espaciais existentes em distâncias inferiores a das observações faz aparecer uma aparente descontinuidade na origem, denominada "efeito de pepita". No caso extremo de ausência total de correlação espacial, temos um "efeito de pepita puro":  $\gamma(h) = C_0$  (valor constante diferente de zero), para todo  $h$ , com  $\gamma(0) = 0$  por definição.

Se a VR estudada apresentar simultaneamente diferentes variabilidades para escalas de observação diferentes (por exemplo, valor observado igual à soma de uma componente de variabilidade regional e de uma componente de variabilidade local), o variograma experimental obtido será a soma dos variogramas individuais, correspondentes a cada escala considerada.

Por outro lado, caso hajam anisotropias, variabilidades diferentes em direções diferentes no espaço (devido, por exemplo, à presença de controles estruturais da mineralização), os variogramas calculados para estas direções serão diferentes, tornando-se possível considerar quantitativamente essas anisotropias.

**Variâncias de estimação e de dispersão** Conhecido o variograma, é possível calcular a variância de estimação - definida como variância do erro - seja da quantidade (valor verdadeiro - valor estimado), que é função do processo ou método utilizado para estimar o valor verdadeiro.

Seja  $z(V)$  o valor médio verdadeiro da VR  $Z$  no suporte, ou domínio,  $V$  (a notação  $V$  é geral,  $V$  pode ser um ponto, uma superfície, um volume etc.) e  $z(v)$  o valor médio de  $Z$  num suporte  $v$  formado por uma ou mais unidades elementares. A variância de estimação  $\sigma^2(v, V)$  ou variância do erro cometido estimando  $z(V)$  por  $z(v)$  pode ser calculada por:

$$\sigma^2(v, V) = 2 (V, v) - (V, V) - (v, v) \quad (7)$$

onde a notação  $(V, v)$  corresponde ao variograma médio obtido, fazendo uma extremidade do segmento  $h$  percorrer todos os pontos do domínio  $V$  e a outra, independentemente, todos os pontos de  $v$ .

É importante notar que  $\sigma^2(v, V)$  depende não apenas da estrutura da VR (variograma), mas também da geometria de  $v$  e de  $V$  e de suas posições relativas no espaço.

Da mesma forma, a variância de dispersão  $D^2(v/V)$ , que corresponde à noção clássica de variância de dispersão ( $s^2$ ), é obtida por:

$$D^2(v/V) = (V, V) - (v, v) \quad (8)$$

**Krigagem linear** Por construção, a krigagem linear é o melhor (no sentido em que fornece uma variância de estimação mínima) estimador linear não enviesado, que permite estimar o valor médio  $z(V)$  a partir de uma combinação linear das  $N$  informações experimentais disponíveis:

$$z_k(V) = \sum_{\alpha=1}^N \lambda_{\alpha} z(v_{\alpha}) \quad (9)$$

Os coeficientes  $\lambda_\alpha$  são calculados de forma a minimizar a variância de estimação de  $z(V)$  por  $z_k(V)$ , o que nos conduz ao seguinte sistema de  $N+1$  equações:

$$\begin{cases} \sum_{\beta=1}^N \lambda_\beta \bar{\gamma}_{\alpha\beta} + \mu = \bar{\gamma}_{\alpha v} & , \alpha = 1 \text{ à } N \\ \sum_{\alpha=1}^N \lambda_\alpha = 1 & , \text{ condição de não-enviesamento} \end{cases} \quad (10)$$

onde:  $\mu$  = parâmetro de Lagrange,  
 $\bar{\gamma}_{\alpha v} = \bar{\gamma}(V_\alpha, V_v)$

A variância de krigagem é dada por:

$$\sigma_k^2 = \sum_{\alpha=1}^N \lambda_\alpha \bar{\gamma}_{\alpha v} + \mu - \bar{\gamma}(V, V) \quad (11)$$

Podemos assim, conhecendo a função variograma, não só calcular o melhor estimador linear (melhor no sentido "variância de estimação mínima"), mas também calcular a respectiva variância de estimação, que pode ser usada como critério para medir a qualidade da estimação.

**ALGUMAS UTILIZAÇÕES SIMPLES NO TRATAMENTO DE DADOS DE PROSPECÇÃO GEOQUÍMICA**

Na prática o variograma experimental é calculado, para um número finito de distâncias  $h$ , múltiplas de uma distância elementar chamada "passo", função da malha de amostragem. Para cada distância  $h$  considerada, o variograma experimental é uma média estatística da quantidade  $(Z(x) - Z(x + h))^2$ , que é caracterizada também por uma certa variância.

Caso não existam valores anômalos entre os dados utilizados, o variograma assim obtido caracterizaria a estrutura espacial do fundo médio geoquímico.

Por outro lado, na presença de valores anômalos, esta estrutura não aparecerá claramente definida pelo variograma experimental. Isto acontece porque o valor médio de  $(Z(x) - Z(x + h))^2$  será modificado em maior ou menor grau pela presença dos valores anômalos, em função da posição espacial das amostras.

Podemos, então, definir como pontos anômalos, aqueles que contribuirão para aumentar muito a variância de  $(Z(x) - Z(x + h))^2$ , para um  $h$  dado. Este critério, associado a uma posterior verificação da distribuição espacial dos pontos considerados anômalos, parece ser melhor que uma simples escolha, mais ou menos arbitrária, de um valor limite separando anomalia do fundo, visto que considera a estrutura e a localização espacial da variável estudada.

Podemos melhorar a definição das zonas anômalas e mesmo fazer aparecer zonas de pequeno contraste anomalia-fundo, usando a técnica chamada de validação cruzada. Para isto, preci-se do variograma que caracteriza a estrutura do fundo geoquímico, que pode ser calculado experimentalmente a partir dos mesmos dados, sem considerar as amostras anteriormente julgadas anômalas.

A técnica de validação cruzada consiste em estimar por krigagem, em cada ponto amostrado, um novo valor para a variável considerada, usando como informação apenas os demais pontos (o valor observado no ponto é momentaneamente esquecido) e calcular a variável residual "valor observado - valor krigado".

Analisando este valor residual associado à sua respectiva variância de estimação, é possível identificar mais facilmente

os pontos anômalos, nos quais o valor residual é muito diferente de zero.

Por outro lado, o simples estudo variográfico pode contribuir muito para o estudo geoquímico pela análise do comportamento do variograma, de parâmetros como sua amplitude, efeito de pepita, das anisotropias que possam existir, das diferentes estruturas que podem aparecer a escalas diferentes, etc. Normalmente, sempre encontramos uma explicação física (geológica, geoquímica) para estes parâmetros.

Naturalmente que, se desejamos trabalhar com mapas de curvas de isovalores para as variáveis analisadas, o método de interpolação a ser utilizado deverá ser a krigagem que, além de conduzir a uma variância de estimação mínima, fornecerá, também, um mapa destas variâncias, o que nos permite, em caso de malha de amostragem irregular, conhecer as regiões onde se tem maior ou menor confiança na posição e valores das curvas.

Um estudo multivariável não muito elaborado pode ser feito neste estágio, comparando os variogramas obtidos para os diversos elementos analisados e verificando os que têm comportamento semelhante.

Embora de forma não sistemática, alguns trabalhos já foram publicados (David & Dagbert 1974, Pereira 1980), usando uma ou várias das "técnicas" que acabaram de ser citadas e mostram que, de fato, se podem obter bons resultados no tratamento de dados de prospecção geoquímica.

**OUTRAS APLICAÇÕES ENVOLVENDO UM FORMALISMO MATEMÁTICO MAIS COMPLEXO**

Matheron (1982) propôs, sob o nome de "análise krigante", um modelo topp-probabilístico que permite estimar, em um ponto dado, variáveis fictícias cujos variogramas correspondem aos diferentes modelos utilizados quando se ajusta o variograma experimental por uma soma de modelos teóricos.

Ora, caso se possa interpretar e dar um significado, sobre anomalia e fundo regional, a cada estrutura definida no modelo, por exemplo, o efeito de pepita representando a parte mais errática (anomalia) e uma estrutura de grande amplitude representando a parte mais continua (fundo), será possível estimar ponto a ponto e mapear cada uma destas duas componentes.

Posteriormente, o modelo permite avaliar em cada ponto experimental a contribuição relativa de cada componente, o que serve de critério para classificar os valores observados em função de sua maior ou menor correlação com as estruturas definidas no modelo.

Esta técnica envolve uma metodologia mais complicada de cálculo numérico e de programação e pressupõe que se possa interpretar corretamente as estruturas mostradas pelo variograma experimental. Ela foi utilizada satisfatoriamente para o tratamento de dados de prospecção geoquímica por Sandjivy(1983).

Da mesma forma que se definiu o variograma, pode-se definir o co- variograma ou variograma cruzado  $\gamma_{ij}(h)$ , o que nos permite um tratamento geoestatístico multivariável:

$$\gamma_{ij}^*(h) = \frac{1}{2N} \sum_{\alpha=1}^N [Z_i(x_\alpha) - Z_i(x_\alpha + h)] [Z_j(x_\alpha) - Z_j(x_\alpha + h)] \quad (12)$$

Nota-se, no entanto, que a interpretação do co-variograma nem sempre é evidente ou fácil, e envolve cálculos mais pesados. Além disto, o ganho de informação sobre o tratamento univariável é, às vezes, muito pequeno ou nulo.

Wackernagel (1988) propôs uma metodologia utilizando uma análise geoestatística multivariável (funções co-variograma e análise krigante multivariável) combinada com uma análise clássica em componentes principais.

A utilização desta metodologia foi exemplificada com um conjunto de dados de prospecção geoquímica já estudado por

outros métodos, e o resultado final foi simplesmente confirmar e evidenciar a interpretação já dada anteriormente, utilizando-se métodos mais simples.

Certamente que se deve trabalhar mais estas técnicas e verificar se realmente existe um ganho significativo de informação em relação a outras técnicas geoestatísticas mais simples de serem aplicadas e menos custosas em tempo de computação.

**UMA APLICAÇÃO POSSÍVEL DE GEOESTATÍSTICA NÃO-LINEAR** Ao procurar zonas anômalas para proceder a um detalhamento, está-se, na realidade, propondo a seguinte indagação: qual a probabilidade de encontrarmos nesta área, concentrações acima de um valor dado?

A resposta a esta questão pode ser obtida utilizando-se técnicas de geoestatística não-linear, por exemplo krigagem disjuntiva. Estes métodos exigem hipóteses fortes (mais restritivas) sobre a VR estudada, o que restringe um pouco sua aplicabilidade. No entanto, verificando-se que o comportamento da VR considerada pode adaptar-se a estas hipóteses, é possível ter um ganho significativo de informação, em relação à geoestatística linear.

A krigagem disjuntiva procura estimar, no interior de um domínio (uma área ou volume) bem definido, a proporção de valores  $z(x)$  superiores a um valor de corte  $z_c$ , a partir das observações experimentais disponíveis. Em outras palavras, estamos estimando a lei de distribuição de  $Z(x)$  condicionalmente aos valores experimentais disponíveis.

Esta técnica, embora envolvendo um formalismo matemático mais complexo e exigindo hipóteses mais fortes, é relativamente simples de ser colocada em prática e responde

diretamente à questão da definição de áreas prioritárias para detalhamento.

Embora o resultado seja melhor, nem sempre os dados de prospecção geoquímica atenderão às hipóteses de base necessárias ao modelo (estacionaridade, permanência de lei de distribuição etc.), mas vale a pena verificar, sempre que possível, sua aplicabilidade.

**CONCLUSÃO** Em prospecção geoquímica, trabalha-se uma informação eminentemente espacial, ou seja, a posição relativa das amostras no espaço é tão importante quanto os valores das variáveis observadas, para efeito de interpretação.

A geoestatística, sendo um método topo-probabilístico, é adequada e ideal para o estudo dos dados de geoquímica. Mesmo considerando apenas um tratamento inicial univariável, os resultados já obtidos, em vários estudos realizados, comprovam este fato e sugerem que se deve recorrer com mais frequência a estas técnicas.

Embora envolvendo um formalismo matemático mais elaborado e exigindo a verificação de algumas hipóteses de base, estas técnicas são fáceis e simples de serem colocadas em prática, sobretudo em se tratando de geoestatística linear. Em alguns casos, utilizando técnicas um pouco mais complexas, podem-se obter resultados extremamente interessantes e realizar também um tratamento multivariável dos dados.

Torna-se necessária, então, uma maior divulgação da geoestatística junto aos profissionais envolvidos em prospecção geoquímica e a realização de maior número de estudos e aplicações práticas, o que conduzirá, certamente, a uma melhor definição das técnicas mais adequadas e à obtenção de melhores resultados.

#### REFERÊNCIAS BIBLIOGRÁFICAS

- CROISSANT, A. 1977. La géostatistique comme outil dans la prospection géochimique. *Sciences de la Terre*, 9:129-144. (Série Informatique).
- DAVID, M. & DAGBERT, M. 1974. Lakeview revisited: variograms and correspondence analysis - new tools for the understanding of geochemical data. In: EXPLORATION GEOCHEMISTRY SYMPOSIUM, 5<sup>th</sup>. *Proceedings...* Amsterdam, Elsevier. p. 163-181.
- JOURNEL, A.G. & HUIJBREGTS, C.J. 1978. *Mining Geostatistics*. New York, Academic Press. 600 p.
- MATHERON, G. 1965. *Les Variables Régionalisées et leur Estimation*. Paris, Masson. 212 p.
- MATHERON, G. 1982. *Pour Une Analyse Krigeante des Données Régionalisées*. Fontainebleau, Centre de Géostatistique, 22 p. (Note Interne N-732).
- PEREIRA, H. 1980. Traitement de données de prospection géochimique base sur l'analyse structurale. *Sciences de la Terre*, 14:143-159. (Série Informatique).
- SANDJIVY, L. 1983. Analyse krigeante de données géochimiques. *Sciences de la Terre*. 18:141-172. (Série Informatique).
- SANDJIVY, L. 1984. The factorial kriging analysis of regionalized data. Its application to geochemical prospecting. In: VERLY, G. et al. eds. *Geostatistics for Natural Resources Characterization*, Part I. Dordrecht, D.Reidel Publ.p.559-571.
- WACKERNAGEL, H. 1988. Geostatistical techniques for interpreting multivariate spatial information. In: CHUNG, C.F. et al. eds. *Quantitative Analysis of Mineral and Energy Resources*. Dordrecht, D. Reidel Publ. p. 393-409.

Manuscrito A667

Recebido em 17 de julho de 1990

Revisão do autor em 11 de julho de 1991

Revisão aceita em 8 de abril de 1992