

# QUIMIOMETRÍA, Una disciplina útil para el análisis químico

## Grupo de Quimiometría y Cualimetría de Tarragona

Departamento de Química Analítica y Química Orgánica

Universitat Rovira i Virgili.

Pl. Imperial Tàrraco, 1 43005-Tarragona España

### Resumen

Se describen en este artículo introductorio las principales características de la quimiometría. Se detalla su finalidad, su posición entre las ciencias experimentales, las herramientas o técnicas más relevantes que utiliza, los objetivos prácticos que persigue, los factores que han ayudado a su implantación y algunos de los objetivos que se plantean para el futuro.

### Introducción

La automatización y computerización de los laboratorios ha llevado consigo diversas consecuencias. Una de ellas es la rápida adquisición de gran cantidad de datos. Ahora bien, sabemos que la posesión de dichos datos dista, muchas veces, de proporcionar respuestas adecuadas. Obtener de datos no es sinónimo de poseer información; debemos interpretarlos y colocarlos en el contexto adecuado para convertirlos en información útil para el usuario. La *quimiometría* es la disciplina que tiene esta finalidad.

La palabra quimiometría, inventada hace aproximadamente treinta años, quiere resumir el concepto que engloba la medida en química. Se podría argumentar que, ciertamente, la medida en química siempre ha sido el campo de actuación de la química analítica. La quimiometría trata, específicamente, de todas aquellos procesos que transforman señales analíticas y datos más o menos complejos en información. La quimiometría utiliza métodos de origen matemático, estadístico y otros procedentes del campo de la lógica formal para conseguir sus fines. Por todo ello, la quimiometría se sitúa en un campo interdisciplinar. Aunque sus métodos y herramientas provienen de otras disciplinas (como, de hecho, ocurre habitualmente en la química analítica), claramente los fines de la quimiometría están ligados a la química y su éxito depende de los problemas químicos que sea capaz de resolver.

## 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 120x7 datos. La representación numérica, en forma de tabla o matriz de datos, de los valores obtenidos se representa en la Figura 1.

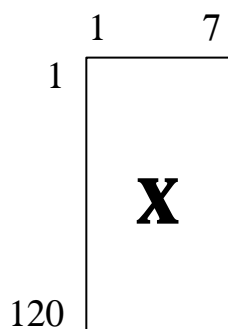


Figura 1. Representación de la matriz de datos **X** formada por 120 muestras u objetos. Cada uno de ellos está caracterizado por 7 parámetros o variables.

Nos puede interesar conocer: a) si todo el conjunto de muestras (llamadas objetos en el lenguaje utilizado en la quimiometría) es homogéneo o no. b) En caso de que no lo sea, si existe alguna muestra particular, discrepante (*outlier*), o conjunto de muestras que se diferencien del resto. c) Si encontramos muestras características, es conveniente conocer cuáles de los parámetros analizados (llamados *variables*) son distintos del resto de las muestras. d) 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.** Siguiendo con el ejemplo anterior, 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-químicos -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 2.

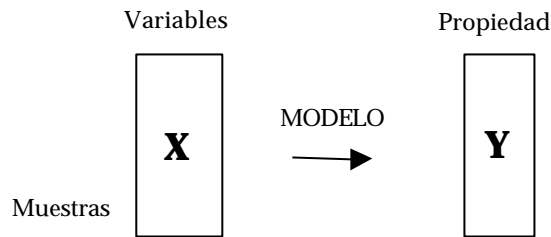


Figura 2. Representación de la relación entre una matriz de datos **X** formada por variables descriptoras de cada uno de los objetos y una matriz **Y** que contiene otro tipo de variables (en el ejemplo, son variables que describen propiedades como la calidad)

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 construidos *'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).

$$\mathbf{Y} = \mathbf{X} \mathbf{B} + \mathbf{E} \quad (1)$$

$$\mathbf{y} = \mathbf{X} \mathbf{b} + \mathbf{e} \quad (2)$$

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.

Normalmente existen diversos factores que intervienen en las experiencias tales como la composición y concentración de reactivos, presencia/ ausencia de catalizador, presión, temperatura, etc. Para llevar a cabo el número mínimo de experimentos –que suele ser un aspecto muy apreciado en ámbitos industriales, dado el coste que representan- deben variarse simultáneamente

los valores que asignamos a las variables experimentales. Esta estrategia, opuesta a la variación de un solo factor en cada experiencia mientras el resto permanece constante, demanda un cierto conocimiento de las técnicas adecuadas pero, a cambio, asegura la presencia de información relevante en los datos y una forma mucho más rápida y fiable de obtenerla.

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.** Siguiendo con el ejemplo propuesto anteriormente, 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 que 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.

Para que las técnicas quimiométricas muestren toda su potencialidad, los datos de entrada deben ser adecuados, es decir, la variación que contienen ha de deberse a las diferencias en la propiedad que quiere medirse. Si los datos contienen otros tipos de variabilidad, como por ejemplo, ruido aleatorio, deriva, correlación entre ellos, etc., entonces deben procesarse antes de someterlos al tratamiento quimiométrico propiamente dicho. El conjunto de métodos que preprocesan los datos reciben el nombre de técnicas de tratamiento de señales.

### ***Algunas claves para el éxito***

El éxito de la quimiometría depende en buena parte, tal como ocurre en otras disciplinas que se centran en el estudio de la medida (econometría, biometría, infometría,...), de la resolución de casos prácticos interesantes. Es conveniente, por tanto, que junto con el desarrollo de nuevos métodos, aquellos que están consolidados se apliquen con éxito a la resolución de problemas reales. En el caso de la quimiometría, los problemas prácticos a solucionar deben estar relacionados con la química. De esta forma, la quimiometría no corre el riesgo de convertirse en una rama de la ciencia alejada de la realidad. Más aún, los practicantes de la quimiometría deben estar atentos al desarrollo constante de la química y generar métodos que se adapten a las nuevas problemáticas que van surgiendo.

Para ello, por una parte, se necesitan investigadores bien formados y atentos a la realidad cambiante de la química. Por otra parte se requieren expertos, conocedores de las distintas técnicas, que sepan aplicar con éxito las herramientas disponibles. Sin embargo, lo más necesario es que exista un buen número de usuarios. Usuarios formados e informados, que conozcan y puedan definir con precisión los problemas químicos y a la vez conozcan el potencial de la quimiometría. Ellos deberían ser los interlocutores obligados de los expertos.

Para que todo ello sea una realidad, es necesario que se consoliden e incrementen los niveles de información, de formación y la accesibilidad al conocimiento y a las herramientas que hacen posible su aplicación.

Hoy en día, el medio más eficaz para adquirir o divulgar información es a través de *Internet*. Entre las numerosas *webs* relacionadas con la quimiometría, sin

duda, la de la Sociedad Española de Quimiometría y Cualimetría [1] representa una buena entrada. Allí mismo se indican muchas otras direcciones de otras tantas sociedades nacionales de quimiometría y de grupos que destacan en la escena internacional. La página de la sección norteamericana de la Sociedad Internacional de Quimiometría es una de las que sobresale por su valiosa información [2]. Además, esta sección es la responsable de mantener con vigor la lista de correo electrónico de la *International Chemometrics Society* [3] que facilita todas las discusiones, comentarios y observaciones que sus miembros consideran llevar a cabo de forma abierta.

En el capítulo de formación es donde podemos sentirnos más orgullosos los españoles. Debe reconocerse la labor de aquellos que, en su día, lograron que seamos de los pocos países, sino el único, que tenemos incorporada la enseñanza de la quimiometría en los planes de estudio de la licenciatura en Química. Este es un factor fundamental para consolidar la gran base de potenciales usuarios y de donde deben salir los expertos e investigadores del futuro. Adicionalmente, los diversos grupos de quimiometría que existen en nuestro país llevan a cabo acciones de formación continua mediante la organización de cursos de postgrado, seminarios o escuelas de quimiometría. De nuevo las distintas páginas *web* son el medio más eficaz para disponer de información sobre estas actividades.

El lector interesado en profundizar en el estudio de la quimiometría dispone de excelentes libros de texto [4-6], de artículos que se publican en las revistas especializadas, en particular, las dos que se dedican exclusivamente a la quimiometría [7,8], y de numerosos tutoriales que se encuentran en distintas páginas *web*.

Finalmente, cada día existen más programas de ordenador disponibles para los usuarios. De nuevo, los investigadores que desarrollan '*soft*' son cada día más conscientes de que la accesibilidad de sus programas redonda en un beneficio para la quimiometría.

Queda todavía queda mucho camino por recorrer para que la quimiometría sea una disciplina familiar al usuario final, pero se realizan pasos importantes. Por ejemplo, recientemente ha finalizado su trabajo una red temática que ha llevado a cabo ejercicios de intercomparación de métodos quimiométricos, y la Comisión Europea ha decidido crear el Instituto Virtual de Quimiometría y Metrología Industrial, que va a poner a disposición de todos los usuarios el conocimiento, las técnicas y metodologías quimiométricas que precisen para resolver sus problemas específicos [9]

## Conclusiones

Los intentos para definir la quimiometría han sido numerosos. Siebert [10] ha proporcionado recientemente una definición sugerente: "La quimiometria es la aplicación de métodos matemáticos y estadísticos así como de los principios de la buena ciencia de la medida para extraer de forma eficiente información útil de datos químicos"

Uno de los paradigmas del análisis químico, la especificidad de la señal analítica, se transforma mediante la quimiometría. En muchos casos ya no es de obligado cumplimiento eliminar las interferencias para obtener un resultado correcto. Con ello, las etapas tediosas de pretratamiento químico de las muestras para aislar los analitos de interés pueden soslayarse dado que la selectividad se consigue mediante métodos quimiométricos, no físico-químicos. La quimiometría contribuye, por tanto, a hacer realidad diversas tendencias de la química analítica actual: rapidez, abaratamiento de costes, miniaturización o transportabilidad de instrumentos.

## Referencias bibliográficas

- [1] <<http://www.ub.es/gesq/spchso/main.htm>>
- [2] <<http://iris4.chem.ohiou.edu/listserv.html>>
- [3] [ICS-L@UMDD.UMD.EDU](mailto:ICS-L@UMDD.UMD.EDU) Para incorporarse a esta lista debe enviarse un mensaje con el siguiente contenido '*Subscribe ICS-L your\_name\_goes\_here*' a la dirección: [listserv@umdd.umd.edu](mailto:listserv@umdd.umd.edu)
- [4] D.L. Massart et al. Handbook of Chemometrics and Qualimetrics. Part A and B. Data handling in science and technology. 20A, 20B. Elsevier, Amsterdam, 1997.
- [5] K.R. Beebe, R.J. Pell, M.B. Seasholtz. Chemometrics: A Practical Guide. Wiley, 1998
- [6] G. Ramis, M.C. García, Quimiometría. Ed. Síntesis, 2001
- [7] Chemometrics and Intelligent Laboratory Systems. Elsevier.  
<http://www.elsevier.nl/locate/chemometrics>
- [8] Journal of Chemometrics. Wiley.  
<http://www.interscience.wiley.com/jpages/0886-9383/>
- [9] Virtual Institute for Chemometrics and Industrial Metrology, VICIM. Thematic network supported under contract of the European Commission nº G7RT-CT-2001-05067
- [10] Karl J. Siebert, Chemometrics in Brewing - A Review. J Am Soc Brew Chem 2001, 59(4), 147-156.

Los autores agradecen todos los comentarios relacionados con los contenidos de este artículo. Pueden dirigirse, mediante mensaje electrónico, a la dirección: [quimio@quimica.urv.es](mailto:quimio@quimica.urv.es). Una versión en soporte electrónico de este artículo e información suplementaria puede encontrarse en: <http://www.quimica.urv.es/quimio>