

## **APLICAÇÃO DA ANÁLISE DISCRIMINANTE À DADOS MULTIVARIADOS REGIONALIZADOS**

**Paulo Milton Barbosa Landim** (Universidade Estadual Paulista, campus de Rio Claro)  
13506-900 Rio Claro(SP)/Brasil - [plandim@rc.unesp.br](mailto:plandim@rc.unesp.br)

**Fernando António Leal Pacheco** (Universidade de Trás-os-Montes e Alto Douro)  
5000 Vila Real/Portugal – [fpacheco@utad.pt](mailto:fpacheco@utad.pt)

### **Introdução**

Normalmente os estudos geológicos são baseados em coleta de amostras, porém os métodos clássicos da análise estatística multivariada não levam em consideração a posição espacial dos pontos de amostragem e, conseqüentemente, suas relações. Por outro lado a metodologia geoestatística multivariada, como a co-krigagem e a krigagem fatorial, apresenta essas propriedades, porém nem sempre é de fácil aplicação. Ferramentas outras se tornam, então, necessárias para ressaltar os padrões de classificação regionalizada por acaso existentes e daí a necessidade de métodos estatísticos que enfoquem a análise espacial de dados geológicos multivariados. Podem ser adaptações de métodos estatísticos multidimensionais procurando modelar espacial ou cronologicamente fenômenos geológicos. Isso, porém, somente é possível se a matriz de dados multidimensionais apresentar também as coordenadas geográficas dos pontos amostrados.

Os dados precisam ser entendidos, tanto espaciais como cronologicamente, em sua estrutura multidimensional e existem a disposição, graças à moderna tecnologia computacional, ferramentas gráficas que permitem a visualização simultânea de amostras no espaço e/ou no tempo e as primeiras idéias a respeito da estrutura multidimensional podem começar a surgir a partir dessas exibições gráficas. Os produtos gráficos obtidos a partir das informações numéricas são avaliados levando em consideração tanto o conhecimento já adquirido com dados similares como com fatos científicos relacionados às variáveis sob estudo. A interpretação da estrutura espacial ou temporal, as associações e as relações casuais entre variáveis devem, então, serem organizadas num modelo que se ajuste. A modelagem, se correta, não apenas descreve o fenômeno nos locais amostrados, mas pode se tornar válida para interpolações em locais ou intervalos de tempo adjacentes, não amostrados, representando um passo além com relação às informações contidas nos dados numéricos. Na verdade este é o grande desafio da análise multivariada de dados espaciais, a estimação de valores para situações de previsão quantitativa.

No presente estudo aplicou-se a técnica estatística multivariada da análise discriminante para classificar e mapear, em termos probabilísticos, as influências principais que teriam agido no controle da composição química de águas subterrâneas de uma região granitóide no centro-leste de Portugal, conhecida como plutonito do Fundão.

## Material e Métodos

A matriz de dados analisada é composta por 160 análises geoquímicas de águas subterrâneas provenientes da citada região com teores, em ml/l, de  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{Ca}^{+2}$ ,  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{NO}_3^-$ ,  $\text{HCO}_3^-$  e  $\text{SiO}_2$ . (Figura 1a). Tais dados já foram analisados por Pacheco (1998) que utilizou a análise das correspondências para determinar as principais influências sobre a composição química dessas águas. Segundo esse estudo, que conseguiu identificar águas poluídas e não poluídas, por regressão linear múltipla, três seriam os fatores controladores: alteração por intemperismo, contaminação agrícola e contaminação doméstica (Figura 1b).

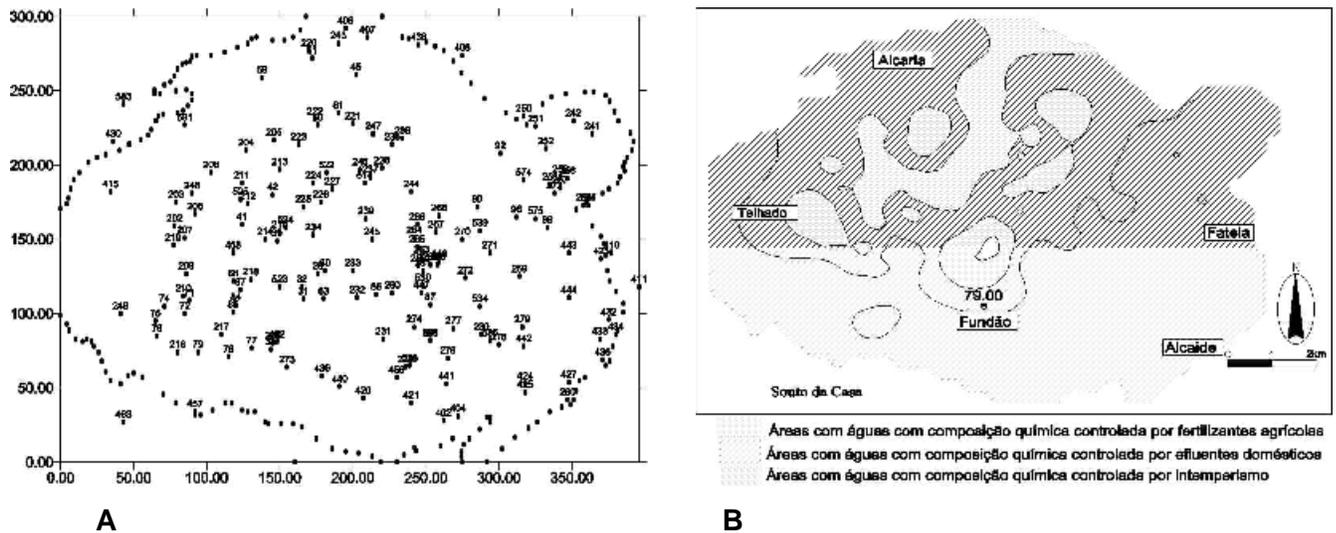


Figura 1. a) Pontos de coleta e delimitação do plutonito; b) distribuição das águas (Pacheco,1998)

Inicialmente aplicou-se uma análise de agrupamentos aos dados. Tal análise é uma técnica numérica cujo propósito fundamental é classificar os valores de uma matriz de dados multidimensionais sob estudo em grupos discretos. O método utilizado foi o do agrupamento hierárquico e o coeficiente de similaridade adotado, para o procedimento aglomerativo, foi o de Ward, que se baseia no agrupamento pela variância mínima, onde o enfoque é sobre a variabilidade que existe dentro de cada caso e os agrupamentos são efetuados ao se determinar que pares de casos, quando tomados em conjunto, apresentam o menor acréscimo de variabilidade.

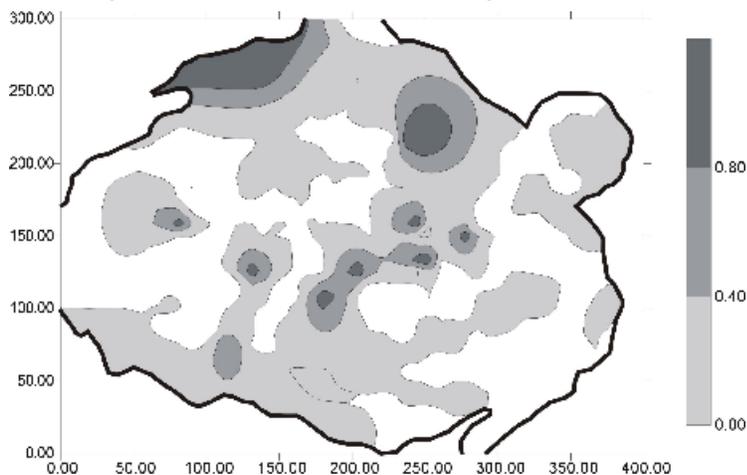
Depois de obtido o dendrograma resultante e verificada a presença de grupos, os mesmos foram submetidos à análise discriminante multigrupos. Nessa análise a matriz inicial de todas as variâncias e covariâncias pode ser parcializada entre categorias ou grupos e verificada a soma total de quadrados, a soma de quadrados entre grupos e a soma de quadrados dentro dos grupos. Isso significa que para a aplicação desta metodologia é essencial que os grupos já estejam previamente

definidos. Como no caso da análise de variância convencional a soma de quadrados entre grupos [E] mais a soma de quadrados dentro dos grupos [D] é igual à soma total de quadrados [T]:  $[T] = [E] + [D]$

Quando a razão  $[E]/[D]$  apresentar um valor alto isto significará que as médias dos grupos são bem diferentes entre si e os valores dentro de cada grupo estão bem concentrados ao redor dos respectivos centróides, ou seja, há uma discriminação significativa entre os grupos definidos. Dentro de cada grupo é possível, portanto, calcular a probabilidade de cada elemento pertencer ao respectivo grupo. De posse dessas probabilidades e associadas às respectivas coordenadas geográficas dos pontos, mapas de distribuição são confeccionados para indicar a distribuição espacial de cada um dos grupos.

## RESULTADOS

A análise de agrupamentos revelou três grupos, que foram interpretados de acordo com o trabalho de Pacheco (1998), como águas não poluídas em que a composição é influenciada fundamentalmente pela alteração das rochas e águas poluídas, seja por efluentes domésticos seja por contaminação de defensivos agrícolas. Os mapas de probabilidade de distribuição de cada um desses grupos são mostrados nas Figuras 2, 3 e 4.

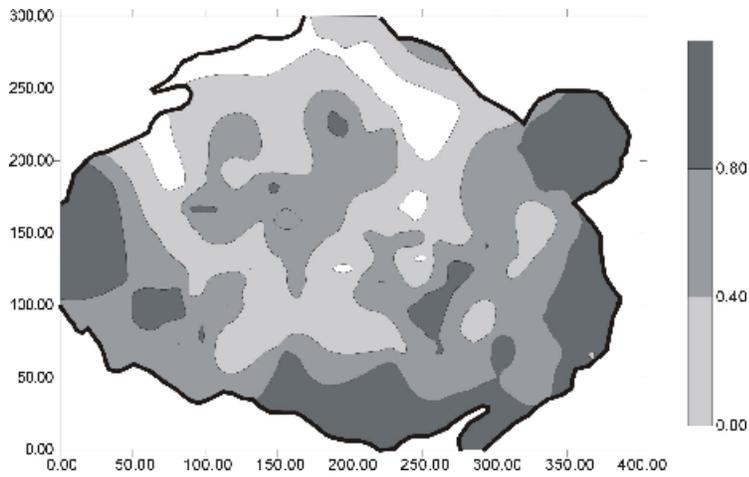


**Figura 2: Probabilidades de distribuição das águas com composição química controlada por fertilizantes agrícolas**

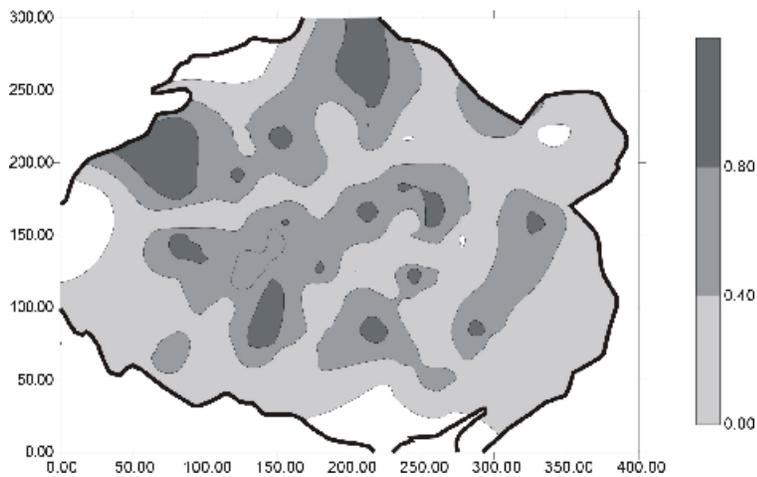
## CONCLUSÕES

Os resultados aqui obtidos condizem perfeitamente com os apresentados por Pacheco (1998), que utilizou de um enfoque diferente deste. Pelo exposto verifica-se o potencial de aplicação da análise discriminante em dados multivariados georreferenciados. É importante ressaltar que a possibilidade de sucesso ao aplicar tal metodologia reside no fato que antes da aplicação dessa análise os grupos já devem ser previamente conhecidos, seja por manipulação matemática, como no

presente caso, seja por um conhecimento “a priori”, que pode incluir, inclusive, dados qualitativos (soft data).



**Figura 3: Probabilidades de distribuição das águas com composição química controlada por intemperismo**



**Figura 4: Probabilidades de distribuição das águas com composição química controlada por efluentes domésticos**

#### REFERÊNCIA BIBLIOGRÁFICA

PACHECO, F. A. L. (1998) – *Application of Correspondence Analysis in the Assessment of Groundwater Chemistry: Mathematical Geology*, **30**(2):129-161