

# Taxonomia

## – um algoritmo simples para a classificação de resultados

António João de Carvalho da Cruz <sup>a</sup>

A interpretação de resultados experimentais - actividade que faz parte integrante da química [1] - requer técnicas de análise ajustadas à natureza dos resultados disponíveis, por um lado, e ao objectivo do estudo, por outro [2].

Por isso, quando mais do que uma propriedade é determinada para um sistema é indispensável a utilização de métodos multi-paramétricos [2].

Entre as vantagens desses métodos, pode referir-se aquelas que dizem respeito à redução do volume de dados que é necessário interpretar e as que resultam da possibilidade de representação gráfica a duas dimensões do que, sem a análise multi-paramétrica, para N propriedades, só pode ser graficamente representado a N dimensões [3, 5-6].

A regressão multi-linear, a análise de componentes principais, a análise de factores e a taxonomia numérica – são alguns dos métodos multi-paramétricos mais correntemente utilizados, para os quais existem diversos algoritmos possíveis [3, 5-6]. À excepção dos métodos de taxonomia numérica, todos os outros recorrem à álgebra matricial (inversão de matrizes, determinação de valores próprios e de vectores próprios, por exemplo), o que exige computadores de grande poder de cálculo – pelo menos, em comparação com as necessidades de alguns dos métodos de taxonomia numérica.

O algoritmo TAXONOMIA, aqui representado, embora considere várias opções, é facilmente transposto para uma linguagem de programação (BASIC, por exemplo) e apresenta como características o facto de não necessitar de grande quantidade de memória nem de tempo de computação elevado. Um micro-computador de pequeno porte como o Zx Spectrum é, em princípio, suficiente para qualquer utilização corrente.

### Métodos de taxonomia numérica

Os métodos de taxonomia numérica ou métodos hierárquicos de classificação (*hierarchical clustering methods*, na bibliografia de expressão inglesa) ou métodos de identificação de padrões (*pattern recognition*) são um conjunto de métodos matemáticos de classificação de resultados ou unidades taxonómicas operacionais (OTU ou *operational taxonomic units*) segundo a semelhança de um conjunto de parâmetros denominados características [3].

Nesse conjunto pode-se distinguir duas grandes classes de métodos: os que procedem à classificação em grupos ou conjuntos previamente conhecidos (*supervised learning methods*) e aqueles que classificam os dados em conjuntos

que parecem relacionados (*unsupervised learning methods*) [4].

A esta segunda classe pertencem os métodos SAHN ou métodos sequenciais, aglomerativos, hierárquicos e de não sobreposição. São sequenciais porque apenas um elemento é classificado de cada vez. Aglomerativos porque partem de unidades taxonómicas individuais que são agrupadas em conjuntos, os quais, por sua vez, dão origem a outros conjuntos. Hierárquicos porque o número de conjuntos diminui com o nível ou avanço da classificação. De não sobreposição porque, para cada nível de classificação, cada unidade taxonómica individual pertence apenas a um conjunto [3]. Destes métodos, os métodos UPGMA (*unweighted pair group method using arithmetic averages*), ou métodos de agrupamento não pesado através de médias aritméticas, são dos mais utilizados – por exemplo para a classificação de objectos, a partir da composição química, segundo a sua origem [1, 7-8]. Muito semelhantes são os métodos WPGMA (*weighted pair...*), que só diferem dos anteriores no facto de usarem médias pesadas no agrupamento. O algoritmo TAXONOMIA permite utilizar um método de cada um dos tipos.

Convém referir que as classificações obtidas não são independentes dos métodos utilizados nem, sequer, do parâmetro empregue como medida de semelhança [2-3, 5-6].

### Medidas de semelhança

A classificação das unidades taxonómicas operacionais faz-se com base num parâmetro que mede a semelhança entre as características. Os dois mais frequentemente usados são o coeficiente de correlação e a distância euclideana [3].

Sendo  $X_{ij}$  o valor da característica j da unidade taxonómica individual i e sendo n o número de unidades taxonómicas individuais e m o número de características, a distância euclideana entre a unidade taxonómica i e a unidade taxonómica k é definida por

$$d_{ik} = \left[ \sum_{j=1}^m (X_{ij} - X_{kj})^2 \right]^{1/2} \quad (1)$$

e o coeficiente de correlação entre as unidades taxonómicas i e k é definido por

<sup>a</sup> Bolseiro da Junta Nacional de Investigação Científica e Tecnológica. Centro de Química-Física e Radioquímica da Universidade de Lisboa, Faculdade de Ciências, 1294 Lisboa Codex

$$r_{ik} = \frac{S_{ik}}{S_i S_k} \quad (2)$$

em que  $S_{ik}$  é a covariância entre as características de  $i$  e as características de  $k$  e  $S_i$  e  $S_k$  o desvio-padrão das características, respectivamente, de  $i$  e de  $k$ , ou seja

$$S_{ik} = 1/(m-1) \sum_{j=1}^m (X_{ij} - \bar{X}_i)(X_{kj} - \bar{X}_k) \quad (3)$$

$$S_i = [1/(m-1) \sum_{j=1}^m (X_{ij} - \bar{X}_i)^2]^{1/2} \quad (4)$$

$$S_k = [1/(m-1) \sum_{j=1}^m (X_{kj} - \bar{X}_k)^2]^{1/2} \quad (5)$$

em que  $\bar{X}_i$  e  $\bar{X}_k$  representam o valor médio das  $m$  características da unidade taxonómica  $i$  e da unidade taxonómica  $k$ , respectivamente. A equação (2) pode escrever-se como

$$r_{ik} = \frac{\sum_{j=1}^m (X_{ij} X_{kj}) - \sum_{j=1}^m X_{ij} \sum_{j=1}^m X_{kj} / (m-1)}{\{[\sum_{j=1}^m X_{ij}^2 - (\sum_{j=1}^m X_{ij})^2 / (m-1)][\sum_{j=1}^m X_{kj}^2 - (\sum_{j=1}^m X_{kj})^2 / (m-1)]\}^{1/2}} \quad (6)$$

a qual, do ponto de vista computacional, é mais útil [3, 9]. A distância euclideana mede o afastamento entre os pontos que representam as unidades taxonómicas operacionais num espaço de dimensão igual ao número de unidades taxonómicas individuais ( $m$ ). Assim, duas unidades taxonómicas idênticas são, nessa representação, coincidentes, pelo que a distância euclideana entre elas é 0. Portanto, quanto menor for a distância, maior é a semelhança.

O coeficiente de correlação é um parâmetro que mede a relação entre variáveis. Em módulo, o seu valor está compreendido entre 0 e 1. Quando o coeficiente de correlação é 1 as variáveis comportam-se de igual modo, enquanto que ao valor 0 corresponde a ausência de qualquer relação. Assim, quanto menor for o coeficiente de correlação, menor é a semelhança.

Portanto, o coeficiente de correlação e a distância euclideana não medem a mesma propriedade: o primeiro compara variações, enquanto o segundo compara valores. Por exemplo, se as variáveis tomarem determinados valores para um sistema e o dobro desses valores para outro sistema, o coeficiente de correlação permite classificar as duas amostras como semelhantes (coeficiente de correlação = 1), mas a distância euclideana permite classificá-las como dissemelhantes (distância euclideana > 0).

### Transformação e normalização dos resultados

A classificação das unidades taxonómicas de acordo com as características, no caso em que cada uma destas toma valores de ordem de grandeza diferente da dos valores das restantes características, não é bem sucedida se os resultados não forem normalizados (*scaling*). O coeficiente de correlação terá sempre um valor próximo de 1, enquanto a distância euclideana dependerá quase só da característica com valores de maior grandeza.

Sendo  $\chi_{ij}$  o valor não normalizado da característica  $j$  da unidade taxonómica  $i$ ,  $\bar{\chi}_j$  o valor médio da característica  $j$  nas  $n$  unidades taxonómicas e  $S_j$  o respectivo desvio-padrão,

$$\bar{\chi}_j = 1/n \sum_{i=1}^n \chi_{ij} \quad (7)$$

e

$$S_j = [1/(n-1) \sum_{i=1}^n (\chi_{ij} - \bar{\chi}_j)^2]^{1/2} \quad (8)$$

O valor normalizado da característica  $j$  da unidade taxonómica  $i$ ,  $X_{ij}$ , é obtido por subtracção do valor médio da característica e divisão pelo desvio-padrão:

$$X_{ij} = (\chi_{ij} - \bar{\chi}_j) / S_j \quad (9)$$

Os valores  $X_{ij}$  dizem-se centrados e expressos em unidades de desvio-padrão.

A situação mais vulgar é aquela em que as unidades taxonómicas individuais são sistemas ou amostras e as características, variáveis ou propriedades. A normalização é feita, portanto, através do cálculo da média de cada uma das propriedades e do respectivo desvio-padrão.

Porém, há casos em que interessa classificar as propriedades (cf. a aplicação adiante apresentada), as quais, nestas circunstâncias, são, portanto, as unidades taxonómicas operacionais. Como, em princípio, são maiores as diferenças entre os valores das propriedades de uma só amostra do que as diferenças entre os valores de uma só propriedade nas diferentes amostras, regra geral, a classificação das propriedades não é significativamente afectada pela normalização. No algoritmo TAXONOMIA a normalização é feita sempre propriedade a propriedade, independente da classificação pretendida ser a das amostras ou a das propriedades. Por outro lado, há situações em que é conveniente transformar os resultados antes da normalização. Por exemplo, para a classificação dos meteoritos, a partir da sua composição em elementos vestigiais, são utilizados os logaritmos das concentrações [1].

Porém, deve sublinhar-se a diferença entre a transformação e a normalização dos resultados. Enquanto a primeira, quando é feita, é-o por razões de significado (significado químico, por exemplo), a normalização dos resultados é justificada, apenas, por razões matemáticas.

### Classificação

O ponto de partida para a classificação é a matriz de semelhança, a qual é uma matriz quadrada, de dimensão  $n \times n$ , com elementos que correspondem à medida de semelhança entre a unidade taxonómica correspondente à linha da matriz e a unidade taxonómica correspondente à coluna. Os elementos diagonais são iguais a 0, se o parâmetro de semelhança for a distância euclideana, ou iguais a 1, se o coeficiente de correlação for a medida de semelhança.

Nos métodos SAHN, nessa matriz inicial é procurado o menor elemento não diagonal ou o maior, conforme o parâmetro de semelhança seja, respectivamente, a distância euclideana ou o coeficiente de correlação. As duas unidades taxonómicas correspondentes a esse parâmetro de semelhança são reunidas, sendo os parâmetros de semelhança entre a nova unidade taxonómica e as restantes calculados a partir dos parâmetros correspondentes às duas unidades taxonómicas antes de agrupadas. Deste modo, a

matriz de semelhança é reduzida de uma linha e de uma coluna. O processo é repetido até se obter uma matriz de dimensão  $2 \times 2$ .

No caso particular dos métodos UPGMA e dos métodos WPGMA, o agrupamento (*linkage*) dos parâmetros de semelhança faz-se por média aritmética. No segundo caso, o dos métodos pesados, o cálculo da média é feito sempre por adição dos dois parâmetros anteriores e divisão por dois, independentemente do facto de esses parâmetros corresponderem a conjuntos com diferentes números de unidades taxonómicas individuais. No primeiro caso, o dos métodos não pesados, o cálculo da média é feito considerando o número de unidades taxonómicas individuais correspondentes a cada um dos parâmetros anteriores. Assim, todas as unidades taxonómicas individuais têm o mesmo peso em qualquer nível da classificação, ao contrário do que sucede com os métodos WPGMA [3].

### Interpretação da classificação

Com os valores dos parâmetros de semelhança para os quais se procede aos agrupamentos, pode-se facilmente construir um dendrograma como os apresentados nas Figuras 1 e 2.

Embora não esteja no âmbito desta nota abordar as questões colocadas pela interpretação das classificações, há dois problemas que se podem referir brevemente.

O primeiro é o do número de conjuntos e resulta do facto de qualquer número entre 1 e o número de unidades taxonómicas individuais poder ser considerado. Como esses dois extremos correspondem a situações triviais (todas as unidades taxonómicas individuais formam  $n$  conjuntos de 1 elemento e 1 conjunto de  $n$  elementos), é um número intermédio aquele que, de facto, interessa considerar. Qual é esse número? — eis a questão.

Em algumas das situações correntes há alguma informação disponível *a priori* que pode ajudar nessa escolha. Por exemplo, no caso da classificação das amostras arqueológicas apresentado adiante, ao qual se refere a Figura 1, existe a observação do perfil estratigráfico, segundo o qual as

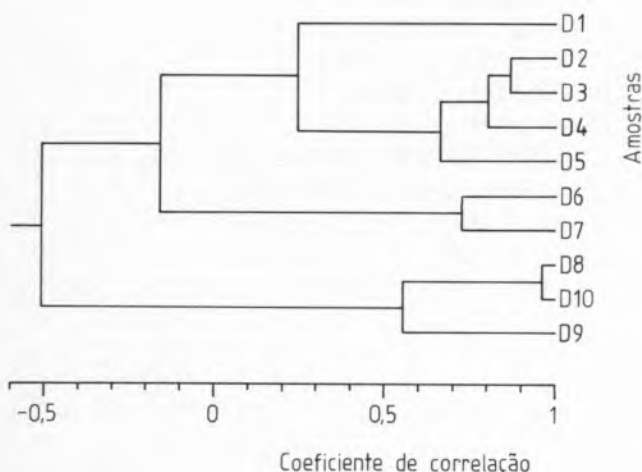


FIGURA 1

Classificação das amostras utilizando o coeficiente de correlação como medida de semelhança

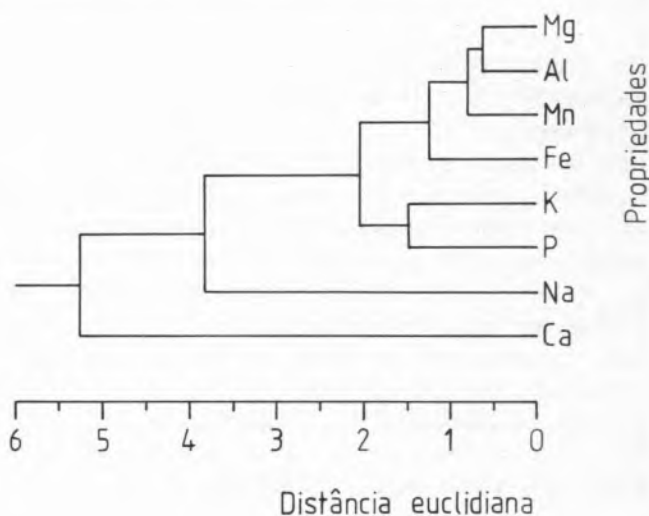


FIGURA 2

Classificação das propriedades utilizando a distância euclidiana como medida de semelhança

amostras D1 a D5 pertencem a uma camada estratigráfica, ainda que complexa e bastante remexida, e as amostras D6 a D10 pertencem a outra camada macroscopicamente distinta da primeira. Nesta situação, considerar apenas dois conjuntos parece ser uma escolha razoável. Para esse nível de classificação, segundo o dendrograma da Figura 1, tem-se um conjunto formado pelas amostras D1 a D7 e um conjunto formado pelas amostras D8 a D10 — o que não está de acordo com as observações macroscópicas, mas que pode ser discutido em função de outras informações [10].

Nos casos em que o coeficiente de correlação é utilizado como medida de semelhança existem aspectos de significado estatístico que podem ser tomados em conta. Assim, como o coeficiente de correlação só é estatisticamente significativo para valores que satisfaçam a equação

$$|r_{ij}| > t / (t^2 + N) \quad (10)$$

em que  $t$  é o parâmetro  $t$  de Student para um dado nível de probabilidade (obtido em tabela) e  $N$  o número de graus de liberdade ( $m-2$ ), ou seja, como só para valores de  $r_{ij}$  que verificam a equação (10) se pode considerar que a correlação obtida, para a probabilidade escolhida, não é devida ao acaso [9], pode-se considerar que só têm significado estatístico os conjuntos definidos a valores do coeficiente de correlação que satisfaçam aquela equação.

Para o exemplo apresentado, para uma probabilidade de 90%, só são estatisticamente significativos valores de  $r_{ij} > 0,6215$ , pelo que, segundo este critério, o número de conjuntos a considerar não deve ser inferior a 5. Nesta situação, têm-se os seguintes conjuntos: D1; D2 a D5; D6 a D7; D9; D8 e D10 (cf. Figura 1). Esta classificação já não se opõe às observações estratigráficas, pois admite, a um nível de classificação mais elevado a separação entre as amostras D1 a D5, por um lado, e as amostras D6 a D10, por outro, mas é necessário interpretar os três conjuntos formados pelas amostras D6 a D10 [10].

Quando não existe qualquer informação *a priori* que possa ajudar a decidir qual o número de conjuntos que deve ser considerado na classificação, este problema torna-se de ainda de mais difícil resolução.



Um segundo problema colocado pela interpretação dos dendrogramas é o que diz respeito ao acaso, à causa e à consequência.

Uma correlação estatisticamente significativa entre duas unidades taxonômicas ou uma pequena distância euclidiana não significam necessariamente uma relação de causa e efeito. Podem ser devidos, apenas, ao acaso. Porém, mesmo que se admita a existência duma relação causal, por um lado, é difícil, por vezes, identificar a causa e o efeito, por outro lado, frequentemente as restantes unidades taxonômicas influem na associação encontrada.

Portanto, a classificação que a taxonomia numérica permite deve ser considerada apenas ao nível das hipóteses, a menos que existam outras indicações independentes [9].

### Explicação do algoritmo

O algoritmo TAXONOMIA, aqui representado, considera várias possibilidades: utilização dum método UPGMA ou dum método WPGMA, classificação das amostras ou das propriedades, transformação e normalização ou não dos resultados introduzidos e utilização de duas medidas de semelhança (coeficiente de correlação e distância euclidiana).

É facilmente transposto para qualquer linguagem de programação e, como já foi mencionado, um micro-computador de pequeno porte como o Zx Spectrum é suficiente para a maioria das aplicações correntes.

No algoritmo pode-se distinguir quatro fases:

- introdução de resultados;
- transformação e normalização dos resultados;
- cálculo da matriz de semelhança;
- classificação dos resultados.

Os resultados são introduzidos depois de definido o número de propriedades (PROP) e o número de amostras (AM). São guardados na matriz A de dimensão AM x PROP. São introduzidos, também, as identificações das amostras e os nomes das variáveis, dados que são guardados nas matrizes de caracteres A\$ e P\$, respectivamente, de dimensão AM x 1 e PROP x 1.

A segunda das fases é opcional, seja no que se refere à transformação, seja no que se refere à normalização dos dados. No caso presente aquela consiste numa transformação logarítmica, mas o algoritmo pode ser facilmente modificado de modo a permitir outra transformação. Porém, é necessário que os resultados suportem a transformação pretendida (no caso da transformação logarítmica é necessário que os resultados sejam maiores que 0). A variável TRANSFORMAR, conforme tenha o valor 1 ou 0, permite seleccionar ou não, respectivamente, essa opção. De igual modo, conforme a variável NORMAL toma o valor 1 ou o valor 0 é feita ou não a normalização dos resultados.

Os resultados, transformados e normalizados no caso de terem sido essas as opções seleccionadas, são transferidos para a matriz M de dimensão AM x PROP. Isso permite que a matriz A aguarde os resultados tal como foram introduzidos.

A partir da matriz M é construída a matriz de semelhança, R, de dimensão AM x AM se se pretender classificar as amostras (variável CLASSE=1) ou de dimensão PROP x

PROP se se pretender classificar as propriedades (variável CLASSE=2). Em qualquer dos casos, a matriz R é simétrica e os seus elementos são calculados através da equação (6), se o parâmetro de semelhança for o coeficiente de correlação (variável PARÂMETRO=1), ou através da equação (1), se o parâmetro de semelhança for a distância euclidiana (variável PARÂMETRO=2).

Finalmente, os resultados são classificados utilizando a média não pesada (variável MÉTODO=1) ou a média pesada (variável MÉTODO=2), a que corresponde, respectivamente, um método UPGMA ou um método WPGMA.

A classificação é apresentada numa tabela semelhante às tabelas II e III, que consistem em pares de amostras ou de propriedades, conforme o caso, e respectivo parâmetro de semelhança. Com esses resultados pode-se traçar um dendrograma como os das Figuras 1 e 2.

### Algoritmo

#### INÍCIO DO ALGORITMO TAXONOMIA

```
{ Introdução dos resultados }
Ler AM, PROP
Para i variando de 1 a PROP repetir
| Ler P$(i)
Para i variando de 1 a AM repetir
| Ler A$(i)
  Para j variando de 1 a PROP repetir
  | Ler A(i,j)
Ler TRANSFORMAR, NORMAL, CLASSE,
PARÂMETRO, MÉTODO
{ Transformação e normalização dos resultados }
Para j variando de 1 a PROP repetir
| Se NORMAL=1 então
| | SX← 0
| | SX2← 0
Para i variando de 1 a AM repetir
| Se TRANSFORMAR=1 então
| | M(i,j)←ln A(i,j)/2.303
| Senão
| | M(i,j)←A(i,j)
| Se NORMAL=1 então
| | SX←SX+M(i,j)
| | SX2←SX2+M(i,j)*M(i,j)
Se NORMAL=1 então
| X←SX/AM
| S←((SX2-SX*SX/AM)/(AM-1))1/2
Para i variando de 1 a AM repetir
| M(i,j)←(M(i,j)-X)/S
{ Cálculo da matriz de semelhança }
Se PARÂMETRO=1 então
| Rii←1
Senão
| Rii←0
Se CLASSE=1 então
| RM←AM
| RN←PROP
Senão
| RM←PROP
| RN←AM
```

```
Para i variando de 1 a RM repetir
Se CLASSE=1 então
  R$(i)←A$(i)
Senão
  R$(i)←P$(i)
R(i,i)←Rii
Para j variando de 1 a i-1 repetir
  SX←0
  Se PARÂMETRO=1 então
    SX2←0
    SY←0
    SY2←0
    SXY←0
  Para k variando de 1 a RN repetir
    Se CLASSE=1 então
      X←M(i,k)
      Y←M(j,k)
    Senão
      X←M(k,i)
      Y←M(k,j)
    Se PARÂMETRO=1 então
      SX←SX+X
      SY←SY+Y
      SX2←SX2+X*X
      SY2←SY2+Y*Y
      SXY←SXY+X*Y
    Senão
      SX←SX+(X-Y)*(X-Y)
  Se PARÂMETRO=1 então
    R(i,j)←(SXY-SX*SY/RN)/((SX2-SX*SX/RN)*(SY2-
    -SY*SY/RN))1/2
  Senão
    R(i,j)←(SX)1/2
  R(j,i)←R(i,j)
{Classificação}
Para i variando de 1 a RM repetir
  S(i)←1
Para k variando de RM a 2 diminuindo 1 repetir
  ri←1
  rj←2
  M←R(1,2)
  Para i variando de 1 a k repetir
    Para j variando de 1 a i-1 repetir
      Se (PARÂMETRO=1 e R(i,j)>M) ou (PARÂMETRO=2
      e R(i,j)<M) então
        M←R(i,j)
        ri←i
        rj←j
  Se ri>rj então
    ri↔rj
  Escrever R$(ri),»,»,R$(rj),»:»,M
Para i variando de 1 a k repetir
  Se i=ri ou i=rj então
    R(ri,i)←Rii
  Senão
    R(ri,i)←(S(ri)*R(ri,i)+S(rj)*R(rj,i))/(S(ri)+S(rj))
Se MÉTODO=1 então
  S(ri)←S(ri)+S(rj)
Para i variando de rj+1 a k repetir
  Se MÉTODO=1 então
```

```
|S(i-1)←S(i)
|R$(i-1)←R$(i)
Para j variando de 1 a k repetir
  |R(i-1,j)←R(i,j)
  |R(i,j-1)←R(i,j)
```

FIM DO ALGORITMO

Aplicação: a coluna de amostragem D da gruta do Caldeirão

No âmbito dum estudo geoquímico de preenchimentos sedi- mentares de grutas e de abrigos com vestígios de ocupação humana pré-histórica [10-12], determinou-se para um con- junto de 10 amostras, correspondentes à coluna de amostra- gem D da gruta do Caldeirão (Tomar), diversas proprieda- des químicas, entre as quais a concentração de oito elementos na fracção solúvel em ácido clorídrico concen- trado, resultados que são apresentados na Tabela 1 [10].

TABELA 1  
Resultados utilizados para a classificação

Amostras	Teor %							
	Na	K	Mg	Ca	Fe	Mn	Al	P
D1	0,0451	0,238	0,238	9,0	2,16	0,086	2,22	0,510
D2	0,0452	0,265	0,258	9,7	2,46	0,099	2,46	0,569
D3	0,0448	0,280	0,258	7,5	2,58	0,105	2,57	0,573
D4	0,0462	0,278	0,263	9,0	2,97	0,104	2,60	0,629
D5	0,0452	0,268	0,262	10,2	2,58	0,096	2,47	0,655
D6	0,0432	0,295	0,251	10,8	2,39	0,094	2,46	0,720
D7	0,0450	0,283	0,251	12,3	2,44	0,098	2,38	0,617
D8	0,0463	0,256	0,228	13,7	2,16	0,085	2,09	0,518
D9	0,0401	0,239	0,209	13,6	1,89	0,073	1,84	0,415
D10	0,0484	0,252	0,223	14,2	2,10	0,080	1,91	0,471

As amostras D1 a D5 correspondem à camada superficial A/ B/C/D, muito remexida e com materiais arqueológicos que vão do Neolítico Médio ou Final ao presente, e as amostras D6 a D10 correspondem à camada E<sub>p</sub>, Solutrense (Paleolítico Superior) [10].

Na Tabela 2 apresenta-se a classificação das amostras – obtida com um programa escrito em BASIC a partir do algoritmo TAXONOMIA. Foi seleccionado o método UPGMA, não se fez qualquer transformação dos resulta- dos mas fez-se a sua normalização. Utilizou-se o coefi- ciente de correlação como medida de semelhança. Com os resultados da Tabela 2 traçou-se o dendrograma da Figura 1, já discutido atrás.

TABELA 2  
Classificação das amostras

Amostras	Coeficiente de correlação
D8 / D10	0,9610
D2 / D3	0,8658
D2 / D4	0,8013
D6 / D7	0,7267
D2 / D5	0,6608
D8 / D9	0,5558
D1 / D2	0,2446
D1 / D6	-0,1581
D1 / D8	-0,5087

Na Tabela 3 apresenta-se uma classificação das 8 propriedades feita nas mesmas condições que a classificação das amostras excepto no facto de se ter utilizado a distância euclideana como medida de semelhança. Com esses resultados traçou-se o dendrograma da Figura 2, o qual sugere que os teores em K, Mg, Mn, Fe, Al e P nessa fracção estão relacionados entre si, ao contrário do que sucede com o Ca e o Na. Esta classificação, idêntica à que se obtém utilizando o coeficiente de correlação como medida de semelhança, pode ser interpretada em termos de diferentes origens dos materiais. Assim, os 6 elementos que surgem associados correspondem a materiais com origem no

TABELA 3  
Classificação das propriedades

Propriedades	Distância euclidiana
Mg / Al	0,6302
Mg / Mn	0,8081
Mg / Fe	1,2548
K / P	1,4782
K / Mg	2,0556
Na / K	3,8175
Na / Ca	5,2659

exterior da gruta, enquanto o Ca existente nos sedimentos corresponde a materiais com origem na própria gruta, que é calcárea [10].

#### Referências

- [1] Massart, Desiré L.; Kaufman, Leonard; Esbensen, Kim H.; *Anal. Chem.*, **54**(6), 911-917 (1982).
- [2] Brown, Steven D.; Barker, Todd Q.; Larivee, Robert J.; Monfre, Stephen L.; Wilk, Harlan R.; *Anal. Chem.*, **60**(12), 252R-273R (1988).
- [3] Massart, Desiré L.; Dijkstra, Auke; Kaufman, Leonard; *Evaluation and Optimization of Laboratory Methods and Analytical Procedures*, Amsterdam, Elsevier (1978).
- [4] Kowalski, B. R.; Schatzki, T. F.; Stross, F. H.; *Anal. Chem.*, **44**(13), 2176-2180 (1972).
- [5] Kowalski, Bruce R.; *Anal. Chem.*, **47**(13), 1152A-1162A (1975).
- [6] Beebe, Kenneth R.; Kowalski, Bruce R.; *Anal. Chem.*, **59**(17), 1007A-1017A (1987).
- [7] Cabral, J. M. P.; Gouveia, M. A.; *Portugal. Phys.*, **10**(1-2), 101-116 (1979).
- [8] Picon, Maurice; Le Mièrre, Marie, in Miskovsky, Jean-Claude (ed.), *Géologie de la Préhistoire: Méthodes, techniques, applications*, Paris, Gêopré, 883-901 (1987).
- [9] Glass, Gene V.; Stanley, Julian C.; *Métodos Estadísticos Aplicados a las Ciencias Sociales*, Madrid, Prentice/Hall (1974).
- [10] Cruz, António João de Carvalho, submetido a *Arqueologia* (1989).
- [11] Cruz, António João de Carvalho, *Arqueologia*, **16**, 130-134 (1987).
- [12] Cruz, António João de Carvalho, *Ler História*, **13**, 143-146 (1988).

## JOHANNES NICOLAUS BRØNSTED (1879-1947)



Brønsted nasceu em Vardø, na Dinamarca, e em 1908 obteve o seu doutoramento em Química na Universidade de Copenhague. Foi, então, nomeado professor de Química-Física nessa Universidade, onde se conservou até à data da sua morte.

O trabalho de Brønsted desenvolveu-se principalmente nos domínios da Termodinâmica e da Cinética de soluções electrolíticas. Em 1923, quase em simultâneo com o químico britânico T. Martin Lowry (1874-1936), propôs definições de ácidos e bases em termos de transferência protónica. Num importante artigo que publicou em 1924 em colaboração com o químico americano Victor K. La Mer

(1895-1966) usou a teoria de Debye-Hückel para obter uma relação entre o coeficiente de actividade de um ião e a força iónica, entidade esta que tinha sido definida em 1921 pelo químico americano Gilbert Newton Lewis (1875-1946).

Com base nas definições de ácidos e bases desenvolveu o conceito de catálise ácido-base geral. Em 1924 propôs relações, actualmente conhecidas como «equações de Brønsted», entre a actividade catalítica de um ácido ou de uma base e a sua força ácida ou básica.