

Cálculo de modelos de clasificación de gasolinas a partir de perfiles cromatográficos

Calculation of gasoline classification models from chromatographic profiles



<https://cu-id.com/2144/v17e06>

[✉]Reinaldo Fernández Fernández*, [✉]Yumirka Comesaña García, Ángel Dago Morales, Roberto Oropesa Rodríguez, [✉]Alicia Romero Hernández

RESUMEN: El control de calidad de las gasolinas exige la determinación de su composición por cromatografía gaseosa capilar de alta resolución; este análisis proporciona los contenidos totales de parafinas, isoparafinas, aromáticos, naftenos y olefinas (método PIANO). El objetivo de este trabajo fue desarrollar modelos de clasificación de gasolinas tomando como variables el perfil cromatográfico completo para extraer la información química relacionada con la covarianza de los diferentes compuestos presentes en el sistema. Se procesaron los cromatogramas de 70 muestras de gasolinas colectadas entre los años 2016 y 2019, en los que fue imprescindible minimizar el desplazamiento de los tiempos de retención y de la línea base que se origina en las diferentes mediciones. Se utilizó como técnica de reconocimiento de patrones el análisis jerárquico de clúster (HCA), el cual permitió definir por similitud dos clases que agrupan en una categoría diferente las muestras del año 2019. A partir de los resultados de este análisis exploratorio se aplicó el método de clasificación de modelado suave independiente por analogía de clases (SIMCA), obteniéndose que la distancia entre las clases y sus proyecciones resultaron adecuadas, sin errores de clasificación. Estas técnicas constituyen herramientas idóneas para detectar adulteraciones y cambios en los procesos de producción.

Palabras clave: cromatografía gaseosa, modelos de clasificación, gasolinas.

ABSTRACT: The quality control of gasoline requires the determination of its composition by high-resolution capillary gas chromatography; this analysis provides the total contents of paraffins, isoparaffins, aromatics, naphthenes and olefins (PIANO method). The objective of this work was to develop gasoline classification models taking the complete chromatographic profile as variables, which allowed extracting the chemical information related to the covariance of the different compounds present in the system. The profiles of 70 gasoline samples collected between 2016 and 2019 were processed, in which it was essential to minimize the shift in retention times and baseline that originates from the different measurements. Hierarchical cluster analysis (HCA) was used as a pattern recognition technique, which allowed defining by similarity two classes that group the samples of the year 2019 in a different category. Based on the results of this exploratory analysis, the Soft Independent Modeling of Class Analogy (SIMCA) classification method was applied, obtaining that the distance between the classes and their projections was adequate, without classification errors. These techniques are ideal tools to detect adulterations and changes in production processes.

Keywords: gas chromatography, classification models, gasolines.

INTRODUCCIÓN

El desarrollo de métodos cuantitativos univariados a partir de datos cromatográficos se basa habitualmente en obtener la altura o el área de un pico en función de la concentración del analito que se quiere cuantificar. Pese a la utilidad que presentan estos métodos de calibración y a su profusa aplicación, en numerosas ocasiones la información más importante no se encuentra en picos cromatográficos analizados de forma aislada,

sino en todo su perfil, en particular, cuando se analizan mezclas complejas. La utilidad de emplear la información de todo el perfil cromatográfico, conjuntamente con las técnicas multivariadas de análisis de datos, está dada por el hecho de que se puede extraer la información química que está relacionada con la covarianza de los diferentes compuestos que presenta el sistema; en muchos casos el compuesto en menor proporción es el que tiene mayor poder de discriminación (Szymanska, 2018).

Recibido: 21/05/2024

Aprobado en su forma original: 07/06/2025

Centro de Investigaciones del Petróleo, CEINPET. Churrucá No.481 e/ Washington y Vía Blanca. Cerro, La Habana. Cuba.

E-mail: yumirka@ceinpet.cupet.cu, dagoangel49@yahoo.com, orodosmil20@gmail.com, alimoy@ceinpet.cupet.cu

*Correo electrónico: reyn@ceinpet.cupet.cu

Conflicto de Intereses: Los autores de este trabajo declaran que no existe conflicto de intereses.

Contribuciones de los autores: **Gestión de Proyecto:** Reinaldo Fernández Fernández, Yumirka Comesaña García y Ángel Dago Morales. **Metodología:** Reinaldo Fernández Fernández y Yumirka Comesaña García. **Investigación:** Reinaldo Fernández Fernández, Yumirka Comesaña García, Ángel Dago Morales y Roberto Oropesa Rodríguez. **Análisis formal:** Reinaldo Fernández Fernández, Roberto Oropesa Rodríguez y Alicia Romero Hernández. **Conservación de datos:** Reinaldo Fernández Fernández, Roberto Oropesa Rodríguez y Alicia Romero Hernández. **Redacción - primera redacción:** Reinaldo Fernández Fernández. **Redacción - revisión y edición:** Reinaldo Fernández Fernández y Yumirka Comesaña García.

Artículo bajo licencia [Creative Commons Reconocimiento-NoComercial 4.0 Internacional \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

Se debe destacar la utilidad del empleo de los perfiles cromatográficos en problemas de tipo forense, por ejemplo: para la detección de adulteraciones de medicamentos (Custer et al., 2016; Deconinck et al., 2012), la definición del origen de determinadas drogas (Martínez-Calvo et al., 2011), y para la identificación de los responsables de derrames de petróleo (Juahir et al., 2017; Ismail et al. 2016; Christensen et al., 2007; Faksness et al., 2002; Daling et al., 2002).

La calidad de los combustibles generalmente se controla mediante las normas ASTM (American Society for Testing Materials); estas normas cubren las especificaciones de los productos, así como los procedimientos analíticos que se emplean en la caracterización fisicoquímica de los mismos. En el caso específico de la gasolina, que es una mezcla de cientos de hidrocarburos obtenida por destilación fraccionada del petróleo, su composición varía en función de la naturaleza del crudo y del propio proceso tecnológico de refinación. Entre los parámetros físico-químicos que exige su control de calidad se encuentra la determinación de su composición por cromatografía gaseosa capilar de alta resolución. Este análisis se realiza acorde a la norma ASTM D6729, y permite la cuantificación de los contenidos totales de parafinas, isoparafinas, aromáticos, naftenos y olefinas, lo cual se conoce como método PIANO. Si bien estos cinco parámetros constituyen una valiosa información para el control del proceso industrial y de la calidad del producto final, son insuficientes para emplearlos en la detección de adulteraciones de combustibles, lo que actualmente constituye un problema generalizado (Dadson et al., 2018; Wiedemann et al., 2005; Doble et al., 2003). Las técnicas quimiométricas de reconocimiento de patrones, entre ellas los métodos de clasificación y de análisis exploratorio de datos, son las herramientas idóneas para la detección de adulteraciones de gasolinas (Flumignan et al., 2007; Carvalho & Dantas, 2014; Skrobot et al., 2007). Para la aplicación de las citadas técnicas es imprescindible minimizar en los cromatogramas el desplazamiento de los tiempos de retención y la co-elución de picos que se origina en las diferentes mediciones (Skov et al., 2006), ya que dificulta el desarrollo de los modelos multivariados y en muchos casos conduce a resultados erróneos (Malmquist & Danielsson, 1994); por lo cual, es necesario emplear previamente procedimientos matemáticos para alinear picos comunes (Engel et al., 2013). En líneas generales las transformaciones que garantizan una adecuada calidad de los resultados del análisis multivariado están enfocadas en minimizar el ruido, corregir la línea base y alinear los picos cromatográficos (Sinkov et al., 2011; Trygg et al., 2009).

El ruido es una señal que se caracteriza porque sus valores en dos instantes de tiempo diferentes no están correlacionados; esta variación aleatoria puede reducirse mediante técnicas de suavizado digital.

La aplicación de las técnicas de suavizado se lleva a cabo comúnmente mediante el filtrado digital (Thekkudan & Rutan, 2009), que es un proceso que elimina las frecuencias atribuidas al ruido y mantiene la información importante de la señal. Uno de los filtros más utilizados para analizar datos espectroscópicos y cromatográficos es el filtro de Savitzky-Golay, ya que no altera prácticamente la información de la señal. Éste se basa en ajustar los puntos medidos en una ventana a un polinomio de grado n , de forma tal que el valor de la respuesta en cada punto se reemplaza por una combinación lineal de los puntos vecinos. La elección del orden del polinomio de ajuste y del ancho de la ventana son parámetros a tener en cuenta, puesto que si el suavizado es mínimo no se consigue eliminar todo el ruido, mientras que si el suavizado es excesivo se corre el riesgo de perder información.

La variación de la línea base es otro factor que deteriora la calidad de la señal y complica la extracción de la información. Mientras que el ruido se produce a altas frecuencias de la señal, la variación de la línea base se produce a bajas frecuencias. El problema causado por esta variación es particular de cada conjunto de datos, lo cual dificulta la elección de las técnicas para la corrección de la señal. El método general, propuesto por Gan y col. (2006), consiste en la estimación aproximada de la línea base mediante un proceso iterativo de ajuste de un polinomio: se calcula una línea base que constituye un umbral a partir del cual solo se mantiene la parte de la señal sobreestimada, o sea, los picos. Si se ha perdido información se vuelve a realizar una iteración con una nueva línea base hasta que el resultado sea aceptable.

El proceso de alineamiento de las señales instrumentales es una técnica que permite alinear picos comunes obtenidos de diferentes mediciones. En el caso de perfiles cromatográficos o espectroscópicos complejos se ha empleado con éxito la técnica de alineación optimizada por correlación (Correlation Optimized Warping, COW), la cual emplea una función de corrección lineal a tramos basada en el modelo comprensión/expansión (Trung & Laukens, 2013). Ésta utiliza una señal o vector de referencia que debe ser representativo de las muestras que se quieren alinear (Kumara, 2018); una opción es utilizar como referencia la media de los cromatogramas, o un cromatograma representativo del conjunto. Los parámetros que se ajustan para maximizar la correlación entre las muestras y el cromatograma de referencia son: el vector a ser usado como referencia, la longitud del segmento y el grado de flexibilidad que controla la cantidad de compresión o estiramiento de cada segmento (Bloemberg et al., 2013; Skov et al., 2006). Para comprobar que los parámetros utilizados en el ajuste de los perfiles fueron adecuados, y que no provocan variaciones considerables en las áreas o en el perfil de los picos individuales se utilizan variados esquemas; el más sencillo es visualizar los perfiles

e ir variando los parámetros de las técnicas utilizadas en su alineación hasta lograr un ajuste apropiado (Sinkov et al., 2011; Trung & Laukens, 2013). Además de la comparación visual, Kai-Tal et al., (2006) y Peiyan et al., (2018) utilizan con efectividad el análisis de la semejanza en la agrupación de las muestras en función de los cambios realizados.

Los métodos de reconocimiento de patrones cubren dos amplias vertientes de trabajo: el análisis exploratorio de datos y la clasificación. La finalidad del análisis exploratorio es examinar los datos experimentales previamente a la aplicación de cualquier técnica estadística; no requiere de un conocimiento previo del sistema y proporciona métodos sencillos para organizar los datos, identificar muestras o variables atípicas y analizar las posibles agrupaciones, lo cual permite definir las posibles clases o categorías del sistema estudiado. Los métodos no supervisados más utilizados son: el Análisis Jerárquico de Clúster (Hierarchical Cluster Analysis, HCA) y el Análisis por Componentes Principales (Principal Components Analysis, PCA).

En un análisis clasificatorio se construyen modelos capaces de pronosticar la pertenencia de una muestra a una determinada clase. La matriz de datos, además de las variables experimentales, incluye una variable categórica que indica la categoría a la que pertenece la muestra, y que constituye la variable que se quiere predecir. Los procedimientos basados en la similaridad asumen que las muestras de una misma categoría se encuentran más agrupadas en el espacio de las variables que las muestras de una categoría diferente. La técnica de Modelado Suave Independiente por Analogía de Clases (Soft Independent Modeling of Class Analogy, SIMCA) ha sido uno de los métodos de clasificación más recurridos en el control de calidad en muchas industrias. Es un método supervisado de reconocimiento de patrones que se basa en el principio de analogía entre las muestras que pertenecen a una misma agrupación, y emplea para el cálculo de las distancias los valores determinados mediante análisis por componentes principales.

El objetivo de este trabajo es desarrollar modelos de clasificación de gasolinas tomando como variables el perfil cromatográfico alineado de muestras colectadas en diferentes periodos de tiempo. Se utiliza como técnica de reconocimiento de patrones el Análisis jerárquico de clúster para definir la existencia de diferentes categorías. A partir de los resultados del análisis exploratorio se aplica como método de clasificación el Modelado suave independiente por analogía de clases.

MATERIALES Y MÉTODOS

Muestras y análisis cromatográfico

Se emplearon los cromatogramas de 70 muestras de gasolinas procedentes de una refinería de la Unión Cubapetróleo, colectadas entre 2016 y 2019, en los periodos que a continuación se detallan:

M1-M10: Año 2019.

M11-M35: Año 2017.

M36-M70: Año 2016.

Las condiciones del análisis cromatográfico se realizaron acorde a la norma ASTM D6729 (2014). Se utilizó un cromatógrafo Agilent 7890A equipado con detector de ionización a la llama, automuestrador G4513A y columna capilar DB-Petro de 100 m, 0,25 mm diámetro interno y 0,5 µm de espesor de película. Los componentes fueron separados en la columna con hidrógeno como gas portador con un flujo de 1,8 ml/min. El horno se programó con cuatro rampas de calentamiento: 35 °C (15 min) hasta 70 °C a 1,5 °C/min; 70 °C (10 min) hasta 150 °C a 3,0 °C/min; 150 °C (10 min) hasta 200 °C a 5,0 °C/min; 200 °C (10 min) hasta 250 °C a 50 °C/min. La estabilidad de la respuesta instrumental se verificó mediante el análisis del patrón Alphagaz PIANO Calibration Standards Supelco 4-4594-U y del material de referencia Sigma WSPA RM-1 de gasolina California fase II. Se utilizó el programa Agilent GC ChemStation para la adquisición y procesamiento de los datos.

Matriz X de perfiles cromatográficos

El intervalo de tiempos de retención que conformó el perfil cromatográfico de las muestras analizadas abarcó desde 5,74 hasta 60,35 min, con un paso de 0,0084; lo cual implicó 6554 variables por muestra: (dimensiones de la matriz $X_{70 \times 6554}$). Para minimizar el ruido instrumental se utilizó el filtro de Savitzky-Golay; en la corrección de la línea base se empleó el método lineal de ajuste por polinomios; y en la alineación de los perfiles, la alineación optimizada mediante correlación. Los parámetros utilizados para el cálculo de las citadas transformaciones se realizaron acorde a las referencias reportadas para cada una de ellas (Kumara, 2018; Bloemberg et al., 2013).

Métodos quimiométricos utilizados

Se utilizó como técnica de reconocimiento de patrones el análisis jerárquico de clúster (HCA), el cual permite definir por similaridad la existencia de diferentes categorías. Los resultados de los clúster jerárquicos son presentados en forma de dendograma, donde el eje vertical representa la medida similaridad a la cual sucesivamente las muestras son unidas al grupo principal.

Para identificar los diferentes tipos de gasolinas se aplicó el método de clasificación de modelado suave independiente por analogía de clases (SIMCA) a partir de los resultados obtenidos del análisis exploratorio. En los cálculos se emplearon los sistemas de programas Pirouette (2003) y MATLAB (2008).

RESULTADOS Y DISCUSIÓN

Los perfiles cromatográficos de las muestras de gasolinas analizadas por el método PIANO se ajustaron mediante el siguiente procedimiento:

- Conformación de la matriz X de trabajo de dimensiones $X_{70 \times 6554}$.
- Suavizado de los datos mediante el empleo del filtro de Savitzky-Golay: se utilizó una ventana de cinco puntos, un polinomio de ajuste de segundo grado y la primera derivada.
- Empleo de la técnica lineal para la corrección de la línea base: se empleó un polinomio de primer grado.
- Alineación de los picos mediante el uso de la técnica COW: la longitud del segmento o tamaño de ventana utilizado fue de cinco; el grado de flexibilidad se fijó en 50. En la selección del cromatograma representativo (vector de referencia) se adoptó la variante automática.

El primer paso para comprobar si el procedimiento de alineación de los picos cromatográficos fue adecuado consistió en visualizar y verificar si hubo

algún tipo de desajuste en los perfiles de sus áreas en todo el intervalo de trabajo. La comparación visual de los perfiles antes y después de las transformaciones es sencilla, pero aporta suficiente información como para descartar transformaciones que no conducen a mejoría en los datos experimentales.

Los resultados demostraron que se logró una adecuada corrección de la línea base, así como de la alineación de los picos cromatográficos (Fig. 1).

Resultados del análisis jerárquico de clúster (HCA)

Dado que el número de clases existentes no se conoce cuando se obtienen perfiles similares como en el caso de las gasolinas, se aplicó el algoritmo de agrupamiento no supervisado HCA para el reconocimiento de estas.

En la figura 2 se expone en forma de dendograma el resultado de las 70 muestras organizadas en filas acorde a su similitud. La longitud de las diferentes ramas es proporcional a las distancias entre los grupos vinculados: a mayor longitud de una rama es menor la similitud entre los objetos. El valor de similitud 1.0 indica un duplicado exacto, y el 0.0 corresponde con la máxima distancia y disimilitud.

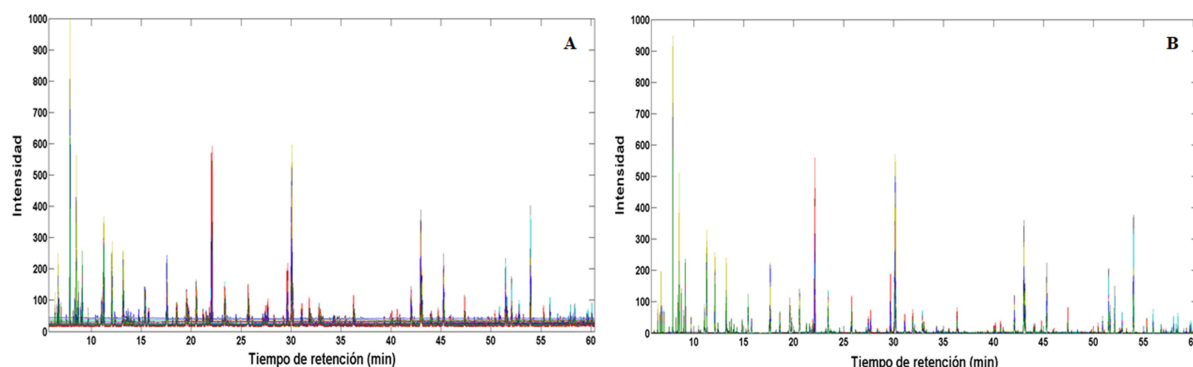


Fig. 1. Perfiles cromatográficos en el intervalo de trabajo: A) antes de alinear, B) después de la alineación

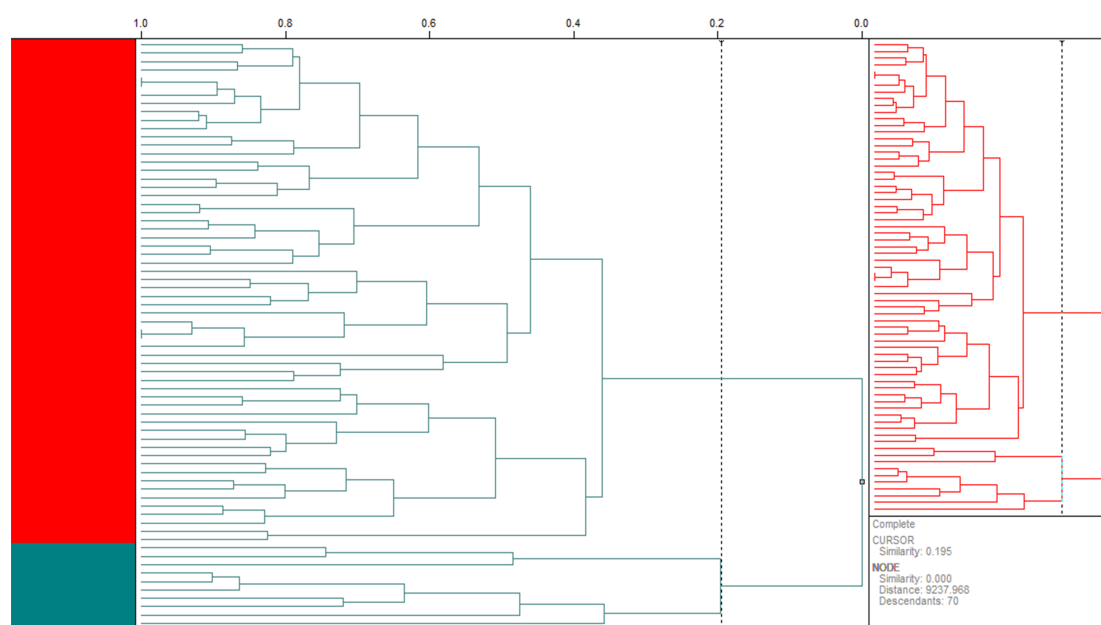


Fig. 2. Modelo HCA con los 70 perfiles cromatográficos

El análisis permitió definir dos clases, donde se separan las muestras colectadas en el año 2019 (M1-M10) del resto. En la **Figura 3** se presenta el dendograma ampliado correspondiente a estas 10 muestras, utilizando como método de conexión entre las mismas el del cálculo de los vecinos más cercanos.

Resultados de la clasificación por el método SIMCA

El modelo SIMCA se calculó sobre la base de las dos clases definidas con anterioridad por HCA, las cuales fueron identificadas como 1 y 2 (clase 1 para las muestras colectadas en el año 2019, y clase 2 para el resto de las muestras). Este método supervisado utiliza un análisis de componentes principales para desarrollar un modelo de cada grupo dentro del conjunto.

Esas categorías se modelan por una serie de estructuras lineales que dependen del número de componentes requeridos para reproducir los datos de la clase.

El modelo se determinó con dos factores para las muestras pertenecientes a la clase 1 y con tres factores para las muestras del grupo 2. La distancia entre las clases fue de 5,7 que es un valor aceptable para este tipo de modelo: la selectividad de un modelo SIMCA se considera adecuada cuando el valor de dicho parámetro es > 3 (Cavado-Osorio et al., 2014).

En el diagrama de Coomans se presentan las distancias entre las clases (**Fig. 4**). Todas las muestras del conjunto de entrenamiento se ubican en los cuadrantes correctos; no hay muestras sin clasificar, ni errores de clasificación.

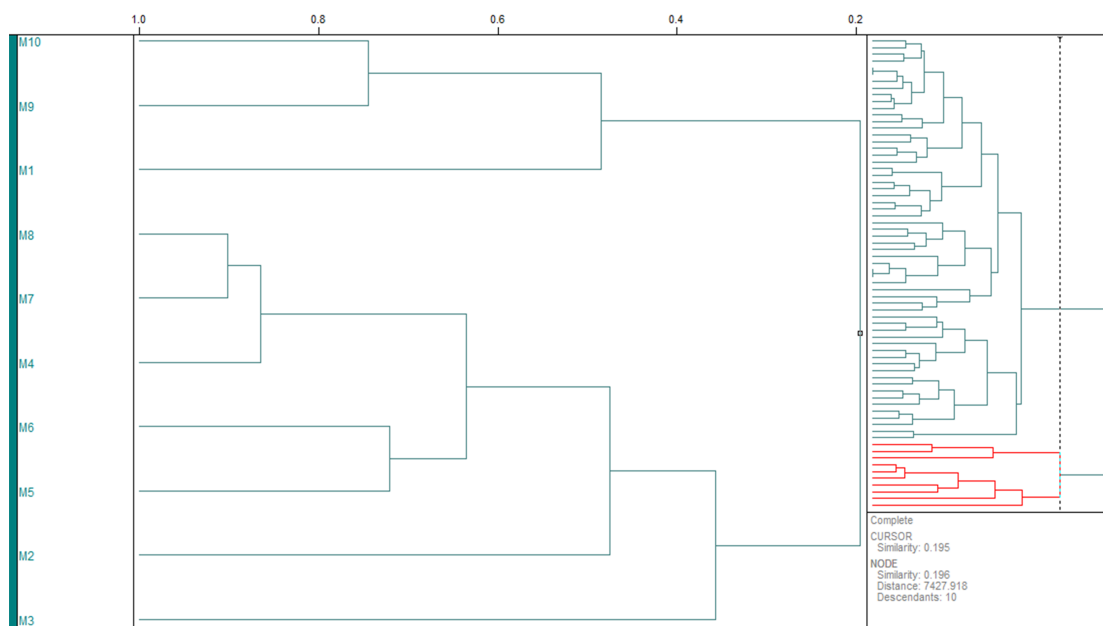


Fig. 3. Modelo HCA de una clase (M1-M10)

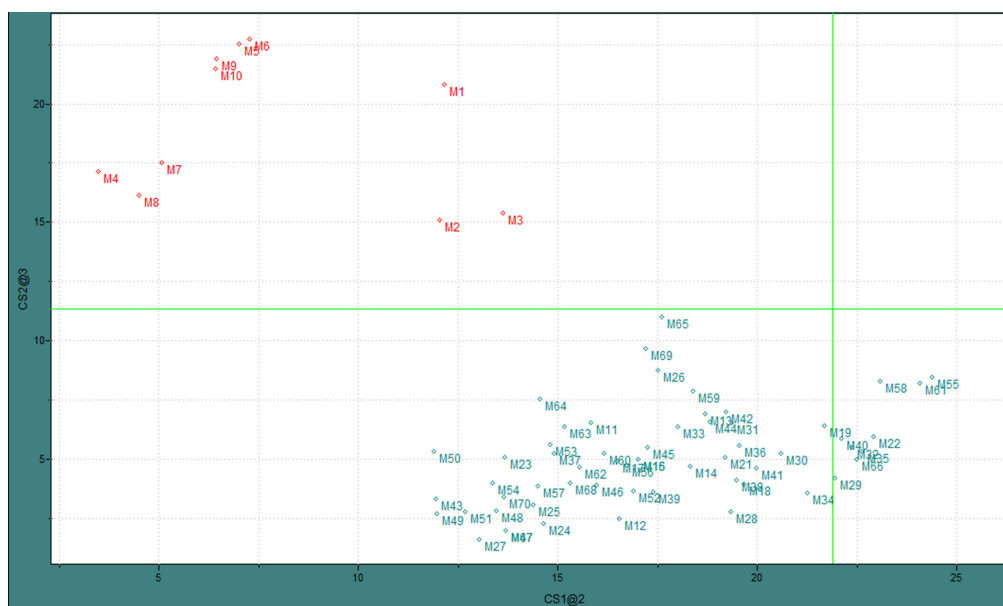


Fig. 4. Modelo SIMCA. Diagrama de Coomans

La proyección tridimensional de las muestras del conjunto de entrenamiento (Fig. 5) ratifica el adecuado grado de separación de los límites calculados para ambas clases. Esta gráfica proporciona una evaluación visual donde las fronteras de las elipses están fundamentadas en la desviación estándar de las puntuaciones en la dirección de cada componente principal.

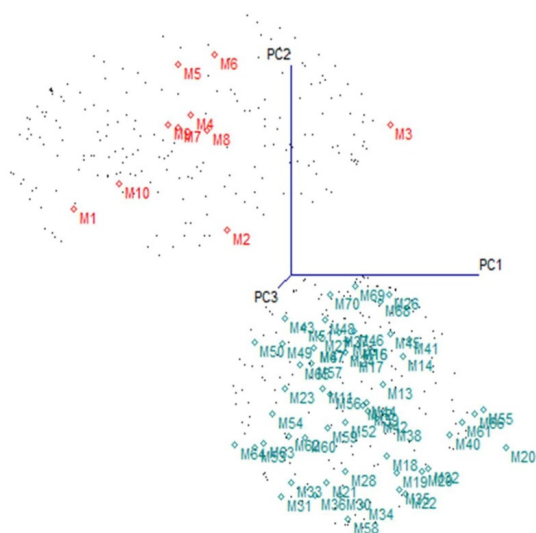


Fig. 5. Modelo SIMCA. Proyección tridimensional de las clases

En las Figuras 6 y 7 se muestran los modelos PCA para las clases 1 y 2, respectivamente. El esquema integra cada una de las categorías y calcula sus límites o fronteras con una probabilidad del 95%.

Los resultados demuestran que es factible calcular modelos robustos de reconocimiento de patrones para el control de calidad de combustibles y la detección de adulteraciones a partir de sus perfiles cromatográficos.

CONCLUSIONES

1. Se desarrollaron modelos de clasificación de gasolinas tomando como variables el perfil cromatográfico alineado de las muestras. Se utiliza como técnica de reconocimiento de patrones el análisis jerárquico de clúster, el cual permite definir por similitud dos clases que corresponden con los periodos de muestreo realizados, agrupando en una categoría diferente las muestras colectadas en el año 2019.
2. A partir de los resultados de este análisis exploratorio se aplica el método de clasificación de modelado suave independiente por analogía de clases, obteniéndose que la distancia entre las clases (5,7) y sus proyecciones resultaron adecuadas, sin errores de clasificación. Estas técnicas constituyen herramientas idóneas para detectar adulteraciones y cambios en los procesos de producción, así como para diferenciar las formaciones geológicas.

BIBLIOGRAFÍA

- Agilent GC ChemStation Software, version B-04-02, 2010.
- ASTM D6729-14. 2014. Standard Test Method for Determination of Individual Components in Spark Ignition Engine Fuels by 100 Metre Capillary High Resolution Gas Chromatography.
- Bloemberg, T. G., Gerretzen, J., Lunshof, A. & Wehrens, R. 2013. Warping methods for spectroscopy and chromatographic signal alignment: A tutorial. *Anal. Chim. Acta*, 781:14-32.
- Carvalho, F. & Dantas Filho, H. 2014. Estudo da qualidade da gasolina tipo A e sua composição química empregando análise de componentes principais. *Quim. Nova*, 37(1): 33-38.

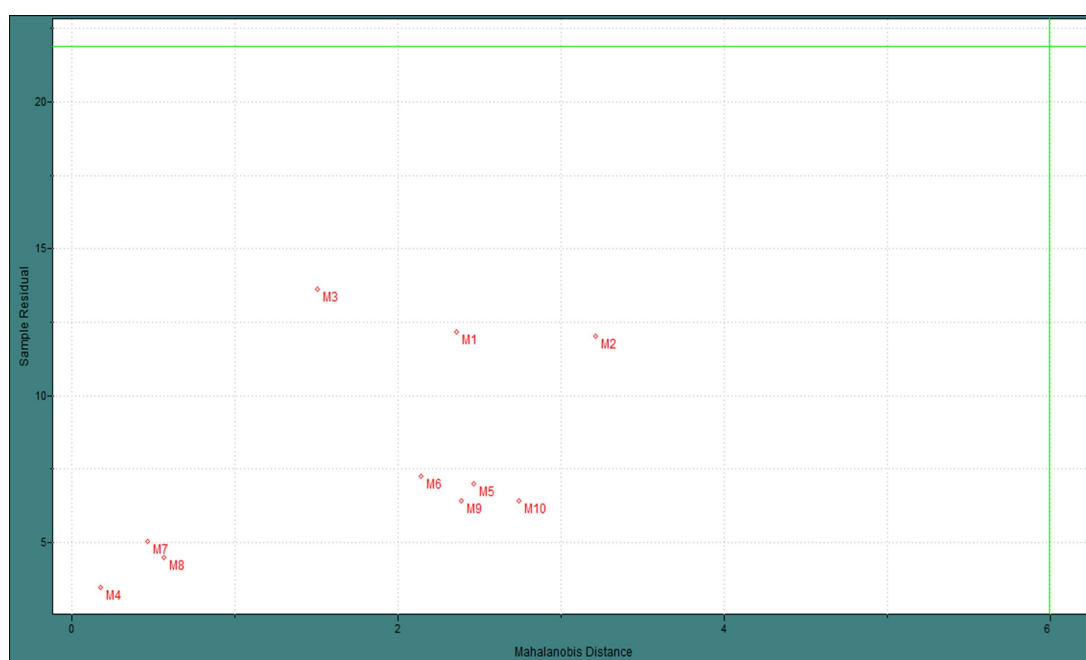


Fig. 6. Modelo SIMCA. Distancia a la clase 1 (con líneas continuas se destaca el límite de la clase para un nivel de significación del 5 %)

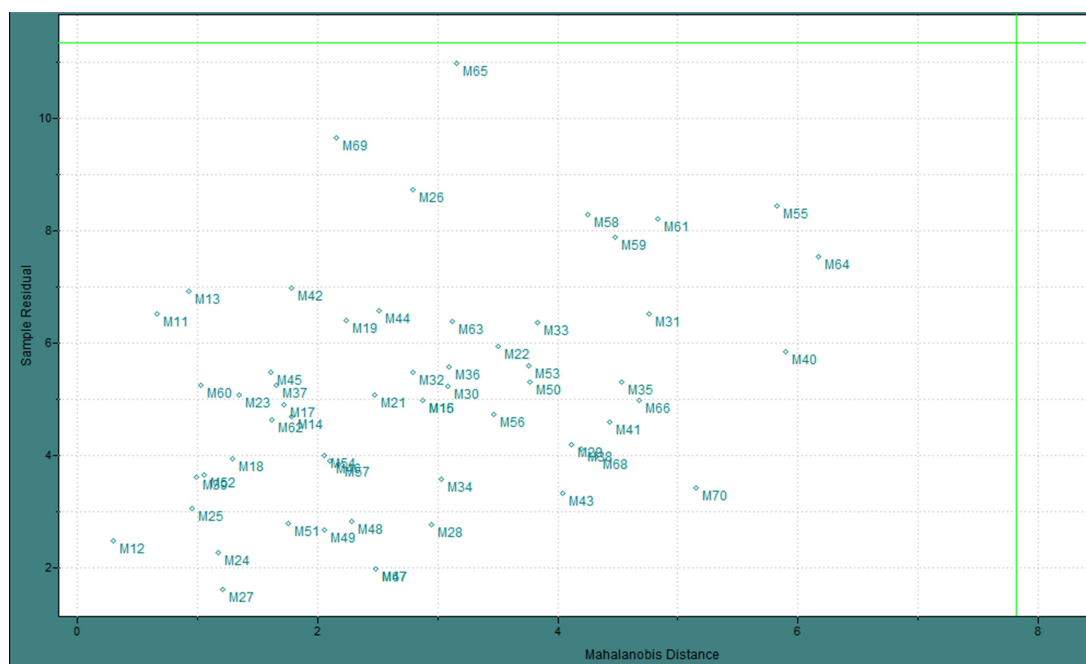


Fig. 7. Modelo SIMCA. Distancia a la clase 2 (con líneas continuas se destaca el límite de la clase para un nivel de significación del 5 %)

- Cavado-Osorio, A., Comesaña, Y., Fernández, R. & Dago-Morales, A. 2014. Técnicas quimiométricas de reconocimiento de patrones para el control de calidad de turbocombustibles Jet A-1 a partir de sus propiedades físico químicas. *Revista CENIC Ciencias Quím.*, 45: 18-26.
- Christensen, J. H. & Tomasi, G. 2007. Practical aspects of chemometric for oil spill fingerprinting. *J. Chromatogr. A*, 1169: 1-22.
- Custers, D., Krakowska, B., De Beer, J.O., Courselle, P., Daszykowski, M., Apers, S. & Deconinck, E. 2016. Chromatographic impurity fingerprinting of genuine and counterfeit Cialis® as a means to compare the discriminating ability of PDA and MS detection. *Talanta*, 146: 540-548.
- Dadson, J., Pandam, S. & Asiedu, N. 2018. Modeling the characteristics and quantification of adulterants in gasoline using FTIR spectroscopy and chemometric calibrations. *Cogent chemistry*, 4: 1-22.
- Daling, P. S., Faksness, L. G., Hansen, A. B. & Stout, S. A. 2002. Improved and standardized methodology for oil spill fingerprinting. *Environ. Forensics*, 3: 263-278.
- Deconinck, E., Sacré, P. Y., Courselle, P. & De Beer, J.O. 2012. Chemometrics and chromatographic fingerprints to discriminate and classify counterfeit medicines containing PDE-5 inhibitors. *Talanta*, 100: 123-133.
- Doble, P., Sandercock, M., Pasquier, E., Petocz, P., Roux, C. & Dawson, M. 2003. Classification of Premium and regular gasoline by gas chromatography/mass spectrometry, principal component analysis and artificial neural networks. *Forensic Science International*, 132: 26-39.
- Engel, J., Gerretzen, J., Szymanska, E., Jansen, J., Downey, G., Blanchet, L. & Buydens Lutgarde, M.C. 2013. Breaking with trends in pre-processing? *Trends Anal. Chem.*, 50:96-106.
- Faksness, L. G., Daling, P. S. & Hansen, A. B. 2002. Round Robin study- Oil spill identification. *Environ. Forensic*, 3:279-291.
- Flumignan, D.L., Tininis, A., Ferreira, F. & De Oliveira, J.E. 2007. Screening Brazilian C gasoline quality: Application of the SIMCA chemometric method to gas chromatographic data. *Anal. Chim. Acta*, 595: 128-135.
- Fernández, R., Dago-Morales, A., Oropesa, R., Comesaña, Y. & Romero, A. 2019. Predicción de propiedades físicoquímicas de gasolinas mediante cromatografía gaseosa y métodos de regresión multivariados. *Avances Cien. Ing.*, 10(3): 21-32.
- Gan, F., Ruan, G. & Mo, J. 2006. Baseline correction by improved iterative polynomial fitting with automatic threshold. *Chemom. Intell. Lab. Syst.*, 82(1): 59-65.
- Ismail, A., Toriman, M. E., Juahir, H., Kassim, A. M., Zain, S. M. & Ahmad, W. K. W. 2016. Chemometric techniques in oil classification from oil spill fingerprinting. *Mar. Pollut. Bull.*, 111: 339-46.
- Juahir, H., Ismail, A. M., Mohamed, S. B., Toriman, M. E., Kassim, A. M. & Zain, S. M. 2017. Improving oil classification quality from oil spill fingerprinting beyond six-sigma approach. *Mar. Pollut. Bull.*, 120:322-32.
- Kai-Tai, F., Yi-Zeng, L., Xiao-Lin, Y., Chan, K. & Guang-Hua, L. 2006. Critical value determination on similarity of fingerprints. *Chemom. Intell. Lab. Syst.*, 82(1-2): 236-240.

- Kumara, K. 2018. Optimizing the process of reference selection for correlation optimised warping (COW) and interval correlation shifting (icoshift) analysis: automating the chromatographic alignment procedure. *Analytical Methods*, 10(2):190-203.
- Malmquist, G. & Danielsson, R. 1994. Alignment of chromatographic profiles for principal component analysis - A prerequisite for fingerprinting methods. *J. Chromatogr., A*, 687: 71-88.
- Martínez-Calvo, A., Rodríguez, D. & Talavera, I. 2011. Clasificación de marihuana en nacional y extranjera empleando cromatografía gaseosa y técnicas de reconocimiento de patrones. *Rev. Cub. Quím.*, 23 (2): 88-96.
- MATLAB. 2008. The Language of Technical Computing. The Math Works, version 7.7.0.471.
- Peiyan, S., Kaiwen, B., Haoshuai, L., Fujuan, L., Xinping, W., Lixin, C., Guangmei, L., Qing, Z., Hongxia, T. & Mutai, B. 2018. An efficient classification method for fuel and crude oil types based on m/z 256 mass chromatography by COW-PCA-LDA. *Fuel*, 222: 416-423.
- Pirouette. Infometrix, Inc. 2003. Versión 3.11.
- Sinkov, N. A., Johnston, B. M., Sandercock, P. & Harynuk, J. 2011. Automated optimization and construction of chemometric models based on highly variable raw chromatographic data. *Anal. Chim. Acta*, 697(1-2): 8-15.
- Skov, T., van den Berg, F., Tomasi, G. & Bro, R. 2006. Automated alignment of chromatographic data. *J. Chemometrics*, 20: 484-497.
- Skrobot, V. L., Castro, E. V. R., Pereira, R. C. C., Pasa, V. M. D. & Fortes, I. C. P. 2007. Use of principal component analysis (PCA) and linear discriminant analysis (LDA) in gas chromatographic (GC) data in the investigation of gasoline adulteration. *Energy Fuels*, 21: 3394-400.
- Szymańska, E. 2018. Modern data science for analytical chemical data - A comprehensive review. *Anal. Chim. Acta*, 1028:1-10.
- Thekkuddan, D. F. & Rutan, S. C. 2009. Denoising and signal-to-noise ratio enhancement: classical filtering (pp.9-24). In *Comprehensive Chemometrics*, Brown S., Taulers R., Walczak B. (Ed). Elsevier, Oxford.
- Trung Nghia, V. & Laukens, K. 2013. Getting your peaks in line: A review of alignment methods for NMR spectral data. *Metabolites*, 3:259-276.
- Trygg, J., Gabrielsson, J. & Lundstedt, T. 2009. Background Estimation, Denoising and Preprocessing. In *Comprehensive Chemometrics*, Brown, S., Tauler, R., Walczak, B. (Ed), Elsevier, Oxford, 2: 1-8.
- Wiedemann, L. S. M., d'Avila, L.A. & Azevedo, D.A. 2005. Adulteration detection of Brazilian gasoline samples by statistical analysis. *Fuel*, 84:467-473.