

CVRM / Centro de Geo-sistemas

*ANÁLISE GEOESTATÍSTICA DE DADOS*

**António Jorge Sousa**

*Janeiro de 2000*

# ANÁLISE GEOESTATÍSTICA DE DADOS

## INTRODUÇÃO

Nas Ciências da Terra é usual a ocorrência de quadros multivariados de dados que podem ser organizados sob a forma de uma matriz de dimensão  $n \times p$  ( $n$  linhas por  $p$  colunas). A matriz de dados contém os valores observados em  $n$  objectos (unidades estatísticas ou amostras<sup>1</sup>). Cada um dos  $n$  objectos é caracterizado pelos valores respeitantes a  $p$  variáveis (atributos). As variáveis podem tomar valores numéricos, contínuos ou discretos, ou valores qualitativos (códigos). Trata-se, portanto, de quadros de dados do tipo dos tratados pelos procedimentos estatísticos multivariados. No entanto, apresentam, em geral, uma característica exclusiva que advém do facto de as variáveis serem medidas ou observadas em amostras recolhidas em pontos (ou volumes<sup>2</sup> centrados nesses pontos) localizados no espaço geográfico. Cada uma das amostras é, assim, caracterizada também por um conjunto de 1, 2 ou 3 coordenadas espaciais a que pode acrescer, em determinado tipo problemas, 1 coordenada temporal. Trata-se, portanto, de dados regionalizados, na medida em que os valores tomados pelas variáveis dependem da localização espacial.

Em geral, os quadros de dados são de grandes dimensões, pelo que é necessário encontrar métodos de análise que permitam extrair a informação relevante da massa *bruta* de dados.

Neste trabalho passa-se em revista diversas técnicas cujo objectivo é a descrição de quadros de dados. Particular atenção será dada aos métodos factoriais, de base geométrica, que permitem a visualização dos dados em espaços de dimensão reduzida, engendrada por um pequeno conjunto de factores que sintetizam os dados.

Na primeira parte do trabalho, relembra-se os fundamentos teóricos da Análise em Componentes Principais (ACP), método incluído na família dos métodos factoriais particularmente adaptado para extrair as características estruturais essenciais de um quadro de dados quantitativos.

No entanto a ACP deixa escapar uma característica crucial deste tipo de dados que é o seu carácter regionalizado. As variáveis apresentam, em geral, uma estrutura espacial que, salvo casos particulares, não é detectada em toda a sua complexidade por este método.

---

<sup>1</sup> A cada um dos elementos da amostra dá-se também, vulgarmente, o nome de amostra.

<sup>2</sup> Cada amostra é também caracterizada pelo volume (e/ou instante de tempo) onde a variável, é medida. A esta característica dá-se o nome de *suporte* da amostra.

Após revisão das principais medidas de continuidade espacial, introduz-se as três técnicas nucleares da Análise Geoestatística de Dados. Em comum baseiam-se na generalização da ACP por introdução da estrutura espacial das variáveis.

A primeira, denominada Análise Factorial de Corregionalizações, é uma técnica puramente descritiva, baseada na análise das medidas de continuidade experimental das variáveis.

A segunda, denominada Análise Factorial do Modelo Linear de Corregionalização, pressupõe a modelação prévia da estrutura espacial das variáveis, e baseia-se na análise das matrizes de corregionalização teóricas. Pretende identificar as características estruturais essenciais, denominados factores regionalizados e componentes espaciais, de um conjunto de dados regionalizados.

A Krigagem Factorial, ao permitir estimar aqueles factores e componentes identificados pela Análise Factorial do Modelo Linear de Corregionalização, possibilita a sua cartografia espacial, facilitando a interpretação.

As potencialidades de cada uma das técnicas apresentadas são evidenciadas através de casos de estudo ilustrativos.

A matriz de dados será representada pela matriz  $\mathbf{Z}$ , cujas colunas, representadas por  $\mathbf{z}_j$ ,  $j=1, \dots, p$ , correspondem às  $p$  variáveis. As coordenadas espaciais serão representadas pela matrix  $\mathbf{X}$ . As coordenadas espaciais da amostra  $i$  serão representados por  $\mathbf{x}_i$ . O elemento genérico de  $\mathbf{Z}$ , representado por  $z_{ij}$  ou  $z_j(\mathbf{x}_i)$ , corresponde ao valor da variável  $j$  medido ou observado na amostra  $i$  recolhida no ponto de coordenadas  $\mathbf{x}_i$ .

## **ANÁLISE EM COMPONENTES PRINCIPAIS**

A Análise em Componentes Principais foi formulada independentemente por Pearson (1901) e Hotteling (1933). Trata-se de uma técnica de análise de dados que pode ser introduzida recorrendo a conceitos geométricos, em que as variáveis originais, tomadas como eixos definindo um espaço onde se posicionam as amostras, são substituídas por outras variáveis (os factores), obtidos à custa de um novo sistema de eixos que se *ajusta melhor* à nuvem constituída pelas amostras. Esta é uma formulação mais próxima da de Pearson.

Em geral, as variáveis de base são muito heterogéneas, apresentando valores médios e variâncias muito diferentes, podendo estar expressas em unidades de medida de natureza diversa. Os resultados da análise destes dados brutos podem reflectir apenas o efeito das diferentes unidades (% , ppm, Kg, m, cm, ...) em que cada uma das variáveis foi medida.

Para reduzir este efeito é conveniente transformar a matriz de dados  $\mathbf{Z}$  numa outra  $\mathbf{T}$ , cujo elemento genérico  $t_{ij}$  é dado por:

$$t_{ij} = \frac{1}{\sqrt{n}} \frac{z_{ij} - \bar{z}_j}{s_j}$$

onde  $\bar{z}_j$  e  $s_j$  são, respectivamente, a média aritmética e o desvio padrão dos valores tomados pela variável  $j$ .

As novas variáveis  $t_j$  são adimensionais com média nula e variância unitária. A matriz  $\mathbf{T}$  pode ser assimilada a duas nuvens de pontos distintas conforme o espaço escolhido:

- Cada linha de  $\mathbf{T}$  é um vector que representa a posição de um indivíduo no espaço das variáveis  $R^p$ .
- Cada coluna de  $\mathbf{T}$  é um vector que representa a posição de uma variável no espaço dos indivíduos  $R^n$

Como já referido, o objectivo genérico dos métodos factoriais de análise de dados, em geral, e da Análise em Componentes Principais, em particular, é o de encontrar um novo sistema de eixos que se ajuste o melhor possível, de acordo com determinado critério, a estas nuvens de pontos.

No caso da análise em  $R^p$ , o critério dos mínimos quadrados conduz, a que os eixos factoriais sejam dados pelos vectores próprios  $\mathbf{u}_\alpha$  da chamada matriz de inércia da nuvem das amostras, dada por  $\mathbf{T}'\mathbf{T}$  (Pereira & Sousa, 1988). Estes novos eixos estão hierarquizados pelos valores próprios  $\lambda_\alpha$  da matriz de inércia.

Analogamente, no caso da análise em  $R^n$ , o critério dos mínimos quadrados conduz, a que os eixos factoriais sejam dados pelos vectores próprios  $\mathbf{v}_\alpha$  da chamada matriz de inércia das variáveis dada por  $\mathbf{T}\mathbf{T}'$  (Pereira & Sousa, 1988). Estes novos eixos estão hierarquizados pelos valores próprios  $\mu_\alpha$  da matriz de inércia.

Mostra-se que, admitindo que  $p < n$ , os  $p$  primeiros valores próprios de  $\mathbf{T}\mathbf{T}'$  são iguais aos valores próprios de  $\mathbf{T}'\mathbf{T}$  enquanto os restantes  $n-p$  são nulos. Por outro lado, prova-se (Pereira & Sousa, 1988) que os  $p$  primeiros eixos factoriais são colineares e estão relacionados pelas relações de transição seguintes:

$$\begin{aligned} \mathbf{v}_\alpha &= \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{T} \mathbf{u}_\alpha \\ \mathbf{u}_\alpha &= \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{T}' \mathbf{v}_\alpha \end{aligned} \tag{1}$$

Os primeiros eixos factoriais, cujo número escolhido de acordo com determinados critérios, deve explicar uma proporção importante da inércia total (Pereira & Sousa, 1988), representam os traços principais presentes no quadro de dados. A análise da projecção das amostras e das variáveis nos planos formados pelos primeiros eixos factoriais, permite evidenciar, através das proximidades e oposições entre variáveis e entre amostras, as principais estruturas presentes. É de notar que as relações de transição permitem a projecção simultânea das amostras e das variáveis nos mesmos planos factoriais. Esta possibilidade, que apresenta vantagens evidentes na interpretação dos resultados, nem sempre é aproveitada devidamente.

Conclui-se, portanto, que a Análise em Componentes Principais se baseia na extracção dos vectores e valores próprios da matriz de inércia  $\mathbf{T}'\mathbf{T}$  cujo elemento genérico é dado por:

$$r_{jj'} = \sum_{i=1}^n t_{ij} t_{ij'} = \frac{1}{n} \sum_{i=1}^n \frac{(z_{ij} - \bar{z}_j)(z_{ij'} - \bar{z}_{j'})}{s_j s_{j'}}$$

A matriz de inércia consiste, portanto, na matriz de correlação entre variáveis.

A Análise em Componentes Principais foi utilizada para descrever a estrutura de um conjunto de 491 amostras do jazigo de Moinho, onde são conhecidas os teores de 6 elementos químicos: S, Fe, Cu, Zn, Pb, As (Sousa, 1989a).

Nas figuras 1 e 2 apresenta-se, respectivamente, as projecções das variáveis e das amostras no primeiro plano factorial

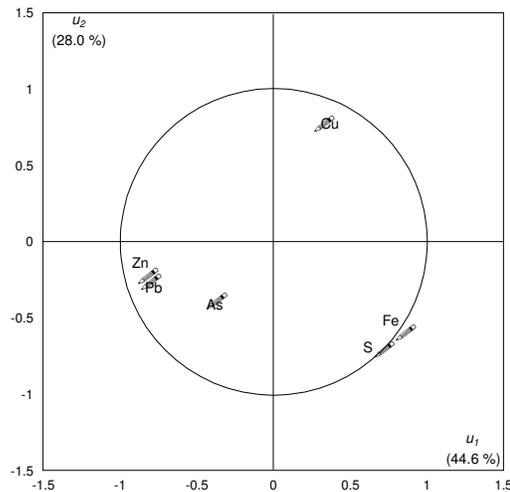


Figura 1 – Projecção das variáveis no 1º plano factorial

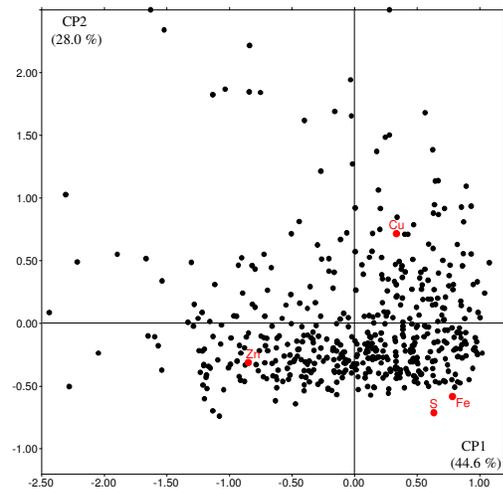


Figura 1 - Projecção das amostras no 1º plano factorial

Verifica-se que o Zn e o Pb aparecem juntos, reflectindo a elevada correlação existente entre eles, o mesmo acontecendo com o S e o Fe. A oposição entre estas variáveis, explicam o eixo 1 (qualquer uma destas variáveis apresenta coordenadas

neste eixo com valores absolutos muito próximo de 1, traduzindo portanto correlações muito elevadas com este eixo). O Cu é o responsável pelo 2º eixo factorial, evidenciando um comportamento independente face aos pares de variáveis referidos anteriormente. Relativamente ao As, mal representado neste plano, nada se pode concluir.

No caso das amostras verifica-se que estas formam uma nuvem contínua de pontos, densa no 4º quadrante, correspondendo a valores elevados de S e Fe (a projecção simultânea das variáveis facilita a leitura do gráfico), dispersando-se ao longo do eixo 1 e do eixo 2.

### MEDIDAS DE CONTINUIDADE ESPACIAL DAS VARIÁVEIS

Salvo em casos muito particulares e raros, a correlação entre pares de variáveis, porque não toma em atenção a sua estrutura espacial, é um parâmetro demasiado pobre para medir a dependência (espacial) entre essas variáveis. É necessário introduzir outras medidas que tomem em conta o carácter regionalizado das variáveis. Usualmente, é costume definir três funções que permitem quantificar a variação conjunta de variáveis regionalizadas: a covariância espacial cruzada, a correlação espacial cruzada e o variograma cruzado.

Estas medidas baseiam-se nos valores das variáveis medidos em pares de amostras separadas geograficamente pelo vector  $\mathbf{h}$  (ver figura 3).

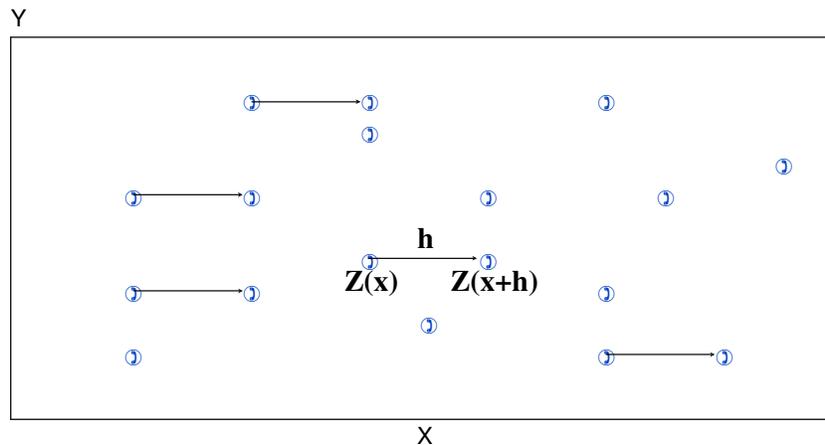


Figura 3 – Representação esquemática de pares de amostras separadas pelo vector  $\mathbf{h}$ .

Para os valores da variável  $j$  medidos nas amostras iniciais de cada par, pode-se calcular os valores médios  $m_j(-\mathbf{h})$  e as variâncias  $s_j^2(-\mathbf{h})$  :

$$m_j(-\mathbf{h}) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} Z(\mathbf{x}_i) \quad s_j^2(-\mathbf{h}) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} [Z_j(\mathbf{x}_i) - m_j(-\mathbf{h})]^2$$

Analogamente, pode-se calcular os valores médios  $m_{j'}(+\mathbf{h})$  e as variâncias  $\sigma_{j'}^2(+\mathbf{h})$  dos valores da variável  $j'$  medidos nas amostras finais de cada par:

$$m_{j'}(+\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} Z_{j'}(\mathbf{x}_i + \mathbf{h}) \quad s_{j'}^2(+\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [Z_{j'}(\mathbf{x}_i) - m_{j'}(+\mathbf{h})]^2$$

A partir destes parâmetros pode-se definir as medidas de continuidade espacial<sup>3</sup> referidas anteriormente:

#### *Covariância Espacial Cruzada*

$$C_{jj'}^*(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} Z_j(\mathbf{x}_i) Z_{j'}(\mathbf{x}_i + \mathbf{h}) - m_j(-\mathbf{h}) m_{j'}(+\mathbf{h})$$

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

#### *Correlação Espacial Cruzada (Correlograma cruzado)*

$$r_{jj'}^*(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} \frac{C_{jj'}^*(\mathbf{h})}{\sqrt{s_j^2(-\mathbf{h}) s_{j'}^2(+\mathbf{h})}}$$

#### *Variograma Cruzado*

$$\gamma_{jj'}^*(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [Z_j(\mathbf{x}_i + \mathbf{h}) - Z_j(\mathbf{x}_i)] [Z_{j'}(\mathbf{x}_i + \mathbf{h}) - Z_{j'}(\mathbf{x}_i)]$$

Verifica-se que, ao contrário do variograma que é sempre uma função simétrica, as duas primeiras funções são, em geral, assimétricas, isto é:  $C_{jj'}^*(\mathbf{h}) \neq C_{jj'}^*(-\mathbf{h})$  e  $r_{jj'}^*(\mathbf{h}) \neq r_{jj'}^*(-\mathbf{h})$ .

Note-se que, fazendo  $j=j'$ , encontra-se as medidas de continuidade espacial simples (monovariadas). Neste caso, é fácil verificar que as três funções são simétricas.

As medidas de continuidade espacial simples e cruzadas são poderosos instrumentos de detecção das estruturas espaciais evidenciados pelos dados. Assim, é possível verificar a presença ou ausência de anisotropias direccionais, a homogeneidade espacial e o grau de continuidade espacial das variáveis (Sousa & Muge, 1990; Goovaerts, 1997).

O coeficiente de correlação clássico corresponde à correlação cruzada quando  $\mathbf{h}=0$ . Estruturalmente é, portanto, uma medida bastante mais pobre, como se ilustra no exemplo seguinte.

Em determinada estação mediu-se, diariamente ao longo de 2 anos, os valores da pluviosidade e da piezometria de um determinado aquífero. Naturalmente a

<sup>3</sup> Utiliza-se o sinal \* para simbolizar que as medidas referidas são amostrais (experimentais).

piezometria é influenciada pela pluviosidade, no entanto o coeficiente de correlação calculado é de apenas 0.344. Na figura 4 apresenta-se o correlograma cruzado das duas variáveis. Verifica-se que existe um atraso de aproximadamente 20 dias no impacto da pluviosidade sobre a piezometria. O valor da correlação é máximo para  $h=20$ , atingindo o valor de 0.843. Verifica-se ainda que a função não é simétrica em torno da origem, embora seja aparente uma simetria relativamente a  $h=20$ .

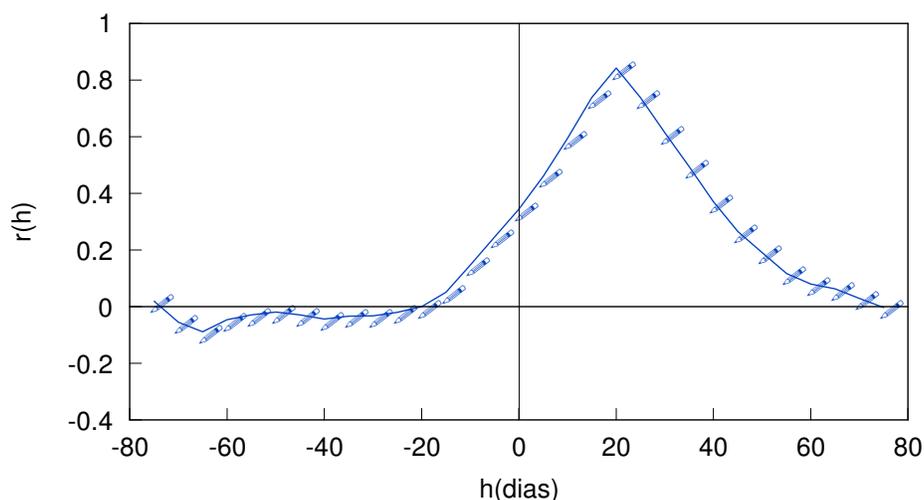


Figura 4 – Correlograma das variáveis piezometria e pluviosidade.

Em geoquímica é vulgar encontrar situações semelhantes, provocada pela diferente mobilidade dos elementos químicos no meio.

Verifica-se, portanto, que a Análise em Componentes Principais, porque baseada na correlação a passo zero ( $h=0$ ) entre as variáveis é insuficiente para sintetizar a estrutura espacial das variáveis, devendo ser complementada com a análise das matrizes de correlação espacial para diferentes passos  $h$ . A esta técnica dá-se o nome de Análise Factorial de Corregionalizações.

A Análise Factorial de Corregionalizações foi utilizada para descrever a estrutura espacial do conjunto de dados de Moinho já referidos. Na figura 5 apresenta-se os resultados obtidos para  $h=10$  e  $h=100$ . Verifica-se que os resultados são semelhantes entre si e aos da Análise em Componentes Principais. A principal diferença reside no facto de quando  $h$  aumenta o Cu tender a se opor ao par formado pelas variáveis Zn e Pb (para  $h=100$  esta oposição é responsável pela formação do 1º eixo, enquanto o par S, Fe explica o eixo 2). Para  $h=10$ , nota-se que o 2º eixo factorial é explicado pelo As, enquanto o Cu está mal representado neste plano. Verifica-se, portanto, que a estrutura das variáveis varia, significativamente, com  $h$ .

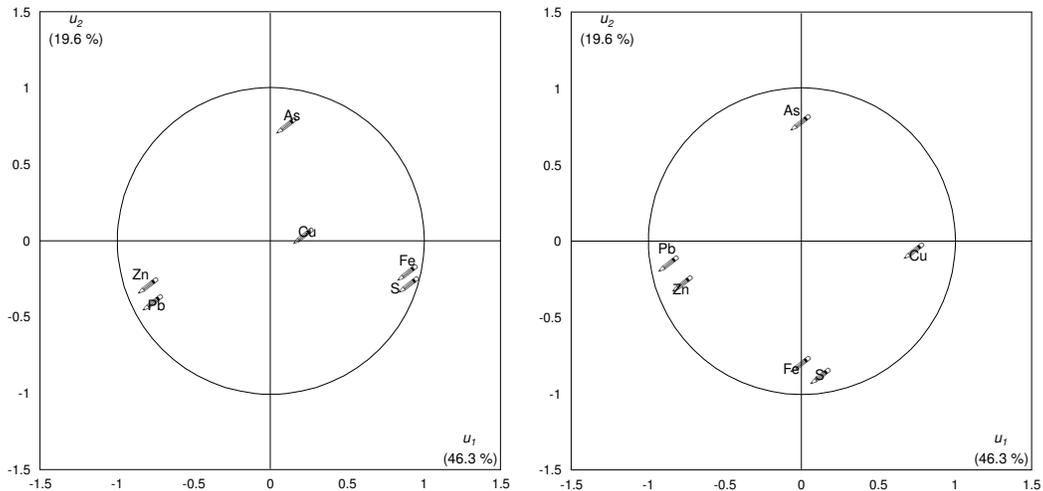


Figura 5 - Análise Factorial das Correlações Espaciais (Moinho) para h=10 e h=100

## INTRODUÇÃO À TEORIA DAS VARIÁVEIS REGIONALIZADAS

Raramente a descrição dos dados regionalizados constitui o objectivo final de um estudo estatístico. Habitualmente, pretende-se caracterizar a população de que o conjunto de dados constitui uma amostra. Torna-se, então, necessário encontrar um modelo que reproduza as principais características dessa população. É habitual considerar as variáveis regionalizadas como uma realização de uma função aleatórias.

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

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

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. Neste trabalho considera-se apenas dois tipos de estacionaridade: estacionaridades de 2ª ordem e intrínseca (Sousa & Muge, 1990; Goovaerts, 1997).

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

$$\begin{aligned} E\{Z_j(x)\} &= m_j \quad \forall x \\ E\{[Z_j(x) - m_j][Z_{j'}(x) - m_{j'}]\} &= C_{jj'}(\mathbf{h}) \end{aligned} \quad (2)$$

$$E\{Z_j(x+\mathbf{h})Z_j(x)\} - m_j^2 = C_j(\mathbf{h}) \quad \forall x$$

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

$$\text{Var}\{Z_j(x)\} = C_j(0)$$

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 correlogramização o seja, pois a covariância cruzada  $C_{jj'}(\mathbf{h})$  pode não depender apenas do vector  $\mathbf{h}$ .

A hipótese intrínseca impõe a estacionaridade dos dois primeiros momentos dos acréscimos espaciais. Assim, sendo os acréscimos a passo  $\mathbf{h}$  dados por:

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

aquelas duas condições de estacionaridade são expressas por:

$$E\{\varepsilon_j(\mathbf{h})\} = 0$$

$$E\{\varepsilon_j(\mathbf{h})\varepsilon_{j'}(\mathbf{h})\} = 2\gamma_{jj'}(\mathbf{h}) \quad \forall x$$

$$E\{[\varepsilon_j(\mathbf{h})]^2\} = 2\gamma_j(\mathbf{h}) \quad \forall x$$

O variograma cruzado entre as variáveis  $Z_j(x)$  e  $Z_{j'}(x)$  é dado por:

$$\gamma_{jj'}(\mathbf{h}) = \frac{1}{2} E\{\varepsilon_j(\mathbf{h})\varepsilon_{j'}(\mathbf{h})\}$$

No caso de estacionaridade de 2ª ordem, o variograma e as covariâncias verificam a seguinte relação, que se obtém desenvolvendo (2):

$$\gamma_{jj'}(\mathbf{h}) = C_{jj'}(0) - \frac{1}{2} [C_{jj'}(\mathbf{h}) + C_{j'j}(\mathbf{h})]$$

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. No entanto, o variograma<sup>4</sup> 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  $\mathbf{h}$ , que tem de ser, também, estimada).

---

<sup>4</sup> O variograma é também uma função mais geral do que a covariância ou a correlação espaciais, na medida em que estas duas só têm existência na hipótese de estacionaridade de 2ª ordem, enquanto o variograma é também definido na hipótese intrínseca. Justifica-se, assim, o uso generalizado do variograma em geoestatística.

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:

$$\gamma_{jj'}^*(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_1^{N(\mathbf{h})} [Z_j(x_i + \mathbf{h}) - Z_j(x_i)] [Z_{j'}(x_i + \mathbf{h}) - Z_{j'}(x_i)]$$

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

## ANÁLISE FACTORIAL DO MODELO LINEAR DE CORREGIONALIZAÇÃO

São frequentes as correionalizações que podem ser modeladas pela soma de várias 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 pode reduzir-se à sobreposição de  $N^e$  estruturas imbricadas, cada uma delas modelada por um esquema teórico (esférico, gaussiano, linear, etc.). Cada variograma é dado por:

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

No quadro do modelo linear de correionalizações (Goovaerts, 1997) para cada escala  $u$ , a forma e a amplitude dos variogramas são as mesmas para todas as variáveis.

Neste modelo, os variogramas simples e cruzados  $\gamma_{jj'}(\mathbf{h})$  são dados por:

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

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  $(p \times p)$ , cuja soma, no caso de estacionaridade de 2ª ordem, reproduz a matriz variância-covariância dos dados:

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

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

$$Z_j(x) = \sum_{u=0}^{N^e-1} \sum_{l=1}^p e_{jl}^u y_l^u(x)$$

em que os coeficientes verificam a relação seguinte:

$$b_{jj'}^u = \sum_{l=1}^p e_{jl}^u e_{j'l}^u$$

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

$$e_{jl}^u = \sqrt{\mu_l} v_{jl}$$

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

O modelo linear conduz à decomposição da correionalização numa série de  $N^e$  estruturas associadas a diferentes escalas. Para cada escala, obteve-se uma matriz variância-covariância, de termo geral  $b_{jj'}^u$ , que dá conta do sistema de interdependências entre as  $p$  variáveis para a escala  $u$ . No caso de estacionaridade de 2ª ordem, pode-se, também, definir a matriz de correlação  $\mathbf{R}^u$ , relativa a cada escala.

Em paralelo com a Análise em Componentes Principais (baseada na matriz das correlações experimentais), pode agora factorizar-se cada uma das matrizes  $\mathbf{R}^u$ , obtendo-se assim uma síntese da estrutura das variáveis, relacionada agora com a escala espacial do fenómeno. A esta técnica, introduzida por Wackernagel (1988), dá-se o nome de Análise Factorial do Modelo Linear de Correionalização.

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.

No caso de estudo do jazigo de Moinho apresentado anteriormente, os variogramas simples e cruzados foram modelados por um modelo linear de correionalização constituído pela sobreposição de 3 estruturas: uma pepítica e duas esféricas, com amplitudes de 36 e 120 metros, respectivamente (Sousa, 1989a). Nas figuras 6 a 9 apresenta-se os resultados da Análise Factorial do Modelo Linear de Correionalização e na figura 10, para comparação, repete-se os resultados da Análise em Componentes Principais.

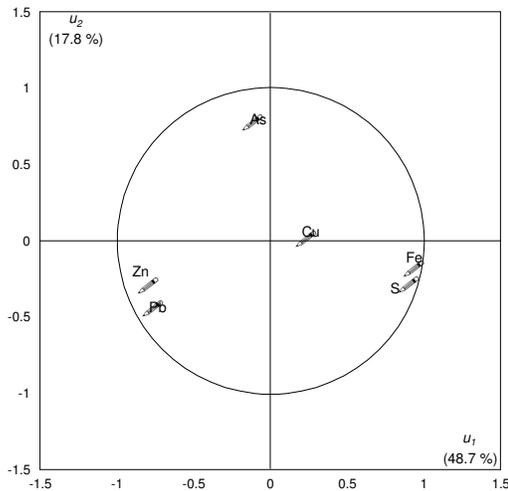


Figura 6 - Projecção das variáveis no 1º plano factorial (efeito de pepita)

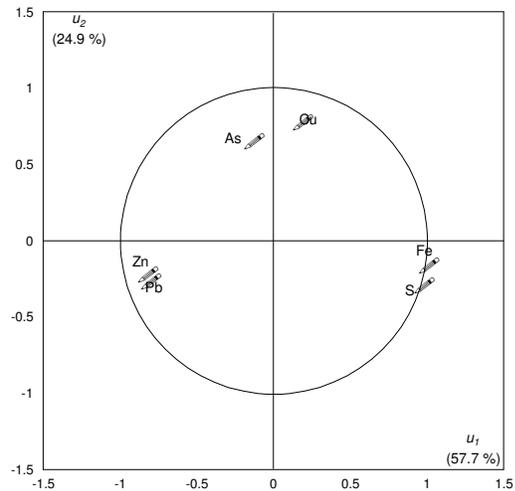


Figura 7 - Projecção das variáveis no 1º plano factorial (1ª estrutura)

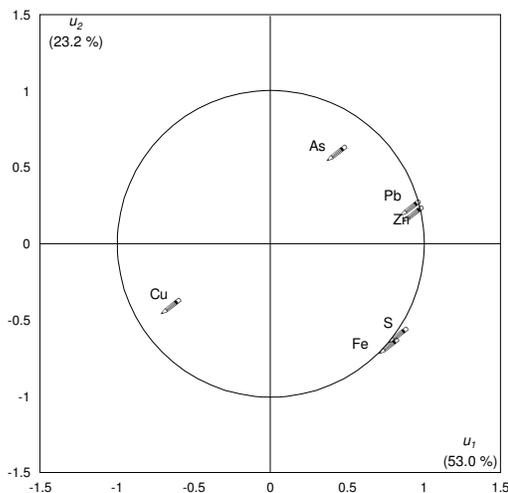


Figura 9 - Projecção das variáveis no 1º plano factorial (2ª estrutura)

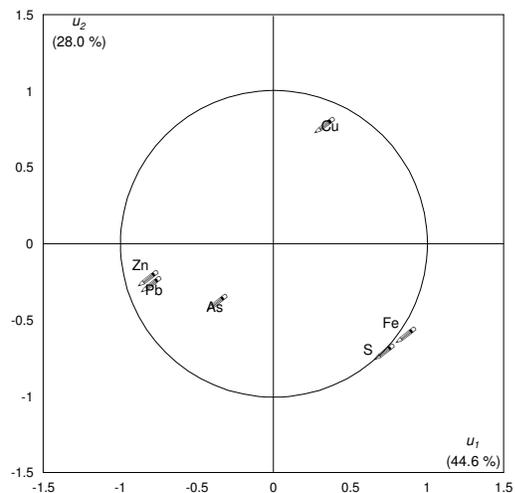


Figura 10 - Projecção das variáveis no 1º plano factorial (ACP)

Os resultados obtidos para cada uma das estruturas mostram, no caso do efeito de pepita e da 1ª estrutura, os mesmos traços gerais evidenciados pela Análise em Componentes Principais (Sousa, 1989a). As projecções das variáveis no 1º plano factorial da 2ª estrutura apresenta, no entanto, uma relação nova: O Cu passa a mostrar, a esta escala, uma elevada correlação negativa com o par constituído pelo Zn e Pb, isto é, na estrutura mais contínua os teores em cobre seguem, de forma inversa, os teores em zinco e chumbo: *o comportamento conjunto destas três variáveis depende da escala considerada*. Esta conclusão, que não pode ser retirada dos resultados da ACP, está de acordo com a ideia generalizada de que, no jazigo de Moinho, o Cu se concentrou preferencialmente a muro enquanto o Zn e Pb ocorrem predominantemente a tecto da massa. No entanto, *localmente*, o Cu tem um comportamento que é independente dos valores de Zn e Pb.

## KRIGAGEM FACTORIAL

De acordo com o modelo linear de correionalização, as variáveis regionalizadas podem ser interpretadas como combinações lineares de um conjunto de factores regionalizados ortogonais:

$$Z_j(x) = \sum_{u=0}^{N_e-1} \sum_{l=1}^p e_{jl}^u y_l^u(x)$$

Estes factores regionalizados podem ser estimados a partir dos valores amostrais pelo seguinte estimador (Wackernagel, 1995):

$$y_l^{u*}(x) = \sum_{j=1}^p \sum_{\alpha=1}^n \lambda_{j\alpha}^{ul} z_{\alpha}(x_{\alpha})$$

utilizando o sistema de Cokrigagem Factorial (Wackernagel, 1995):

$$\begin{cases} \sum_{j=1}^p \sum_{\beta=1}^{n_j} \lambda_{j\beta}^{ul} \gamma_{j\beta}^{u'}(x_{\alpha} - x_{\beta}) + \mu_{j'} = e_{jl}^u \gamma^{u'}(x_{\alpha} - x_0) & \forall \alpha = 1, \dots, n; \forall j' = 1, \dots, p \\ \sum_{\beta=1}^{n_j} \lambda_{j\beta}^{ul} = 0 & \forall j = 1, \dots, p \end{cases}$$

Por sua vez, evidenciando os variograma a presença de  $N_e$  estruturas, é natural modelar cada uma das variáveis  $Z_j(x)$  pela soma de  $N_e$  componentes espaciais, independentes entre si, cada uma das quais corresponde a uma determinada escala, a que se sobrepõe a média:

$$Z_j(x) = \sum_{u=0}^{N_e-1} Z_j^u(x) + m_j$$

A krigagem factorial (Goovaerts, 1997) permite estimar linearmente aquelas componentes, em cada ponto  $x_0$ :

$$Z^{u*}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i)$$

em que  $n$  é o número de amostras utilizadas na estimação de  $Z^{u*}(x_0)$ . Os ponderadores  $\lambda_{\alpha}$  são obtidos resolvendo o sistema de krigagem factorial:

$$\begin{cases} \sum_{i=1}^n \lambda_{\alpha} \gamma(x_i - x_{i'}) + \mu = \gamma^{u'}(x_{i'} - x_0) \\ \sum_{i=1}^n \lambda_{\alpha} = 0 \end{cases}$$

em que  $\gamma(x_\alpha - x_\beta)$  representa o valor do variograma entre  $x_\alpha$  e  $x_\beta$ .

As componentes espaciais  $Z^u(x_i)$  assim estimadas podem ser cartografadas separadamente, permitindo identificar espacialmente as zonas com maior afinidade com cada uma das escalas de variabilidade.

A krigagem factorial, foi utilizada no tratamento dos teores em Zn e Cu nos solos, obtidos em diversas campanhas de prospecção efectuadas pelo Serviço de Fomento Mineiro numa área com cerca de 50 km<sup>2</sup> localizada no Alentejo entre as vilas de Moura e Ficalho.

A análise dos variogramas mostra, que podem ser modeladas pela soma de efeito de pepita e duas estruturas esféricas (Sousa, 1995). Na figura 11 e 12 estão representadas, as componentes espaciais correspondentes às varias estruturas do Zn e do Cu, respectivamente.

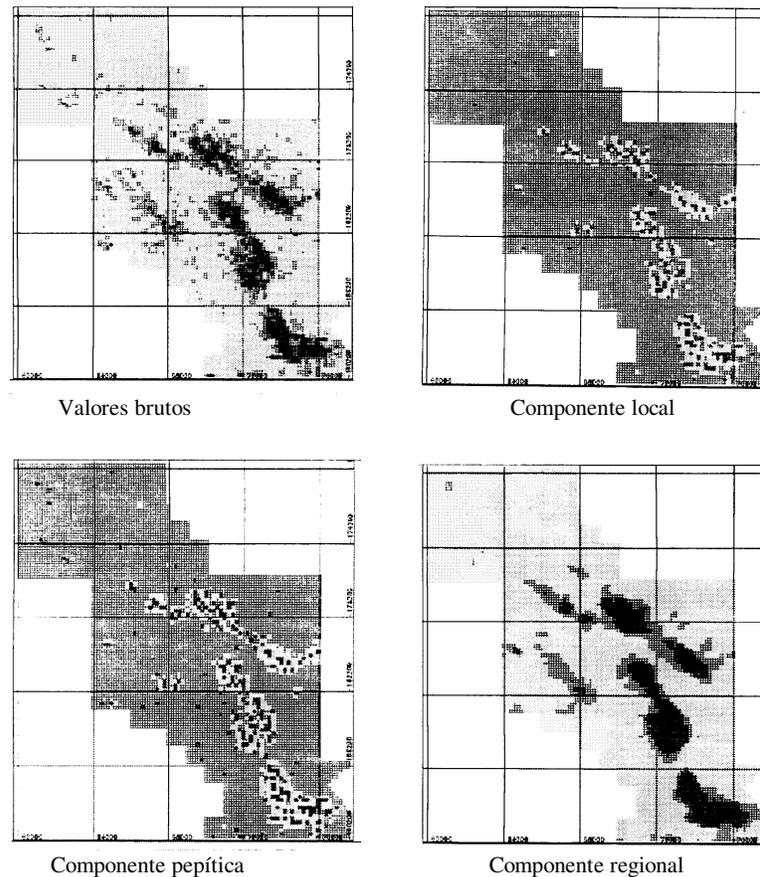


Figura 11 - Cartografia do zinco

Verifica-se que a cartografia da componente regional do Zn marca claramente as principais estruturas e litologias carbonatadas presentes; a distribuição regional do Zn

está relacionada principalmente com as dolomias câmbrias que apresentam fundos muito altos, contrastando com os baixos fundos da cobertura cenozoica e fundos intermédios de outras litologias. As cartografias das componentes local e pepítica do Zn não evidenciam relação nítida com a litologia e definem bem as estruturas mineralizadas ricas em zinco e chumbo (Preguiça, Vila Ruiva, Carrasca SE, Vale de Grou e Merlinha). A mineralização de Enfermarias é assinalada apenas na cartografia da componente pepítica.

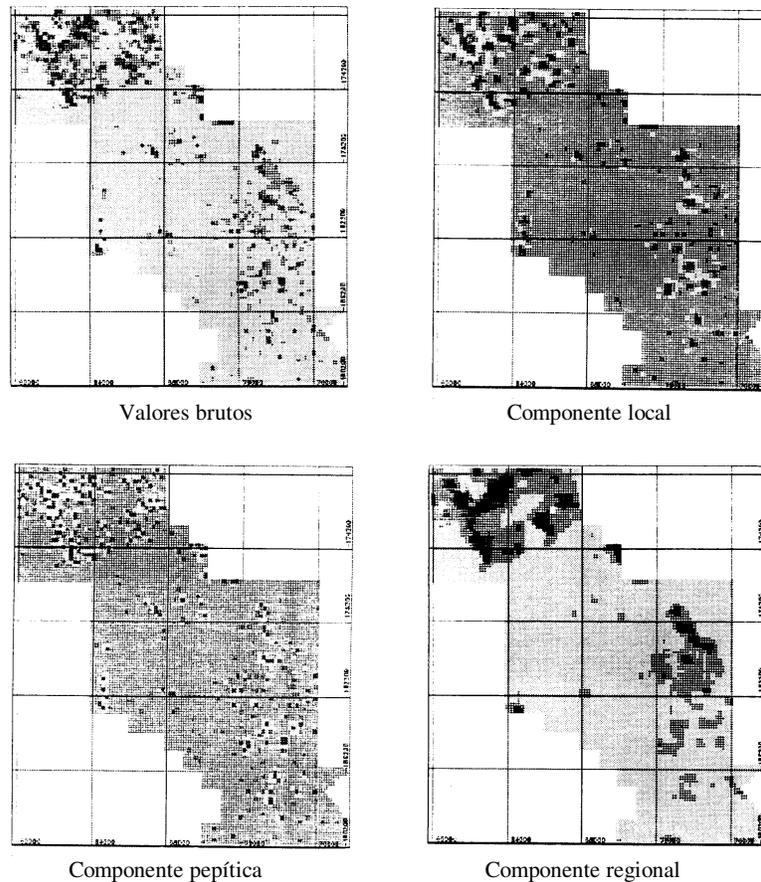


Figura 12 - Cartografia do Cobre

A cartografia da componente regional do Cu marca muito bem as litologias a que o Cobre anda associado (rochas vulcânicas e xistos do Silúrico) assim como a estrutura das Enfermarias. As minas da Preguiça e de Rui Gomes estão bem marcadas nas cartografias da componente local e pepítica.

## BIBLIOGRAFIA

- Cressie, N. (1991) - *Statistics for spatial data analysis*. Wiley-Interscience, 900 pp.
- Goovaerts, P. (1997) - *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York, 561 pp.

- Pereira, H. G. & Sousa, A. J. (1988) - *Análise de Dados para o Tratamento de Quadros Multidimensionais*. CVRMUTL, IST, 115p.
- Sousa, A. J. (1989a) - Geostatistical Data Analysis: An Application to Ore Typology. M. Armstrong (Ed.), *Geostatistics*, 851-860, Kluwer Academic Publishers.
- Sousa, A. J. (1989b) - *Geoestatística Multivariada*. LMPM, IST, 12p.
- Sousa, A. J. & Muge, F. (1990) - *Elementos de Geoestatística*. LMPM, IST, 63p.
- Sousa, A. J. (1994) - Contribuição da Krigagem Factorial para a Cartografia Geoquímica. *Geo-Sistemas*, **3**, 133-142, 1994.
- 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. (1995) - *Multivariate Geostatistics*. Springer, 220 p.