

Aplicación de la espectroscopia infrarroja y técnicas de reconocimiento de patrones en la caracterización de aditivos de la industria del petróleo

Application of infrared spectroscopy and pattern recognition techniques in the characterization of additives of petroleum industry

Karelia Aldana-Jiménez^a, Yumirka Comesaña-García^a, Frank Alexis Ochoa-Valladares^a, Roberto Oropesa-Rodríguez^a, Ángel Dago-Morales^{a,*}

^aCentro de Investigación del Petróleo, Cuba.

*adago@ceinpet.cupet.cu

Recibido: 31 de enero del 2018; **Aceptado:** 20 de junio del 2018

RESUMEN

Los aditivos químicos son ampliamente utilizados en la industria del petróleo; éstos varían de acuerdo al proceso tecnológico en los que están involucrados según su funcionalidad. En su caracterización se determinan un conjunto de propiedades físico-químicas para certificar que cumplen con las especificaciones de calidad establecidas. Los métodos de ensayo que se emplean son laboriosos, consumen tiempo y volúmenes de muestra. En el presente trabajo se demuestra la factibilidad de emplear los datos que se obtienen de la espectroscopia infrarroja de rango medio conjuntamente con los métodos de reconociendo de patrones para caracterizar dos aditivos: uno antiestático utilizado para la transportación y carga de combustibles y el otro un inhibidor de corrosión. Se utilizó el método de análisis por componentes principales para la exploración de los datos, con vistas a visualizar la presencia de agrupamientos de similar comportamiento en las muestras de estudio. Para el proceso de clasificación se utilizó el modelo blando independiente por analogías de clase, los diferentes modelos de pruebas se construyeron a partir de un conjunto de entrenamiento de muestras representativas. Para comprobar el desempeño de los mismos y minimizar los problemas relacionados con ajustes, se validaron a partir de conjuntos de muestras independientes. La sensibilidad y selectividad obtenidas en las etapas de entrenamiento y validación fueron iguales a 1; los modelos no presentaron errores de clasificación por lo que pueden ser utilizados para el control de calidad de los aditivos estudiados.

Palabras clave: reconocimiento de patrones; clasificación; SIMCA; aditivos

ABSTRACT

Chemical additives are widely used in the oil industry; these vary according to their functionality as well as according to the technological process in which they are involved. A physical-chemical characterization has to be carried out, in order to certify the required quality specifications. The test methods used are laborious, consuming time and an important volume of samples are needed. The aim of this work is to demonstrate the feasibility of combining medium range infrared spectroscopy data together with pattern recognizing methods to characterize two additives, an antistatic used during the loading and transportation of fuels and a corrosion inhibitor. The principal component analysis method was used to explore the data, with a view to visualizing the presence of groupings of similar behavior in the study samples. For the classification process the independent soft model was used by class analogies, the different test models were

constructed from a training set of representative samples. Selectivity and sensitivity obtained in the training and validation steps were equal to 1, the models did not present classification errors that its suggested our model could be employed for the quality control of the additives reported in this study.

Keywords: pattern recognition; classification; SIMCA; additives

INTRODUCCIÓN

La industria del petróleo utiliza en sus diferentes procesos tecnológicos una amplia gama de aditivos; se emplean en la perforación, extracción, recuperación mejorada del petróleo, refinación; así como en el transporte y almacenamiento de los diferentes productos obtenidos en cada una de las diferentes etapas productivas. En particular, para mejorar las propiedades de los combustibles obtenidos en la refinación se emplean dispersantes, detergentes, inhibidores de corrosión, antiestáticos, entre otros.

Estos aditivos, por lo general, son mezclas complejas y requieren de un control de calidad adecuado para verificar que cumplen con las propiedades especificadas por los fabricantes, y detectar además los posibles problemas ocasionados por su vencimiento o adulteración. En su caracterización se utilizan métodos estandarizados, a partir de los cuales, se puede verificar si las propiedades físico químicas obtenidas coinciden con las reportadas para el producto. Estos métodos son laboriosos, consumen tiempo y volúmenes de muestra. La espectroscopía infrarroja con transformada de *Fourier* (FT-MIR, por sus siglas en inglés) se ha estado utilizando, en apoyo a los ensayos físico químico, como un método para el cribaje de los aditivos que no cumplen con las especificaciones del fabricante (Maldonado *et al.* 2018; Ferreira *et al.* 2014; Pacheco, 2014; ASTM D5845, 2001). Sin embargo, en algunos aditivos, por su compleja formulación, la simple inspección visual de sus espectros no aporta la información deseada. En estos casos resulta conveniente el empleo combinado de los métodos quimiométricos de reconocimiento de patrones con los datos que brinda la técnica FT-MIR.

La literatura especializada define el reconocimiento de patrones como la zona del conocimiento que se ocupa del desarrollo de teorías, métodos, técnicas y dispositivos computacionales para la realización de procesos ingenieriles, computacionales y matemáticos, relacionados con objetos físicos o abstractos, que tienen el propósito de extraer información que permita establecer propiedades y vínculos entre conjuntos de los objetos estudiados sobre la base de los cuales se realiza una tarea de identificación o clasificación. Como se destaca en esta definición general el reconocimiento de patrones abarca dos grandes aplicaciones: análisis exploratorio de datos y clasificación (Shulcloppe, 2013; Brereton, 2015, 2007).

Durante la última década, el reconocimiento de patrones ha sido una de las especialidades de la quimiometría que se ha desarrollado y aplicado más profusamente en el ámbito de la química industrial (Cai *et al.* 2015; Klessler, 2013; Pomerantsev *et al.* 2012; Chian y Colegrove, 2007), química analítica (Kumar *et al.* 2014; Olivieri y Rivas, 2011; Miller, 2010), química forense (Materazzi *et al.* 2017; Risoluti *et al.* 2016; Renterghem *et al.* 2013; Talavera *et al.* 2012, 2011) y química ambiental (Hopke, 2015; Gredilla *et al.* 2013; Mas *et al.* 2010). Se han utilizado ampliamente en el control de calidad de materias primas y productos terminados (Pérez *et al.* 2017; Carvalho *et al.* 2014; Bassabasi y Oussama, 2013; Khanmohammadi *et al.* 2013; Flumignan *et al.* 2010;

Sheng y Lian, 2012), y en la detección eficaz de productos adulterados (Nespeca *et al.* 2018; Vempatapu y Kamaujia, 2017; Shand *et al.* 2017; Mendes y Barbiera, 2013; Coelho *et al.* 2011).

Una de las finalidades que se persigue cuando se analizan muestras caracterizadas por numerosos análisis (resultados multidimensionales) es la de encontrar semejanzas entre ellas, que permitan establecer agrupaciones en base a los análisis efectuados. Las agrupaciones encontradas pueden servir posteriormente para clasificar objetos desconocidos de una forma rápida y objetiva. La industria alimentaria y la industria farmacéutica han sido pilares en el uso de estas herramientas quimiométricas (Olivieri, 2017; Gabrielsson *et al.* 2002). En estas investigaciones se han empleado, conjuntamente las técnicas multivariadas de reconocimiento de patrones con diferentes técnicas analíticas: cromatografía gaseosa, cromatografía líquida de alta resolución, espectroscopía infrarroja y análisis de trazas de elementos metálicos (Brereton, 2007).

El análisis por componentes principales (PCA, por sus siglas en inglés) y el análisis jerárquico de clúster (HCA, por sus siglas en inglés) se utilizan como métodos de análisis exploratorio de datos, los cuales permiten definir las agrupaciones presentes en los sistemas estudiados. El modelado suave independiente de analogías de clase (SIMCA, por sus siglas en inglés) es un método de clasificación que tiene la ventaja que se puede utilizar como un clasificador de una sola clase, o de varias clases; otros clasificadores, tales como: análisis discriminante lineal, análisis discriminante por mínimos cuadrados parciales y máquinas de vectores soporte requieren que los sistemas estudiados estén definidos al menos por dos clases (Brereton, 2009). Para detectar anomalías o adulteraciones de un producto el clasificador idóneo es el de una sola clase, que es la clase definida a través de un conjunto de entrenamiento con las muestras que son definidas como muestras de referencia.

El objetivo del presente trabajo es ejemplificar la utilidad del empleo de los métodos de reconocimiento de patrones PCA y SIMCA en conjunción con la técnica espectroscópica FT-MIR, en la caracterización de dos aditivos empleados en la industria del petróleo: uno utilizado como antiestático para la transportación y carga de combustibles y el otro como inhibidor de corrosión, que denominaremos Anti_est; y Inh_corr, respectivamente.

MATERIALES Y MÉTODOS

Aditivos objeto de estudio y condiciones del registro de los espectros infrarrojos

Se estudiaron dos tipos de aditivos: uno antiestático empleado para el manejo y transportación de combustibles y el otro un inhibidor de corrosión, que denominamos Anti_est y Inh_corr, respectivamente.

El registro de los espectros infrarrojos se realizó en un equipo FT-MIR modelo Frontier de la firma Perkin Elmer. Se utilizó la técnica de reflectancia total atenuada horizontal con una porta muestra de selenuro de zinc. El registro se realizó en por ciento de transmitancia, en el intervalo de número de onda entre 4000 a 650 cm^{-1} , con 16 barridos y una resolución de 4 cm^{-1} . En la adquisición y procesamiento de los espectros se empleó el programa Spectrum (PerkinElmer Spectrum, 2015).

Análisis quimiométrico: reconocimiento de patrones

En el análisis exploratorio de datos se utilizó como método de reconocimiento de patrones no supervisado el PCA. El número óptimo de componentes principales (CP) se seleccionó teniendo en cuenta el porcentaje de varianza que cada CP acumuló. Se utilizaron los estadígrafos Q residual y T^2 de Hotelling para definir la presencia de muestras o variables discrepantes. Estos estadígrafos permiten detectar si la variación incluida en los componentes principales considerados es más grande que la que correspondería si solo influyeran variaciones aleatorias.

Para el cálculo de los modelos de clasificación se empleó el método supervisado de reconocimiento de patrones SIMCA, el cual tiene la ventaja de ser un clasificador de una o varias clases.⁴¹ En el presente trabajo se utilizó como un clasificador de una sola clase, el cual permite detectar las muestras discrepantes a la clase óptima definida para el sistema. El método se basa en el principio de analogía entre las muestras que pertenecen a una misma clase. Este método calcula un modelo PCA para cada clase o categoría del sistema; posteriormente, integra cada una de las clases y calcula sus límites o fronteras con una probabilidad dada. En el cálculo de las distancias emplea las puntuaciones de las muestras obtenidas en el análisis PCA. El modelo se calcula y refina mediante un conjunto de muestras de entrenamiento (conjunto de entrenamiento); posteriormente, para comprobar su desempeño y minimizar los problemas relacionados con un exceso de ajuste se valida con muestras externas (conjunto de validación). Al final, el modelo desarrollado se utiliza para predecir las clases de muestras desconocidas; la asignación de clases se realiza cuando la distancia residual es menor que el límite estadístico definido para cada una de las clases; para una mayor información sobre este método el método SIMCA recomendamos consultar las siguientes referencias (Brereton, 2015, 2007, 2009; Esbensen, 2001).

Para la evaluación de los resultados del modelo SIMCA se utilizaron los diagramas de Coomans, en los cuales se reportan las distancias de las muestras a la clase definida por el modelo; se fijó en 5 % el nivel de significación de los límites y fronteras que definen la clase. El número de muestras clasificadas correctamente por el modelo en las etapas de entrenamiento y validación es una medida de la calidad del mismo (Westad y Marini, 2015; Brereton, 2006). Para una categoría determinada, la sensibilidad es la proporción de casos que son correctamente predichos como miembros de esta clase (verdaderos positivos) dividido entre la suma de la cantidad de verdaderos positivos y de falsos negativos. La selectividad se define como la proporción de casos de otras clases que son correctamente predichos (verdaderos negativos) dividido entre la suma de verdaderos negativos y falsos positivos. En el procesamiento matemático y estadístico se emplearon los programas Pirouette versión 3.11 (Pirouette, 2003) y el PLS-Toolbox versión 6.7.1 (PLS-Toolbox, 2012).

- Conjunto de adiestramiento y validación del aditivo antiestático: Anti_est

El conjunto de entrenamiento se conformó con los espectros de 15 muestras del aditivo antiestático que cumplían con las especificaciones técnicas. La notación utilizada para las muestras de entrenamiento fue E1, E2, ... E15. Los modelos se calcularon en el intervalo de longitudes de onda 2000 - 650 cm^{-1} . La dimensión de la matriz X de entrenamiento fue de 15 x 1351 (15 espectros y 1351 variables experimentales). Para minimizar el corrimiento de la línea base se aplicó la transformación primera derivada (se utilizó un

polinomio de segundo orden y un filtro de 15 puntos); como preprocesamiento de los datos se utilizó el centrado en la media. Para validar el modelo SIMCA se empleó un conjunto de validación que incluyó los espectros de 15 muestras del aditivo Anti_est, cuya notación fue V1, V2,...V15; y los espectros de 3 muestras de aditivo que se adulteraron “*exprofeso*” con un 25 y 50% de diesel, que es uno de los componentes del producto; la notación utilizada fue: A1_25, A1_50, A2_25, A2_50 y A3_25, A3_50; lo cual nos permite verificar la capacidad del modelo para detectar muestras discrepantes. La dimensión de la matriz de validación fue 21 x 1351.

- Conjunto de adiestramiento y validación del aditivo inhibidor de corrosión: Inh_corr

El cálculo de los modelos se realizó en el intervalo de longitudes de onda 1889 - 778 cm^{-1} . El conjunto de entrenamiento se conformó con los espectros de 20 muestras del aditivo Inh_corr que cumplieran con las especificaciones técnicas, las cuales se tomaron como referencia: la notación utilizada fue: E1, E2,...E20. La dimensión de la matriz del conjunto de entrenamiento fue de 20 x 1112 (20 espectros y 1112 variables espectrales). Para minimizar el efecto del corrimiento de la línea base se utilizó la corrección multiplicativa de la señal (MSC, por sus siglas en inglés). Como técnica de preprocesamiento se empleó el centrado en la media. Para validar el modelo SIMCA se confeccionó una matriz de validación conformada por los espectros de siete aditivos de referencia: la notación utilizada fue: V1, V2,...V7; a la cual se le agregó los espectros de una muestra considerada como atípica, la cual se registró por triplicado: en el mismo equipo, en días y operadores diferentes: se le dio la notación P1, P2 y P3.

RESULTADOS Y DISCUSIÓN

Aditivo antiestático: Anti_est

El modelo PCA calculado con dos componentes principales explicó 89 por ciento de la variabilidad del sistema. Las puntuaciones de las muestras del conjunto de entrenamiento reflejan que el sistema está formado por una sola clase que en este caso es el conjunto de muestras Anti_est que cumplen las especificaciones y están en el marco del control de calidad (Figura 1). Los resultados de los estadígrafos Q residual y T^2 de Hotelling no arrojaron la existencia de muestras discrepantes (Figura 2).

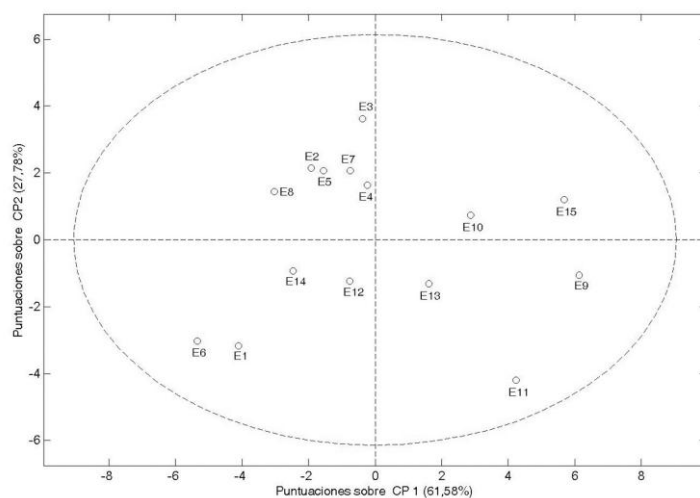


Figura 1. Modelo PCA: gráfico con las puntuaciones en el plano PC 1 vs PC 2 del conjunto de entrenamiento.

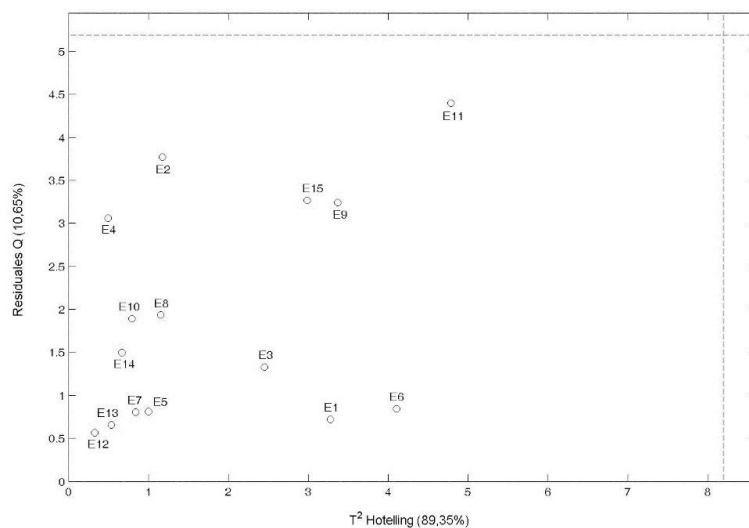


Figura 2. Modelo PCA: estadígrafos Q residual y T² de Hotelling (con líneas discontinuas se destacan los límites de cada uno de los estadígrafos para un nivel de significación del 5 %).

El modelo SIMCA se calculó con dos componentes principales para la clase definida anteriormente mediante el análisis PCA. Todas las muestras del conjunto de entrenamiento clasificaron correctamente (15 verdaderos positivos); no hubo muestras sin clasificar (Figura 3).

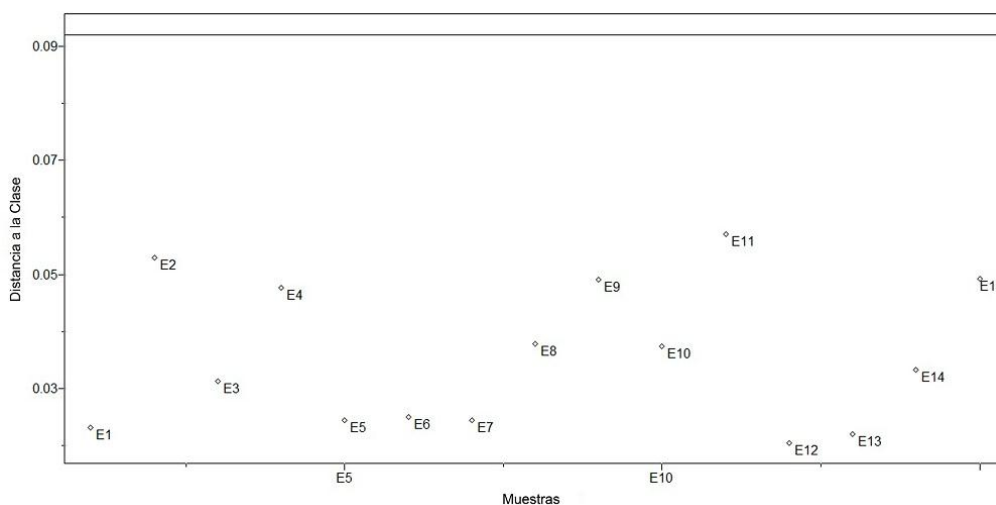


Figura 3. Modelo SIMCA (etapa de entrenamiento): diagrama de Coomans: distancias de las muestras del conjunto de entrenamiento con respecto a la clase Anti_est (con líneas discontinuas se destacan los límites de las clases para un nivel de significación del 5 %).

El resultado de la validación del modelo SIMCA predijo correctamente las 15 muestras del aditivo que cumplían las especificaciones técnicas (15 VP), y no clasificó las 6 muestras adulteradas (6 VN) (Figura 4). El modelo no presentó errores de clasificación en las etapas de entrenamiento y validación: la sensibilidad y selectividad de ambas etapas fue de 1. El mayor poder discriminatorio de las variables estuvo comprendido entre las frecuencias 1600 - 1450 cm⁻¹ que se corresponden con las vibraciones de tensión C=C aromáticos y de doblaje C-H (en el plano) y entre 900 - 650 cm⁻¹ con las vibraciones de doblaje C-H (fuera del plano).

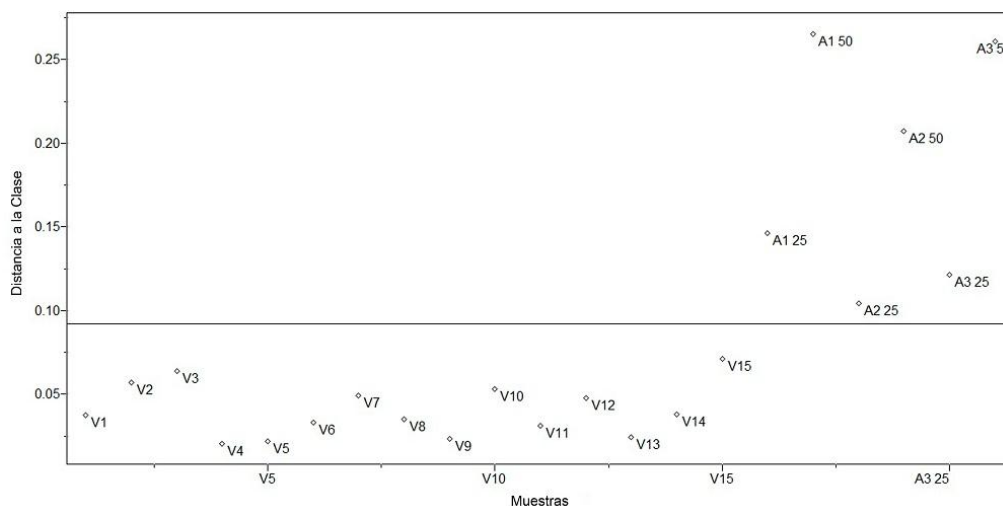


Figura 4. Modelo SIMCA (etapa de validación): diagrama de Coomans: distancias de las muestras del conjunto de validación con respecto a la clase Anti_est (con línea discontinua se destaca el límite de la clase para un nivel de significación del 5 %).

Sin embargo, resulta difícil discriminar las muestras adulteradas mediante simple inspección visual de sus espectros infrarrojos (Fig.5), cosa muy común en el caso de los aditivos químicos, que por lo regular son mezclas extremadamente complejas; y por ende, la similitud de sus espectros no implica una identidad de la composición química de la muestra.

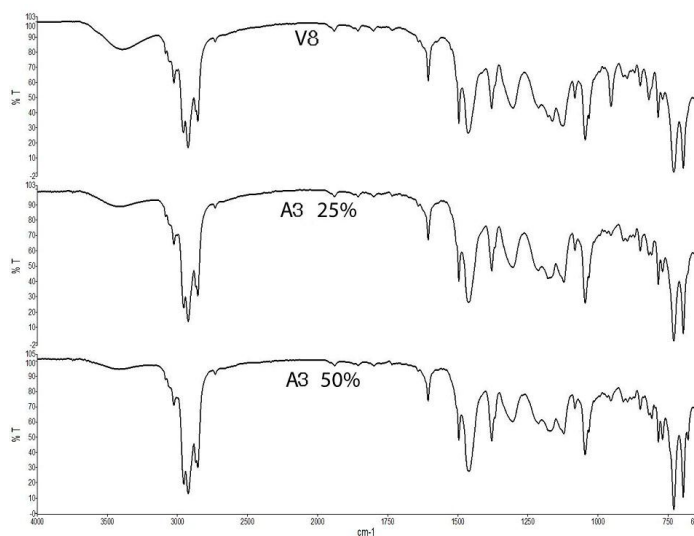


Figura 5. Espectros FT-MIR de la muestra de validación V8 y de la muestra adulterada con diesel A3 al 25 y 50 %.

Aditivo inhibidor de corrosión Inh_corr

El modelo PCA se calculó con 2 componentes principales que explicaron el 98 % de la variabilidad del sistema. Se debe destacar que el aporte del CP2 es pequeño (0,58 %), la mayor variabilidad se recoge en el CP1 con un 98,5 %. Otro aspecto de interés con respecto al modelo calculado está relacionado con las puntuaciones de las muestras a lo largo de CP1; las muestras de referencia denominadas E4, E5, E6, E10, E11 y E12 se ubican en la parte positiva de CP1, y el resto las muestras en la parte negativa (Figura 6). El modelo fue capaz de detectar pequeñas diferencias en sus espectros producto de las

variaciones que pueden estar presentes en este tipo de aditivo, cuya composición puede variar sin dejar de cumplir con las especificaciones técnicas y de calidad. Los estadígrafos Q residual y T^2 de Hotelling (Figura 7) indican que las puntuaciones de todos los espectros se ajustan adecuadamente al modelo.

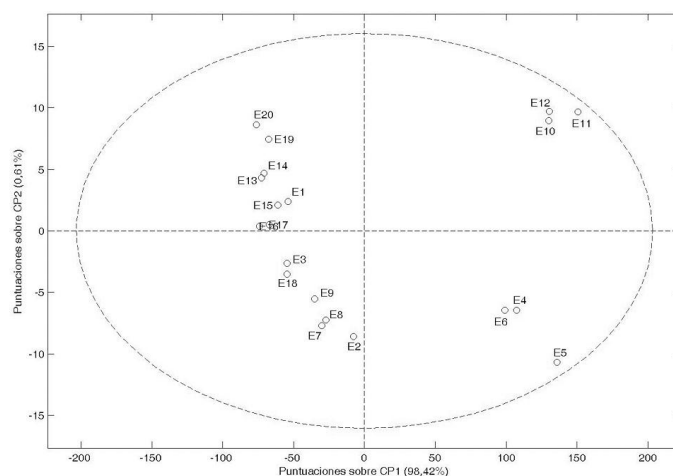


Figura 6. Resultados del PCA: puntuaciones de las muestras sobre los dos componentes principales (con líneas discontinuas se destacan los límites de la elipse para un nivel de significación del 5 %).

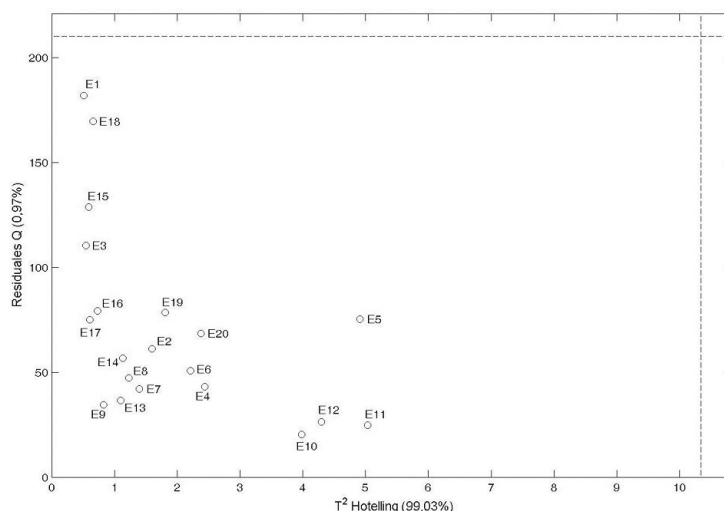


Figura 7. Resultados del modelo PCA: estadígrafos Q residual y T^2 de Hotelling (con líneas discontinuas se destacan los límites de cada uno de los estadígrafos para un nivel de significación del 5 %).

Teniendo en cuenta los resultados del PCA el modelo SIMCA se calculó para una sola clase. Los diagramas de Coomans de los conjuntos de entrenamiento y validación reflejan una correcta clasificación del sistema estudiado (Fig. 8 y 9). En la etapa de validación el modelo SIMCA predijo correctamente las 7 muestras del aditivo que cumplían con las especificaciones técnicas y discriminó los espectros de la muestra discripante. El estadígrafo Q residual de la muestra problema y de una muestra de referencia (Fig.10) refleja que el mayor poder discriminatorio de las variables estuvo comprendido entre $1350-800\text{ cm}^{-1}$; que se corresponden con vibraciones de tensión C-O-C y de doblaje C-H, C-N y C-O.

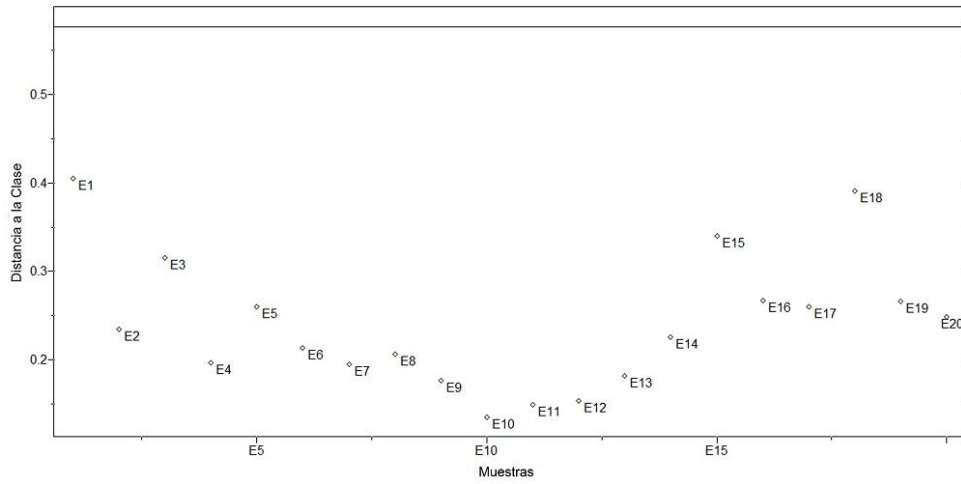


Figura 8. Modelo SIMCA (etapa de adiestramiento): diagrama de Coomans (con línea continua se destaca la distancia límite de la clase para un nivel de significación del 5 %).

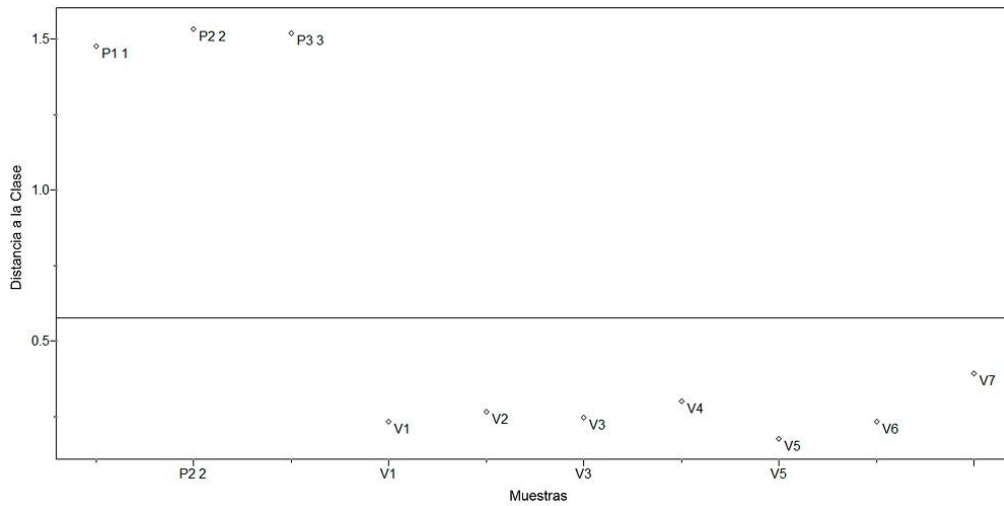


Figura 9. Modelo SIMCA (etapa de validación): diagrama de Coomans (con línea continua se destaca la distancia límite de la clase para un nivel de significación del 5 %).

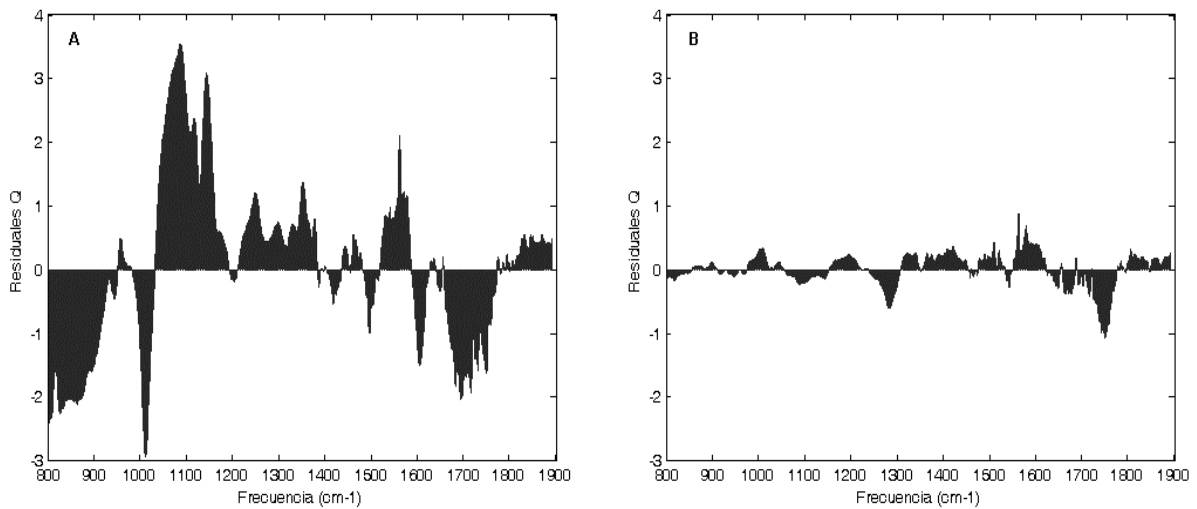


Figura10. Estadígrafo Q residual de la muestra problema P (A) vs muestra de referencia (B).

Cuando se compara el espectro infrarrojo de la muestra problema con el de una muestra de referencia se observa que mediante la inspección visual resulta difícil discriminar la muestra problema (Figura 11).

Los resultados de este trabajo ejemplificados mediante el estudio de dos aditivos utilizados en la industria del petróleo demuestran la utilidad de aplicar las técnicas de reconocimiento de patrones cuando se pretenda utilizar la técnica FT-MIR en el control de calidad de mezclas complejas.

CONCLUSIONES

Los aditivos que comúnmente se emplean en la industria del petróleo son mezclas de elevada complejidad; y, por ende, utilizar solo la comparación visual de sus espectros infrarrojos como método de control rutinario para verificar las especificaciones de calidad de este tipo de producto no es recomendable. Sobre la base del estudio de dos aditivos; uno antiestático y el otro un inhibidor de corrosión, se demostró la utilidad de combinar los resultados de la espectroscopía FT-MIR con las técnicas quimiométricas de reconocimiento de patrones. El método de clasificación SIMCA logró discriminar en la etapa de validación las muestras discrepantes de los modelos desarrollados: los modelos no tuvieron falsos positivos, ni falsos negativos. Si se define un conjunto de entrenamiento adecuado el uso combinado de ambas herramientas puede ser de gran utilidad en el control rutinario de los aditivos empleados en la industria del petróleo.

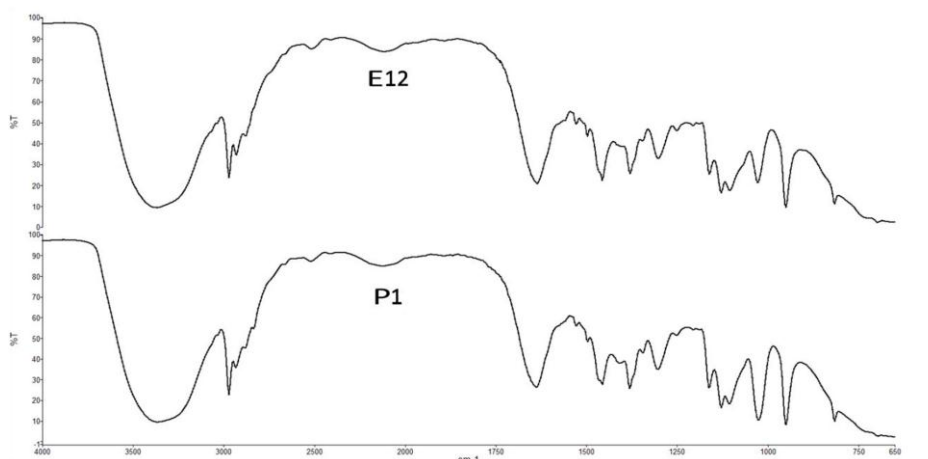


Figura 11. Espectros IR de la muestra de referencia E12 y de la muestra problema P1 registrados mediante la técnica HATR.

REFERENCIAS BIBLIOGRÁFICAS

ASTM D5845-01: Standard test method for determination of MTBE, ETBE, TAME, DIPE, Methanol, Ethanol and tert-Butanol in gasoline by infrared spectroscopy.

Bassbasi H. A., Oussama M. A. (2013). Clasificación and quality control of lubricants oils by infrared spectroscopy and chemometric. *International Journal of Advanced Technology and Engineering Research*, 3(4), 59-62.

Brereton R. G. (2006). Consequences of sample size, variable selection, and model validation and optimization, for predicting classification ability form analytical data. *Trends Analytical Chemistry*, 25(11), 1103-1111.

Brereton R. G. (2007). Pattern Recognition. Chapter 5, In *Applied Chemometrics for Scientists* (pp.145-191). England: John Wiley and Sons, Ltd.

Brereton R. G. (2009). One Class Classifiers Chapter 6. In *Chemometrics for Pattern Recognition* (pp.233-288). England: John Wiley and Sons, Ltd.

Brereton R. G. (2015). Pattern recognition in chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 149, Part B, 90-96.

Cai E., Liu D., Liang L., Xu G. (2015). Monitoring of chemical industrial processes using integrated complex network theory with PCA. *Chemometrics and Intelligent Laboratory Systems*, 140, 22-35.

Carvalho F. I. M., Dantas-Fiho H. A. (2014). Study of type A gasoline quality and its chemical composition using principal component analysis. *Quimica. Nova*, 37(1), 33-38.

Chian H., Colegrove L. F. (2007). Industrial implementation of on-line multivariate quality control. *Chemometrics and Intelligent Laboratory Systems*, 88(2), 143-53.

Coelho Pontes M. J., Pereira C. F., Pimentel M. F., Cruz Vasconcelos F. V., Brito Silva A. G. (2011). Screening analysis to detect adulteration in diesel/ biodiesel blends using near infrared spectrometry and multivariate classification. *Talanta*, 85(4), 2159-2165.

Esbensen K. H. (2001). SIMCA: An Introduction to Classification Chapter 15. In *Multivariate Data Analysis in Practice* (pp. 335-351). Norway: Process AS.

Ferreira da Silva M. P., Brito L.R., Honorato F. A., Silveira Paim A. P., Pimentel M. F. (2014). Classification of gasoline as with or without dispersant and detergent additives using infrared spectroscopy and multivariate classification. *Fuel*, 116(15), 151-157.

Flumignan D. L., Boralle N., Eduardo de Oliveira J. (2010). Screening Brazilian comercial gasoline quality by hydrogen nuclear magnetic resonance spectroscopic fingerprintings and pattern-recognition multivariate chemometric analysis. *Talanta*, 82(1), 99-105.

Gabrielsson J., Lindberg N. O., Lundstedt T. (2002). Multivariate methods in pharmaceutical applications: Review. *Journal of Chemometrics*, 16(3), 141-160.

Gredilla A., Fdez-Ortiz de Vallejuelo S., Amigo J. M., de Diego A., Madariaga J. M. (2013). Unsupervised pattern-recognition techniques to investigate metal pollution in estuaries. *Trends in Analytical Chemistry*, 46, 59-69.

Hopke P.K. (2015). Chemometrics applied to environmental systems. *Chemometrics and Intelligent Laboratory Systems*, 149, Part B, 205-214.

Kessler R.W. (2013). Perspectives in process analysis. *Journal of Chemometrics*, 27(11), 369-378.

Khanmohammadi M., Garmarudi A. B., de la Guardia M. (2013). Feature selection strategies for quality screening of diesel samples by infrared spectrometry and linear discriminant analysis. *Talanta*, 104, 128-134.

Kumar N., Bansal A., Sarma G. S., Rawal R. K. (2014). Chemometrics tools used in analytical chemistry: An overview. *Talanta*, *123*, 186-199.

Maldonado M., Barreiro P., Gutiérrez R., Vergara G. (2018). Mid- infrared uncooled sensor for the identification of pure fuel, additives and adulterants in gasoline. *Fuel Processing Technology*, *171*, 287-292.

Mas S., de Juan A., Tauler R., Olivieri A. C., Escandar G. M. (2010). Application of chemometric methods to environmental analysis of organic pollutants: A review. *Talanta*, *80*(3), 1052-1067.

Materazzi S., Gregori A., Ripani L., Apriceno A., Risoluti R. (2017). Cocaine profiling: Implementation of a predictive model by ATR-FTIR coupled with chemometrics in forensic chemistry. *Talanta*, *166*, 328-335.

Mendes G., Barbeira P. J. S. (2013). Detection and quantification of adulterants in gasoline using distillation curves and multivariate methods. *Fuel*, *112*, 163-71.

Miller J. N., Miller J. C. (2010). Multivariate analysis Chapter 8. In: *Statistics and Chemometrics for Analytical Chemistry* (pp. 221-250). England: Prentice Hall, Pearson Education Limited.

Nespeca M. G., Villarrubia J. F., Munhoz L., Flumignan D. L., Oliveira J. E. (2018). Rapid and sensitive method for detecting adulterants in gasoline using ultra-fast gas chromatography and Partial Least Square Discriminant Analysis. *Fuel*, *215*, 204-211.

Olive P. (2017). Class-modelling in food analytical chemistry: Development, sampling, optimisation and validation issues - A tutorial. *Analytica Chimica Acta*, *982*, 9-19.

Olivieri A. C., Rivas G. A. (2011). La química analítica del siglo XXI. *Asociación Ciencia Hoy*, *21*(124), 51-56. Recuperado de [https://es.scribd.com/document/355044602/Química Analítica](https://es.scribd.com/document/355044602/Química-Analítica).

Pacheco-Galindo L. A. (2014). Aplicaciones de la espectrometría de infrarrojo en la caracterización y tipificación de aditivos para la industria petrolera. (Tesis inédita de ingeniero químico industrial). Escuela Superior de Ingeniería Química e Industrias Extractivas, México DF.

Pérez-Caballero G., Andrade J. M., Olmos P., Molina Y., Jiménez I., Durán JJ, *et al.* (2017). Authentication of tequilas using pattern recognition and supervised classification. *Trends in Analytical Chemistry*, *94*, 117-129.

PerkinElmer Spectrum, version 10.4.4.449, copyright 2015 PerkinElmer, Inc.

Pirouette, Tutorial of Multivariate Data Analysis. (2003). Infometrix, Inc., version 3.11. Chapter 6 Classification Methods, 12-25.

PLS-Toolbox 6.7.1 for use with MATLAB. (2012). Eigenvector Research, Inc., version 6.7.

Pomerantsev A. L., Rodionova O. Y. E. (2012). Process analytical technology: a critical view of the chemometricians. *Journal of Chemometrics*, *26*(6), 299-310.

Renterghem P. V., Pierre-Edouard S., Saugy M., Van Eenoo P. (2013). Statistical discrimination of steroid profiles in doping control with support vector machines. *Analytica Chimica Acta*, 768, 41-48.

Risoluti R., Materazzi S., Ripani L. (2016). Early detection of emerging street drugs by near infrared spectroscopy and chemometrics. *Talanta*, 153, 407-413.

Shand C. A., Wendler R., Dawson L., Yates K., Stephenson H. (2017). Multivariate analysis of Scotch whisky by total reflection x-ray fluorescence and chemometric methods: A potential tool in the identification of counterfeits. *Analytica Chimica Acta*, 976, 14-24.

Sheng L., Lian-kui D. (2012). Classification of gasoline brand and origin by Raman spectroscopy and a novel R-weighted LSSVM algorithm. *Fuel*, 96, 146-152.

Shulcopper J. L. (2013). Acerca del surgimiento del Reconocimiento de Patrones en Cuba. *Revista Cubana de Ciencias Informáticas*, 7(2), 169-192.

Talavera-Bustamante I., Armentero-Madrado I. M. (2012). Clasificación de los suelos con fines periciales mediante microscopía electrónica de barrido con analizador de rayos X y el empleo de herramientas quimiométricas para el procesamiento multivariante de los datos. *Revista Cubana de Química*, 24(1), 32-39.

Talavera-Bustamante I., Rodríguez-González D., Martínez-Calvo A. (2011). Clasificación de marihuana nacional y extranjera empleando cromatografía gaseosa y técnicas de reconocimiento de patrones. *Revista Cubana de Química*, 23(2), 88-96.

Vempatapu B. P., Kamaujia P. K. (2017). Monitoring petroleum fuel adulteration: A review of analytical methods. *Trends in Analytica Chemistry*, 92, 1-11.

Westard F., Marini F. (2015). Validation of chemometric models- a tutorial. *Analytica Chimica Acta*, 893, 14-24.