Laboratório de Mineralurgia e Planeamento
Mineiro

GEOESTATÍSTICA MULTIVARIADA

António Jorge Sousa

Outubro de 1989

GEOESTATÍSTICA MULTIVARIADA

1 - INTRODUÇÃO

A teoria e os princípios básicos da Geoestatística Multivariada, cujo objectivo principal é a estimação simultânea de um conjunto de variáveis espacialmente correlacionadas (variáveis corregionalizadas), são conhecidos desde os primórdios da Geoestatística. No entanto, as aplicações têm sido escassas, a menos que a variável de maior interesse económico, por apresentar custos de amostragem substancialmente superiores, seja conhecida apenas num pequeno número de pontos. Esta situação deve-se, provavelmente, à complexidade computacional e de notação introduzidas. No entanto, esta situação tem-se vindo a modificar progressivamente, com várias aplicações nos últimos anos em domínios tão variados como sejam a indústria mineira, geoquímica, pedologia e geohidrologia — v.g. Guarascio (1976), Magri (1972), Wackernagel (1987), Aboufirassi e Mariño (1984), Ahmed e Marsily (1987) e Ribeiro e Muge (1989).

A Geoestatística Multivariada apresenta também potencialidades enormes no âmbito da análise estrutural, conforme descrito por Wackernagel (1988, 1989) e Sousa (1988b, 1989). A análise dos variogramas simples e cruzados das diversas variáveis, enriquece os resultados dos métodos clássicos de Análise de Dados, que não tomam em conta a estrutura espacial das variáveis, justificando, só por si, o estudo conjunto das variáveis regionalizadas. Neste trabalho, procura-se, sempre que pertinente, fazer o paralelismo entre a Geoestatística Multivariada e a Análise de Dados (em particular a Análise em Componentes Principais).

2. VARIOGRAMAS CRUZADOS

Seja um conjunto de p variáveis regionalizadas conhecidas em n pontos amostrais, com coordenadas x_i:

$$\left\{ Z_j(x_i);\ i = 1,\ \ldots\ n;\ j = 1,\ \ldots,\ p \right\}$$

Conceptualmente, pode tomar-se as p variáveis regionalizadas $z_j(x)$ como uma realização de uma Função Aleatória p-dimensional $Z_j(x)$. Esta realização é, habitualmente, única, pelo que a inferência estatística só é possível se a Função Aleatória verificar algumas hipóteses de estacionaridade. Na Geoestatística mineira é habitual considerar dois tipos de estacionaridade (Cf. Journel e Huijbregts, 1978, p. 32-33): estacionaridades de 2ª ordem e intrínseca.

Na hipótese de estacionaridade de 2ª ordem os dois primeiros momentos são constantes, não dependendo da localização espacial:

$$\mathrm{E}\left\{ Z_j (x) \right\} = m_j \qquad \forall x$$

$$\mathrm{E}\left\{ \left[Z_j \ (x+\mathbf{h}) - m_j \right] \left[Z_{j'} (x) - m_{j'} \right] \right\} = C_{jj'} \ (\mathbf{h}) \qquad \forall x \tag{1}$$

$$\mathrm{E}\left\{ Z_j \ (x+\mathbf{h}) \ Z_j (x) \ \right\} - m_j^2 = C_j \ (\mathbf{h}) \qquad \forall x$$

A função C_j (**h**) é a covariância espacial da variável $Z_j(x)$, que depende apenas do vector **h**. A estacionaridade da covariância espacial implica a estacionaridade da variância:

$$\mathrm{Var}\left\{ Z_j (x) \right\} = C_j \ (0)$$

Note-se que a estacionaridade da covariância cruzada é mais forte do que a estacionaridade da covariância simples, isto é, mesmo que cada uma das variáveis $Z_j(x)$ seja estacionária de segunda ordem, nada garante que a corregionalização o seja, pois a covariância cruzada $C_{jj'}$ (**h**) pode não depender apenas do vector **h**.

A hipótese intrínseca impõe a estacionaridade dos dois primeiros momentos dos acréscimos espaciais. Assim, sendo os acréscimos a passo **h** dados por:

$$e_j (\mathbf{h}) = Z_j \ (x+\mathbf{h}) \text{ - } Z_j (x)$$

temos que aquelas duas condições de estacionaridade são expressas por:

$$\mathrm{E}\left\{ e_j (\mathbf{h}) \right\} = 0$$

$$\mathrm{E}\left\{ e_j (\mathbf{h}) e_{j'} (\mathbf{h}) \right\} = 2g_{jj'} \ (\mathbf{h}) \qquad \forall x$$

$$\mathrm{E}\left\{ \left[e_j (\mathbf{h}) \right]^2 \right\} = 2g_j \ (\mathbf{h}) \qquad \forall x$$

O segundo momento dos acréscimos é, por definição, o variograma da variável $Z_j(x)$:

$$g_j \ (\mathbf{h}) = \frac{1}{2} \mathrm{E}\left\{ \left[e_j (\mathbf{h}) \right]^2 \right\}$$

Do mesmo modo se pode definir o variograma cruzado entre as variáveis $Z_j(x)$ e $Z_{j'}(x)$:

$$g_{jj'} \ (\mathbf{h}) = \frac{1}{2} \mathrm{E}\left\{ e_j (\mathbf{h}) e_{j'} (\mathbf{h}) \right\} \tag{2}$$

Convém salientar que, contrariamente ao caso monovariado, na hipótese mais restritiva (na medida em que abrange uma menor variedade de fenómenos) de estacionaridade de 2ª ordem, o variograma e a covariância espacial não são equivalentes. Neste caso o variograma e as covariâncias verificam a seguinte relação, que se obtém desenvolvendo (1):

$$g_{jj'}(\mathbf{h}) = C_{jj'}(0) - \frac{1}{2}\left[C_{jj'}(\mathbf{h}) + C_{j'j}(\mathbf{h})\right] \qquad (3)$$

Verifica-se assim que o variograma é uma função mais *pobre* do que a covariância sendo equivalentes apenas no caso (frequente) em que estas são simétricas, isto é, em que se verifique a seguinte relação entre as duas funções:

$$C_{jj'}(\mathbf{h}) = C_{j'j}(\mathbf{h})$$

Pode-se definir, no entanto, uma relação de simetria generalizada, que resulta directamente de (1):

$$C_{jj'}(\mathbf{h}) = C_{j'j}(-\mathbf{h})$$

No entanto, o variograma põe problemas de estimação menos delicados do que as covariâncias (simples e cruzadas), pois estas não podem ser estimadas sem enviezamento (provocado pelo desconhecimento da média a passo **h**, que tem de ser estimada).

A expressão (3) mostra que o variograma cruzado $\gamma_{jj'}(\boldsymbol{h})$ é, pelo contrário, uma função simétrica. Verifica-se também que pode tomar valores negativos, enquanto os variogramas simples só tomam valores positivos. Fazendo $j = j'$ na expressão (3), obtém-se a relação, já conhecida da Geoestatística monovariada, entre o variograma e a covariância espacial, no caso da hipótese de estacionaridade de 2ª ordem:

$$g_j(\mathbf{h}) = C_j(0) - C_j(\mathbf{h})$$

O cálculo do variograma cruzado experimental pode fazer-se pelo estimador seguinte, em que a esperança do produto dos acréscimos é substituída pela média espacial:

$$g_{jj'}^{*}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})}\sum_{1}^{N(\mathbf{h})}\left[Z_j(x_i + \mathbf{h}) - Z_j(x_i)\right]\left[Z_{j'}(x_i + \mathbf{h}) - Z_{j'}(x_i)\right]$$

onde N(**h**) é o número de pares de pontos separados de **h**.

Para $j = j'$, obtém-se a expressão clássica do variograma experimental para uma única variável regionalizada:

$$g_j^{*}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})}\sum_{1}^{N(\mathbf{h})}\left\{\left[Z_j(x_i + \mathbf{h}) - Z_j(x_i)\right]^2\right\}$$

Para o cálculo dos variogramas cruzados experimentais existe uma expressão que permite simplificar os cálculos, no caso em que as variáveis foram igualmente amostradas. Sendo $z_s(x) = z_j(x) + z_{j'}(x)$, pode-se demonstrar que:

$$g_{jj'}(\mathbf{h}) = \frac{1}{2}\left[g_s(\mathbf{h}) - g_j(\mathbf{h}) - g_{j'}(\mathbf{h})\right]$$

O variograma cruzado $\gamma_{jj'}(\mathbf{h})$ pode ser calculado à custa dos variogramas simples das variáveis $z_j(x)$, $z_{j'}(x)$ e da variável soma.

$$g_{jj'}{}^*(\mathbf{h}) = \frac{1}{2}\left[g_s{}^*(\mathbf{h}) - g_j{}^*(\mathbf{h}) - g_{j'}{}^*(\mathbf{h})\right]$$

3 - MODELOS DE CORREGIONALIZAÇÃO

Os variogramas experimentais podem, para cada classe de distância **h**, ser dispostos segundo uma forma matricial Γ_k^*, representando cada elemento o valor do variograma cruzado das variáveis j e j', para cada classe de distância **h** (ver figura 1). A variância das combinações lineares autorizadas devem ser positivas, pelo que os modelos dos variogramas $\gamma_{jj'}(\mathbf{h})$ devem ser condicionalmente definidos positivos e as matrizes Γ_k^* definidas positivas para qualquer k.

Do ponto de vista da Análise de Dados, para cada classe de distância e direcção (dada pelo vector **h**) *pode calcular-se uma matriz de variância - covariância dos acréscimos* ***e*(h)**. *Sendo K o número de classes de distância estatisticamente significativas (número máximo de passos do variograma), obtém-se uma pilha de K matrizes variância-covariância (Figura 1). Do ponto de vista da Geoestatística Multivariada essa pilha de matrizes é constituída por um conjunto de p variogramas simples (fazendo variar l até K para a diagonal principal das sucessivas matrizes variância-covariância) e p(p-1)/2 variogramas cruzados (para os elementos não diagonais). Estes variogramas são funções ordinárias do passo* **h** *e constituem a ferramenta estrutural básica da Geoestatística Multivariada.*

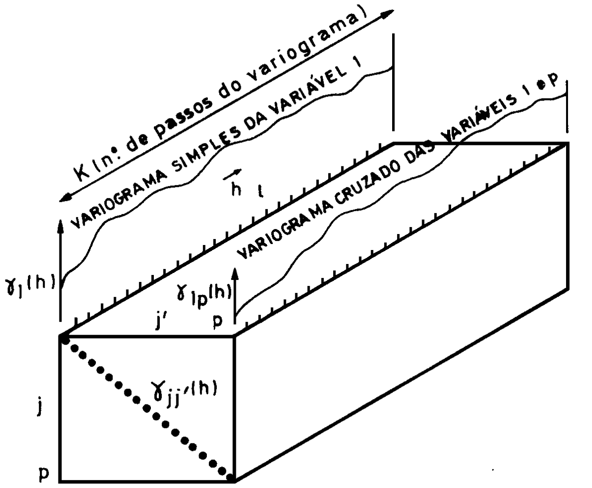

Figura 1 - Matriz dos variogramas simples e cruzados.

Em paralelo com a geoestatística monovariada, pode por-se a hipótese de que uma corregionalização se pode traduzir por uma soma de estruturas imbricadas, revelando uma sucessão de fenómenos de transição (a diferentes escalas espaciais) que actuam de modo análogo para as *p* variáveis. Então cada um dos variogramas simples e cruzados da figura 1 pode reduzir-se à sobreposição de N^e estruturas imbricadas, cada uma delas modelizada por um esquema teórico (esférico, gaussiano, linear, etc.). Cada variograma é dado por:

$$g_{jj'}(\mathbf{h}) = \sum_{u=0}^{N^e-1} b_{jj'}^{u} g_{jj'}^{u}(\mathbf{h})$$

3.1 - Modelo Linear

No quadro do modelo linear de corregionalizações (Journel e Huijbregts, 1978, p. 171), para cada escala *u*, a forma e a amplitude dos variogramas são as mesmas para todas as variáveis, podendo, assim, simplificar-se a pilha de matrizes da figura 1 representando os variogramas simples e cruzados através da consideração de, apenas, N^e matrizes (N^e << *K*), cada uma das quais reflecte uma estrutura imbricada, válida para todas as *p* variáveis (v.g. figura 2).

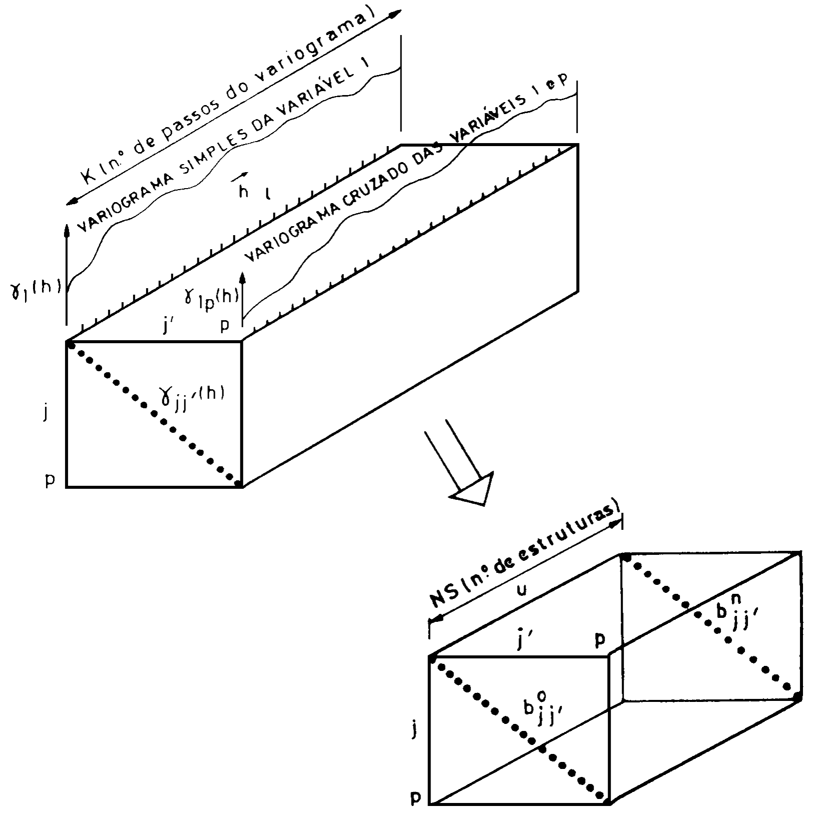

Figura 2 - Matriz dos variogramas simples e cruzados (modelo linear)

Nesta formulação a pilha de variogramas simples e cruzados $\gamma_{jj'}(\mathbf{h})$ é dada por:

$$g_{jj'}(\mathbf{h}) = \sum_{u=0}^{N^e-1} b_{jj'}^u g^u(\mathbf{h})$$

onde $b_{jj'}^u$ é o patamar do variograma cruzado para as variáveis j e j' e para a estrutura u. Em geral, N^e é da ordem de 2 ou 3, o que permite reduzir o problema da descrição estrutural das variáveis correlacionadas espacialmente através de 2 ou 3 matrizes $\mathbf{B}^u$, definidas positivas, de dimensão (p x p), cuja soma reproduz a matriz variância-covariância dos dados:

$$s_{jj}^{\prime 2} = \sum_{u=0}^{N^e-1} b_{jj'}^u$$

Neste caso, as variáveis podem ser expressas como combinações lineares de factores ortogonais $y_a^u(x)$:

$$Z_j(x) = \sum_{u=0}^{N^e-1} \sum_{a=1}^{p} e_{ja}^u y_a^u(x)$$

Os coeficientes devem verificar a relação seguinte:

$$b_{jj'}^u = \sum_{a=1}^{p} e_{ja}^u e_{j'a}^u$$

Estes coeficientes podem ser obtidos por diagonalização das matrizes de corregionalização $\mathbf{B}^u$:

$$e_{ja}^u = \sqrt{m_a}\, v_{ja}$$

em que m_a é o valor próprio de ordem a da matriz $\mathbf{B}^u$ e v_{ja} é a coordenada j do vector próprio que lhe está associado.

3.2 - Modelo intrínseco

No modelo linear admite-se que as amplitudes são idênticas (em cada estrutura) para todos os variogramas. Pode ainda acontecer que, simultaneamente, as matrizes $\mathbf{B}^u$ sejam todas proporcionais a uma matriz $\mathbf{B}$. Trata-se do modelo de corregionalização intrínseca, significativamente mais simples do que o modelo linear, embora mais difícil de ajustar aos dados, pois o número de parâmetros a estimar é menor. Em Pereira e Soares (1989) encontra-se uma aplicação da geoestatística a variáveis piscícolas, que verificam uma corregionalização deste tipo.

Neste caso os variogramas $g_{jj'}(\mathbf{h})$ são dados por:

$$g_{jj'}(\mathbf{h}) = \sum_{u=0}^{N^e-1} b_{jj'}^u g^u(\mathbf{h})$$

com

$$b^u_{jj'} = b_{jj'} \boldsymbol{b}^u$$

sendo $\boldsymbol{b}^u$ uma constante de proporcionalidade

A pilha de variogramas simples e cruzados, no quadro do modelo intrínseco reduz-se a uma única matriz **B**, cuja diagonalização conduz a factores ortogonais independentes de **h**.

A análise multivariada reduz-se, no caso do modelo intrínseco, à análise monovariada dos factores resultantes de uma Análise em Componentes Principais. Trata-se de uma simplificação importante pois permite reduzir a cokrigagem a algumas (poucas) krigagens.

4 - PROBLEMAS DE AJUSTAMENTO DO MODELO LINEAR

Calculada a matriz variância-covariância dos dados e analisados os variogramas experimentais para cada variável (o que permite ter uma primeira ideia sobre o número N^e de estruturas imbricadas presentes na regionalização), põe-se agora o problema de decompor a matriz variância - covariância na soma de N^e matrizes que respeitem os variogramas experimentais simples e cruzados e que sejam do tipo definido positivo. Esta última condição, ao mesmo tempo que é exigida pela coerência interna do método, pois conduz a combinações lineares autorizadas, assegurando variâncias positivas (Cf. Matheron, 1982, p. 6), leva a que, do ponto de vista da Análise de Dados, seja possível efectuar a factorização dessas sub-matrizes (paralelamente à Análise em Componentes Principais), conforme proposto por Wackernagel (1985a).

Uma matriz **B** diz-se definida positiva se, para todo o vector **y** não nulo, se verificar a seguinte relação (Dias Agudo, 1972, p.116):

$$\mathbf{yBy} > 0$$

De entre os diversos critérios que permitem verificar se uma matriz é definida positiva (cf. Demidovich e Maron, 1976, pp. 388-389), destacam-se os dois utilizados no algoritmo de ajustamento semi-automático do modelo linear de corregionalizações, descrito em Sousa (1988) :

i) A matriz **B** é definida positiva se todos os valores próprios forem positivos ou nulos.
ii) Uma matriz real e simétrica **B** é definida positiva se todos os menores principais forem positivos ou nulos.

Como consequência do último critério vem que:

$$\left| b^u_{jj'} \right| \leq \sqrt{b^u_{jj} \; b^u_{j'j'}}$$

e

$$b_{jj}^{u} \geq 0$$

Das duas relações anteriores pode-se concluir que se

$$b_{jj}^{u} = 0$$

ou

$$b_{j'j'}^{u} = 0$$

então

$$b_{jj'}^{u} = 0$$

Toda a estrutura presente num variograma cruzado tem, necessariamente, que aparecer nos variogramas simples das variáveis correspondentes. A conclusão contrária não é verdadeira: uma estrutura pode aparecer num variograma simples sem que se manifeste nos variogramas cruzados.

Descreve-se seguidamente um procedimento iterativo que permite ajustar as matrizes $\mathbf{B}^{u}$, no caso de uma corregionalização linear:

i) Determinação visual do número de estruturas elementares presentes nos variogramas experimentais simples, bem como dos parâmetros de forma respectivos (amplitudes dos modelos de variograma com patamar).
ii) Ajustamento dos parâmetros de forma dos variogramas cruzados, sabendo que uma estrutura não pode aparecer no variograma cruzado se não estiver também presente nos variogramas simples correspondentes.
iii) Ajustamento visual dos patamares de cada uma das estruturas presentes nos variogramas experimentais simples e cruzados.
iv) Diagonalização das matrizes $\mathbf{B}^{u}$. Se os valores próprios forem todos positivos, estas matrizes são definidas positivas. Se alguma das matrizes é indefinida pode-se anular os valores próprios negativos (normalmente desprezáveis face ao traço da matriz) e reconstituir a matriz, bem como as outras matrizes de corregionalização, de modo a assegurar a coerência com os valores experimentais.

5 - COKRIGAGEM

Conforme já referido o objectivo principal da Geoestatística multivariada consiste na estimação simultânea de um conjunto de variáveis em pontos ou domínios de dimensão v.

Trata-se de estimar o valor z (x), à custa de uma combinação linear da informação disponível em todas as variáveis:

$$Z^{*}_{j_0}(x_0) = \sum_{j=1}^{p}\sum_{i=1}^{n_j} l_{ji} Z_j(x_i)$$

A esperança do erro de estimação é dada por:

$$E\left\{Z_{j_0}(x_0) - Z^{*}_{j_0}(x_0)\right\} = \left\{Z_{j_0}(x_0)\right\} - \sum_{i=1}^{n_{j0}} l_{j_0 i} E\left\{Z_{j_0}(x_i)\right\} \sum_{j \neq j_0}^{p} l_{ji} E\left\{Z_j(x_i)\right\}$$

$$E\left\{Z_{j_0}(x_0) - Z^{*}_{j_0}(x_0)\right\} = m_{j_0}\left[1 - \sum_{i=1}^{n_{j0}} l_{j_0 i}\right] - \sum_{j \neq j_0} m_j \sum_{i=1}^{n_j} l_{ji}$$

A condição de não enviezamento traduz-se então pelas *p* equações:

$$\sum_{i=1}^{n_{j0}} l_{j_0 i} = 1 \qquad (4)$$

e

$$\sum_{i=1}^{n_j} l_{ji} = 0 \qquad \forall = 1,\ldots p; \quad jj \neq j_0$$

A minimização da variância de krigagem $E\left\{\left[z_{j_0}(x_0) - z^{*}_{j_0}(x_0)\right]^2\right\}$, sujeita aos constrangimentos de não enviezamento conduz ao sistema de cokrigagem:

$$\begin{cases} \sum_{j=1}^{p}\sum_{i=1}^{n_j} l_{ji} g_{jj'}(x_i - x_{i'}) + m_j = g_{j_0 j'}(x_0 - x_{i'}) & \forall i' = 1,\ldots,n = \sum n_j;\ j' = 1,\ldots,p \\ \sum_{i=1}^{n_{j0}} l_{j_0 i} = 1 & \\ \sum_{i=1}^{n_j} l_{ji} = 0 & \forall j = 1,\ldots p; \quad j \neq j_0 \end{cases}$$

A relação (4) exige $n_j \neq 0$, isto é,.a variável primária deve ser conhecida em pelo menos uma das amostras estimantes.

6 - ANÁLISE GEOESTATÍSTICA DE DADOS

6.1 - Análise factorial de corregionalizações

O fenómeno corregionalizado global foi decomposto numa série de N^e estruturas associadas a diferentes escalas. Para cada escala, obteve-se uma matriz variância-covariância, de termo geral $b^u_{jj'}$, que dá conta do sistema de interdependências entre as p variáveis para a escala u. Em paralelo com a Análise em Componentes Principais da matriz variância - covariância global $\boldsymbol{B}$, pode agora factorizar-se cada uma das matrizes $\boldsymbol{B^u}$, obtendo-se assim uma síntese da estrutura das variáveis, relacionada agora com a escala espacial do fenómeno.

Supondo que cada uma das estruturas espaciais imbricadas foi interpretada com base nos seus variogramas, o que conduziu à decomposição da variância total em N^e termos - o efeito de pepita, o patamar da estrutura 1, o patamar da estrutura 2, etc. - pode agora interpretar-se cada uma das estruturas espaciais *per si*, do ponto de vista da Análise de Dados. Assim os factores resultantes da estrutura pepítica estão relacionados com determinado conjunto de variáveis, os que se obtêm a partir da 1ª estrutura com outro conjunto de variáveis, etc.

Enriquece-se deste modo a análise factorial global através da consideração da factorização das N^e estruturas diferentes, cada uma delas ligada a uma escala do fenómeno regionalizado.

Conforme já referido anteriormente (Cf. 3.2), quando as variáveis se apresentam em corregionalização intrínseca a diagonalização das N^e matrizes é equivalente à Análise em Componentes Principais da matriz variância-covariância.

REFERÊNCIAS

Aboufirassi, M., Mariño, M. A. (1984) - Cokriging of aquifer transmissivities from field measurements of transmissivity and specific capacity. *Mathematical Geology*, **16**(1):19-35.

Ahmed, S., Marsily, G. (1987) - Some Applications of Multivariate Kriging in Ground Water Hydrology. *Sciences de la Terre*.

Demidovich, B. P., Maron, I. A. (1976) - *Computational Mathematics*. Mir Publishers (Trad. inglesa), Moscow, 692 pp..

Dias Agudo, F. R. (1978) - *Introdução à Álgebra Linear e Geometria Analítica*. Livraria Escolar Editora, Lisboa, 147 p..

Guarascio, M. (1976) - Improving the Uranium Deposits Estimation (The Novazza Case). M. Guarascio *et al* (eds.), *Advanced Geostatistics in the Mining Industry,* Nato ASI Series C24, Reidel, Dordrecht, pp. 351-367.

Journel, A., Huijbregts, Ch. J. (1978) - *Mining Geostatistics*. Academic Press, London, 600 pp..

Magri, E. J. (1982) - Calculations of Grade and Tonnage for two co - products from a projected South African Gold Mine. *J. S. Afr. Inst. Min. Metal.*, Mar. 1982.

Matheron, G. (1982) - *Pour une Analyse Krigeante des Données Regionalisées*. Les Cahiers du CGMM, N. 732, Fontaineblau, 21 pp..

Pereira, H. G., Soares, A. O. (1989) - Application of geostatistics to groundfish survey data. M. Armstrong (ed.), *Geostatistics*, Kluwer Academic Publishers, Dordrecht, pp. 459-467.

Ribeiro, L. T., Muge, F. H. (1989) - A Geostatistical Approach to the Modelling of a Piezometric Field. M. Armstrong (ed.), *Geostatistics*, Kluwer Academic Publishers, Dordrecht, pp. 651-660.

Sousa, A. J. (1988a) - Factorial Kriging as a Method to include Spatial Structure into Classification. A Case Study on a Sulphide Orebody. *Quantitative Analysis of Mineral and Energy Resources*, NATO ASI, Series C 223, 385-392, Reidel, Dordrecht.

Sousa, A. J. (1988b) - *Análise de Dados e Geoestatística Multivariada. Aplicação à tipologia de minérios*. Tese, IST, Lisboa.

Sousa, A. J. (1989) - Geostatistical data analysis: an application to ore typologie. M. Armstrong (ed.), *Geostatistics*, Kluwer Academic Publishers, Dordrecht, pp. 851-860.

Wakernagel, H. (1985) - *L' Inférence d' un Modéle Lineaire en Géostatistique Multivariable*. Tese Docteur de 3eme Cycle, ENSMP, 100 pp..

Wackernagel, H. (1988) - Geostatistical Techniques for Interpreting Multivariate Spatial Information. *Quantitative Analysis of Mineral and Energy Resources*, NATO ASI Series, C 223, 393-409, Reidel, Dordrecht, pp. 393-409.

Wackernagel, H., Petitgas, P., Touffait, Y. (1989) - Overview of methods for co-regionalization analysis. M. Armstrong (ed.), *Geostatistics*, Kluwer Academic Publishers, Dordrecht, pp. 409-420.