

# The application of mathematical and statistical methods in industry

**Carlos Alciaturi<sup>1,2</sup>, Marcos Escobar<sup>2,3</sup>, Iván Esteves<sup>1</sup>, Zoilabet Duque<sup>1</sup>**

<sup>1</sup>Fundación Instituto Zuliano de Investigaciones Tecnológicas (INZIT),  
Dirección de Investigación y Desarrollo, Maracaibo, Venezuela.  
alciaturi@gmail.com. Tel-Fax: (58) 261-7915371.

<sup>2</sup>División de Estudios para Graduados, Facultad de Ingeniería, Universidad del Zulia.

<sup>3</sup>CARBOZULIA, Av. 4 Esquina Calle 60, Edif. Marfi, Torre A, Maracaibo, Venezuela.

## Abstract

Chemometrics has been defined as “The discipline that uses mathematical and statistical methods for selection or design of optimal experimental conditions, and to provide relevant information from chemical data”. This is an area of intense activity, with ample applications in the chemical, energy, and process industry, and in environmental studies. The fast progress in its applications has received impulse from the advances in electronics and computers, which have made possible the acquisition, transmission and processing of data with great efficiency. This paper explains the advantages provided by coupling chemometrics with process analytical chemistry (PAT). Some examples are presented of the use of chemometric methods for material characterization, detection of deviations, process optimization, and instrumental data interpretation.

**Palabras clave:** Chemometrics; Process Analytical Technology (PAT).

# Aplicación de métodos matemáticos y estadísticos en la industria

## Resumen

La Quimiometría ha sido definida como “la disciplina que utiliza métodos matemáticos y estadísticos para seleccionar o diseñar procedimientos y experimentos óptimos, y proveer información relevante a partir de datos químicos”. Esta es un área de intensa actividad, por sus amplias aplicaciones en la industria química, de procesos, energética y en estudios ambientales. El rápido progreso en sus aplicaciones se ha visto impulsado por los avances en electrónica y computación, los cuales han hecho posible la adquisición, transmisión y procesamiento de datos con gran eficiencia. En este artículo se explica las ventajas del acoplamiento de la Quimiometría con la tecnología analítica de procesos (PAT). Además, se presentan ejemplos del uso de algunos métodos quimiométricos, para la caracterización de materiales, detección de desviaciones, optimización de procesos, e interpretación de datos instrumentales.

**Palabras clave:** Quimiometría; Tecnología Analítica de Procesos (PAT).

## Introducción

No existe una sola definición de Quimiometría. Entre las usadas más frecuentemente cabe mencionar: “La disciplina que utiliza métodos matemáticos y estadísticos para seleccionar o di-

señar procedimientos y experimentos óptimos, y proveer información relevante a partir de datos químicos” [1]; “La Quimiometría es la ciencia de relacionar las medidas hechas a un proceso o sistema químico con el estado del sistema por medio de la aplicación de métodos matemáticos o esta-

dísticos" (definición de la Sociedad Internacional de Quimiometría (ICS) [2]); "La Quimiometría es el uso de métodos matemáticos y estadísticos para manejar, interpretar, y predecir data química." [3]; "El proceso por el cual la data (por ejemplo, números en una tabla) es transformada en la información usada para la toma de decisiones." [4]; "La Quimiometría es la disciplina química que usa matemática, estadística, y lógica formal para: (a) diseñar o seleccionar procedimientos experimentales óptimos; (b) proveer la máxima posible información química a partir del análisis de la data química; y (c) obtener conocimiento acerca de los sistemas químicos." [5]. Una lista de artículos, libros y un tutorial en Internet están incluidos en la bibliografía [6-23]. Entre las áreas de mayor crecimiento se encuentran: monitoreo de reacciones; análisis instrumental; análisis clínico; caracterización de materiales; tecnología de fermentación; optimización de síntesis y procesos; ciencia forense [5, 9]; geoquímica y ciencias ambientales [10-13].

### **1. El tema común: uso de computadoras conectadas a instrumentos de medición**

Actualmente, instrumentos de medición situados en el campo, en los centros de producción o en los laboratorios, son conectados (por diversos medios, tales como fibras ópticas, e inclusive por Internet) a computadoras para el registro, transmisión e interpretación de los datos. Estos últimos deben ser analizados por métodos matemáticos y estadísticos (quimiométricos). Una parte de estos métodos incluyen el llamado análisis multivariado, en el que se consideran simultáneamente múltiples variables para cada unidad bajo observación o estudio" [7]. El término Tecnología Analítica de Procesos (conocida como PAT, Process Analytical Technology) describe el conjunto de tecnologías que incluye herramientas analíticas, físicas, químicas y matemáticas usadas para caracterizar procesos químicos y biológicos [24-27]. El énfasis está en la realización de medidas por medio de sensores durante el desarrollo de los correspondientes procesos, con el fin de caracterizar propiedades de intermedios y productos, complementando o sustituyendo la información proporcionada por los métodos analíticos tradicionales, los cuales presen-

tan con frecuencia el inconveniente (para su aplicación en procesos de planta) de la necesidad de preparación de muestra y lentitud para la obtención de respuestas. Sin embargo, la implementación de la Tecnología Analítica de Procesos presenta dificultades; en particular, la creación de una normativa adecuada y la educación del personal en las disciplinas necesarias (espectroscopia, quimiometría, control de procesos) y su integración [27].

### **2. Vinculación con otras disciplinas**

La quimiometría utiliza métodos originados en otras disciplinas, para interpretar resultados, desarrollar modelos, y proponer mejoras en los procesos químicos. Se puede mencionar Estadística Multivariada, Teoría del Control, Análisis Numérico, Investigación de Operaciones, Análisis de Series de Tiempo, y otras.

Por ejemplo, la estadística multivariada permite interpretar las señales producidas por sensores en un proceso industrial y utilizarlas para predecir el rendimiento y la calidad de los productos. El reconocimiento de patrones permite usar las señales de los instrumentos para clasificar automáticamente los productos de un proceso. El análisis numérico se utiliza para predecir las condiciones óptimas de un proceso a través de un cálculo de regresión de los datos históricos de una planta. El análisis de series de tiempo permite detectar cambios y predecir tendencias [28, 29].

### **3. Aplicaciones de la Quimiometría en Ingeniería Química, Ingeniería de Procesos y Química Analítica**

Las aplicaciones incluyen: análisis en tiempo real de la calidad de productos; optimización del rendimiento de procesos; detección de variaciones en los parámetros de procesos; predicción de las propiedades y rendimientos de productos en base a los parámetros de operación. Algunos ejemplos son: monitoreo en tiempo real de la fermentación del vino [30] y la detección de cambios sutiles en la composición de productos farmacéuticos [31], caracterización físico-química de

mezclas complejas de hidrocarburos en la industria de refinación del petróleo [32] y en petroquímica [33].

Debe señalarse que la aplicación más amplia y la de mayor importancia en términos económicos la tiene la espectroscopia de infrarrojo cercano (NIR), debido a sus múltiples aplicaciones en monitoreo industrial y ambiental [27, 30-33].

#### 4. El análisis multivariado

El Análisis Multivariado permite hacer clasificaciones y realizar cálculos cuantitativos y predicciones cuando existen múltiples variables independientes (presión, temperatura, voltaje aplicado...) y una o múltiples variables dependientes (calidad del producto, rendimiento...) [7]. El análisis multivariado se puede dividir en análisis de patrones (ej. análisis de componentes principales, PCA (utilizado para diagnóstico y clasificación [8, 9]); y regresión (utilizados para cuantificación y predicción): Regresión de Componentes Principales (PCR) [8, 9, 34]; Mínimos Cuadrados Parciales, PLS [18, 19, 35]. Al disminuir la dimensionalidad, estos métodos permiten la interpretación y representación gráfica de la información proporcionada por los métodos instrumentales de análisis químico. La tendencia es a realizar la presentación de los resultados en forma gráfica, en tiempo real.

##### 4.1. PLS aplicado a la predicción de la composición del maíz

La regresión de mínimos cuadrados parciales, en la literatura referida como PLS o PLSR, es uno de los algoritmos más usados en quimiometría [17-20, 35]. Aunque originada en el área de la economía [17,18] su mayor aplicación está actualmente en química analítica e ingeniería química [8, 9]. La suposición básica es que el sistema o proceso estudiado depende de un número pequeño de "variables latentes". Este concepto es similar al de componentes principales. Las variables latentes son estimadas como combinaciones lineales de las variables observadas, donde se maximiza la covarianza entre las variables dependientes (Y) y las independientes (X). Dadas dos matrices X y Y, PLS no solamente da un modelo de regresión sino que modela también las es-

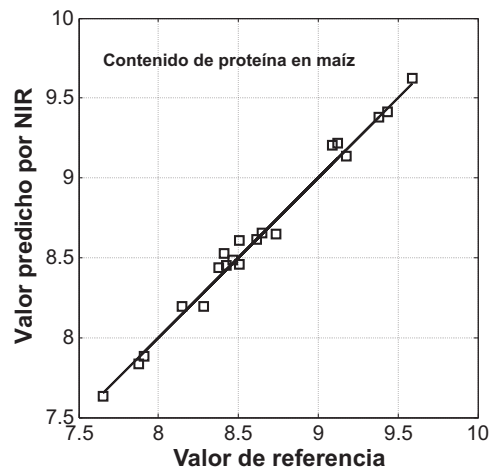


Figura 1. Valores predichos vs. Valores de referencia para el contenido de proteína en maíz.

estructuras de X y Y. El modelo es llamado PLS1 cuando se tiene una sola variable independiente, y PLS2 cuando existe más de una variable independiente [15-17]. Una discusión de los aspectos prácticos, ventajas, y dificultades de la aplicación del modelo PLS se encuentra en [16]. Se ha publicado una revisión en español [35], donde se encuentran referencias adicionales. Los datos del ejemplo que se muestra a continuación fueron tomados de la página web de Eigenvector Research, Inc. [36]. Se trata de 80 muestras de maíz, a las cuales se les midió por métodos de referencia los contenidos de humedad, aceite, proteína y almidón. Espectros de infrarrojo cercano (NIR) fueron medidos en el intervalo 1100-2498 nm, usando 700 canales. La regresión PLS fue realizada por los autores utilizando el algoritmo publicado por Geladi y Kowalski [17, 18]. Para determinar el número óptimo de variables latentes (aquel que minimiza el error de predicción) se utilizó validación cruzada [17, 18, 35]. Los resultados mostrados en la Figura 1 y la Tabla 1 muestran que las predicciones obtenidas son de precisión.

#### 5. Tendencias a futuro

Durante las próximas décadas, existe la oportunidad y el desafío de aplicar los principios y métodos de la quimiometría en todas las áreas de la química. Si bien estos métodos son útiles y necesarios, la parte más difícil del desafío podría

Tabla 1  
Resultados del PLS en la predicción de propiedades de muestras de maíz, a partir de datos espectroscópicos de NIR [36]

Propiedad, intervalo y desviación estándar (D.E) de las muestras	Número de variables latentes	Desviación estándar (D.E.) de las predicciones
Humedad (% en peso); 9,38-10,94. D.E. 0,3535	25	0,0060
Aceite (% en peso); 3,105-3,787. D. E. 0,1795	25	0,0251
Proteína (% en peso); 7,654-9,592 D.E. 0,488	30	0,0617
Almidón (% en peso); 62,826-65,841 D.E. 0,7873	23	0,0907

ser educar a los químicos e ingenieros químicos acerca de los principios y aplicaciones de la quimiometría. Por ejemplo, pueden incluirse métodos quimiométricos en cursos de análisis instrumental, u ofrecer cursos en programas de postgrado de ingeniería química, tecnología de alimentos, o ambiente. Los avances en la electrónica y la óptica, permitirán cada vez más disponer de una variedad de sensores para la obtención de data en tiempo real. Estos avances, junto con el uso de computadoras para el procesamiento de datos, resultarán en un aumento de la eficiencia de los procesos, disminución de la contaminación, y un desarrollo sustentable. Las empresas que apliquen los métodos avanzados de diseño experimental y análisis de datos, aumentarán su eficiencia y competitividad. Algunas áreas de gran impacto serán el monitoreo de procesos en las industrias de alimentos y farmacéutica, y el monitoreo ambiental. Por consiguiente, es seguro decir que el futuro de la quimiometría es promisorio.

### Referencias bibliográficas

1. Sitio web de Wikipedia: <http://es.wikipedia.org/wiki/Quimiometría>.
2. Capítulo norteamericano de la International Chemometrics Society: <http://www.namics.nysaes.cornell.edu/>.
3. Malinowski E.R.: Factor Analysis in Chemistry, Segunda Edición, 1991.
4. Beebe K.R., Pell R.J. y Seasholtz M.B.: Chemometrics: A Practical Guide: John Wiley & Sons, Inc., New York, 1998.
5. Massart D.L., Vandeginste B.G.M., Buydens L.M.C., de Jong S., Lewi P.J. y Smeyers-Verbeke J.: Data Handling in Science and Technology 20A: Handbook of Chemometrics and Qualimetrics: Parts A and B., Elsevier, Amsterdam, 1998.
6. Otto M.: Chemometrics: Statistics and Computer Applications in Analytical Chemistry, Wiley- VCH, Weinheim, 1998.
7. Dillon W.R. y Goldstein M.: Multivariate Analysis, John Wiley & Sons Inc., London, 1984.
8. Brereton R.G.: Chemometrics: data analysis for the laboratory and chemical plant. Chichester, UK: John Wiley & Sons, 2003.
9. Brereton R.G. "Introduction to multivariate calibration in analytical chemistry". Analyst, Vol. 125 (2000) 2125-2154.
10. Einax J.W., Zwanziger H.W. y Geiss, S.: Chemometrics in environmental analysis. VCH-Wiley, Germany, 1997.
11. Terrado M., Lavigne M.P., Tremblay S., Duchesne S., Villeneuve J.P., Rousseau A.N., Barceló D. y Tauler R. "Distribution and assessment of surface water contamination by application of chemometric and deterministic models". J. Hydrology, Vol. 369 (2009) 416-426.
12. Bech J., Tume P., Sokolovska M., Reverter F., Sanchez P., Longan L., Bech J., Puente A. y Oliver T., "Pedogeochemical mapping of Cr, Ni, and Cu in soils of the Barcelona Province (Catalonia, Spain): relationships with soil physico-chemical characteristics". J. Geochem. Explor. 96, (2008) 106-116.
13. Tušar M. y Norviè M. "Data exploration on standard asphalt mix analyses". J. Chemometrics, Vol. 23, (2009) 283-293.

14. Jackson J.E.: A User's Guide to Principal Components. New York: John Wiley & Sons, 1991.
15. Kowalski B.R (Ed.), Chemometrics: Theory and Application. ACS Symposium Series 52. (American Chemical Society: Washington, DC, 1977).
16. Beebe K.R. y Kowalski B.R. "An Introduction to Multivariate Calibration and Analysis." *Anal. Chem.*, 59:1007A-1017A, 1987.
17. Geladi P. y Kowalski B.R. "Partial least-squares regression: a tutorial." *Anal. Chim. Acta.* 1986; 185:1-17.
18. Geladi P. y Kowalski B. R.: "An Example of 2-Block Predictive Partial-Least Squares Regression with Simulated data". *Anal. Chim. Acta*, Vol. 185, (1986) 19-32.
19. Wold S., Sjostrom M., y Eriksson L. "PLS-regression: a basic tool for chemometrics". *Chem. Intell. Lab. Syst.* Vol. 58 (2001) 109-130.
20. Ramis G. y García M.C.: "Quimiometría" Editorial Síntesis (Madrid) 2002
21. Fernández C.M.: "Quimiometría" Universidad de Valencia 2005.
22. Miller J. y Miller, J.: "Estadística y Quimiometría para Química Analítica", Pearson Educación, 4ª Ed. 2002.
23. Sitio web de la Universitat Rovira i Virgili: [www.urv.es](http://www.urv.es)
24. Lopes J.A., Costa P.F., Alves T.P. y Menezes J. C.: "Chemometrics in bioprocess engineering. Process analytical technology (PAT) applications". *Chem. Intell. Lab. Syst.* Vol. 74, No. 3 (2004) 269-275.
25. Workman J., Koch M. y Veltkamp D.: "Process analytical chemistry". *Anal. Chem.* Vol. 77 (2005) 3789-3806.
26. Codgill R.P., Anderson, C.A. y Drennen, J.K.: "Using NIR spectroscopy as an integrated PAT tool". *Spectroscopy* Vol. 19 (2004) 104-109.
27. Davies, T. en What is PAT ? [www.spectroscopyeurope.com/TD\\_16\\_2.pdf](http://www.spectroscopyeurope.com/TD_16_2.pdf)
28. Vandeginste, B.G.M.: *Analytica Chimica Acta*, 150 (1983) 199-206.
29. Mark H. y Workman J.: "Chemometrics in Spectroscopy". Academic Press, London, 2007.
30. Cozzolino D., Parker M., Damberg R., Herderich M. y Gishen M.: "Chemometrics and Visible-Near Infrared Spectroscopic Monitoring of Red Wine Fermentation in a Pilot Scale". *Biotechnology and Bioengineering*, Vol. 95, No. 6 (2006) 1101-1107.
31. Frasson S. y Pasquini, C.: "Identification of counterfeit drugs using near infrared spectroscopy". *Analyst* Vol. 126 (2001) 2218-2224.
32. Jonathan P.: "On the statistical calibrations of analytical fingerprints for rapid characterization of complex hydrocarbon mixtures". British Region Meeting on Chemometrics. International Biometrics Society, York, Marzo 2006.
33. Macho S. y Larrechi M.S.: "Near-infrared spectroscopy and multivariate calibration for the quantitative determination of certain properties in the petrochemical industry". *TrAC Trends in Analytical Chemistry*, Vol. 21, No. 12 (2002) 799-806.
34. Alciaturi C.E., Escobar M.E. y Vallejo R.: "Prediction of coal properties by derivative DRIFT spectroscopy", *Fuel* Vol. 75, No. 4 (1996) 491-499.
35. Alciaturi C.E., Escobar M.E., De La Cruz C. y Rincón C.: "Partial least squares (PLS) regression and its application to coal analysis", *Rev. Téc. Ing. Univ. Zulia* Vol. 26, No. 3 (2003) 197-204.
36. Sitio web de Eigenvector Research, Inc.: <http://software.eigenvector.com>.

Recibido el 30 de Julio de 2009

En forma revisada el 22 de Febrero de 2010