

## Las Matemáticas y la Química: Quimiometría y Cualimetría



Miguel Á. Rodríguez Delgado

Profesor Titular de Química Analítica  
Departamento de Química Analítica, Nutrición y Bromatología  
Universidad de La Laguna

### 1. Definición y evolución histórica de la Quimiometría y Cualimetría

En 1975, *The Chemometrics Society* definió la **Quimiometría** como “la disciplina química que utiliza métodos matemáticos y estadísticos para diseñar o seleccionar procedimientos de medida y experimentos óptimos, y para proporcionar la máxima información química mediante el análisis de datos químicos”. D.L. Massart (1997) define la Quimiometría como “la disciplina química que utiliza la Matemática, la Estadística y la lógica formal para diseñar o seleccionar procedimientos experimentales óptimos, proporcionar la máxima información química relevante a partir del análisis de datos químicos y obtener conocimiento a partir de sistemas químicos”. Respecto a la primera, esta definición añade el uso de la lógica formal, base de la inteligencia artificial, y matiza dos hechos: que la información buscada debe ser “relevante”, y que la información sobre el sistema químico debe transformarse en conocimiento global sobre el problema en estudio.

La Quimiometría mejora el rendimiento del proceso analítico en todas sus etapas. Siguiendo el esquema de la Figura 1, en las etapas de toma de muestra y diseño de los experimentos se pretende conseguir datos de una determinada calidad con un mínimo esfuerzo experimental. Una vez obtenidos los datos, el tratamiento de los mismos permite aumentar la calidad de los resultados, facilitar su interpretación, y fundamentar las conclusiones con rigor.

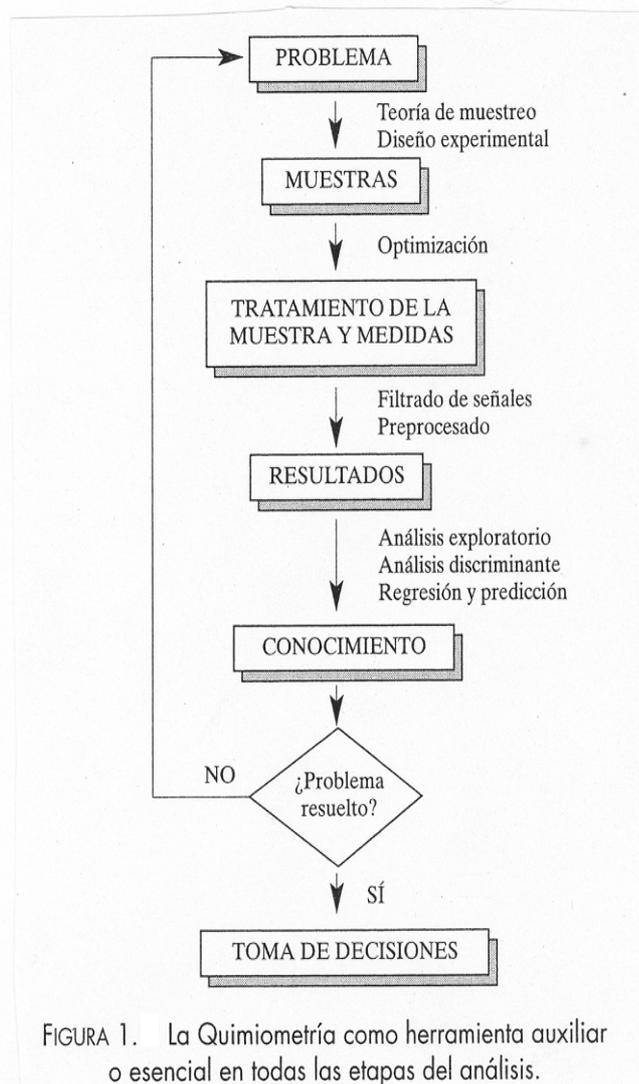
La Quimiometría se puede considerar como una rama aplicada y especializada de ambas, la Química Analítica y la Estadística Aplicada, con una función análoga a la que desempeñan disciplinas como la Biometría, Sociometría, Econometría o Psicometría en relación a la Biología, Sociología, Economía o Psicología, entre otras. Las diferencias entre estas disciplinas metrológicas y la Quimiometría son debidas a las particulares características de los datos de origen químico:

- a) El limitado número de factores (concentraciones, pH, fuerza iónica, presencia de luz, etc.) que influyen sobre los resultados.
- b) La posibilidad de control de la mayoría de dichos factores por parte del experimentador. Puesto que los factores son escasos, y en su mayoría controlables, es fácil repetir los experimentos en las condiciones de trabajo deseadas.
- c) La incertidumbre relativamente pequeña (entre un 0.3% y un 3%) que afecta a la mayoría de las medidas analíticas.

En Biometría, Sociometría, etc., en general, se tienen muchos más factores, y muchos de ellos no son controlables, por lo que es difícil o imposible repetir los experimentos en las mismas condiciones. Además, con frecuencia, la incertidumbre con la que se conocen los datos es elevada. En general, el número total de factores puestos en juego crece en la forma:

Física y Química [ $<10^2$ ] < Bioquímica [ $>10^2$ ] < Biología [ $>10^3$ ] < Ciencias Humanas [ $>10^4$ ]

La mayoría de las diversas ramas de la Estadística se desarrollaron en el primer tercio del siglo XX. De esta época datan los primeros números de revistas de investigación tales como *Biometrika* (1901), *Econometrica* (1933) y *Psicometrika* (1936). En cambio, el desarrollo de la Quimiometría como disciplina científica independiente no se inició hasta 1969, con los trabajos de Jurs, Kowalski e Isenhour. Años después aparecieron las primeras revistas de investigación en Quimiometría: *Chemometrics and Intelligent Laboratory Systems* (1986) y *Journal of Chemometrics* (1987).



El desarrollo de la Quimiometría fue tardío como consecuencia de la lenta evolución de la instrumentación científica. En la primera mitad del siglo XX era difícil y costoso obtener, tratar y almacenar datos químico-analíticos. Se disponía de pocos datos, y su examen permitía al

analista extraer toda la información significativa. Esta situación comenzó a cambiar hacia 1960, debido al fuerte impulso que experimentó el desarrollo de nueva instrumentación. Sin embargo, en la década 1960-1970 la recopilación de los datos y su tratamiento manual con algoritmos matemáticos eran procesos lentos y tediosos. Por ello, los algoritmos con cierta complejidad de cálculo (incluida la regresión lineal simple) no tuvieron interés práctico hasta la popularización de las calculadoras electrónicas hacia 1975. Por otra parte, la idea de que los datos químicos generados por los nuevos instrumentos podían transformarse en información significativa mediante la aplicación de técnicas estadísticas estaba poco extendida. Como observó C.F. Bender (1973), “durante años, los científicos experimentalistas han llenado [de datos] los cuadernos de laboratorio, y a menudo los han descartado por falta de técnicas de interpretación adecuadas”.

El impulso definitivo de la Quimiometría como disciplina científica independiente fue debido a la popularización del microordenador, y su incorporación a los instrumentos científicos, a mediados de la década 1980-1990. Actualmente, un laboratorio analítico automatizado es capaz de generar, tratar y almacenar en corto tiempo muchos miles de datos, referentes a muestras que han sido analizadas secuencialmente mediante técnicas de enorme poder resolutivo: cromatografía, electroforesis, entre otras, acopladas a diversos detectores simples, multi-canales y de barrido. Además, muchas técnicas quimiométricas se encuentran incorporadas al *software* de la instrumentación analítica, y se aplican de un modo habitual por los profesionales del laboratorio químico analítico.

La liberalización del comercio en áreas geográficas cada vez más grandes ha provocado que la comparabilidad de los resultados analíticos sea una de las metas más importantes a lograr. Para alcanzarla, los organismos internacionales competentes han desarrollado un sistema de garantía de calidad que se basa fundamentalmente en el ámbito voluntario, la normalización, la certificación, la acreditación y la calibración. Los conocimientos y herramientas que sirven para desarrollar y evaluar la calidad del análisis químico, fundamentalmente, pero que se extienden también a otros campos, han sido agrupadas bajo el nombre de **Cualimetría**.

La calidad considerada genéricamente es un concepto abstracto que tiene muchas implicaciones, por lo que no es de extrañar que se encuentren un sinfín de definiciones. La Real Academia Española define la calidad como “Propiedad o conjunto de propiedades inherentes a una cosa, que permiten apreciarla como igual, mejor o peor que las restantes de su especie”. Así un producto alimenticio A es de más calidad que otro B porque tiene menos aditivos o contaminantes. Un resultado analítico A será mejor que otro B si el primero se acerca más al verdadero valor y presenta menos dispersión (mayor precisión). La ISO define la calidad como “la totalidad de los rasgos y características de un producto, proceso o servicio que inciden en su capacidad de satisfacer necesidades reguladas o implícitas”. Teniendo en cuenta los matices que la palabra calidad tiene en relación con el laboratorio y sus relaciones entre sí, se puede definir la calidad del laboratorio analítico como “conjunto de características de la información que satisfacen las demandas/exigencias del organismo público/privado del que depende y/o del cliente o usuario”. Las propiedades analíticas pueden considerarse divididas en dos grupos según su importancia relativa: las denominadas básicas como exactitud, precisión, sensibilidad, selectividad y rapidez; y las complementarias tales como coste, grado de participación humana (automatización), robustez, seguridad para el personal, etc. En la evaluación de esta calidad no cabe duda que las herramientas básicas son las matemáticas pues hay que generar información al usuario con un cierto grado de credibilidad, por lo que todas las etapas de calibración, validación, etc., son fundamentales para asegurar una calidad integral.

## 2. Máquinas y programas utilizados en Quimiometría

En Quimiometría, los ordenadores de gran capacidad de cálculo no son habitualmente necesarios. Los ordenadores personales (tipo estándar, de sobremesa o portátiles) tienen memoria y velocidad más que suficientes para manejar los datos (normalmente por debajo de unos cientos de miles) y realizar los cálculos necesarios para aplicar las técnicas estadísticas más complejas. Por su parte, las calculadoras de bolsillo programables (o con funciones preprogramadas) son también útiles para aplicar herramientas simples sobre unos pocos datos distribuidos a lo largo de una o dos variables. Las colecciones de programas o paquetes estadísticos más populares del mercado se encuentran disponibles en versiones MacIntosh y Windows. Pueden distinguirse varios tipos:

- a) Los paquetes estadísticos de aplicación general, como SAS (*Statistical Analysis Systems*), Statistica, S-Plus y Statgraphics. Son entornos de trabajo con opciones que abarcan un gran número de herramientas estadísticas preprogramadas de todo tipo. Dentro de esta categoría pueden incluirse también algunos paquetes de programas originalmente desarrollados para aplicaciones médicas, sociológicas, etc., si bien actualmente tienen un enfoque amplio y general, como el BMDP (*Bio-Medical Data Processing*), o el SPSS (*Statistical Package for Social Sciences*).
- b) Los paquetes de Quimiometría especialmente diseñados para el tratamiento de datos químico-analíticos, como *The Unscrambler*, *Trajans*, *Pirouette* y *PARVUS*, entre otros. Estos programas no contemplan todo tipo de herramientas estadísticas, sino que centran su atención en las opciones más utilizadas por los quimiometristas. Si bien se pierde el enfoque general, las herramientas de mayor aplicación en Quimiometría se utilizan con mayor rapidez y comodidad, y los resultados se presentan directamente en la forma esperada por el profesional del laboratorio. En algún caso, incorporan también herramientas estadísticas específicas, fruto de la investigación en Quimiometría y poco extendidas entre las otras especialidades de la Estadística Aplicada.
- c) Los entornos asistidos para el desarrollo de herramientas estadísticas, como MATLAB. En un entorno de este tipo se tienen opciones y herramientas básicas (como vectores y matrices) para que el usuario programe su propio *software* estadístico sin partir de cero. Se necesita cierto entrenamiento para utilizar MATLAB, pero dominar un entorno de este tipo posibilita la investigación en Quimiometría, esto es, el desarrollo de nuevas herramientas o la modificación de las ya existentes.
- d) Las hojas de cálculo como *Lotus*, *Excel* y *Quattro Pro*, entre otras. Estos programas, habituales en los paquetes de ofimática (como *Microsoft Office* o *Corel Office*), contienen una extensa relación de herramientas estadísticas. Por esta razón, y por la comodidad con la que permiten realizar cálculos, son valiosas herramientas auxiliares en Quimiometría, especialmente para obtener descriptores estadísticos básicos y hacer representaciones gráficas. También son útiles como programa “puente”, para transformar datos generados por los instrumentos, o procedentes de otras fuentes, a un formato interpretable por el editor de datos del paquete estadístico.

### 3. Algunas áreas de interés

**Reconociendo tendencias en los datos.** Supongamos que una empresa (farmacéutica, de pinturas, agroalimentaria, de química básica, ...) produce diversos lotes de producto comercializable a la semana. Sus clientes desean recibir un producto de calidad constante y, por tanto, la empresa debe controlar la homogeneidad entre lotes. Si suponemos que por cada uno de los 10 lotes fabricados semanalmente elegimos 12 muestras (que sean representativas de las fases inferior, media y superior del reactor donde se han producido, por ejemplo), tendremos un conjunto de 120 muestras. Si cada muestra individual se caracteriza por 7 parámetros distintos (que sean un reflejo de su composición, forma, tamaño u otras propiedades que la caractericen), ya tenemos un conjunto de datos considerable formado por  $120 \times 7$  datos. La representación numérica, en forma de tabla o matriz de datos, de los valores obtenidos se ilustra en la Figura 2:

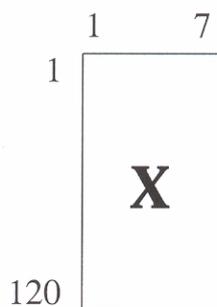


Figura 2. Representación de la matriz de datos  $X$  formada por 120 muestras u objetos. Cada uno de ellos está caracterizado por 7 parámetros o variables.

Nos puede interesar conocer:

- Si todo el conjunto de muestras (llamadas objetos en el lenguaje utilizado en la Quimiometría) es homogéneo o no.
- En caso de que no lo sea, si existe alguna muestra particular discrepante (*outlier*), o conjunto de muestras que se diferencien del resto.
- Si encontramos muestras características, es conveniente conocer cuáles de los parámetros analizados (llamados variables) son distintos del resto de las muestras.
- Siendo un poco ambiciosos, también nos gustaría saber si la información contenida en alguna de las variables es suplementaria del resto de variables y, por tanto, si realizamos demasiado trabajo analítico con el consiguiente gasto innecesario en tiempo y reactivos.

Este tipo de información se obtiene aplicando una serie de técnicas quimiométricas que se han denominado genéricamente como *pattern recognition*. En español se ha traducido por técnicas de reconocimiento de modelos, de formas, de pautas, etc. Concretamente, se aplican técnicas de agrupación o análisis *cluster* y técnicas de representación. Estas últimas tienen por finalidad representar cada uno de los 120 objetos – en el ejemplo escogido –, no en el espacio heptadimensional de las variables medidas (espacio que el ser humano es incapaz de visualizar), sino en un espacio reducido que sea, como máximo, tridimensional. El secreto, claro está, consiste en trasladar la máxima información contenida en el espacio multidimensional original al espacio de dimensionalidad reducida. De la misma forma que una buena foto – que es

bidimensional – conserva la información tridimensional de los aspectos retratados, las técnicas de representación han de hacer lo mismo con espacios de dimensionalidad elevada.

De hecho, la capacidad de combinar adecuadamente las variables medidas para formar otras nuevas – denominadas variables latentes o factores – que contengan la información original forma el núcleo de numerosas técnicas quimiométricas. Con ello se consigue extraer y representar información útil de un conjunto de datos multidimensional (multivariante en lenguaje quimiométrico, dado que usualmente las múltiples variables medidas varían simultáneamente) que, sin duda, contiene la información original pero de forma muy poco transparente y accesible.

Finalmente, también en este campo, nos podría interesar conocer si alguno de los nuevos lotes producidos (por nosotros mismos en otro periodo de tiempo o un producto similar de la competencia) es parecido o no a los producidos caracterizados anteriormente. Las técnicas a aplicar ahora son las técnicas de clasificación, y algunos métodos de este campo, como el análisis discriminante lineal o el método SIMCA, han adquirido cierta popularidad.

**Relaciones entre conjuntos de datos.** Supongamos ahora que nuestro interés se centra en relacionar la composición de las muestras con su calidad. El término “calidad” debe estar claramente identificado. A menudo consiste en la medición de parámetros físico-químico como el índice de octano para gasolinas, la suavidad al tacto en aprestos para tejidos o resultados de cata para vinos. Pueden existir una o más variables que quieran relacionarse con las variables que describen la composición, por lo que, ahora, el objetivo que se persigue es la relación entre las matrices X e Y representadas en la Figura 3.

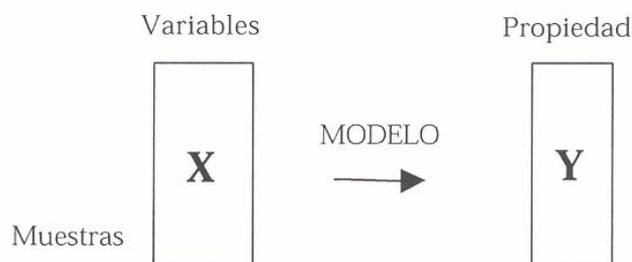


Figura 3. Representación de la relación entre una matriz de datos X formada por variables descriptoras de cada uno de los objetos y una matriz Y que contiene otro tipo de variables (en el ejemplo, son variables que describen propiedades como la calidad).

La relación entre matrices se establece mediante una ecuación o modelo matemático. Frecuentemente no existen modelos que relacionen X e Y y que obedezcan a leyes bien establecidas (como la ley de Lambert-Beer, la ley de Ohm o la de Fick). Por ello, los modelos que se construyen son, a menudo, modelos *ad hoc* que se adaptan a la estructura de los datos. Siendo así, es extremadamente importante que estos modelos sean válidos en el campo de trabajo establecido. La validación de los modelos es, pues, una etapa fundamental, y nunca debe obviarse en Quimiometría.

La ecuación (1) representa un modelo en el que las variables Y se relacionan con las variables X mediante la matriz de coeficientes B. Dado que cada valor medido de las variables

de la matriz  $Y$  no resulta exactamente de la combinación lineal de las variables de la matriz  $X$ , existe una matriz que contiene los errores,  $E$ . En caso de relacionar las variables predictoras de la matriz  $X$  con una sola variable, ésta toma la forma de vector,  $y$ , representándose mediante letras minúsculas, como en la ecuación (2).

$$Y = XB + E \quad (1)$$

$$y = Xb + e \quad (2)$$

Podemos estar interesados en dos tipos de informaciones que proporciona la parte determinista de los modelos. Por un lado, en el valor de los coeficientes,  $B$ . Estos coeficientes expresan la importancia de la relación entre las variables  $y$ , en muchos casos, explican el comportamiento de la variación existente en los datos.

Frecuentemente, sin embargo, estamos interesados en el carácter predictivo de los modelos. Una vez éstos han sido validados, nos interesa conocer el valor de la variable  $y$ , que ahora recibe el nombre de variable respuesta, toda vez que se han medido los valores de las variables  $x$  para el objeto analizado. A nadie se le escapa que si las variables respuesta de la matriz  $Y$  son difíciles, caras o tediosas de medir mientras que las variables de la matriz  $X$  se miden fácilmente (como por ejemplo un espectro en el infrarrojo medio), obtendremos una gran ventaja al poder predecir las primeras a partir de las segundas. El conjunto de técnicas que se utilizan para esta finalidad recibe el nombre de regresión multivariante. El lector interesado en estos temas reconocerá la Regresión mediante Componentes Principales (PCR, *Principal Component Regression*) o el análisis mediante Mínimos Cuadrados Parciales (PLS, *Partial Least Squares*) como alguno de los métodos más conocidos.

En las áreas de química orgánica, farmacología y otras, existe un gran interés en establecer relaciones entre la estructura de los compuestos sintetizados, sus características espectrales y su reactividad química o actividad farmacológica. Este es un campo muy activo que frecuentemente se generaliza con el nombre de QSAR, *Quantitative Structure Activity Relationships*.

**¿Dónde y cómo realizar las experiencias?** Los científicos con cierta experiencia en el análisis de datos reales saben que, frecuentemente, éstos no contienen la información suficiente. A menudo, los datos registrados no son representativos del fenómeno que quiere estudiarse, no contienen suficiente variabilidad, la parte aleatoria es más relevante que la parte correspondiente a la variación sistemática o existen otras razones por las que las técnicas quimiométricas no puedan extraer información útil sobre el conjunto de datos.

Si no existe información es evidente que es imposible extraerla, aunque sea con las técnicas quimiométricas más potentes. Quizás motivado por el hecho de que la información raramente se manifiesta de forma explícita a través de los datos registrados, este principio, que parece tan obvio, se ignora a menudo. Las consecuencias son desastrosas para la Quimiometría, dado que puede achacarse el fracaso del análisis a las técnicas utilizadas cuando en realidad es debido a la deficiencia de los datos registrados. Debemos asegurarnos, por tanto, que la información se encuentra efectivamente en los datos a analizar. La única forma de hacerlo es programando con detalle las experiencias que conducen a ellos. El diseño de experiencias o experimentos es la parte de la Quimiometría que estudia dónde, cómo y cuando deben realizarse las experiencias para que contengan la información necesaria. El empleo de los distintos tipos de diseños experimentales (factoriales completos o fraccionales, de Hadamard, etc.), de técnicas de optimización de resultados o de superficies de respuesta constituyen algunos de los aspectos estudiados por esta extensa área de la Quimiometría.

**Análisis de sistemas en evolución.** Supongamos que la empresa quiere monitorizar y controlar la reacción química mediante la cual se genera el producto en un reactor discontinuo. Conocer el nivel en que se encuentran presentes los productos intermedios o el tiempo en el que se obtiene el máximo rendimiento del producto final son aspectos que tienen implicaciones económicas importantes. Hasta ahora, normalmente la monitorización de estas reacciones se lleva a cabo mediante sondas que miden la temperatura, el flujo u otras variables físicas. Si se quiere obtener información sobre la composición, se debe extraer una porción de la mezcla de la reacción. Esta alícuota se analiza mediante cromatografía u otras técnicas instrumentales que proporcionen la suficiente selectividad. Este proceso analítico conlleva tiempo, con lo que no es posible seguir de forma suficientemente rápida el progreso de la reacción en el reactor industrial. Estamos ante un caso típico de un sistema en evolución que puede seguirse perfectamente *in situ* si se utilizan las llamadas técnicas multivariantes de resolución de curvas. Registrando, por ejemplo, sucesivos espectros NIR de la mezcla de la reacción (con un tiempo de registro aproximado por espectro de sólo unos segundos), puede conseguirse la información que se describía al principio del párrafo casi a tiempo real.

**Otras áreas de interés.** Existen muchas otras áreas en donde se aplican distintas técnicas quimiométricas. Por ejemplo, los algoritmos genéticos se han mostrado muy útiles cuando es necesario optimizar procesos en los que existen múltiples respuestas situadas en máximos locales.

En aquellos otros casos en donde las variables registradas (a menudo son datos instrumentales) tienen una relación no lineal con la respuesta de interés (una concentración, un índice de calidad, etc.), las redes neuronales se han mostrado muy eficaces.

### **Bibliografía**

- [1] M. Blanco, R. Boque, R. Cela, J. Coello, S. Paspocho, M.C. Ortiz, J. Riba, F.X. Rius, A. Ruiz, L.A. Sarabia y X. Tomás: *Avances en Quimiometría Práctica*. Universidad de Santiago de Compostela, 1991.
- [2] J.C. Miller, J.N. Miller: *Estadística y Quimiometría para Química Analítica* (4ª edición). Ed. Prentice Hall, 2002.
- [3] G. Ramis Ramos, M.C. García Álvarez-Coque: *Quimiometría*. Ed. Síntesis, 2001.
- [4] M. Valcárcel, A. Ríos: *La calidad en los laboratorios analíticos*. Ed. Reverté, 1992.

### **En Internet**

<http://www.ub.es/gesq/spchso>

*Sociedad Española de Quimiometría y Cualimetría*

<http://iris4.chem.ohiou.edu>

*North American Chapter of the International Chemometrics Society*

<http://argo.urv.es/quimio/cat/maincat.html>

*Grupo de Quimiometría y Cualimetría de la Universidad Pública de Tarragona*

Sociedad,Ciencia,Tecnología y Matemáticas

"Las Matemáticas y la Química: Quimiometría y Cualimetría"

Miguel A. Rodríguez Delgado

Profesor Titular de Química Analítica de la  
Universidad de La Laguna

## Quimiometría

"la disciplina química que utiliza métodos matemáticos y estadísticos para diseñar o seleccionar procedimientos de medida y experimentos óptimos, y para proporcionar la máxima información química mediante el análisis de datos químicos".

## Quimiometría

"la disciplina química que utiliza la Matemática, la Estadística y la lógica formal para diseñar o seleccionar procedimientos experimentales óptimos, proporcionar la máxima información química relevante a partir del análisis de datos químicos y obtener conocimiento a partir de sistemas químicos" (1997).



1) permite diseñar y realizar experimentos más eficaces, y extraer más información a partir de los datos

2) potencia la incorporación del químico, y en general, del profesional del laboratorio, a la cadena de control del proceso productivo

## QUIMIOMETRÍA QUÍMICA

BIOMETRÍA  
BIOLOGÍA

SOCIOMETRÍA  
SOCIOLOGÍA

ECONOMETRÍA  
ECONOMÍA

PSICOMETRÍA  
PSICOLOGÍA

### DATOS DE ORIGEN QUÍMICO

A) El limitado número de factores (concentraciones, pH, fuerza iónica, presencia de luz, etc.) que influyen sobre los resultados

B) La posibilidad de control de la mayoría de dichos factores por parte del experimentador. Puesto que los factores son escasos, y en su mayoría controlables, es fácil repetir los experimentos en las condiciones de trabajo deseadas

C) La incertidumbre relativamente pequeña (entre un 0.3% y un 3%) que afecta a la mayoría de las medidas analíticas.

*Física y Química [ $<10^2$ ]*

*Bioquímica [ $>10^2$ ]*

*Biología [ $>10^3$ ]*

*Ciencias Humanas [ $>10^4$ ]*

*.- Chemometrics and Intelligent Laboratory Systems (1986)*

*.- Journal of Chemometrics (1987)*



*C. F. Bender*

*"durante años, los científicos experimentalistas han llenado [de datos] los cuadernos de laboratorio, y a menudo los han descartado por falta de técnicas de interpretación adecuadas" (1973)*

### PROGRAMAS UTILIZADOS EN QUIMIOMETRÍA PAQUETES ESTADÍSTICOS DE APLICACIÓN GENERAL

*SAS (Statistical Analysis Systems)*

*Statistica*

*Statgraphics.*

*S-Plus*

*SPSS (Statistical Package for Social Sciences)*

*BMDP (Bio-Medical Data Processing)*

## PROGRAMAS UTILIZADOS EN QUIMIOMETRÍA

### PAQUETES ESTADÍSTICOS PARA TRATAMIENTO DE DATOS QUÍMICO-ANALÍTICOS

The Unscrambler

Trajans

Pirouette

PARVUS

## PROGRAMAS UTILIZADOS EN QUIMIOMETRÍA

### ENTORNOS ASISTIDOS PARA EL DESARROLLO DE HERRAMIENTAS ESTADÍSTICAS

MATLAB



INVESTIGACIÓN EN QUIMIOMETRÍA

## PROGRAMAS UTILIZADOS EN QUIMIOMETRÍA

### OTROS PROGRAMAS

LOTUS

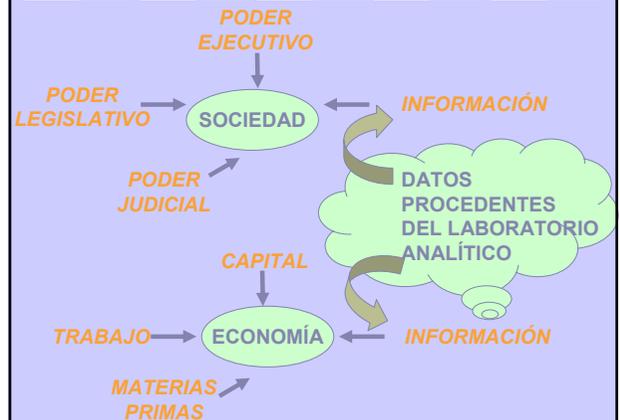
HERRAMIENTAS AUXILIARES

EXCEL



QUATTRO PRO

TRANSFORMACIÓN DE DATOS



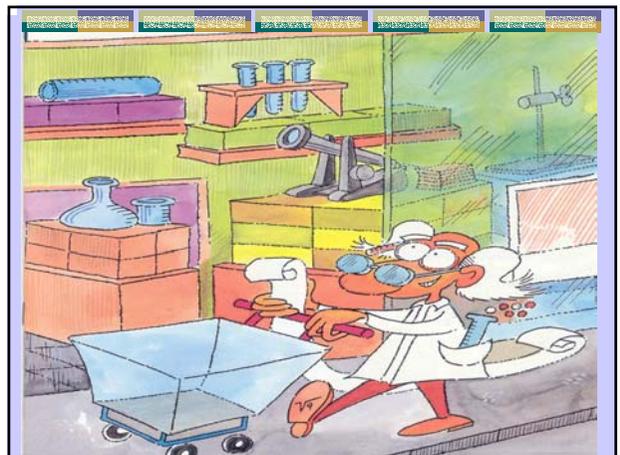
## CALIDAD

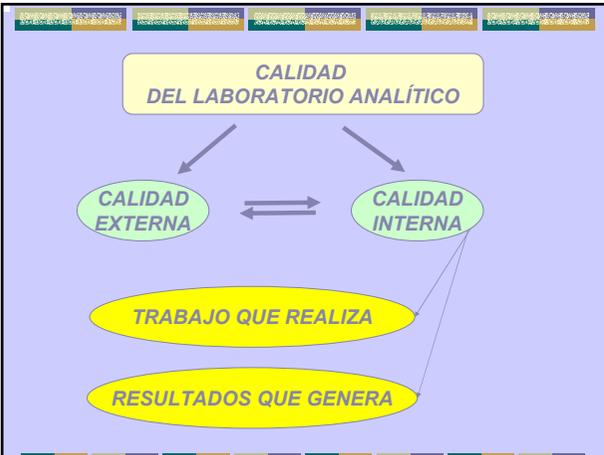
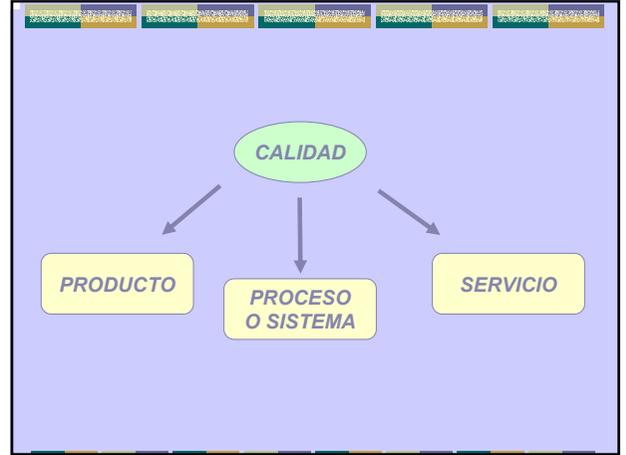
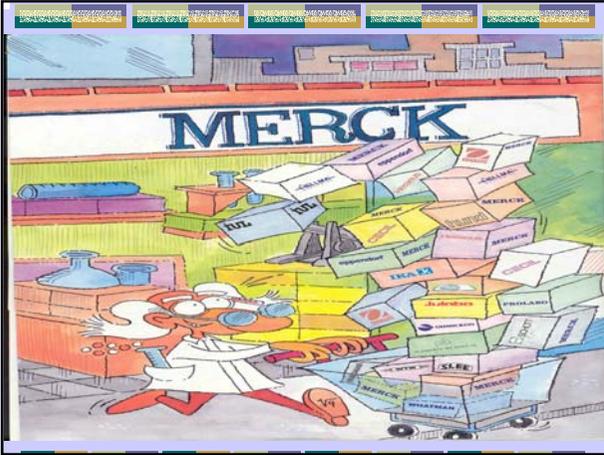
### La Real Academia Española

“ Propiedad o conjunto de propiedades inherentes a una cosa, que permiten apreciarla como igual, mejor o peor que las restantes de su especie”

### ISO (International Organization for Standardization)

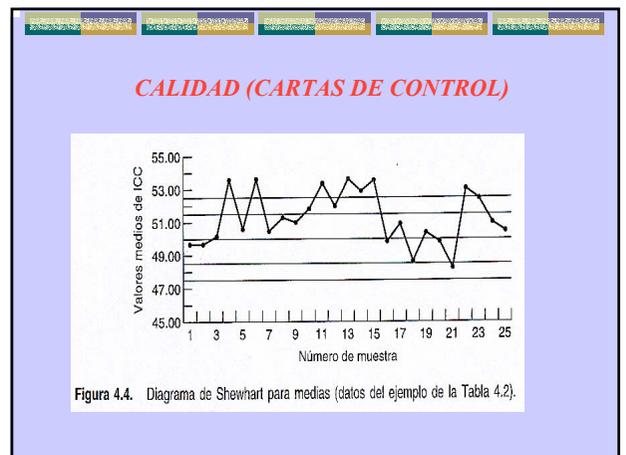
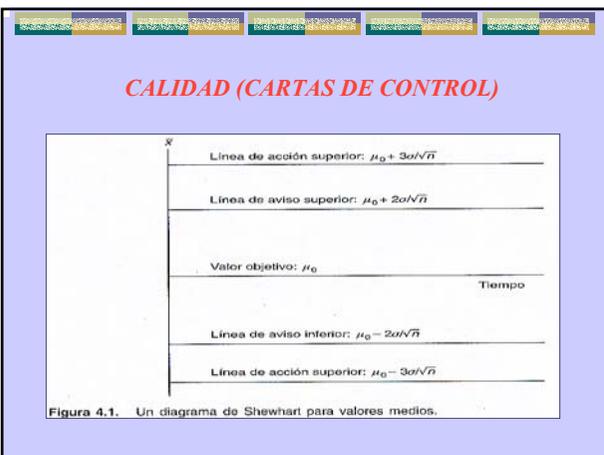
“ La totalidad de los rasgos y características de un producto, proceso o servicio que inciden en su capacidad de satisfacer necesidades reguladas o implícitas”





**CUALIMETRÍA**

*“Los conocimientos y herramientas que sirven para desarrollar y evaluar la calidad del análisis químico pero que se extiende también a otros campos”*



## OPTIMIZACIÓN Y DISEÑO EXPERIMENTAL

Sir Ronald A. Fisher

métodos estadísticos conjuntamente  
con el diseño de experimentos



DISEÑO EXPERIMENTAL Y OPTIMIZACIÓN

F. Yates (1934, 1937)

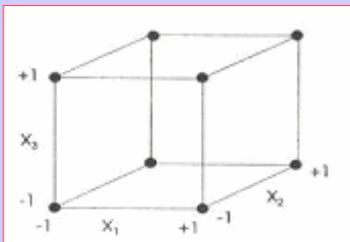
G.E.P. Box

Otros investigadores

O. Kempthorne (1952)

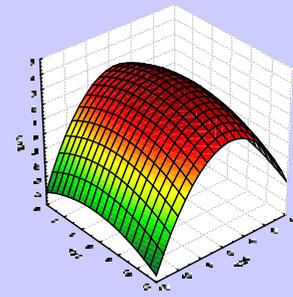
W.G. Cochran (1947, 1957).

## Diseños de screening



Diseño factorial completo a dos niveles.  $x_1, x_2$  y  $x_3$  representan los factores

## Diseños de superficie de respuesta



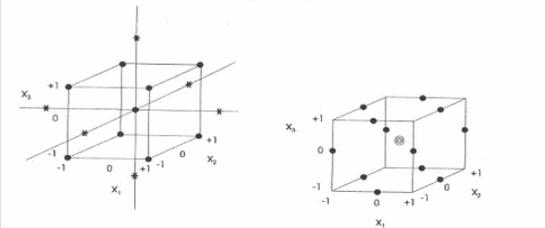
Ejemplo de superficie de respuesta

## Diseños de superficie de respuesta

### DISEÑOS ÓPTIMOS

Diseño compuesto central o  
Central Composite Design (CCD)

Diseño de Box-Behnken (DBB)



## ETAPAS DE UN PROCESO DE OPTIMIZACIÓN

.- Selección de una función objeto

.- Selección de los factores más importantes

.- Optimización

## REDES NEURONALES ARTIFICIALES

## LAS REDES NEURONALES ARTIFICIALES

*“Estructuras compuestas por elementos procesados de forma adaptada y simple, interconectados entre sí de una manera muy densa y compleja, capaces de llevar a cabo cálculos computacionales de gran envergadura para el procesamiento y representación de dichos datos”*

*(Hecht-Nielsen, 1990; Schalkoff, 1997).*

## LAS REDES NEURONALES ARTIFICIALES

*Capacidad para procesar información como lo hacen los sistemas biológicos*

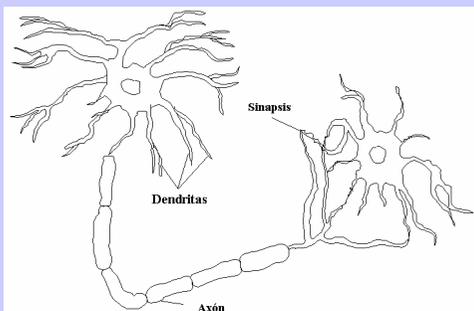
- con un procesamiento no lineal  
lo que permite un mejor ajuste de los datos*
- con una gran robustez  
lo que hace que presenten una insensibilidad al ruido*
- un gran paralelismo  
que se traduce en un rápido procesamiento de los datos*
- con una capacidad de aprendizaje y adaptación  
que permite al sistema modificar su estructura interna*

## LAS REDES NEURONALES ARTIFICIALES

*Capacidad para procesar información como lo hacen los sistemas biológicos*

- una gran habilidad  
para manejar información imprecisa y no clara.*
- con una capacidad para generalizar,  
que hacen que estos modelos puedan aplicarse al modelado de datos no conocidos*

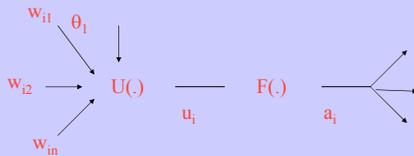
## Representación de una neurona biológica



## Datos de tejido cortical humano (las cifras son aproximadas)

Número de neuronas	100 billones
Número de sinapsis por neurona	1.000
Número total de sinapsis	100.000 billones
Operaciones por neurona	100
Número total de operaciones	10.000 trillones/sec
Volumen de la mente humana	300 cm <sup>3</sup>
Peso de la mente humana	1,5 kg
Longitud dendrítica	1 cm

### Unidad de proceso típica



### Función de propagación

#### Función de base radial (RBF)

$$U_i(w, x) = \sum_{j=1}^n w_{ij} x_j$$

#### Función lineal de base (LBF)

$$U_i(w, x) = \sqrt{\sum_{j=1}^n (x_j - w_{ij})^2}$$

### Función de activación o transferencia

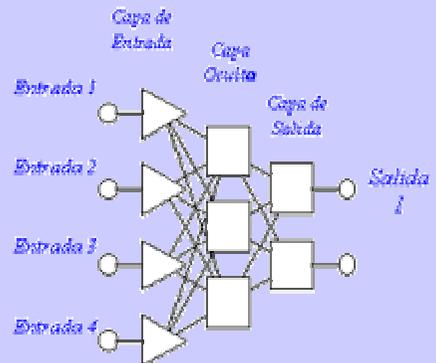
#### Función sigmoial

$$F(U_i) = \frac{1}{1 + e^{-U_i/\sigma}}$$

#### Función gaussiana

$$F(U_i) = ce^{-U_i/\sigma^2}$$

### Ejemplo de Red Neuronal Artificial.



### LAS ANNS SE PUEDEN CLASIFICAR

#### .-DE APRENDIZAJE SUPERVISADO

Implica un entrenamiento de la red con los outputs perfectamente conocidos y el uso del error del output predicho por la red con respecto al verdadero para determinar la cantidad en la que debe ajustarse cada peso

#### .-DE APRENDIZAJE NO SUPERVISADO

No requiere un output correcto para los inputs, sin embargo, explorando la estructura de los datos y las relaciones existentes entre los inputs, la red agrupa los inputs en clusters (categorías) basándose en su similitud (redes de Kohonen)

### REGLAS DE APRENDIZAJE

#### Regla de aprendizaje por corrección de error

Se utiliza en aprendizaje supervisado en el que la diferencia aritmética (error) entre la solución de las ANNs en cualquier momento (ciclo) del transcurso del entrenamiento de esa red y su correspondiente respuesta deseada, es utilizado para modificar los pesos hasta reducir gradualmente el error

#### Regla de aprendizaje de Boltzmann

Cada neurona genera una salida o output basado en una distribución estadística de Boltzmann (Jain et al., 1996) lo que lleva a un aprendizaje extremadamente lento

## REGLAS DE APRENDIZAJE

### Regla de aprendizaje de Hebb (Hebb, 1949)

*Si las neuronas a ambos lados de la sinapsis son activadas de forma sincronizada y repetidamente, la fuerza de la sinapsis se ve incrementada de manera selectiva”*

### El aprendizaje competitivo

*En el que todas las neuronas están forzadas a competir entre ellas de manera que sólo una será activada como resultado del ajuste de los pesos (Jain, 1996)*

## OTRAS CLASIFICACIONES DE LAS ANNs

- 1) *La función que va a desempeñar la ANN (clasificación, optimización...)*
- 2) *El grado (parcial o completo) de conexión de las neuronas de la red.*
- 3) *La dirección del flujo de información en la red (recurrente o no recurrente), siendo las redes recurrentes sistemas dinámicos en los que el estado en cualquier momento es dependiente de un estado previo.*

## OTRAS CLASIFICACIONES DE LAS ANNs

- 4) *El tipo de algoritmo de aprendizaje utilizado y que representa una serie de ecuaciones sistemáticas que utiliza las salidas o outputs obtenidas por la red, junto con una medida la actuación de la red para ajustar la estructura interna de la ANN.*
- 5) *La regla de aprendizaje (el motor del algoritmo de aprendizaje).*
- 6) *El grado de aprendizaje supervisado necesario para el entrenamiento de la red.*

## ALGORITMOS DE APRENDIZAJE

### RETROPROPAGACIÓN (BACK-PROPAGATION)

### CONJUGATE GRADIENT DESCEND

### QUASI-NEWTON

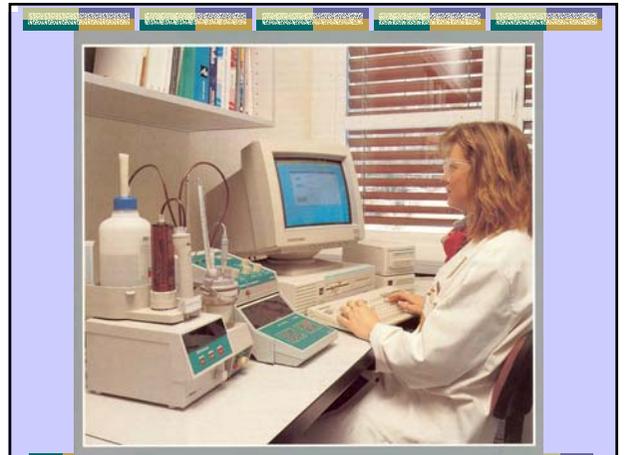
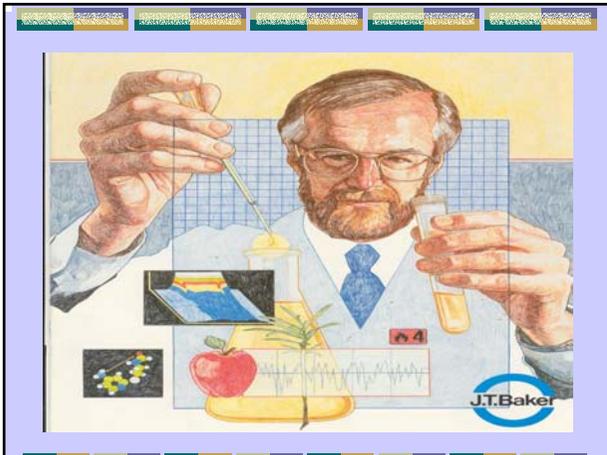
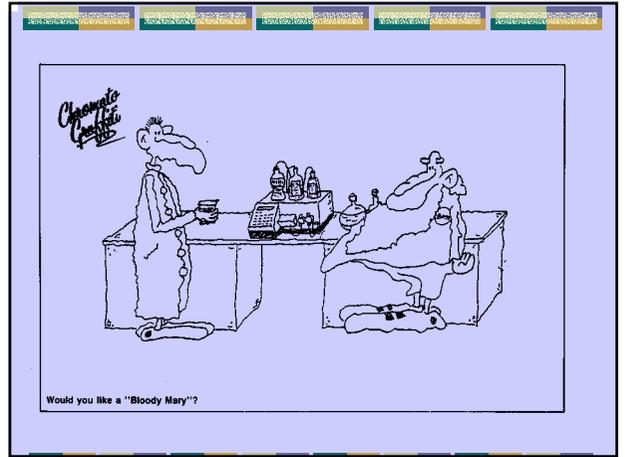
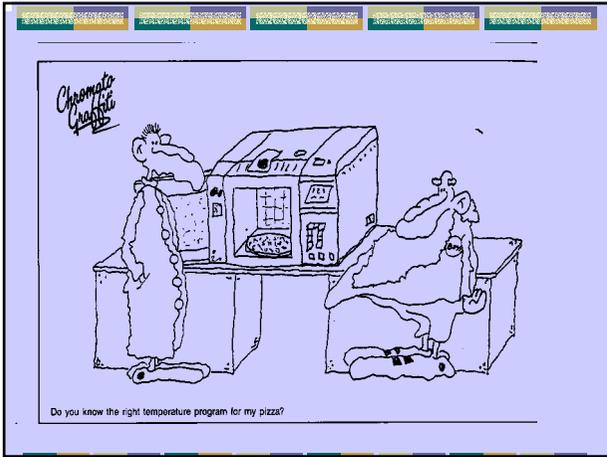
### LEVENBERG-MAROUARDT

## PROCESO DE SELECCIÓN DE UNA ANN

- 1) Búsqueda de la red neuronal que describa y se adapte mejor al problema en cuestión
- 2) Visualización de si la red es capaz de predecir datos que no ha visto o que no han sido utilizados en el entrenamiento
- 3) Utilización de esa red para predecir el resultado de aquellos datos deseados y de los cuales no se conoce su output

## APLICACIONES DE LAS ANNs

- 1) ASOCIACIÓN
- 2) CLASIFICACIÓN
- 3) MODELADO O MODELING
- 4) OPTIMIZACIÓN





## TIPIFICACIÓN DE LOS VINOS DULCES Y SECOS DE LANZAROTE, LA PALMA Y EL HIERRO

Memoria de Licenciatura  
Sergio Frías García

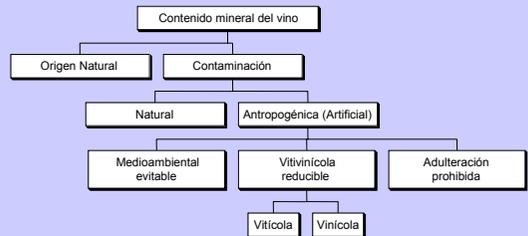
## AGRADECIMIENTOS

- Consejo Regulador de la Denominación de Origen de La Palma
- Consejo Regulador de la Denominación de Origen de El Hierro
- Consejo Regulador de la Denominación de Origen de Lanzarote
- ICIA

## IMPORTANCIA DE LOS ELEMENTOS MINERALES EN EL VINO

- Interés nutricional
- Interés toxicológico
- Quebras férricas y cúpricas
- Caracterización (fraude)

## ORIGEN DE LOS ELEMENTOS MINERALES EN EL VINO

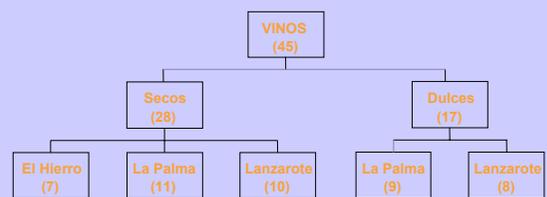


## OBJETIVOS

### ◆ TIPIFICACIÓN DE VINOS DE CANARIAS

- Estudiar el efecto matriz.
- Contenido (Fe, Cu, Mn, Zn, Sr, Li, Rb, Hg y Se) en los vinos dulces y secos de las denominaciones de origen de Lanzarote, La Palma y El Hierro.
- Análisis estadístico:
  - univariado
  - multivariado

## MUESTRAS



### EFFECTO MATRIX

	t <sub>calculado (dulces)</sub>	t <sub>calculado (secos)</sub>
<i>Fe</i>	28,9	15,7
<i>Cu</i>	10,4	21,1
<i>Mn</i>	65,8	13,6
<i>Zn</i>	38,8	19,2
<i>Sr</i>	4,9	14,5
<i>Li</i>	23,5	30,9
<i>Rb</i>	72,7	8,4
<i>Hg</i>	1,83	1,23

t<sub>critico</sub>=2,45

### EFFECTO MATRIX

DIGESTIÓN				
	Vino (ml)	HNO <sub>3</sub> (ml)	Tiempo (min)	Potencia (%)
Dulces	10	6	15	50
Secos	20	4	15	40
CONCENTRACIÓN				
	Tiempo (min)		Potencia (%)	
Dulces	30		100	
Secos	30		80	

### ELIMINACIÓN DEL EFECTO MATRIX

	t <sub>calculado (dulces)</sub>	t <sub>calculado (secos)</sub>
<i>Fe</i>	1,85	2,23
<i>Cu</i>	1,17	1,75
<i>Mn</i>	0,56	1,16
<i>Zn</i>	1,68	2,30
<i>Sr</i>	2,09	1,12
<i>Li</i>	1,81	1,69
<i>Rb</i>	1,01	2,01
<i>Hg</i>	0,65	0,35
<i>Se</i>	2,24	1,91

t<sub>critico</sub>=2,45

### REPETITIVIDAD

Elemento	CV %	
	SECOS (n=6)	DULCES (n=6)
<i>Fe</i> (mg/l)	0,44	1,13
<i>Cu</i> (mg/l)	0,94	1,31
<i>Mn</i> (mg/l)	0,47	1,13
<i>Zn</i> (mg/l)	1,23	1,55
<i>Sr</i> (mg/l)	0,86	0,50
<i>Rb</i> (mg/l)	1,70	0,10
<i>Li</i> (µg/l)	1,61	0,92
<i>Hg</i> (µg/l)	0,20	0,10
<i>Se</i> (µg/l)	0,39	0,12

### REPRODUCIBILIDAD

Elemento	CV %	
	SECOS (n=6)	DULCES (n=6)
<i>Fe</i> (mg/l)	2,21	2,71
<i>Cu</i> (mg/l)	3,54	3,12
<i>Mn</i> (mg/l)	1,53	1,96
<i>Zn</i> (mg/l)	2,45	2,54
<i>Sr</i> (mg/l)	1,74	1,23
<i>Rb</i> (mg/l)	1,65	1,78
<i>Li</i> (µg/l)	2,21	2,02
<i>Hg</i> (µg/l)	3,13	3,32
<i>Se</i> (µg/l)	2,69	2,97

### VINO DULCE CERTIFICADO

Elemento	Valores certificados		Valores obtenidos	
	Media	Intervalo de confianza de la media	Media	Intervalo de confianza de la media
<i>Fe</i> (mg/l)	6,7	0,78	6,3	0,1
<i>Cu</i> (mg/l)	1,15	0,15	1,20	0,01
<i>Zn</i> (mg/l)	0,32	0,09	0,30	0,00
<i>Mn</i> (mg/l)	1,1	0,2	1,3	0,1
<i>Li</i> (µg/l)	7	-	7,4	0,3
<i>Rb</i> (µg/l)	480	-	487	8
<i>Hg</i> (µg/l)	2,8	-	2,9	0,02

# VINOS DULCES

## CONTENIDO MEDIO Y DESVIACIÓN ESTÁNDAR

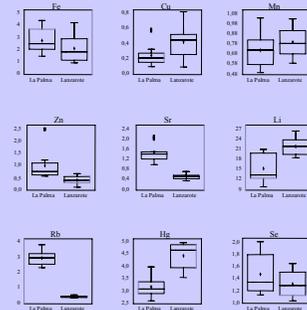
La Palma: Rb > Fe > Sr > Zn > Mn > Cu > Li > Hg > Se  
 Lanzarote: Fe > Mn > Cu > Sr > Rb > Zn > Li > Hg > Se

	Dulces (mg/l)	
	La Palma	Lanzarote
<b>Rb</b>	2,88 (0,45)	0,35 (0,06)
	2,23 – 3,73	0,25 – 0,46
<b>Fe</b>	2,6 (1,0)	2,0 (1,1)
	1,4 – 4,3	0,8 – 4,1
<b>Sr</b>	1,40 (0,35)	0,46 (0,11)
	0,93 – 2,02	0,29 – 0,65
<b>Zn</b>	0,9 (0,6)	0,3 (0,2)
	0,5 – 2,4	0,05 – 0,6
<b>Mn</b>	0,70 (0,17)	0,78 (0,15)
	0,48 – 1,02	0,57 – 1,01
<b>Cu</b>	0,23 (0,14)	0,40 (0,22)
	0,08 – 0,5	0,08 – 0,80
<b>Li*</b>	14,9 (4,0)	21,3 (2,5)
	9,4 – 20,5	17,9 – 25,8
<b>Hg*</b>	3,11 (0,47)	4,39 (0,53)
	2,55 – 3,93	3,49 – 4,92
<b>Se*</b>	1,45 (0,32)	1,29 (0,22)
	1,11 – 1,99	1,01 – 1,62

## ANOVA POR D.O.

	F	p	Dif. sig.
<b>Fe</b>	3,25	0,0809	No
<b>Cu</b>	5,80	0,0220	Sí
<b>Mn</b>	1,92	0,1750	No
<b>Zn</b>	15,14	0,0005	Sí
<b>Sr</b>	105,24	0,0000	Sí
<b>Li</b>	31,07	0,0000	Sí
<b>Rb</b>	495,82	0,0000	Sí
<b>Hg</b>	55,44	0,0000	Sí
<b>Se</b>	2,89	0,0990	No

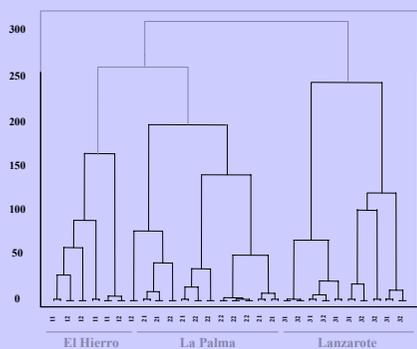
## CRÁFICOS DE CAJA Y BIGOTE POR D.O.



# VINOS SECOS

## ANÁLISIS ESTADÍSTICO MULTIVARIANTE

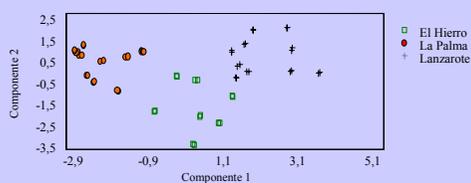
### ANÁLISIS DE AGLOMERADOS



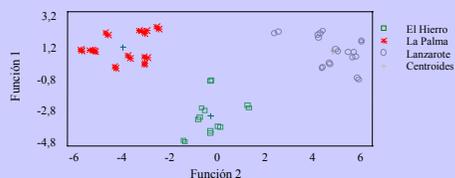
### ANÁLISIS DE COMPONENTES PRINCIPALES

	Componente 1	Componente 2	Componente 3
Fe	0,27	0,17	0,66
Cu	0,07	-0,65	0,15
Zn	0,30	-0,41	0,41
Mn	0,43	0,43	0,01
Sr	-0,37	0,25	0,35
Li	0,30	0,50	0,02
Rb	-0,44	-0,08	0,29
Hg	0,34	0,22	0,17
Se	-0,33	0,12	0,37
% var. exp.	44,98	18,66	13,63
% var. ac.	44,98	63,63	77,26

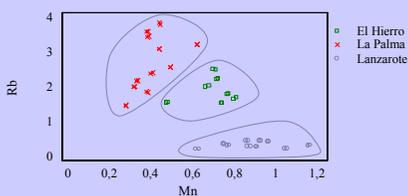
### ANÁLISIS DE COMPONENTES PRINCIPALES



### ANÁLISIS DISCRIMINANTE LINEAL POR D.O.



### ANÁLISIS DISCRIMINANTE LINEAL PASO-A-PASO



**Content of aliphatic hydrocarbons in limpets as a new way for classification of species using Artificial Neural Networks**

J. Hernández-Borges<sup>1</sup>, R. Corbella-Tena<sup>1</sup>, M. A. Rodríguez-Delgado<sup>1</sup>,

F. J. García-Montelongo<sup>1</sup> and J. Havel<sup>2</sup>

<sup>1</sup>Department of Analytical Chemistry, Nutrition and Food Science, Faculty of Chemistry, University of La Laguna, 38701-La Laguna, Spain

<sup>2</sup>Department of Analytical Chemistry, Faculty of Science, Masaryk University,

Kotlářská 2, 611 37 Brno, Czech Republic

Contact: J. Hernández-Borges. E-mail: jhborges@ull.es

## Abstract

It is demonstrated that biological species like limpets can be classified according to their level of n-alkanes when artificial neural networks are applied. Marine intertidal and subtidal limpets of the Canary Islands (Spain) *Patella piperata*, *Patella candei crenata* and *Patella ulyssiponensis aspera*, were selected as bioindicator organisms. Concentration of n-alkanes in the soft tissues of the limpets has been determined by gas chromatography. Data were treated with artificial neural networks (ANNs) and it was found that using suitable architecture of a supervised artificial neural network, the limpets can be successfully distinguished (classified) up to 98%.

## Introduction

Methods used to quantify pollutants in marine environments involve the study of the levels of pollutants in seawater, sediments or a member of the biota [1-3]. Due to their geographical situation and to the very dense tanker traffic around the Canary Islands ecosystem is threatened by accidental and chronic releases of hydrocarbons.

The most useful bioindicator organisms (mussels and oysters) do not occur along the Canary Islands coasts, so it has been necessary to look at other species present in the coastal areas as possible sentinel organisms [4,5]. That is the reason why intertidal and subtidal gastropods (limpets) were chosen.



Canary Islands

The determination of aliphatic hydrocarbons in bioindicator organisms is of great importance to precise the origin and destiny of these widely spread organic compounds throughout the environment. The content of aliphatics from n-C<sub>14</sub> up to n-C<sub>30</sub> is of special interest. Using them to obtain several indexes [6] the biogenic or antropogenic origin of these hydrocarbons can be established.

The aim of this work was to study whether these limpets could be classified according to the aliphatic hydrocarbons they accumulate, using ANNs [7,8]. If such a classification of the species of limpets or any other pollution bio-indicators could be done, that would reveal a different behaviour in the way they assimilate and/or metabolise pollutants. Moreover, the species could even be identified by its pollutants content.

## Experimental

### Data

Experimental data for n-alkanes were taken from [9], where three species of limpets were considered: *Patella piperata*, *Patella candei crenata* and *Patella ulyssiponensis aspera*. These species differ in their location to the coast, as *Patella ulyssiponensis aspera* is an intertidal limpet, while *Patella piperata* and *Patella candei crenata* are subtidal limpets.

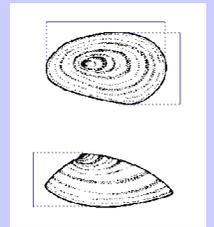


Fig. 2. General aspect of the shell of a limpet

The species were collected at four locations on the coasts of Fuerteventura every two months depending on the sea conditions, by hand at mid-tide, and their aliphatic content (from n-C<sub>14</sub> till n-C<sub>30</sub> including pristine and phytane) was determined by gas chromatography [4,9]. Totally 140 samples were analysed.

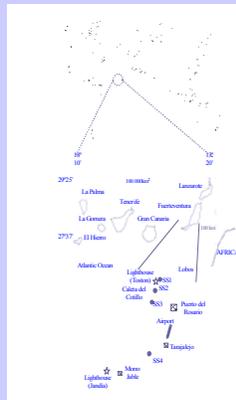
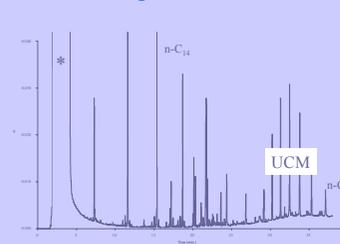


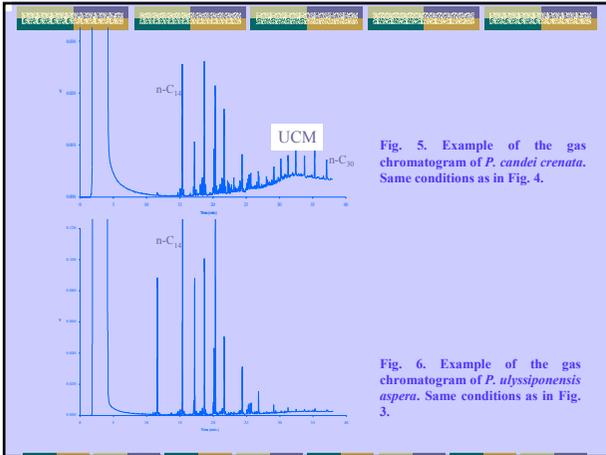
Fig. 3. Location of the Canary Islands and sampling stations (SS).

## Gas chromatograms



At the second hand-part of the chromatograms the UCM (unresolved complex mixture) can clearly be seen which is an indicator of a petrogenic origin of the aliphatic hydrocarbons of these limpets.

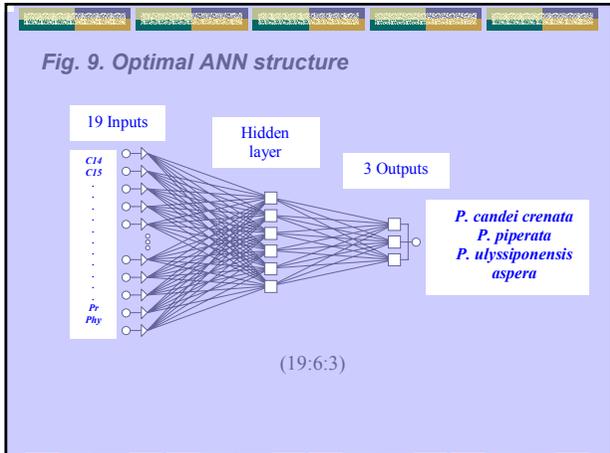
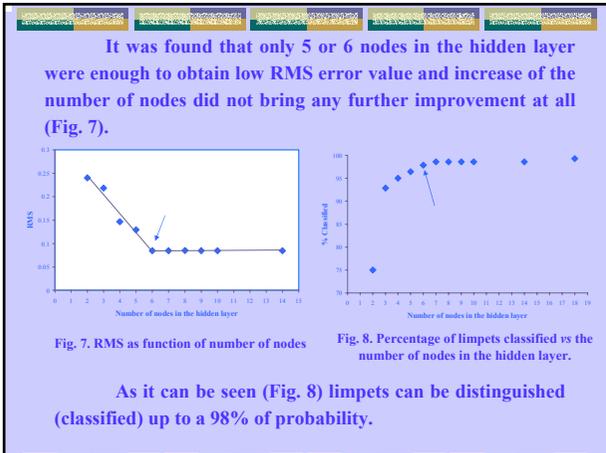
Fig. 4. Example of the gas chromatogram of *P. piperata*. Flame ionisation detector (Varian 3400, Walnut Creek, CA, USA) provided with a Varian 8200 Autosampler, on a SPB-5 32mm x 30m column (He carrier 1.2 ml min<sup>-1</sup>, N<sub>2</sub> make-up 30 ml min<sup>-1</sup>, splitless/split injection, 60°C for 3 min, 8°C min<sup>-1</sup> until 280°C, held for 9.5min; Injector 250°C; Detector 320°C). \* Peak of the solvent



## Results and discussion

### Modelling by ANN

Data were auto-scaled and processed on a Pentium III personal computer using Back Propagation (BP), Quick Propagation (QP) and/or Conjugate Gradient Descent (CGD) minimization methods. The data set was divided into the training set (140 patterns) and the verification set (12 patterns, four for each species of a limpet). Several architectures of the network were investigated and optimal one was searched for. Due to the complexity of the data, BP neural networks method did not usually reach an acceptable root mean square (RMS) value. For this reason combination of the three methods mentioned above was used. Usually, we started first with BP, then QP and CGD methods were applied.



### Prediction of species recognition

Sample	Classified as	RMS × 100	Sample	Classified as	RMS × 100	Sample	Classified as	RMS × 100
Cre 2	Cre	1.68	Pip 1	Pip	0.008	Uly 10	Uly	0.431
Cre 7	Cre	0.032	Pip 4	Uly	53.51	Uly 15	Uly	0.000
Cre 14	Cre	0.649	Pip 10	Pip	0.093	Uly 18	Uly	0.011
Cre 24	Cre	0.055	Pip 18	Pip	0.006	Uly 18	Uly	0.067
Cre 32	Cre	1.208	Pip 24	Pip	0.952	Uly 27	Uly	0.000
Cre 34	Cre	0.119	Pip 48	Pip	0.000			

Table 1. Part of the results of classification concerning *Patella piperata* (pip), *Patella candei crenata* (cre) and *Patella ulysiponensis aspera* (uly) with ANNs.

In this case one from 140 samples was taken out and 139 samples were used for modeling the optimal network (19:6:3) till a suitable RMS value was reached. The sample eliminated was introduced as a test one and it was classified. This process was repeated for other samples in the same way. Table 4 shows the results of the prediction for selected data. It can be concluded from Table 1 that the prediction ability of the network is excellent. Almost all of the cases were classified with prediction error less than 1.7 %.

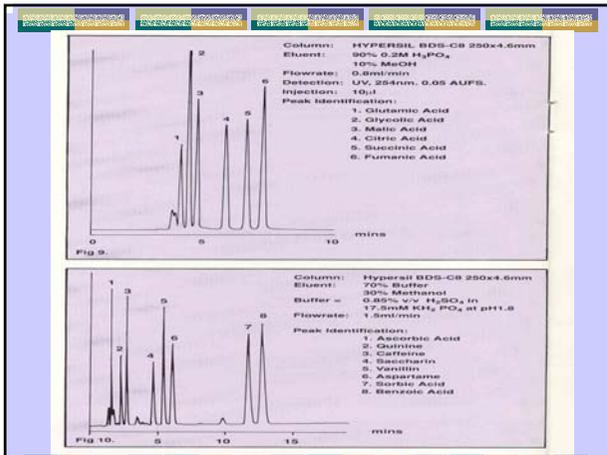
## Conclusions

A new way of biological species classification according to the level of pollutants in the tissues applying ANNs was proposed.

ANN was found suitable for classification of limpet species using analysis of aliphatic hydrocarbons as input data. Optimal ANN structure (19:6:3) permits correct classification of the species up to 98 % with a low prediction error (< 1.7 %).

The classification proposed could be applied when it is difficult to distinguish species only from morphological changes.

Moreover, classification using other pollutants content is under examination.



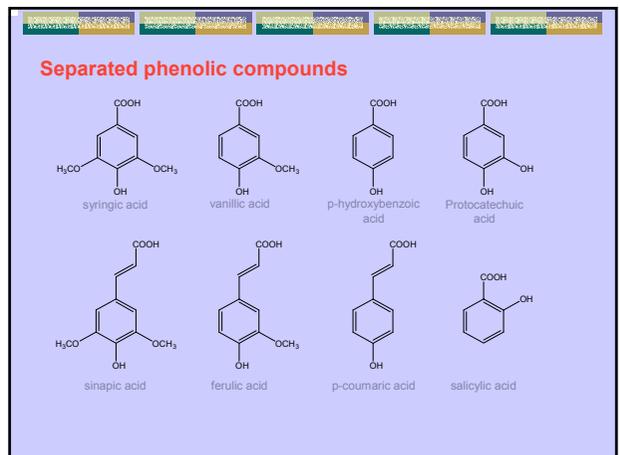
## EXPERIMENTAL DESIGN AND ARTIFICIAL NEURAL NETWORKS IN THE OPTIMISATION OF THE SEPARATION OF ANTIOXIDANTS IN FOOD SAMPLES BY CAPILLARY ELECTROPHORESIS

Vlastimil Dohnal<sup>1</sup>, Guillermo González<sup>2</sup>, Juan Pedro Pérez<sup>2</sup>, Francisco J. Garcia Montelongo<sup>2</sup>, Miguel Angel Rodriguez Delgado<sup>2</sup>

<sup>1</sup>Department of Analytical chemistry, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic  
<sup>2</sup>Department of Analytical Chemistry, Nutrición and Food Science, University of La Laguna, 38071 La Laguna, Tenerife, Sp

### Introduction

Plant phenolic compounds such as hydroxycinnamic acids are very important constituents of the human diet due to these group of compounds show biological properties directly related with beneficial effects on our health. Besides these group of compounds contributing to the sensory characteristics of foods, especially astringency and bitterness. A new capillary electrophoretic method for their determination in natural samples such as cereals, wine and beer was developed using combination of experimental design (ED) and artificial neural networks (ANNs) for their optimisation. With this approach the resolution of peaks, the efficiency of protocatechuic acid peak and time of analysis were improved in comparison with "classical" step-by-step approach. A central composite design with four variables (pH, concentration of phosphate, borate and  $\alpha$ -cyclodextrin) and the number of theoretical plates of protocatechuic acid peak (NTP) as a response was used along all optimisation process. Data from these experiments were used for adaptation of feed-forward back-propagation ANN being selected as the best architecture the following: 4 input neurons, 5 hidden neurons and 1 output neuron. Using this ANN architecture the predicted values of NTP and the corresponding conditions to the highest NTP were selected. The experiments performed under these conditions showed highly increased NTP for the most of compounds with short analysis times (less than 5 min.). Finally the developed method was applied for analysis of synthetic mixtures and real food samples – beer, wine and gofio (roasted cereals) with results according literature data



### Sample preparation

#### Beer & Wine

- Degassed on ultrasonic bath (10 min)
- pH adjusted to 2.0
- Solid phase microextraction

#### Cereals (gofio)

- Extraction of 4g by 18 mL MeOH on ultrasonic bath (10 min)
- Centrifugation 5 min (3 000 rpm)
- Filtration (0.45  $\mu$ m filter)
- Evaporation
- Dissolution in 1mL of MeOH

### CE Condition Analysis

Beckman CE System 5500  
 capillary i.d. 75  $\mu$ m, 40 (33) cm  
 separation voltage -15 kV  
 temperature 25°C  
 detection 214 nm

For analysis of samples the procedure was as follow. Prior to use, the capillary was washed with 1 M NaOH for 30 min, water 15 min. Each measuring day it was washed with 1M NaOH 10 min and for 5 min with water.

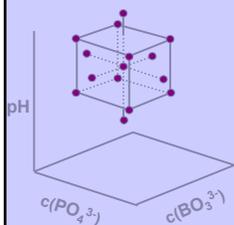
Method for analysis was as follow: 1 min wash with buffer, 2 min buffer fill (from different vial), 5 min analysis, 2 min wash with methanol and 1 min with 1 M NaOH. With this procedure the perfect reproducibility was obtained.

## Planning of experiments

Screening experiments

The reduced factorial design was used for the factor importance investigation.

Central composite design



27 experiments

Factors:	pH	6-8
	c( $\alpha$ -cyclodextrin)	1-5 mg/mL
	c( $\text{PO}_4^{3-}$ )	0-50 mM
	c( $\text{BO}_3^{3-}$ )	50-200 mM

Response: NTP of peak of protocatechuic acid's peak

## The Use of Artificial Neural Network (ANN)

Experimental data

Adaptation of ANN Prediction of the optimal conditions

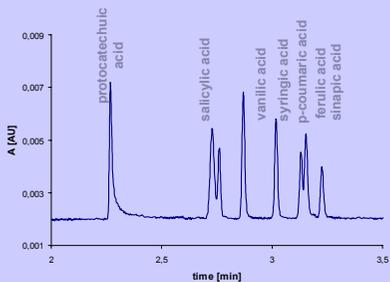
- learning set (20 points)
- test set (5 points)

- searching for the „best“ architecture

- using the „best“ ANN

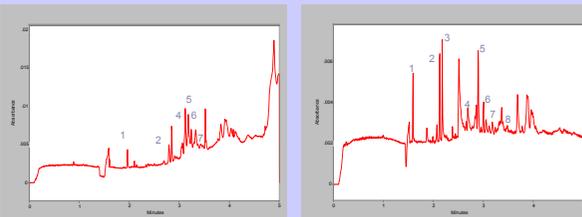


## Measurement under predicted the „best“ conditions



Predicted optimal conditions: pH 7.9, 3.0 mg/mL  $\alpha$ -cyclodextrin, 49 mM  $\text{PO}_4^{3-}$ , 125 mM  $\text{BO}_3^{3-}$ , -15 kV, 20°C, 0.1 mg/mL, 2 s injection

## Real samples analysis



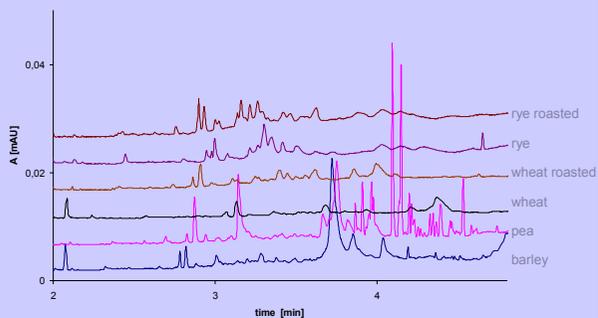
cereals (gofio)

wine

Separation conditions: pH 7.9, 3.0 mg/mL  $\alpha$ -cyclodextrin, 49 mM  $\text{PO}_4^{3-}$ , 125 mM  $\text{BO}_3^{3-}$ , -15 kV, 20°C, 0.1 mg/mL, 2 s injection

Gofio = flour made from different kinds of cereals

## Electropherograms of different types of gofio



## Conclusion

- new rapid method
- analysis of real samples (beer, wine and gofio)
- fingerprints of gofio for identification of origin




## Optimisation of electrophoretic separation of Humic Acids using Artificial Neural Networks and Experimental Design

J. Hernández-Borges<sup>1</sup>, M. L. Pacheco<sup>2</sup> and J. Havel<sup>2</sup>

<sup>1</sup>Department of Analytical Chemistry, Nutrition and Food Science, Faculty of Chemistry, University of La Laguna, 38701-La Laguna, Spain  
<sup>2</sup>Department of Analytical Chemistry, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic

Contact: Josef Havel. E-mail: havel@chemi.muni.cz

### Abstract

Capillary zone electrophoresis has been demonstrated to be a useful technique for the characterization and separation of Humic Acids (HAs). In this work a multicomponent background electrolyte (BGE) consisting of boric acid, Tris and EDTA was optimized using a combination of Experimental Design (ED) and Artificial Neural Networks (ANNs). The optimal composition of BGE was found to be 115 mM boric acid, 95 mM Tris and 0.6 mM EDTA (pH 8.4). A separation voltage of 20 kV, 20°C, 5 sec injection and detection at 214 nm were used. The optimized conditions were applied to characterize several HAs originated from various countries.

### Introduction

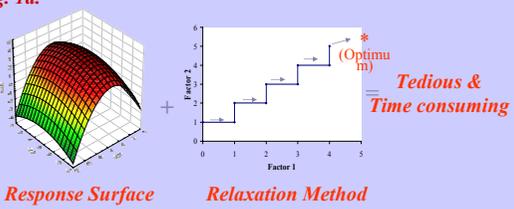
HAs constitute an important part of the organic carbon present in soils and surface waters [1]. The characterization of HAs provides important information to evaluate the interaction of humic substances with metal ions and different classes of organic compounds.

Despite the enormous number of publications in the last few years concerning the structure, chemistry and behavior of humic acids, there is still a problem of efficient separation.

Capillary zone electrophoresis (CZE) technique has proven to be very useful for the separation and characterization of HAs because of their polyelectrolytic properties [2]. In fact, it was CZE the technique by which it was found that humic acids are mixture of many compounds [3].

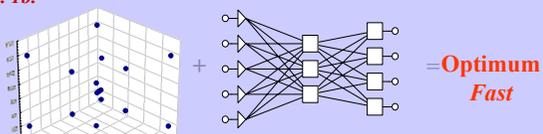
Single factor optimisation is still widely used in Chemistry. However, this step by step approach called Relaxation Method (Fig. 1a) is time consuming and is not always able to reach a minimum. On the contrary, using a combination of ED and ANNs one can predict the minimum easily with higher accuracy (Fig. 1b).

**Fig. 1a.**



*Response Surface*      *Relaxation Method*

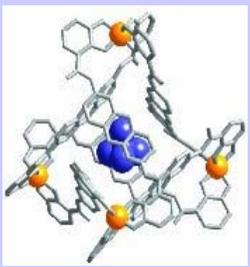
**Fig. 1b.**



*An Experimental Design*      *ANNs*      = **Optimum Fast**

Recently three component buffer consisting of Borate, Tris and EDTA was found efficient for the separation of peptides and proteins [4,5]. In a previous work Pacheco *et al.* [6] developed a multicomponent buffer to separate HAs. Because of the complex interactions between buffer components and also between buffer and HAs fractions, the optimisation of the separation is extremely difficult. In this case, a step by step optimisation is very tedious and time consuming. Therefore, the efficiency of a combination of ED and ANNs should be studied to reduce the number of experiments.

The aim of this work was therefore to examine chemometrical tools like ED and soft modeling of ANNs for optimisation, in order to find a good separation of HAs fractions.



## Results and discussion

A buffer consisting of boric acid, Tris and EDTA was used here as it was found an optimal one to obtain well reproducible electropherograms [6]. The presence of sodium is strongly interfering the separation patterns of HAs [6] and, therefore, this buffer has been also selected because it is sodium free.

When analyzing humic acids with other buffers, including borate alone, the typical humic acid hump appears (Fig. 3).

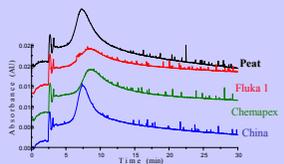


Fig.3. Analysis with 50mM  $\text{Na}_2\text{B}_4\text{O}_7$  as BGE.

However, when using a three component buffer consisting of Borate, Tris and EDTA the result is different (Fig. 4). That is why the concentrations of the three components should be optimized.

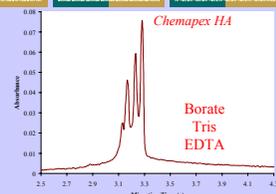


Fig.4. Electropherogram obtained using as buffer a mixture of borate, Tris and EDTA

After the concentrations of Tris, EDTA and boric acid were changed in single-factor experiments, multi-factorial experiments were performed by applying the shown ED (Fig. 5).

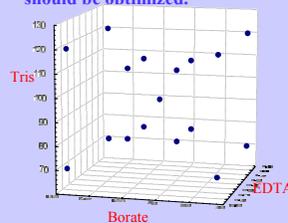


Fig.5. Experimental Design.

## Optimal ANN structure

Concentrations of Borate, Tris and EDTA were used as input values, while the number of peaks and the height of the highest peak were used as outputs. Fig. 6 shows optimal ANN structure.

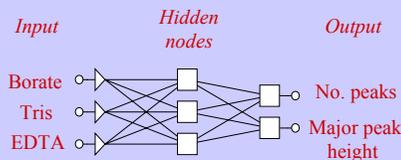


Fig. 6. The optimal structure of the ANN (3:3:2)

Table 1. Summarizes the experimental input values and the number of peaks and heights obtained with the trained network. It can be seen that a good agreement with experiments was reached.

INPUT			Exp. No. of peaks	Predicted No. of peaks	Exp. peak height	Predicted peak height
0.1	135	120	4	6.00	0.02352	0.0406
0.1	135	70	5	4.63	0.06239	0.0686
0.1	45	120	4	4.29	0.07365	0.0693
0.1	45	70	2	1.96	0.08278	0.0706
1.5	135	120	3	3.67	0.04379	0.0505
1.5	135	70	2	2.28	0.01993	0.0219
1.5	45	120	3	2.36	0.02027	0.0192
1.5	45	70	1	1.36	0.01154	0.0206
0.8	90	95	5	4.97	0.05736	0.0577
0.5	105	110	6	6.00	0.04367	0.0439
0.5	105	80	4	3.49	0.05906	0.0692
0.5	75	110	6	5.92	0.04954	0.0565
0.5	75	80	2	2.31	0.06483	0.0680
1.1	105	110	5	2.76	0.07021	0.0571
1.1	105	80	1	1.01	0.04465	0.0568
1.1	75	110	2	1.66	0.05364	0.0481
1.1	75	80	3	2.53	0.03529	0.0266

When giving to the network an extensive set of inputs in order to predict the corresponding output, it was found that the optimal concentration that yields in higher number of peaks and higher peak height is 115 mM Boric Acid, 95 mM Tris and 0.6 mM EDTA (Fig. 7).

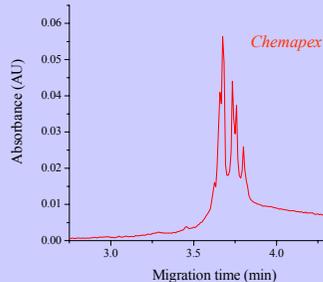


Fig. 7. Electropherogram of Chemapex HA, obtained with 115 mM Boric Acid, 95 mM Tris and 0.6 mM EDTA as a BGE, 20 kV, 20°C, 20 sec injection and 214 nm.

Once the composition of the buffer was optimized, the rest of the parameters (voltage, temperature...) were optimized separately. When reducing the injection time from 20 seconds to 5 an excellent separation of the fractions of humic acids takes place (Fig. 8).

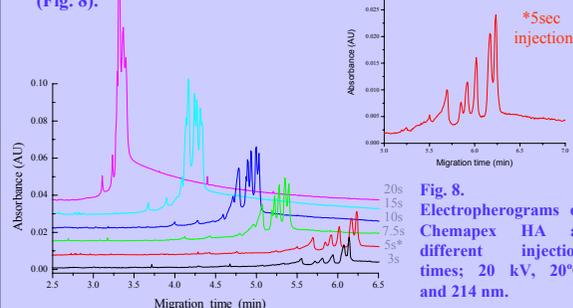


Fig. 8. Electropherograms of Chemapex HA at different injection times; 20 kV, 20°C and 214 nm.

Optimal analysis conditions were found to be 20 kV, 20°C and 5 sec injection. The next step was to compare the behaviour of different HAs in the optimal buffer and conditions (Fig. 9).

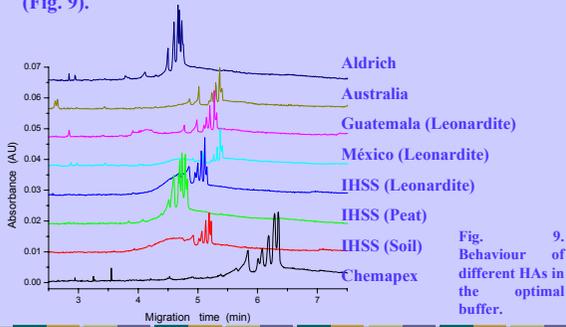


Fig. 9. Behaviour of different HAs in the optimal buffer.

## Conclusions

The use of ED with a combination of ANNs was found to be efficient optimisation approach.

A buffer consisting of Borate, Tris and EDTA was optimised by applying Experimental Design and Artificial Neural Networks. Optimal ANN structure was found to be (3:3:2).

The use of this complex buffer yields well-resolved and reproducible electropherograms that confirm the presence of several fractions in HAs which are similar or perhaps identical in most of the humic acids under study.

