

RECIENTES DESARROLLOS EN CALIBRACIÓN ANALÍTICA EMPLEANDO DATOS INSTRUMENTALES MULTI-VÍA

Alejandro César Olivieri

Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Instituto de Química de Rosario (IQUIR-CONICET), Suipacha 531, S2002LRK Rosario, Argentina, E-mail: olivieri@iquir-conicet.gov.ar

Resumen

El presente trabajo discute los avances que han tenido lugar en los últimos años en el campo del análisis de datos instrumentales multi-vía, es decir, datos que pueden ser organizados en un objeto matemático de al menos tres dimensiones, cuando se registran para un conjunto de muestras diferentes. La atención está focalizada en calibraciones analíticas cuantitativas, discutiéndose la generación de nuevos datos multi-vía, así como el desarrollo de algoritmos de procesamiento, y la estimación de cifras de mérito.

Palabras clave: calibración multi-vía, generación de datos, algoritmos, cifras de mérito

Abstract

Recent developments in analytical calibration employing multi-way instrumental data. The present work discusses the advances which have taken place in the last years in the field of multi-way instrumental data analysis, e.g., data which can be organized into a mathematical object of at least three dimensions, when they are registered for a group of different samples. Attention is focused on quantitative analytical calibrations, discussing the generation of new multiway data, as well as the development of processing algorithms and estimation of figures of merit.

Key words: multi-way calibration, data generation, algorithms, figures of merit

Introducción

Un único dato (escalar) medido para una dada muestra analítica se clasifica como *univariado*, mientras que múltiples datos instrumentales registrados para una única muestra se consideran *multivariados*. Cuando los datos multivariados para una muestra pueden organizarse en una tabla o matriz de datos con dos dimensiones diferentes, o cuando los datos para un conjunto de varias muestras se pueden organizar en un arreglo tri-dimensional, se dice que pertenecen a la clase multi-vía [1-2,3,4,5,6]. La relación entre datos univariados, multivariados y multi-vía se ilustra en la Figura 1.

La medición y procesamiento de datos multi-vía conlleva, desde el punto de vista de la química analítica, una serie de ventajas: (1) mayor sensibilidad, derivada del efecto de promediado del ruido que proviene de múltiples medidas de datos redundantes, (2) mayor

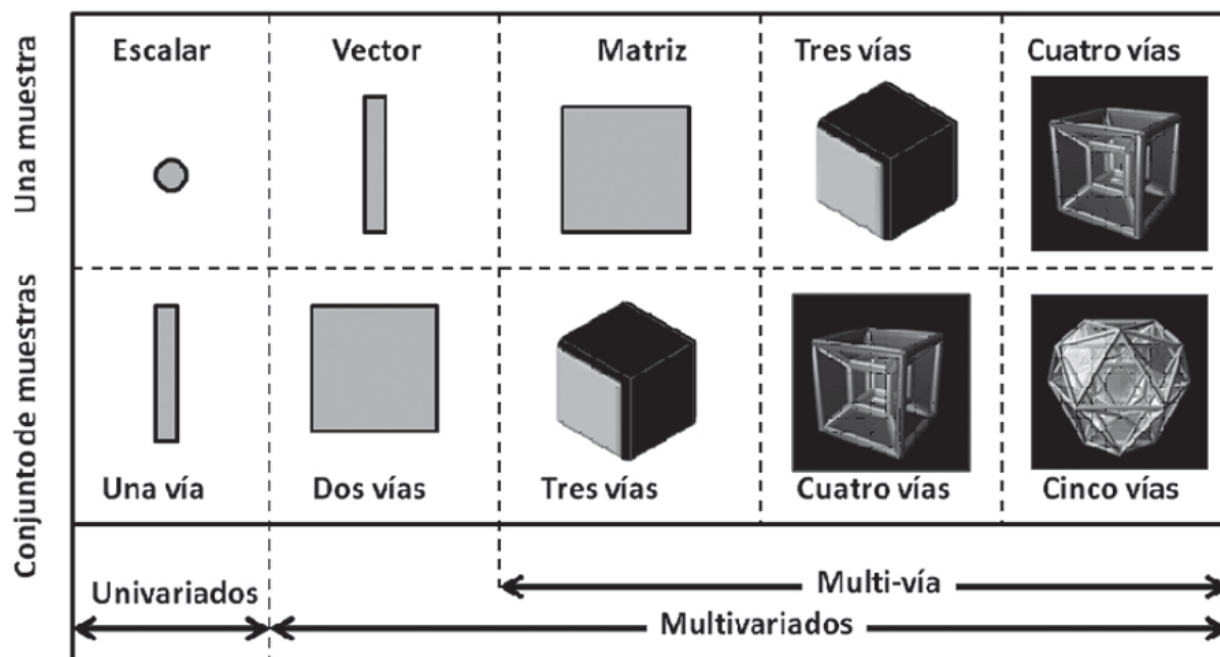


Fig. 1. Ilustración de los diferentes arreglos de datos que pueden construirse para una única muestra y para un conjunto de muestras, y la nomenclatura usualmente empleada para su clasificación.

selectividad, ya que cada nueva dimensión de los datos provee un grado adicional de selectividad parcial, y, probablemente lo más destacado, (3) modelado de la contribución del analito a la señal total y su determinación cuantitativa en presencia de interferencias potenciales desconocidas, ausentes en las muestras de calibración [7]. Es difícil exagerar la importancia de esta última ventaja; quizás sea la mayor contribución de la investigación de datos multi-vía a la química analítica. En términos sencillos, implica que luego de registrar datos multi-vía para muestras de un analito puro, su concentración puede determinarse en cualquier muestra futura, sin importar su complejidad o cuántos interferentes potencialmente contiene. Esta afirmación puede resultar excesivamente optimista, y quizás existan limitaciones de índole práctica para alcanzar esta ventaja; se trata de cuestiones que probablemente exijan intensos estudios teóricos y experimentales en el futuro cercano.

La Figura 2 muestra la distribución aproximada del número de trabajos publicados sobre análisis cuantitativos aplicando calibración multivariada (y multi-vía) a datos registrados mediante una variedad de técnicas. La mayor parte de estos trabajos se basan en mediciones de espectroscopía molecular, lo cual es razonable, debido a que la quimiometría analítica moderna comenzó cuando la espectroscopía de infrarrojo cercano (NIR) fue combinada con la regresión en cuadrados mínimos parciales (PLS) en la década de 1970 [8]. Otras medidas espectrales moleculares tales como absorción en el UV-visible e infrarrojo medio, luminiscencia, etc. fueron gradualmente incorporadas en los años subsiguientes. La generación de datos luminiscentes multi-vía es de particular importancia por una variedad de razones: (1) la luminiscencia es más sensible y selectiva que la absorción, (2) las matrices de excitación-emisión de luminiscencia pueden medirse convenientemente en un único instrumento, y (3) la estructura de los datos luminiscentes es más sencilla que la de otros datos multi-vía (véase más adelante). El campo de investigación espectral está aún en expansión, y es probable que en el futuro cercano la espectroscopía multi-vía domine la porción del gráfico de la Figura 2 en la zona etiquetada como "espectroscopía molecular".

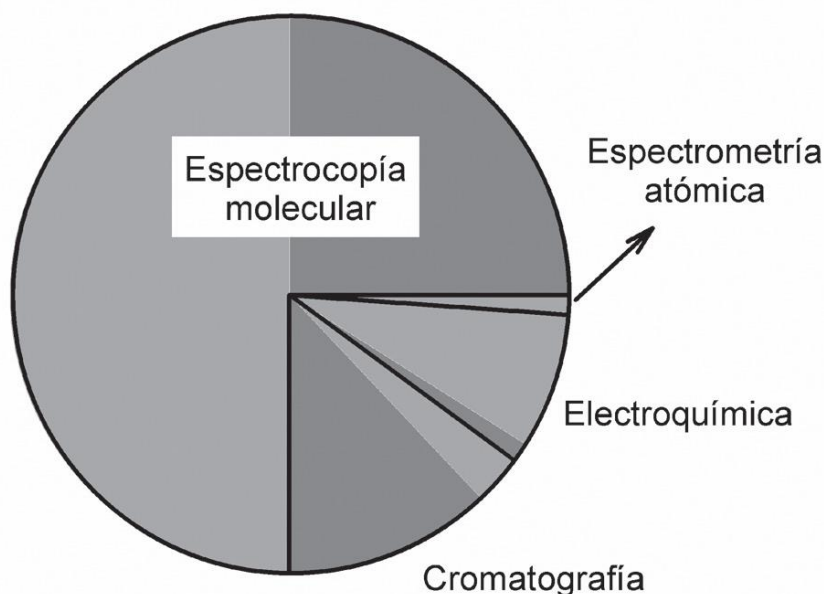


Fig. 2. Gráfico de torta que muestra el número relativo de trabajos publicados sobre calibración multivariada, clasificados de acuerdo con el tipo de datos registrados: espectroscopía molecular, cromatografía, electroquímica y espectrometría atómica. Dentro de cada porción, el color más oscuro corresponde a datos multi-vía. Nótese que los datos cromatográficos pueden involucrar detección multivariada espectral (o electroquímica), pero se clasifican como cromatográficos.

En el otro extremo del gráfico de torta de la Figura 2, la porción más pequeña corresponde, lógicamente, a la espectrometría atómica. La razón es simple: la probabilidad de superposición de líneas atómicas espectrales es muy baja, y por lo tanto hay poco espacio para métodos de calibración multivariada en el mundo atómico. Se conocen muy pocos ejemplos multivariados [9], y específicamente no hay trabajos publicados que se basan en información espectral atómica multi-vía.

El registro de barridos electroquímicos ha dado lugar a un cierto número de trabajos multivariados, los que ocupan una porción significativa del gráfico de torta de la Figura 2 [10].

Entre estos trabajos, no obstante, hay muy pocas determinaciones basadas en el registro de datos multi-vía.

La cromatografía representa una nueva frontera en la generación y procesamiento de datos multi-vía [11-12,13,14,15]. Los términos “*cromatografía*” (*chromatography*) [15] y “*cromatometría*” (*chromametrics*) [16] han sido acuñados para describir la conjunción de cromatografía y quimiometría. La aproximación tradicional para mejorar la selectividad en cromatografía ha sido y sigue siendo principalmente experimental (nuevas fases estacionarias, gradientes de pH, solvente o temperatura, derivatización química, etc.). Sin embargo, han aparecido nuevas posibilidades a medida que la cromatografía se ha vuelto bi-dimensional en sí misma, (LC-LC ó GC-GC) [17,18], y la detección ha aportado dimensiones adicionales [detección con arreglo de diodos (DAD) o espectrométrica de masas (MS)]. Esta será probablemente un área de activa investigación en la generación y análisis de datos multi-vía en los próximos años, de modo que la porción de la torta de la Figura 2, hoy ocupada por los datos cromatográficos, aumentará considerablemente de tamaño.

Nomenclatura

Componentes y muestras

Es importante definir diferentes tipos de constituyentes que pueden existir en una muestra bajo estudio. Particularmente relevantes son aquellos componentes que generan señales que se superponen con la del analito de interés, y por lo tanto pueden considerarse como potenciales interferentes. Los constituyentes presentes en el conjunto de muestras de calibrado se clasifican como “esperados”, debido a que se los incluye en dicho conjunto para que este sea suficientemente representativo. A su vez, pueden dividirse en “calibrados” y “no calibrados”, en referencia a si se conocen o no, respectivamente, sus concentraciones en las muestras de calibrado. Por otro lado, los constituyentes que solamente están presentes en una muestra desconocida se conocen normalmente como “inesperados”. Estos últimos podrían no producir una verdadera interferencia, en el sentido de generar un error sistemático en la determinación del analito [19]. Si la interferencia será real o sólo permanecerá en el terreno de lo potencial depende del tipo de datos que se analizan y del éxito logrado por la aproximación multi-vía que se aplique.

Tipos de calibración

Dependiendo del arreglo instrumental que se emplee, son posibles varias estructuras para los datos instrumentales registrados (Figura 1). Estas estructuras pueden describirse de acuerdo con su número de vías, que es numéricamente igual al número de dimensiones de un arreglo compuesto por datos registrados para un conjunto de muestras. Así, la clásica calibración univariada opera con datos de una vía (un escalar por muestra, un vector o arreglo de una vía para un conjunto de muestras), mientras que la bien conocida calibración NIR/PLS procesa datos de dos vías (un vector espectral por muestra, una matriz o arreglo de dos vías para un conjunto de muestras). Estructuras más complejas que estas corresponden al campo multi-vía: los datos de tres vías para un conjunto de muestras provienen de la medición de matrices por muestra, los de cuatro vías provienen de arreglos tri-dimensionales por muestra, y los de cinco vías de arreglos tetradimensionales por muestra (Figura 1).

Existe una descripción alternativa basada en el concepto de “orden”, que se basa en el número de dimensiones del arreglo de datos colectado para cada muestra. La calibración univariada de una vía es equivalente a la calibración de orden cero, y lo propio es cierto para los pares de calibraciones de dos vías/primer orden, tres vías/segundo orden, cuatro vías/tercer orden, cinco vías/cuarto orden, etc. El concepto de orden se emplea en la expresión popular “ventaja de segundo orden” [7], que es la propiedad antes mencionada que permite determinar analitos en presencia de constituyentes inesperados en una muestra.

Algoritmos multi-vía

Tres vías

Los algoritmos disponibles para el análisis de datos de tres vías pueden clasificarse de acuerdo a su estructura interna, en las siguientes clases: (1) solución de un problema de autovalores, (2) regresión por cuadrados mínimos directos, (3) cuadrados mínimos alternantes (ALS) trilineales, (4) ALS no trilineales, o (5) regresión en variables latentes. En la Tabla I se muestra un resumen adecuado.

Para la correcta aplicación de estas metodologías al análisis de datos de tres vías, es importante establecer si existen relaciones específicas entre los perfiles de los diferentes componentes de las muestras a lo largo de cada una de las dimensiones de los datos. Una propiedad de gran importancia en este sentido es la así llamada trilinealidad, que puede definirse de manera no matemática como sigue: un conjunto de datos de tres vías se considera trilineal si el comportamiento de un dado constituyente se puede describir mediante una relación lineal con cada uno de sus perfiles en las dimensiones de los datos, y dicho perfil es único y común a todas las muestras. Un ejemplo popular de datos trilineales son los de excitación-emisión de luminiscencia.

La violación del principio enunciado anteriormente produce desviaciones de la trilinealidad [4]. La causa más común de pérdida de la trilinealidad es la variación de los perfiles de los componentes en una de las dimensiones de los datos entre muestra y muestra, un fenómeno usual en datos multi-vía de origen cromatográfico (ver más adelante). Un caso particular lo constituyen los datos de espectrometría de masas bidimensionales (MS-MS), que son intrínsecamente no trilineales, porque el espectro de masa de un fragmento depende de su posición en el espectro madre.

Como se resume en la Tabla I, algoritmos que requieren trilinealidad de los datos son los que pertenecen a los siguientes grupos: (1), aniquilación generalizada de rango (GRAM) [20] y descomposición trilineal directa (DTLD) [21], (2) cuadrados mínimos bilineales (BLLS) [22], y (3) análisis de factores paralelos (PARAFAC) [23] y sus variantes, como la descomposición trilineal (TLD) alternante auto-ponderada (SWATLD) [24] y la TLD alternante penalizada (APTLTLD) [25].

Tabla I. Clasificación de algoritmos de procesamiento para juegos de datos de tres y cuatro vías.

Grupo	Estructura interna	Multi-linealidad	Algoritmo*	Vías	Ref.
(1)	Autovalores	Trilineal	GRAM	3	22
			DTLD	3	23
(2)	Cuadrados mínimos directos	Multi-lineal	BLLS/RBL	3	24
			TLLS/RTL	4	38
(3)	Cuadrados mínimos alternantes trilineales	Multi-lineal	PARAFAC	3, 4 y 5	25
			SWATLD	3	26
			APTLTLD	3	27
			APQLD	4	36
			AWRCQLD	4	37
(4)	Cuadrados mínimos alternantes no trilineal	No-trilineal	MCR-ALS	3	28
			PARAFAC2	3	29
(5)	Variables latentes	No-multilineal	U-PLS/RBL	3	32,33
			N-PLS/RBL	3	32
			U-PLS/RTL	4	38
			N-PLS/RTL	4	38
			U-PLS/RQL	5	38

* Ver el texto para más detalles sobre las siglas.

La pérdida de la trilinealidad puede ser tolerada por algoritmos alternativos. Cuando se producen cambios en los perfiles, las alternativas disponibles son las del grupo (4), resolución multivariada de curvas acoplada con ALS (MCR-ALS) [26] y PARAFAC2 [27]. Para

otros sistemas analíticos que muestran desviaciones de la trilinealidad, está disponible el grupo (5): PLS desdoblado y multi-vía (U-PLS y N-PLS) [28,29] Es importante notar que los algoritmos pertenecientes a los grupos 1, 3 y 4 logran en forma natural la ventaja de segundo orden, ya que sus modelos internos son capaces de descomponer la contribución de los potenciales interferentes de la de los analitos a la señal total. En los grupos 2 y 5, sin embargo, el logro de esta ventaja es una actividad post-calibración. La muestra incógnita se somete a un procedimiento conocido como de bilinealización residual (RBL), que separa la porción de la señal que puede explicarse mediante la calibración, de la contribución de los potenciales interferentes, dando lugar a la metodología híbrida BLLS/RBL [24], y a las más flexibles U-PLS/ RBL y N-PLS/RBL [30,31].

Existen métodos adicionales para estudiar sistemas con relaciones no lineales entre señal y concentración, pero son de uso más limitado [32,33].

Cuatro vías

En este caso las metodologías disponibles pertenecen a dos clases diferentes, que pueden homologarse a los grupos 3 y 5 de la sección anterior: cuadrados mínimos alternantes cuadrilineales (grupo 3) y modelos de variables latentes no quadri-lineales (grupo 5) (ver Tabla I). El modelo quadri-lineal PARAFAC es en realidad de tipo multi-vía, pudiendo aplicarse a cualquier número de vías. Existen también variantes complementarias de PARAFAC, tales como la descomposición quadri-lineal alternante penalizada (APQLD) [34] y la alternante ponderada con restricción residual (AWRCQLD) [35]. Estas metodologías pueden aplicarse sólo cuando los datos cumplen con la condición de quadri-linealidad, que implica que las señales de los constituyentes de las muestras pueden describirse por una relación lineal con los perfiles en cada dimensión, y que estos perfiles son comunes a todas las muestras.

Para tratar ciertas desviaciones de la quadri-linealidad, se han desarrollado modelos basados en estructuras latentes. Si se pretende lograr la ventaja de segundo orden, estos modelos deben acoplarse con la tri-linealización residual (RTL), que es la extensión natural de la RBL hacia una dimensión adicional, dando lugar a las regresiones en cuadrados mínimos desdoblados y multi-vía acopladas a RTL (U-PLS/RTL, N-PLS/ RTL) [38]. También existe la técnica híbrida de regresión en cuadrados mínimos trilineales acoplada a RTL (TLLS/RTL) [36].

Cinco vías

Como se explicó anteriormente, PARAFAC es un modelo generalizado que puede aplicarse a datos con cualquier número de vías, en el presente caso con la salvedad de que los datos deben ser penta-lineales. Un trabajo muy reciente describe un algoritmo alternativo que combina UPLS con quadri-linealización residual (RQL), procedimiento desarrollado como extensión de RBL y RTL hacia una dimensión más [37]. Es natural esperar que en los próximos años se produzcan nuevos desarrollos en este campo, motorizados por el progreso en el acoplamiento en tándem de instrumentos. Por lo tanto, es muy probable que la Tabla I se expanda para acomodar nuevos algoritmos.

Software

Existen pocos programas comerciales para análisis multi-vía. El software más poderoso, sin embargo, está disponible libremente en internet, aunque se encuentra fundamentalmente en forma de códigos de MATLAB [25], cuya implementación práctica requiere cierta habilidad en programación. En este sentido, se han desarrollado interfaces gráficas, que son útiles para hacer disponibles las herramientas informáticas a la comunidad analítica [38-39,40], y constituyen la mejor manera de popularizar la calibración multivía en el mercado analítico. La Tabla II muestra una variedad de códigos libres de MATLAB que pueden emplearse para el procesamiento de datos multi-vía.

Generación de datos

Tres vías

En este caso la aproximación clásica, todavía hoy en boga, consiste en registrar matrices de excitación-emisión de fluorescencia. Estas mediciones pueden efectuarse convenientemente en un único instrumento, las señales son sensibles y selectivas, y los datos son trilineales, de modo que todos los algoritmos disponibles pueden aplicarse para su estudio [41].

Tabla II. Software libre para procesamiento de datos multi-vía.

Nombre	Algoritmo(s)	Ref.	Página webb
The N-way toolbox	PARAFAC PARAFAC2 GRAM DTLD U-PLS N-PLS	25	http://www.models.life.ku.dk/algorithms
MCR graphical interface	MCR-ALS	40	http://www.mcrals.info/
GUIPRO graphical interface	MCR-ALS	41	http://personal.ecu.edu/gemperlinep
MVC2 graphical interface	PARAFAC APTLD SWATLD BLLS/RBL U-PLS/RBL N-PLS/RBL	42	www.chemometry.com
MVC3 graphical interface	PARAFAC APQLD AWRCQLD TLLS/RTL U-PLS/RTL N-PLS/RTL	^a	www.chemometry.com

^a En preparación.

^b Todas las páginas fueron visitadas en Diciembre de 2011.

Los avances más recientes en el registro de datos luminiscentes de tres vías que vale la pena mencionar se han dado merced al empleo de: (1) fases sólidas como soporte para los analitos, tales como membranas de nylon [42,43], o (2) técnicas de inyección en flujo, en las que el flujo se detiene cuando la muestra alcanza la celda de detección, y se miden matrices de excitación-emisión de luminiscencia. Dado que en este último caso no hay separación física de los componentes, la selectividad es provista por el algoritmo de calibración de tres vías. Se ha aplicado esta metodología usando diferentes fases sensoras dentro de la celda, tales como polímeros de impronta molecular [44], partículas de sílica gel C18 [45] y polvo de nylon [46,47].

Otros datos de tres vías que pueden generarse en un único instrumento consisten de espectros de absorción UV-visible (usualmente mediante DAD) o de emisión de fluorescencia, modulados en función de: (1) el tiempo de reacción en experimentos cinéticos, o (2) pH en experimentos de gradiente de pH [1,6]. Este tipo de datos no ha sido explotado exhaustivamente, pero constituye un modo sencillo de generar arreglos de tres vías.

La búsqueda de nuevas señales espectroscópicas de tres vías ha continuado, con interesantes posibilidades tales como el decaimiento temporal de luminiscencia sensibilizada de lantánidos [48] y la evolución temporal de quimioluminiscencia [49]. El reciente desarrollo de detectores del tipo DAD en el rango espectral NIR [50] puede impulsar la generación de datos de tres vías de interés industrial.

Otra alternativa para la generación de datos de tres vías es el acoplamiento de instrumentos en tándem. Un cromatógrafo con detección DAD, de fluorescencia de barrido rápido o espectrométrica de masas produce datos de tres vías: para una dada muestra, la primera dimensión es el tiempo de retención provisto por el cromatógrafo, y la segunda es la dimensión de la detección multivariada [4].

En este campo aparece un interesante problema quimiométrico: cuando los perfiles cromatográficos de los componentes cambian de experimento en experimento en la dimensión del tiempo de retención (ya sea en forma y/o en posición del máximo), los datos dejan de ser trilineales. Este problema puede atacarse de dos modos diferentes; (1) corrigiendo los cambios temporales y sincronizando los cromatogramas con un algoritmo apropiado, o (2) aplicando una metodología de calibración capaz de modelar dichos cambios, tales como MCR-ALS o PARAFAC2 [4].

La detección fluorescente de barrido rápido en cromatografía es conveniente debido a su mayor sensibilidad respecto de la absorción UV-visible. Datos de tres vías de LC-fluorescencia se han empleado para la determinación de: (1) diez hidrocarburos policíclicos aromáticos en muestras acuosas en presencia de interferentes mediante MCR-ALS [51], (2) ocho fluoroquinolonas en muestras con y sin interferentes usando PARAFAC y N-PLS/RBL luego de sincronizar los cromatogramas [52], y (3) los marcadores neopterina, biopterina, pterina, xantopterina e isoxantopterina en muestras de orina humana en presencia de interferentes usando MCR-ALS [53].

Finalmente, se han publicado pocos ejemplos electroquímicos de tres vías. Estos se refieren a la determinación de: (1) pesticidas mediante medidas cinéticas de adsorción en voltamperometría de onda cuadrada en el modo *stripping* [54], (2) plomo en muestras con interferentes mediante datos temporales potencial-pulso usando voltamperometría diferencial de pulso [55], y (3) tioles por cromatografía líquida con detección mediante rampas lineales de potencial [56]. En este campo se observarán avances significativos a medida que la detección electroquímica se haga más popular en cromatografía.

Cuatro vías

El registro de la evolución temporal de matrices de excitación-emisión de luminiscencia es una manera conveniente de producir datos de cuatro vías en un único instrumento. Un ejemplo es la medición de matrices de fosforescencia en función del tiempo de decaimiento en muestras enfriadas a la temperatura de helio líquido, y su posterior modelado mediante PARAFAC de cuatro vías para la determinación de 2, 3, 7, 8-tetraclorodibenzo-*p*-dioxina en muestras ambientales complejas [57].

El seguimiento de cinéticas de reacción ha sido más estudiado. Por ejemplo, la oxidación con permanganato ha dado lugar a las siguientes determinaciones: metotrexato y leucovorina en orina mediante PARAFAC [58], TLLS/RTL y U-PLS/RTL [38], ácido fólico y metotrexato en orina mediante PARAFAC y N-PLS [59], y en suero mediante U-PLS/RTL [60], y ácido fólico en presencia de sus dos metabolitos, los ácidos tetrahidrofólico y 5-metiltetrahidrofólico mediante U-PLS/RTL y N-PLS/RTL [61]. Hidrólisis en medio alcalino permitió análisis similares de cuatro vías: procaína y su metabolito el ácido *p*-amino benzoico en suero equino mediante NPLS/ RTL [62] y una variante de PARAFAC [36], carbaril en efluentes [63] y carbaril y 1-naftol en muestras de agua en presencia de potenciales interferentes [64]. La reacción de Hantzsch se utilizó para cuantificar malonaldehído en aceites de oliva, usando una variante no lineal de PARAFAC y una red neuronal acoplada con RTL [65]. Finalmente, las cinéticas de la inducción fotoquímica de fluorescencia y su degradación fotocatalítica fueron procesadas por PARAFAC de cuatro vías para el análisis de varios pesticidas no fluorescentes [66] y de mezclas ternarias de hidrocarburos policíclicos aromáticos respectivamente [67].

El acoplamiento en tándem de cromatografía bidimensional con detección multivariada da también lugar a datos de cuatro vías. En los experimentos de GC, los corrimientos de los tiempos de retención son poco comunes, de modo que los algoritmos cuadri-lineales pueden aplicarse directamente. A modo de ejemplo, puede citarse la medición de la composición porcentual de mezclas de biodiésels y diésels convencionales, mediante modelos PLS de creciente grado de complejidad aplicados a datos de GC con detección MS: (1) datos GC unidimensionales con detección de un único ión (dos vías), (2) datos bidimensionales GC-MS y GC-GC con detección de un único ión (tres vías) y (3) datos completos GC-GC-MS (cuatro vías) [68]. Las cifras de mérito mejoraron notablemente al pasar de los modelos de dos vías a los de tres vías, pero no cambiaron en forma significativa al pasar a los de cuatro vías.

En algunos casos se requirieron correcciones a los tiempos de retención. Por ejemplo, durante el estudio de extractos metabólicos aislados de células de levadura que metabolizan glucosa por fermentación, o etanol por respiración, con el objeto de obtener relaciones de concentraciones de metabolitos [69], se aplicó un algoritmo para la alineación de tiempos de retención basado en lograr coincidencias entre los picos de un dado ión [70].

En el caso de la cromatografía líquida bidimensional (LC-LC), los problemas de corrimientos en los tiempos de retención pueden resolverse dividiendo el cromatograma en pequeñas regiones temporales, empleando luego algoritmos cuadri-lineales. Esta estrategia se utilizó en análisis metabólicos de datos de LC-LC-DAD de cuatro vías mediante PARAFAC, permitiendo el estudio cuantitativo de un número importante de compuestos indólicos en plántulas de maíz [71].

En un reciente estudio metabólico mediante LC-LC-DAD, los datos originales de cuatro vías se desdoblaron en datos de tres vías, concatenando las matrices LC-LC en vectores

a cada longitud de onda. Estas matrices se procesaron luego mediante MCR-ALS, que permite cambios en los tiempos de retención [72].

En el futuro cercano, la generación de datos de cuatro vías, particularmente en cromatografía, recibirá un fuerte impulso. El potencial de la cromatografía bidimensional, junto con el poder de resolución de la detección basada en MS será seguramente explotado en muestras de complejidad creciente.

Merece notarse que el desdoblado de datos de cuatro a tres vías lleva a una disminución de la sensibilidad en comparación con el procesamiento de los datos en su estructura original. Esto requiere, previamente, sincronizar los cromatogramas bidimensionales en las dos dimensiones temporales. Es evidente que se necesitan estudios adicionales en el alineamiento de cromatogramas multi-dimensionales.

Cinco vías

Se ha publicado un único trabajo en este campo, basado en medidas de la evolución cinética y en función del pH de matrices de excitación-emisión de fluorescencia durante la hidrólisis del pesticida carbaril para producir 1-naftol [39]. Estos datos permitieron la resolución de mezclas de carbaril y potenciales interferentes no calibrados, mediante un nuevo algoritmo de estructura latente llamado U-PLS/RQL. Para la aplicación de PARAFAC, la naturaleza de los datos requirió que los arreglos de cinco vías originales fueran desdoblados a arreglos de cuatro vías, concatenando las dimensiones de pH y tiempo, debido a que los datos no puede describirse usando perfiles únicos e independientes para cada componente en estas dos dimensiones, ya que la cinética de hidrólisis es pH-dependiente. En este caso particular, la mayor sensibilidad se obtuvo mediante el algoritmo U-PLS/RQL, que mantuvo la estructura original de los datos.

Es importante disponer de algoritmos para el procesamiento de este tipo de datos, así como conocer sus ventajas y desventajas relativas. En el futuro, se verán nuevos desarrollos en la generación de datos cromatográficos multi-vía, acoplando cromatografía y detección multi-dimensionales. Por ejemplo, la cromatografía LC-LC ó GC-GC con detección MS-MS produciría datos de cinco vías. Dado que los datos MS-MS son intrínsecamente no bilineales, es evidente que se requerirán nuevos algoritmos no multi-lineales para procesar estos juegos de datos.

Cifras de mérito

En calibración de una y dos vías, la estimación de cifras de mérito está firmemente establecida, como lo documentan informes técnicos oficiales de la Unión Internacional de Química Pura y Aplicada (IUPAC) [73,74]. Probablemente la cifra de mérito más importante sea la sensibilidad, ya que esta constituye un ingrediente básico en las expresiones que permiten la estimación de las restantes cifras: sensibilidad analítica, selectividad, límite de detección, límite de cuantificación e incertidumbre en la predicción [75]. La sensibilidad puede definirse como el cambio en respuesta (neta) para un dado cambio en concentración de analito. Mientras en la calibración univariada la sensibilidad está dada por la pendiente de la recta de calibrado [73], en calibración de dos vías se ha desarrollado una expresión análoga basada en el concepto de señal neta del analito (NAS) [76]. La NAS es la porción de la señal total que puede adjudicarse al analito de interés de manera unívoca, y por lo tanto la pendiente de la

recta de calibrado pseudo-univariada (NAS vs. concentración) constituye una definición razonable de la sensibilidad.

Sin embargo, en el campo de los datos de tres vías, se intentó inicialmente aplicar el concepto de NAS, pero se encontraron dificultades en su interpretación, debido al desarrollo de definiciones conflictivas de la NAS. Por ejemplo, en el marco del modelo PARAFAC, se dedujeron dos expresiones diferentes, una por Messick, Kalivas y Lang (MKL) [77] y otra por Ho, Christian y Davidson (HCD) [78]. Estas expresiones condujeron a dos valores de sensibilidad diferentes, sin razón aparente alguna. La diferencia fue resuelta al demostrarse que tanto la definición MKL como la HCD son casos especiales de una definición general de la sensibilidad de PARAFAC (FO, por Faber y Olivieri) [79]:

$$SEN_{FO} = s_n \{[(\mathbf{B}_{cal}^T \mathbf{P}_{B,unx} \mathbf{B}_{cal}) * (\mathbf{C}_{cal}^T \mathbf{P}_{C,unx} \mathbf{C}_{cal})]^{-1}\}_{nn}^{-1/2} \quad (1)$$

donde s_n es la señal del analito puro a concentración unitaria, \mathbf{B}_{cal} y \mathbf{C}_{cal} son dos matrices que contienen los perfiles de los analitos calibrados en las dos dimensiones de los datos, '*' es el producto matricial de Hadamard, el subíndice 'nn' denota el n ésimo elemento diagonal de una matriz, y $\mathbf{P}_{B,unx}$ y $\mathbf{P}_{C,unx}$ son las siguientes matrices de proyección:

$$\mathbf{P}_{B,unx} = \mathbf{I} - \mathbf{B}_{unx} \mathbf{B}_{unx}^+ \quad (2)$$

$$\mathbf{P}_{C,unx} = \mathbf{I} - \mathbf{C}_{unx} \mathbf{C}_{unx}^+ \quad (3)$$

donde \mathbf{I} es una matriz unitaria, \mathbf{B}_{unx} y \mathbf{C}_{unx} contienen los perfiles de los componentes inesperados, y el superíndice '+' denota la operación inversa generalizada. En ausencia de potenciales interferentes, las matrices de las ecuaciones (2) y (3) son unitarias, y la expresión (1) se reduce a la ecuación MKL:

$$SEN_{MKL} = s_n \{[(\mathbf{B}_{cal}^T \mathbf{B}_{cal}) * (\mathbf{C}_{cal}^T \mathbf{C}_{cal})]^{-1}\}_{nn}^{-1/2} \quad (4)$$

Por otro lado, cuando sólo existe un único analito calibrado en presencia de interferentes potenciales, la ecuación (1) se reduce a la expresión HCD:

$$SEN_{HCD} = s_n \{[(\mathbf{B}^T \mathbf{B})^{-1}]_{nn} [(\mathbf{C}^T \mathbf{C})^{-1}]_{nn}\}^{-1/2} \quad (5)$$

donde las matrices \mathbf{B} y \mathbf{C} contienen todos los perfiles, tanto de componentes calibrados como de inesperados:

$$\mathbf{B} = [\mathbf{B}_{cal} \mid \mathbf{B}_{unx}] \quad (6)$$

$$\mathbf{C} = [\mathbf{C}_{cal} \mid \mathbf{C}_{unx}] \quad (7)$$

La exactitud de la ecuación (1) ha sido probada mediante simulaciones exhaustivas de Monte Carlo, habiéndose comprobado también que provee valores realistas en sistemas experimentales. Los cálculos de sensibilidad basados en la ecuación (1) han sido incorporados en una interfaz gráfica para calibración de tres vías [42].

Para el algoritmo PARAFAC, la sensibilidad permite definir en forma precisa la selectividad, la incertidumbre en la concentración y los límites de detección y cuantificación. La cuestión no está completamente resuelta en el caso de MCR-ALS o en las metodologías híbridas que emplean RBL, aunque se han producido algunos avances [33].

La extensión de la exitosa aproximación de tres vías hacia cuatro o más vías ha sido problemática [80]. Sólo recientemente se ha desarrollado una expresión para PARAFAC, que parece funcionar razonablemente bien en el caso de cuatro vías (y presumiblemente para mayor número de vías), pero requiere ser comprobada en sistemas simulados y experimentales adicionales [81]. Restan determinar las consecuencias referentes a las restantes cifras de mérito.

Finalmente, se conoce muy poco acerca de la sensibilidad de los algoritmos que emplean RTL o RQL, aunque seguramente se verán nuevos desarrollos en los próximos años.

Referencias

- [1] G. M. Escandar, N. M. Faber, H. C. Goicoechea, A. Muñoz de la Peña, A. C. Olivieri & R. J. Poppi, *Trends Anal. Chem.* **26**, 752 (2007).
- [2] G. M. Escandar, P. C. Damiani, H. C. Goicoechea & A. C. Olivieri, *Microchem. J.* **82**, 29 (2006).
- [3] A. C. Olivieri, *Anal. Chem.* **80**, 5713 (2008).
- [4] A. C. Olivieri, G. M. Escandar & A. Muñoz de la Peña, *Trends Anal. Chem.* **30**, 607(2011).
- [5] R. Bro, *Crit. Rev. Anal. Chem.* **36**, 279 (2006).
- [6] V. Gómez & M. P. Callao, *Anal. Chim. Acta* **627**, 169 (2008).
- [7] K. S. Booksh & B. R. Kowalski, *Anal. Chem.* **66**, 782A (1994).
- [8] H. Martens & M. Martens, *Multivariate analysis of quality: an introduction*, Wiley, Chichester, U.K., 2000.
- [9] F. Schmidt, L. Cornejo-Ponce, M. I. M. S. Bueno & R. J. Poppi, *J. Chemometrics* **32**, 423 (2003).
- [10] V. Pravdová, M. Pravda & G. G. Guilbault, *Anal. Lett.* **35**, 2389 (2002).
- [11] D. R. Stoll, X. Li, X. Wang, P. W. Carr, S. E. G. Porter & S. C. Rutan, *J. Chromatogr. A* **1168**, 3 (2007).
- [12] M. C. Ortiz & L. Sarabia, *J. Chromatogr. A* **1158**, 94 (2007).
- [13] A. C. Duarte & S. Capelo, *J. Liq. Chromatogr. Rel. Technol.* **29**, 1143 (2006).
- [14] T. Skov & R. Bro, *Anal. Bioanal. Chem.* **390**, 281 (2008).
- [15] J. M. Amigo, T. Skov & R. Bro, *Chem. Rev.* **110**, 4582 (2010).
- [16] V. G. van Mispelaar, *Chromometrics*, Tesis Doctoral, Amsterdam, Holanda, 2005.
- [17] I. François, K. Sandra & P. Sandra, *Anal. Chim. Acta* **641**, 14 (2009).
- [18] M. Adahchour, J. Beens & U. A. Th. Brinkman, *J. Chromatogr. A* **1186**, 67 (2008).
- [19] W. E. Van der Linden, *Pure Appl. Chem.* **61**, 91 (1989).
- [20] E. Sanchez & B. R. Kowalski, *Anal. Chem.* **58**, 496 (1986).
- [21] E. Sanchez & B. R. Kowalski, *J. Chemometrics* **1**, 29 (1990).
- [22] M. Linder & R. Sundberg, *J. Chemometrics* **16**, 12 (2002).
- [23] R. Bro, *Chemom. Intell. Lab. Syst.* **38**, 149 (1997).
- [24] Z. P. Chen, H. L. Wu, J. H. Jiang, Y. Li & R. Q. Yu, *Chemom. Intell. Lab. Syst.* **52**, 75 (2000).
- [25] A. L. Xia, H. L. Wu, D. M. Fang, Y. J. Ding, L. Q. Hu & R. Q. Yu, *J. Chemometrics* **19**, 65 (2005).
- [26] R. Tauler, *Chemom. Intell. Lab. Syst.* **30**, 133 (1995).
- [27] H. A. L. Kiers, J. M. F. Ten Berge & R. Bro, *J. Chemometrics* **13**, 275 (1999).
- [28] S. Wold, P. Geladi, K. Esbensen & J. Öhman, *J. Chemometrics* **1**, 41 (1987).
- [29] R. Bro, *J. Chemometrics* **10**, 47 (1996).
- [30] J. Öhman, P. Geladi & S. Wold, *J. Chemometrics* **4**, 79 (1990).
- [31] A. C. Olivieri, *J. Chemometrics* **19**, 253 (2005). [32] A. C. Olivieri, *J. Chemometrics* **19**, 615 (2005).

- [33] A. García Reiriz, P. C. Damiani & A. C. Olivieri, *Chemom. Intell. Lab. Syst.* **100**, 127 (2010).
- [34] A. L. Xia, H. L. Wu, S. F. Li, S. H. Zhu, L. Q. Hu & R. Q. Yu, *J. Chemometrics* **21**, 133 (2007).
- [35] H. Y. Fu, H. L. Wu, Y. J. Yu, L. L. Yu, S. R. Zhang, J. F. Nie, S. F. Li & R. Q. Yu, *J. Chemometrics* **25**, 408 (2011).
- [36] J. A. Arancibia, A. C. Olivieri, D. Bohoyo Gil, A. Espinosa Mansilla, I. Durán Merás & A. Muñoz de la Peña, *Chemom. Intell. Lab. Syst.* **80**, 77 (2006).
- [37] R. M. Maggio, A. Muñoz de la Peña & A. C. Olivieri, *Chemom. Intell. Lab. Syst.* **109**, 178 (2011).
- [38] J. Jaumot, R. Gargallo, A. de Juan & R. Tauler, *Chemom. Intell. Lab. Syst.* **76**, 101(2005).
- [39] P. J. Gemperline & E. Cash, *Anal. Chem.* **75**, 4236 (2003).
- [40] A. C. Olivieri, H. L. Wu & R. Q. Yu, *Chemom. Intell. Lab. Syst.* **96**, 246 (2009).
- [41] C. M. Andersen & R. Bro, *J. Chemometrics* **17**, 200 (2003).
- [42] S. A. Bortolato, J. A. Arancibia & G. M. Escandar, *Anal. Chim. Acta* **613**, 218 (2008).
- [43] S. A. Bortolato, J. A. Arancibia, G. M. Escandar, *Anal. Chem.* **80**, 8276 (2008).
- [44] A. Valero-Navarro, P. C. Damiani, J. F. Fernández- -Sánchez, A. Segura-Carretero & A. Fernández- -Gutiérrez, *Talanta* **78**, 57 (2009).
- [45] S. A. Bortolato, J. A. Arancibia & G. M. Escandar, *Environ. Sci. Technol.* **45**, 1513 (2011).
- [46] G. Piccirilli & G. M. Escandar, *Anal. Chim. Acta* **601**, 196 (2007).
- [47] G. Piccirilli & G. M. Escandar, *Anal. Chim. Acta* **646**, 90 (2009).
- [48] V. A. Lozano, G. A. Ibañez & A. C. Olivieri, *Anal. Chim. Acta* **610**, 186 (2008).
- [49] V. A. Lozano, G. A. Ibañez & A. C. Olivieri, *Anal. Chem.* **82**, 4510 (2010).
- [50] D. Bertrand & V. Baeten, Instrumentation, en *La spectroscopie infrarouge et ses applications analytiques*, 2da. ed., D. Bertrand & E. Dufour, Eds.; TEC & DOC, París, 2006, pp 247-305.
- [51] S. A. Bortolato, J. A. Arancibia & G. M. Escandar, *Anal. Chem.* **81**, 8074 (2009).
- [52] F. Cañada Cañada, J. A. Arancibia, G. M. Escandar, G. A. Ibañez, A. Espinosa Mansilla, A. Muñoz de la Peña & A. C. Olivieri, *J. Chromatogr. A* **1216**, 4868 (2009).
- [53] A. Mancha de Llanos, M. M. de Zan, M. J. Culzoni, A. Espinosa Mansilla, F. Cañada Cañada, A. Muñoz de la Peña & H. C. Goicoechea, *Anal. Bioanal. Chem.* **399**, 2123 (2011).
- [54] T. Galeano-Díaz, A. Guiberteau-Cabanillas, A. Espinosa-Mansilla & M. D. López-Soto, *Anal. Chim. Acta* **618**, 131 (2008).
- [55] H. Abdollahi & M. Kooshki, *Electroanalysis* **22**, 2245 (2010).
- [56] O. González García, C. Ariño, J. M. Díaz Cruz & M. Esteban, *Electroanalysis* **23**, 140 (2011).
- [57] H. C. Goicoechea, S. Yu, A. C. Olivieri & A. D. Campiglia, *Anal. Chem.* **77**, 2608 (2005).
- [58] A. C. Olivieri, J. A. Arancibia, A. Muñoz de la Peña, I. Durán Merás & A. Espinosa Mansilla, *Anal. Chem.* **76**, 5657 (2004).
- [59] A. Muñoz de la Peña, I. Durán Merás & A. Jiménez Girón, *Anal. Bioanal. Chem.* **385**, 1289 (2006).
- [60] A. Muñoz de la Peña, I. Durán Merás, A. Jiménez Girón & H.C. Goicoechea, *Talanta* **72**, 1261 (2007).
- [61] A. Jiménez Girón, I. Durán Merás, A. Espinosa Mansilla, A. Muñoz de la Peña, F. Cañada Cañada & A. C. Olivieri, *Anal. Chim. Acta* **622**, 94 (2008).
- [62] P. C. Damiani, I. Durán Merás, A. G. García Reiriz, A. Jiménez Girón, A. Muñoz de la Peña & A. C. Olivieri, *Anal. Chem.* **76**, 6949 (2007).
- [63] S. H. Zhu, H. L. Wu, A. L. Xia, J. F. Nie, Y. C. Bian, C. B. Cai & R. Q. Yu, *Talanta* **77**, 1640 (2009).
- [64] R. M. Maggio, P. C. Damiani & A. C. Olivieri, *Anal. Chim. Acta* **677**, 97 (2010).
- [65] A. G. García Reiriz, P. C. Damiani, A. C. Olivieri, F. Cañada Cañada & A. Muñoz de la Peña, *Anal. Chem.* **80**, 7248 (2008).

- [66] M. L. Nahorniak, G. A. Cooper, Y. C. Kim & K. S. Booksh, *Analyst* **130**, 85 (2005).
- [67] Y. C. Kim, J. A. Jordan, M. L. Nahorniak & K. S. Booksh, *Anal. Chem.* **77**, 7679 (2005).
- [68] K. M. Pierce & S. P. Schale, *Talanta* **83**, 1254 (2011).
- [69] R. E. Mohler, K. M. Dombek, J. C. Hoggard, E. T. Young & R. E. Synovec, *Anal. Chem.* **78**, 2700 (2006).
- [70] K. J. Johnson, B. W. Wright, K. H. Jarman & R. E. Synovec, *J. Chromatogr. A* **996**, 141 (2003).
- [71] S. E. G. Porter, D. R. Stoll, S. C. Rutan, P. W. Carr & J. D. Cohen, *Anal. Chem.* **78**, 5559 (2006).
- [72] H. P. Bailey & S. C. Rutan, *Chemom. Intell. Lab. Syst.* **106**, 131 (2011).
- [73] K. Danzer & L. A. Currie *Pure Appl. Chem.* **70**, 993 (1998).
- [74] A. C. Olivieri, N. M. Faber, J. Ferré, R. Boqué, J. H. Kalivas & H. Mark, *Pure & Appl. Chem.* **78**, 633 (2006).
- [75] A. C. Olivieri & N. M. Faber, Validation and error, en *Comprehensive Chemometrics*, S. Brown, R. Tauler & B. Walczak, Eds, Elsevier, Amsterdam, 2009, Vol. 3, pp. 91-120.
- [76] A. Lorber, *Anal. Chem.* **58**, 1167 (1986).
- [77] N. J. Messick, J. H. Kalivas & P. M. Lang, *Anal. Chem.* **68**, 572 (1996).
- [78] C. N. Ho, G. D. Christian & E. R. Davidson, *Anal. Chem.* **52**, 1071 (1980).
- [79] A. C. Olivieri & N. M. Faber, *J. Chemometrics* **19**, 583 (2005).
- [80] A. C. Olivieri, *Anal. Chem.* **77**, 4936 (2005).
- [81] A. C. Olivieri & N. M. Faber, *Anal. Chem.* (in press, DOI: <http://dx.doi.org/10.1021/ac202268k>).

Manuscrito recibido el 23 de enero de 2012.

Aceptado el 20 de marzo de 2012.