

DESARROLLO DE UN MODELO SIMCA PARA LA CLASIFICACIÓN DE KEROSINAS MEDIANTE EL EMPLEO DE LA ESPECTROSCOPIA INFRARROJA

Ángel Dago Morales*, Alberto Cavado Osorio, Reinaldo Fernández Fernández y Ernesto Linchenat Dennes

Centro de Investigaciones del Petróleo, Washington 169 esquina a Churrucá, Municipio Cerro, Ciudad de la Habana, Cuba

Recebido em 16/10/07; aceito em 29/11/07; publicado na web em 26/8/08

DEVELOPMENT OF A SIMCA MODEL FOR CLASSIFICATION OF KEROSENE BY INFRARED SPECTROSCOPY. In the petroleum refining industry, the use of crude from several origins is frequent. This leads to a product of variable chemical composition during refining, hindering quality control. Therefore, it is important to develop classification models that help to better characterize those products. The objective of this study is to develop a SIMCA recognition pattern to classify kerosene using infrared spectroscopy data. The model permits to differentiate two kerosene groups with different chemical compositions, which was corroborated by mass spectrometry.

Keywords: SIMCA model; kerosene classification; infrared spectroscopy.

INTRODUCCIÓN

En el control de la calidad de los destilados medios del petróleo se emplean procedimientos de ensayo que han sido oficialmente establecidos como procedimientos de referencia por diferentes organizaciones internacionales; como por ejemplo; la American Society for Testing Materials (ASTM), la ISO (Internacional Standard Organization) y el IP (The Institute of Petroleum). Es conocido que estos métodos de referencia aunque son exactos emplean un tiempo de medición considerable y requieren, por lo general, de apreciable cantidad de muestra.

Los métodos quimiométricos han tenido gran aplicación en la industria de refinación del petróleo; por ejemplo, los métodos de calibración multivariada para el desarrollo de procedimientos alternativos para la predicción simultánea de propiedades físico químicas en derivados como gasolinas, kerosinas, turbocombustibles y diesel.¹⁻⁹ La exactitud de estos procedimientos alternativos es comparable con la exactitud de los métodos de referencia; presentan además las ventajas de utilizar pequeñas cantidades de muestras y de que el tiempo invertido en el análisis es pequeño. Por otra parte los métodos quimiométricos de clasificación y reconocimiento de patrones han sido ampliamente utilizados en la detección de adulteraciones de combustibles, así como en el control de calidad de los procesos productivos.¹⁰⁻¹³

En nuestro laboratorio se han desarrollado modelos de regresión multivariados para la predicción simultánea de propiedades físico químicas de kerosinas mediante el empleo de la espectroscopía infrarroja de rango medio (FT-MIR) y el método de mínimos cuadrados parciales (PLS). El establecimiento de estas metodologías analíticas alternativas no culmina con el desarrollo y posterior validación del modelo; es necesario establecer un procedimiento de control de su validez con respecto al tiempo, principalmente en la industria de refinación del petróleo, donde ocurren con frecuencia variaciones de la materia prima y de los propios procesos productivos, que originan cambios en la composición química de los productos terminados; y por ende, problemas con la predicción de los modelos.¹⁴⁻¹⁶ Para minimizar estas dificultades relacionadas con la detección de situaciones anómalas originadas por muestras atípicas, o no contempladas en el

proceso de calibración, es necesario desarrollar modelos de clasificación que permitan identificar las muestras que se encuentran fuera de especificación y definir además cuando es necesario realizar una actualización de las calibraciones, y de esta manera lograr modelos más robustos.^{16,17}

El método SIMCA de reconocimiento de patrones (*Soft Independent Modeling of Class Analogy*) desde su introducción por Svante Wold en 1976¹⁸ ha sido uno de los métodos de clasificación más utilizados en el control de calidad en las industrias alimenticia¹⁹ y farmacéutica.²⁰ En la industria del petróleo se ha utilizado para la detección de adulteraciones en gasolinas.^{10,13} Es un método supervisado de reconocimiento de patrones que se basa en el principio de analogía entre las muestras que pertenecen a una misma clase, y emplea para el cálculo de las distancias los *scores* determinados mediante análisis por componentes principales (PCA). El método SIMCA calcula un modelo PCA para cada clase o categoría presente en el sistema objeto de estudio, posteriormente integra cada una de las clases y calcula sus límites o fronteras con una probabilidad dada, comúnmente del 95%.²¹

En caso de que no esté disponible la información externa, las categorías o clases se pueden definir a partir de una evaluación de las agrupaciones mediante el análisis exploratorio de datos (reconocimiento de patrones no supervisado) que brindan métodos como: el Análisis Jerárquico de Grupos (*Hierarchical Cluster Analysis, HCA*) y el Análisis por Componentes Principales (*Principal Components Analysis, PCA*). Una vez definidas las categorías, se construye y refina un modelo para el conjunto de muestras de entrenamiento conocidas *training set*; y después éste se utiliza para predecir las clases de nuevas muestras desconocidas *test set*. Una muestra desconocida o muestra problema puede asignarse a una categoría, a más de una, o a ninguna; y se calcula la probabilidad de cada asignación.

El objetivo de nuestro trabajo es desarrollar un modelo SIMCA de reconocimiento de patrones para clasificar los cortes de producción de kerosinas.

MATERIALES Y MÉTODOS

Se utilizaron 60 muestras de kerosinas provenientes de una refinería de la Unión Cuba Petróleo (CUPET). El análisis preliminar de componentes principales evidenció que las muestras se dividen en dos grupos diferentes, las cuales se identificaron como C y V.

*e-mail: adago@ceinpet.cupet.cu

Los espectros de masas de muestras representativas de cada grupo se obtuvieron en un cromatógrafo de gases con detector másico HRGC-MS, modelo 4000B-Q12 de la firma Konik-Tech, con programación del horno desde 50 hasta 320 °C, rampa de calentamiento de 3 °C/min y columna DB-5 de 30 m. El equipo emplea fuente de impacto electrónico a 70 eV, con rango de masas de 29 a 600 Dalton. En el registro y procesamiento de los datos se utilizó el programa Konikrom Plus versión 2.3.²²

Los espectros infrarrojos de las 60 muestras de kerosinas objeto de estudio se obtuvieron en un espectrómetro FT-MIR Avatar 360Esp, de la firma Nicolet Instrument Corp. Estos se midieron en absorbancia mediante la técnica de transmisión en el rango entre 4000-400 cm⁻¹ con una resolución de 4 cm⁻¹ y 32 barridos por punto. Se utilizó celda fija con ventanas de cloruro de sodio y un paso de luz de 0.03 mm. En el registro y procesamiento de los espectros se usó el sistema de programas Omnic versión 5.2 a.²³

Para el desarrollo del modelo de clasificación se empleó el método SIMCA y los resultados de la espectroscopía infrarroja. El cálculo del modelo se realizó en base a las dos clases definidas en el análisis PCA. Se utilizaron 40 muestras de kerosinas para conformar el conjunto de calibración, 26 de ellas de la clase C y 14 de la clase V. Para validar el modelo se utilizaron 30 muestras: 13 de la clase C, 7 de la V y se agregaron 10 muestras de turbo-combustibles para verificar la selectividad del modelo. Los cálculos del modelo se llevaron a cabo en el rango espectral entre 1750-690 cm⁻¹ para un total de 553 variables por muestra. Las matrices X de los conjuntos de calibración y validación tienen dimensiones 40x553 y 30x553, respectivamente. Para minimizar el efecto del corrimiento de la línea base los espectros fueron corregidos mediante el procedimiento de corrección multiplicativa de la señal (*Multiplicative Scatter Correction, MSC*). Se utilizó como preprocesamiento el centrado en la media (mean centered). El número óptimo de componentes principales se seleccionó teniendo en cuenta el por ciento de varianza que cada componente principal logra explicar.

Para la evaluación de los modelos se utilizaron los diagramas Coomans y los estadígrafos Q residual y T² de Hotelling. Se fijó en un 5% el nivel de significación de los límites de las fronteras que definen cada clase. En el procesamiento matemático y estadístico se emplearon los programas PLS_Toolbox 3.5 soportado en Matlab^{24,25} y Pirouette versión 3.11.²⁶

RESULTADOS Y DISCUSIÓN

En la Figura 1 se presentan los espectros infrarrojos de las 60 muestras de kerosinas (conjunto global) después de aplicar la corrección MSC; se observa que los espectros presentan una adecuada dispersión de la línea base.

A partir de estos datos espectrales (matriz X de dimensiones 60x553) y la técnica quimiométrica de reconocimiento de análisis por componentes principales (PCA) se definió que en el proceso industrial se originan kerosinas de dos grupos diferentes. En la Figura 2 se reportan los scores sobre los dos primeros componentes principales, en la cual se observa que a lo largo del primer componente principal las muestras de kerosinas se separan en dos grupos o clases. En uno de estos grupos -identificado como C- las muestras se concentran en la parte negativa del primer componente principal; y en el otro, -identificado como V- las muestras se concentran en la parte positiva del primer componente principal.

Para confirmar la causa de la separación en dos clases diferentes se seleccionó una muestra representativa de cada una de ellas, y se analizaron por cromatografía gaseosa acoplada a masas. Los resultados de la cromatografía gaseosa demuestran que la separación en dos clases está dada por apreciables diferencias en la composición

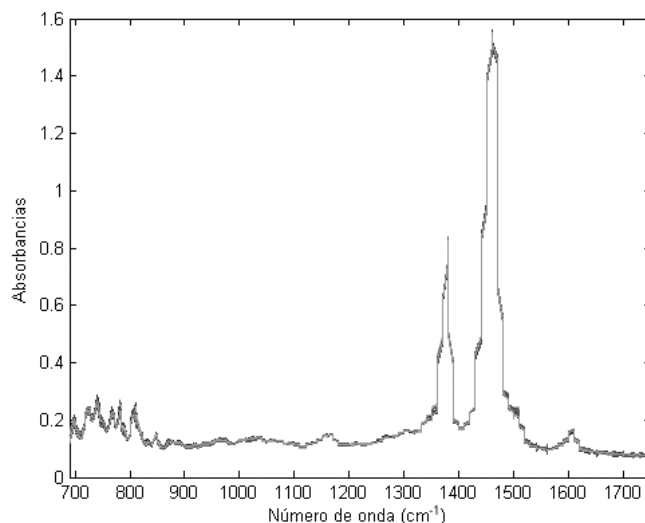


Figura 1. Espectros infrarrojos de las muestras de kerosinas después de aplicar la corrección MSC

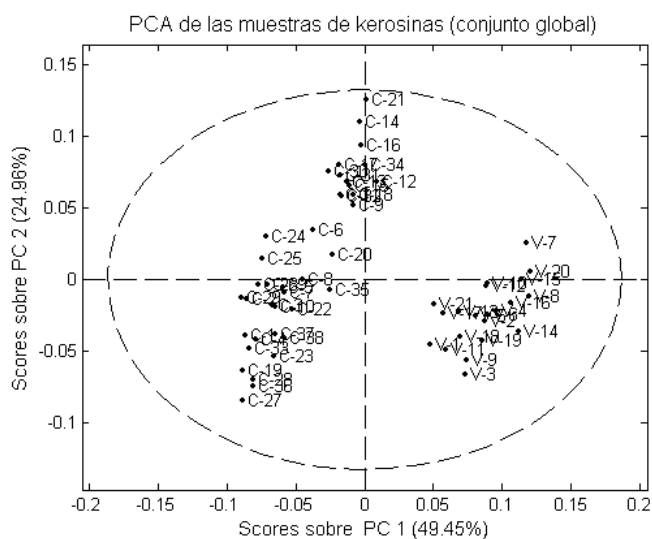


Figura 2. Modelo PCA del conjunto global de muestras de kerosinas

química. En la Figura 1S (Material suplementar) se reportan los cromatogramas de iones totales (TIC) de las muestras representativas de ambas clases; la kerosina clase V (en rojo) presenta componentes más pesados -con mayor número de átomos de carbono- que la de la clase C, y está menos enriquecida en compuestos ligeros.

De igual forma, los fragmentogramas obtenidos en el espectrómetro de masas corroboraron que la muestra clasificada como clase V presenta los iones de mayor peso molecular. El fragmentograma del ión 57 tipifica la riqueza en hidrocarburos saturados alifáticos de la muestra de la clase V en comparación con la muestra de la clase C.

Desarrollo del modelo de clasificación SIMCA

El modelo SIMCA se calculó en base a las dos clases definidas con anterioridad por PCA y CG-EM. De las 60 muestras objeto de estudio 39 forman parte del grupo C y 21 del V. El conjunto de calibración se conformó con 40 muestras (26 de la clase C y 14 de la clase V). Un resumen de los resultados del modelo se presenta en la Tabla 1. En la Figura 3 se reporta el diagrama de Coomans, en el cual se presentan las distancias ortogonales de todas las muestras del grupo de calibración con respecto a las dos clases. En el diagrama se

destaca que se logra la correcta clasificación de todas las muestras y que no se presentan falsos negativos. Se obtuvo una distancia entre las clases de 3.80, la selectividad del modelo se considera adecuada cuando el valor de este parámetro es mayor de 3.^{21,26}

Tabla 1. Resumen de los resultados del cálculo del modelo SIMCA

Clases	Nº de muestras	Nº de factores	% de varianza explicada	Clasificados correctamente	Clasificados incorrectamente	No clasificados
C	26	5	93.4	26	-	-
V	14	5	88.4	14	-	-

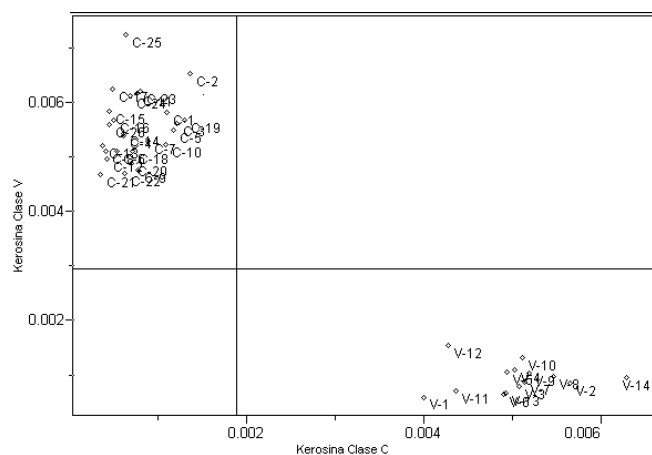


Figura 3. Modelo SIMCA: Distancias de las muestras del conjunto de calibración con respecto a las clases C y V (con líneas continuas se destacan los límites de las clases para un nivel de significación del 5%)

Validación del modelo SIMCA

El grupo de validación se conformó con 13 muestras de la clase C, 7 de la clase V y 10 muestras de turbocombustibles (T1-T10). Las muestras de turbocombustibles son ajenas al modelo desarrollado, y fueron incluidas en el grupo de validación para verificar la selectividad del mismo; es decir, para comprobar si el modelo es capaz de reconocerlas como que no pertenecen a ninguna de las clases.

En la Figura 4 se presenta el diagrama de Coomans, en el cual se reflejan las distancias de las muestras del conjunto de validación con respecto a las clases definidas por el modelo. Se observa como las muestras de validación de las clases C y V clasifican correctamente en sus respectivas clases y las muestras de turbocombustibles caen fuera de los límites que definen las fronteras de las clases. En la Figura 5 se reportan los estadígrafos Q residual y T² de Hotelling calculados para la clase V: se observa como las muestras de validación de la clase V caen dentro de los límites de confianza del 5% y las muestras pertenecientes a la clase C y las muestras de turbocombustibles caen fuera de los límites establecidos.

Los resultados de la validación fueron satisfactorios; el modelo desarrollado clasificó correctamente a las muestras del grupo de validación: todas las muestras fueron clasificadas en su clase respectiva, no se presentaron casos de falsos negativos, las 10 muestras de turbocombustibles fueron correctamente clasificadas como no pertenecientes a ninguna clase.

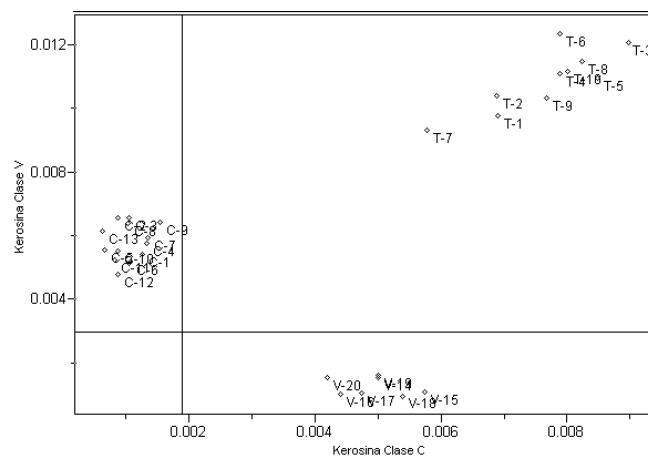


Figura 4. Validación del modelo SIMCA: Distancias de las muestras del grupo de validación con respecto a las clases C y V (con líneas continuas se destacan los límites de las clases para un nivel de significación del 5%)

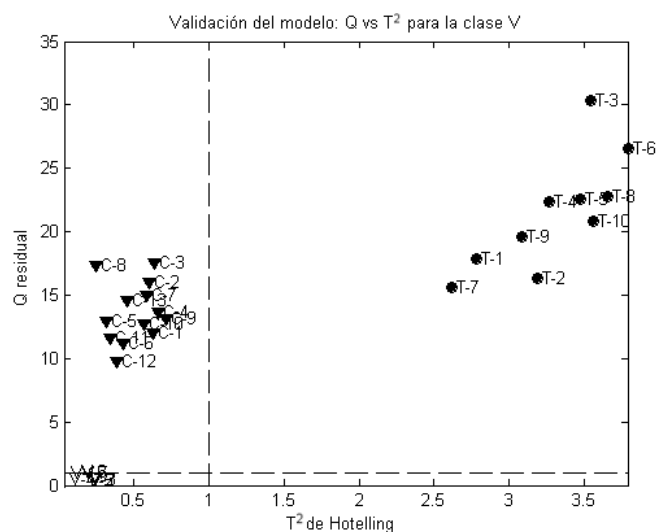


Figura 5. Validación del modelo SIMCA. Descriptores T² y residuos Q de la clase V (con líneas discontinuas se destacan los límites de la clase V para un nivel de significación del 5%)

CONCLUSIONES

Se demostró, mediante la aplicación de técnicas quimiométricas y cromatográficas, que en el proceso de producción se originan dos tipos de cortes de kerosinas, cuya diferencia fundamental está dada en la proporción de los componentes más pesados. Se desarrolló un modelo SIMCA de reconocimiento de patrones supervisado que permite la clasificación de las dos clases de kerosinas a partir de los datos de sus espectros infrarrojos. El modelo desarrollado posee una sensibilidad y selectividad adecuadas para el control de calidad del proceso productivo.

MATERIAL SUPLEMENTAR

A Figura 1S está disponibilizada gratuitamente en la página web <http://quimicanova.sbq.org.br> en la forma de un archivo .PDF.

AGRADECIMIENTOS

Los autores agradecen a la Dirección Técnica de CUPET el financiamiento de este trabajo.

REFERENCIAS

1. Fodor, G. E.; Mason, R. A.; Hutzler, S. A.; *Appl. Spectrosc.* **1999**, *53*, 1292.
2. Wentzell, P. D.; Andrews, D. T.; Walsh, J. M.; Cooley, J. M.; *Can. J. Chem.* **1999**, *77*, 391.
3. Andrade, J. M.; Garrigues, S.; de la Guardia, M.; Gómez-Carracedo, M.; Prada, D.; *Anal. Chim. Acta* **2003**, *482*, 115.
4. Vianney Santos, O. Jr.; Flavia Oliveira, C. C.; Daniella Lima, G.; Andrea Petry, C.; Garcia, E.; Paulo Suarez, A. Z.; Joel Rubim, C.; *Anal. Chim. Acta* **2005**, *547*, 188.
5. Côcco, L. C.; Yamamoto, C. I.; von Meien, O. F.; *Chem. Intel. Lab. Syst.* **2005**, *76*, 55.
6. Pasadakis, N.; Sourligas, S.; Foteinopoulos, Ch.; *Fuel* **2006**, *85*, 1131.
7. Brudzewski, K.; Kesik, A.; Kolodziejczyk, K.; Zborowska, U.; Ulaczyk, J.; *Fuel* **2006**, *85*, 553.
8. Dago-Morales, A.; Fernández, F. R.; Ruiz-Martínez, M. D.; Balmayor, M. M.; Laza, N. M.; Ross, J. E.; Otero, Z. M.; Simeón, A. L.; *Rev. CENIC, Ciencias Químicas* **2006**, *37*, 3.
9. Balabin, R. M.; Safieva, R. Z.; Lomakina, E. I.; *Chem. Intel. Lab. Syst.* **2007**, *88*, 183.
10. Santos de Oliveira, F.; Gomes Teixeira, L. S.; Ugulino Araujo, M. C.; Korn, M.; *Fuel* **2004**, *83*, 917.
11. Wiedemann, L. S. M.; d' Avila, L. A.; Azevedo, D. A.; *Fuel* **2005**, *84*, 467.
12. Pasadakis, N.; Kardamakis, A.; *Anal. Chim. Acta* **2006**, *578*, 250.
13. Flumignan, D. L.; Tininis, A. G.; Ferreira, F. O.; de Oliveira, J. E.; *Anal. Chim. Acta* **2007**, *595*, 128.
14. Andrade, J. M.; García, M. V.; López-Mahia, P.; Prada, D.; *Talanta* **1997**, *44*, 2167.
15. Swierenga, H.; de Weijer, A. P.; Van Wijk, R. J.; Buydens, L. M.; *Chemom. Intell. Lab. Syst.* **1999**, *49*, 1.
16. García-Mencía, M. V.; Andrade, J. M.; López-Mahía, P.; Prada, D.; *Fuel* **2000**, *79*, 1823.
17. Daszyłowski, M.; Kaczmarek, K.; Vander Heyden, Y.; Walczak, B.; *Chem. Intell. Lab. Syst.* **2007**, *85*, 203.
18. Wold, S.; *Pattern Recognition* **1976**, *8*, 127.
19. Forina, M.; Lanteri, S. In *Chemometrics. Mathematics and Statistics in Chemistry*; Kowalski, B. R., ed.; D. Reidel Publishing Company: Reidel, Dordrecht, 1984, p. 305.
20. Gabrielsson, J.; Lindberg, N. O.; Lundstedt, T.; *J. Chem.* **2002**, *16*, 141.
21. Ferreira, M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O.; *Quim. Nova* **1999**, *22*, 724.
22. *Konikrom Plus Software, version 2.3*, Konik-Tech Corporation, 2004.
23. *Nicolet's OMNIC Spectroscopy Software, version 5.2a*, Nicolet Instrument Corporation, Copyright (1992-2000).
24. *PLS_Toolbox 3.5 for use with MATLAB*, Eigenvector Research, Inc., 2004, version 3.5.
25. *MATLAB. The Language of Technical Computing*, The Math Works, 2003, version 6.5.
26. *Pirouette*, Infometrix Inc., Woodinville WA., 2003, version 3.11.