

PROCEDIMIENTO PARA LA ALINEACIÓN DE PERFILES CROMATOGRÁFICOS APLICADO EN LA CARACTERIZACIÓN DE GASOLINAS

Procedure for the alignment of chromatographic profiles applied in the characterization of gasolines.

Yumirka Comesaña García^a, Reinaldo Fernández Fernández^a, Roberto Oropesa Rodríguez^a, Alicia Romero Hernández^a, Ángel Dago-Morales^a,

^{a,*} Centro de Investigación del Petróleo. Churrucá No. 481 e/ Washington y Vía Blanca, Cerro, La Habana, Cuba. adago@nauta.cu

Recibido: 06 de febrero de 2020.

Aceptado: 19 de mayo de 2020.

RESUMEN

La gasolina es una mezcla de cientos de hidrocarburos obtenida del petróleo por destilación fraccionada, su composición varía en función de la naturaleza del crudo y del propio proceso tecnológico. Entre los parámetros fisicoquímicos que exige el control de calidad se encuentra la determinación de la composición por cromatografía gaseosa capilar de alta resolución. Este análisis proporciona los contenidos totales de parafinas, isoparafinas, aromáticos, naftenos y olefinas y constituye una valiosa información para el control del proceso industrial y de la calidad del producto final, pero son insuficientes para emplearlos en problemas de tipo forense. Las técnicas de reconocimiento de patrones son las herramientas idóneas para estos fines, pues permiten extraer la información química que está relacionada con la covarianza de los diferentes compuestos presentes en el sistema; pero para su aplicación es imprescindible minimizar en los cromatogramas el desplazamiento de los tiempos de retención que se origina en las diferentes mediciones. El objetivo de este trabajo fue desarrollar un procedimiento para la alineación de los perfiles cromatográficos de muestras de gasolinas mediante el empleo de las transformaciones: filtro de Savitzky-Golay, mínimos cuadrados asimétricos y alineación optimizada por correlación. Se comprobó mediante las técnicas de análisis por componentes principales y de regresión por mínimos cuadrados parciales que los resultados no difieren de los reportados cuando se utilizan los porcentajes de áreas de los 70 picos más representativos de los cromatogramas. Las varianzas explicadas en los modelos de análisis por componentes principales fueron similares: 92 y 91 %. Los errores residuales medios de los modelos de regresión de mínimos cuadrados parciales para la estimación del número de octano experimental fueron de 1,0 y 1,1; y en ambos casos estuvieron acorde a la reproducibilidad del método de referencia.

Palabras clave: gasolina; mínimos cuadrados asimétricos; alineación optimizada por correlación; técnicas de reconocimiento de patrones.

ABSTRACT

Gasoline is a mixture of hundreds of hydrocarbons obtained from petroleum by fractional distillation, its composition varies depending on the nature of the crude and the technological refining process. Among the physical chemical parameters of quality control the determination of the composition by high-resolution capillary gas chromatography is required. This analysis provides the total contents of paraffins, isoparaffins, aromatics, naphthenes and olefins and constitute valuable information for the control of the industrial process and the quality of the final product, but they are insufficient to be used in forensic problems, specifically, to detect problems related to adulteration. Chemometric pattern recognition techniques are suitable tools for these purposes; but for the application of the techniques it is essential to minimize in the chromatograms the displacement of the retention times that originate in the different measurements. The objective of the work was to develop a chromatographic profiles alignment procedure of gasoline samples using the transformations: Savitzky-Golay filter asymmetric least squares and correlation-optimized alignment. Principal components analysis and partial least squares regression techniques were used to verified that the results are not different from those reported when using the percentages of areas of the 70 most representative peaks of the chromatograms. The variances explained in the principal component analysis models were similar: 92 and 91 %. The mean residual errors of the partial least squares regression models for the estimation of the experimental octane number were 1.0 and 1.1; and in both cases they were in accordance with the reproducibility of the reference method. The chromatographic profile alignment procedure for gasoline can be used in the development of chemometric pattern recognition techniques for solving forensic problems.

Keywords: gasoline; asymmetric least squares; correlation optimized warping; pattern recognition techniques.

INTRODUCCIÓN

El procedimiento habitual para el desarrollo de métodos cuantitativos univariados a partir de datos cromatográficos consiste 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 ya sea de datos espectroscópicos o cromatográficos, 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, o sea, cómo covarían unos con respecto a otros: en muchos casos el compuesto en menor proporción es el que tiene mayor poder de discriminación (Szymanska, 2018). En el estudio de los sistemas metabólicos la información de interés no se limita a un metabolito específico, sino a como covarían entre sí los diferentes metabolitos presentes como, por ejemplo, cuando se desea analizar o controlar la calidad de determinados productos naturales a través de la denominada huella metabolómica (Jian et al., 2015; Villamizar, 2013; Coral Martínez, 2013; Lucio Gutiérrez, 2012). Para continuar con la utilidad del empleo de los perfiles cromatográficos cabe destacar su aplicación 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; Barakat 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 del petróleo por destilación fraccionada, 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: en el argot industrial 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 problemas de tipo forense, fundamentalmente, en la detección de adulteraciones de combustibles, 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 que se origina en las diferentes mediciones (Skov et al., 2006). La existencia de estos desplazamientos 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 la señal de mayor frecuencia 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). El filtrado 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 Savitky-Golay, esto se debe en parte a que el suavizado realizado por este filtro es muy tenue y 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 es importante: un orden bajo puede afectar la altura y el ancho de un pico, en cambio un polinomio de orden superior preserva ambas; la selección del ancho de la ventana es un parámetro también 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.

Además del ruido, otro factor que deteriora la calidad de la señal y complica la extracción de la información es la variación de la línea base. 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. Pese a la eficacia y simplicidad de este método, cuando las señales son muy complejas y presentan una gran cantidad de picos, la aplicación de una estimación por ajuste de polinomio puede resultar insuficiente. Eilers (2003) propuso una alternativa para la corrección de la línea base de señales complejas que consiste en su estimación por medio de mínimos cuadrados asimétricos (AsLs, por sus siglas en inglés). Por su rapidez este método puede ser aplicado a grandes conjuntos de señales; además cabe destacar su sencillez ya que la línea base se ajusta a partir de tres parámetros que se seleccionan según las particularidades de cada señal: orden del polinomio, criterio de convergencia y número de iteraciones (Eilers, 2004).

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 (COW, por sus siglas en inglés). La técnica COW 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 para alinear todas las muestras (Kumara, 2018). El vector usado como referencia debe ser representativo de las muestras que se quieren alinear: una opción es utilizar como referencia la media de los cromatogramas, o un cromatograma representativo del conjunto. Los parámetros que se ajustan para la aplicación de este método de alineación 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. Una vez definido el tamaño del segmento, las señales de la referencia y de las muestras se subdividen acorde a la longitud del segmento seleccionado; y de forma iterativa, se expanden o comprimen según el grado de flexibilidad propuesto. El objetivo final es el de maximizar la correlación entre las muestras y el cromatograma de referencia (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 (Trung & Laukens, 2013). Además de la comparación visual, Kai-Tal et al., (2006) y Peiyan et al., (2018) utilizan con efectividad el análisis por componentes principales para lo cual tienen en cuenta la varianza explicada, así como la semejanza en la agrupación de las muestras en función de los cambios realizados.

El objetivo de este trabajo consistió en desarrollar un procedimiento para alinear los perfiles cromatográficos obtenidos por el método PIANO. Para verificar que el procedimiento propuesto resultó adecuado se utilizó el análisis por componentes principales (PCA, por sus siglas en inglés) y el método de regresión de mínimos cuadrados parciales (PLS, por sus siglas en inglés). Se compararon los resultados del modelo PCA calculado con los datos de los perfiles alineados, con los resultados del modelo PCA calculado a partir de los porcentajes de áreas de los 70 picos cromatográficos más representativos del método PIANO reportado por Fernández et al., 2019. A su vez, se utilizó el método de regresión PLS para la predicción del número de octano experimental (RON, por sus siglas en inglés) empleando ambas matrices de datos y se compararon los errores de estimación de ambos modelos en las etapas de calibración y de validación cruzada.

MATERIALES Y MÉTODOS

Muestras, análisis cromatográfico y valores del número de octano experimental

Se emplearon los cromatogramas y los valores del RON de 70 muestras de gasolinas procedentes de una refinería de la Unión Cuba Petróleo, colectadas entre 2017 y 2019. Las condiciones del análisis cromatográficos se realizaron acorde a ASTM D6729 (2014): se utilizó un cromatógrafo Agilent 7890A equipado con detector de ionización a la llama, automuestreador 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. Para la identificación y cuantificación de los picos cromatográficos se utilizó el programa Agilent GC ChemStation.

Los valores del RON de referencia se determinaron acorde a la norma ASTM D2699 (2016) en un equipo Wakesha de la serie G-45765: la reproducibilidad reportada por la citada norma es de 1,0 en el valor del RON. Los resultados del RON fueron los empleados para conformar la matriz Y utilizada en el método de regresión PLS.

Matriz X de datos con los 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 de mínimos cuadrados asimétricos AsLS; y en la alineación de los perfiles, la alineación optimizada mediante correlación COW. 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).

Matriz X de datos con los 70 picos cromatográficos del método PIANO

Esta matriz se conformó utilizando como variables los porcentajes de áreas de los 70 picos cromatográficos más significativos del método PIANO (Fernández et al., 2019): en el anexo 1 se reporta la relación de los 70 picos seleccionados. Para la transformación de la matriz $X_{70 \times 70}$ se utilizó la normalización a área unidad; y como preprocesamiento se empleó el autoescalado de las variables.

Métodos quimiométricos utilizados

Para verificar si los resultados del empleo de los perfiles cromatográficos, alineados mediante las transformaciones propuestas, conservan similitud con respecto al procedimiento que utiliza los porcentajes de áreas de los 70 picos más representativos del cromatograma, se utilizaron los métodos PCA y PLS. El número óptimo de componentes principales del PCA se seleccionó teniendo en cuenta el porcentaje de varianza explicada por cada componente. En el método de mínimos cuadrados parciales el número óptimo de variables latentes se determinó mediante el procedimiento de validación cruzada con el criterio del objeto excluido "leave one out". Para el cálculo y el análisis de los resultados de los modelos PCA y PLS se tuvo en cuenta lo descrito en la norma ASTM E2617. Los estadígrafos utilizados para la comparación de los resultados de los modelos de regresión PLS fueron la raíz cuadrada del error medio de la estimación del RON en las etapas de calibración y en validación cruzada y el coeficiente de determinación múltiple. En los cálculos se emplearon los sistemas de programas PLS_Toolbox 6.7 soportado en Matlab.

Raíz cuadrada del error medio de estimación del modelo de regresión en la etapa de calibración (RMSEC, por sus siglas en inglés).

$$RMSEC = \sqrt{\frac{\sum_{i=1}^n (y_{ref} - y_{est})^2}{n}}$$

Donde

y_{ref} son los valores de RON, y_{est} son los valores de RON estimados por el modelo de regresión PLS en la etapa de calibración y n es el número de muestras empleadas en el cálculo del modelo PLS.

Raíz cuadrada del error medio de la estimación del modelo de regresión en la etapa de validación cruzada (RMSECV, por sus siglas en inglés)

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_{ref} - y_{est})^2}{n}}$$

Donde:

y_{est} son los valores de RON estimados por el modelo de regresión PLS en la etapa de validación cruzada.

Coefficiente de determinación múltiple R^2 .

$$R^2 = SS_{reg} / S_{tot}$$

Donde:

$$SS_{reg} = \sum (y_{est} - \bar{y}_{ref})^2, S_{tot} = \sum (y_{ref} - \bar{y}_{ref})^2 \text{ y } \bar{y}_{ref} \text{ es la media de los valores de RON}$$

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 70×6554 .
- Suavizado de los datos mediante el empleo del filtro de Savitky-Golay: se utilizó una ventana de siete puntos, un polinomio de ajuste de segundo grado y la primera derivada.
- Empleo de la técnica AsLS para la corrección de la línea base: se empleó un polinomio de segundo grado, el criterio de convergencia delta fue de 1×10^{-9} , y el número de iteraciones se fijó en 100.
- Alineación de los picos mediante el uso de la técnica COW: La longitud del segmento utilizado fue de cinco, el grado de flexibilidad se fijó en 50. En la selección del cromatograma de referencia (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 (Sinkov et al., 2011). El análisis demostró que se logró una adecuada corrección de la línea base, así como de la alineación de los picos cromatográficos (Figs. 1 y 2).

Resultados del análisis por componentes principales (PCA)

Independientemente de que número de variables, así como las transformaciones y preprocesamientos realizados a las matrices X fueron muy diferentes, se observó que el porcentaje de varianza explicada por los modelos PCA fue similar en ambos casos. El modelo PCA calculado con los datos de los perfiles cromatográficos se ajustó con seis

componentes principales que explicaron el 92 % de la varianza de los datos; el calculado con los 70 picos cromatográficos se ajustó con siete componentes principales que explicaron el 91 % de la varianza de los datos (Tabla 1). Otro aspecto interesante que reflejó que la alineación realizada fue correcta está relacionado con la semejanza en la agrupación de las muestras a lo largo de los dos primeros componentes principales, que son los que explicaron el mayor porcentaje de varianza para ambas variantes de cálculo (Figs. 3 y 4).

Resultados del método de regresión por mínimos cuadrados parciales (PLS)

El modelo PLS desarrollado a partir los perfiles cromatográficos se calculó con 7 variables latentes que explicaron 90,5 % de la varianza en la matriz X y 95,4 % en la Y, mientras que el calculado con los 70 picos cromatográficos se ajustó con 6 variables latentes que explicaron el 87,3 % y 94,7 % de la variabilidad de las matrices de datos X y Y, respectivamente. La correspondencia entre los valores estimados de RON y los obtenidos por el método de referencia fue apropiada, lo cual se evidencia a través de los coeficientes de determinación múltiple (R^2) que en los dos modelos fue de 0,95 (Figs. 5, 6 y Tabla 2). Los errores residuales medios de calibración (RMSEC) para la estimación del RON fueron similares y estuvieron ambos en el entorno de la reproducibilidad del método de referencia reportado en la norma ASTM D2699 lo cual demostró un adecuado ajuste en la estimación del RON (Szymańska, 2018).

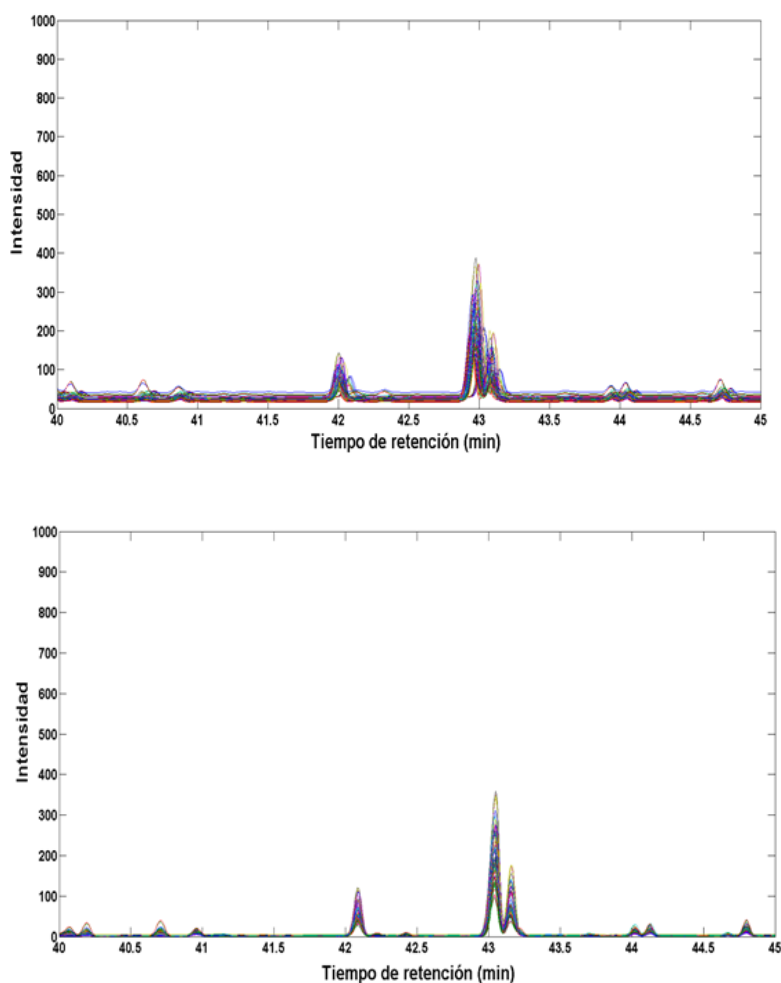


Fig. 1 Perfiles cromatográficos en el intervalo de tiempo de retención entre 40 y 45 min: A) antes de alinear, B) después de la alineación.

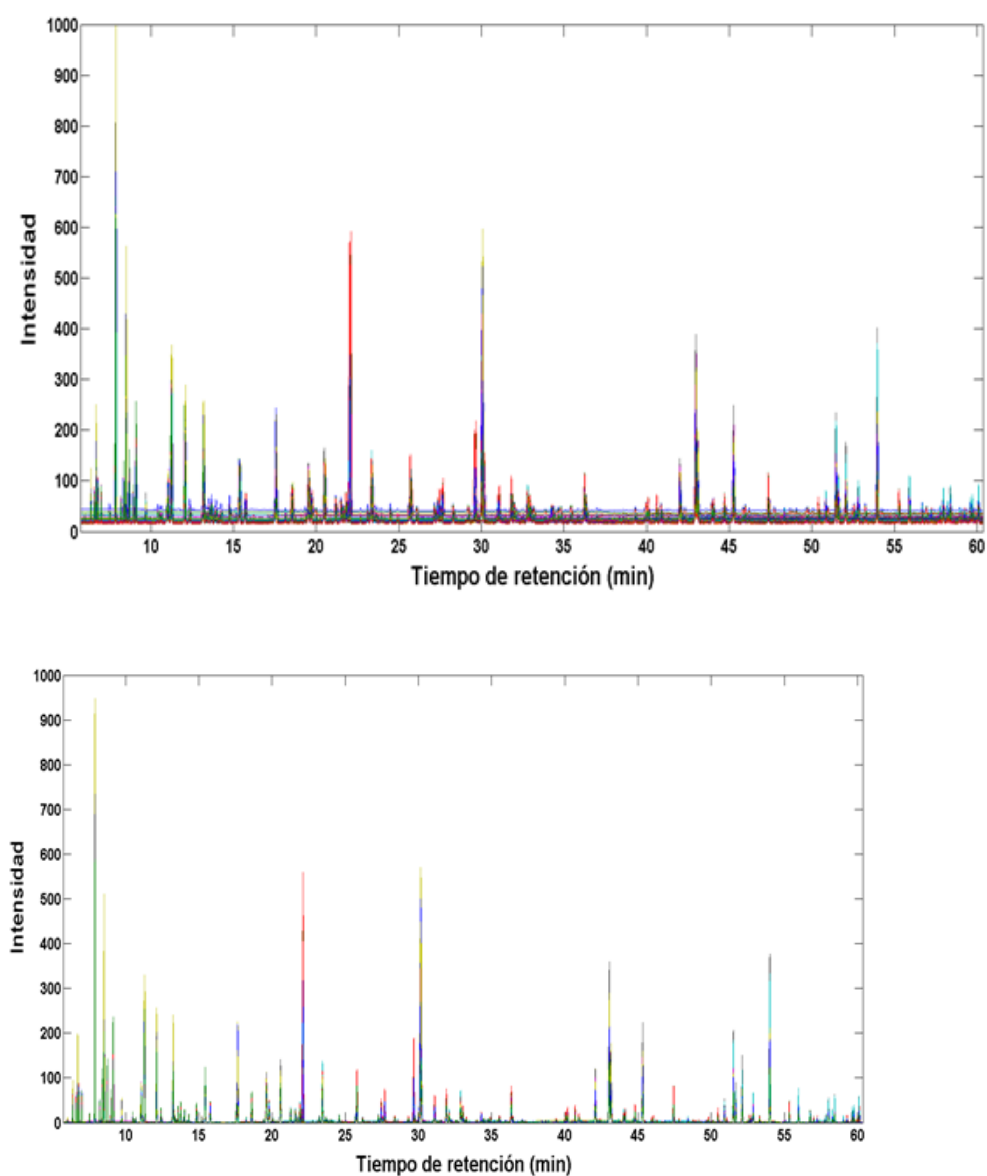


Fig. 2. Perfiles cromatográficos en todo el intervalo de trabajo: A) antes de alinear, B) después de la alineación.

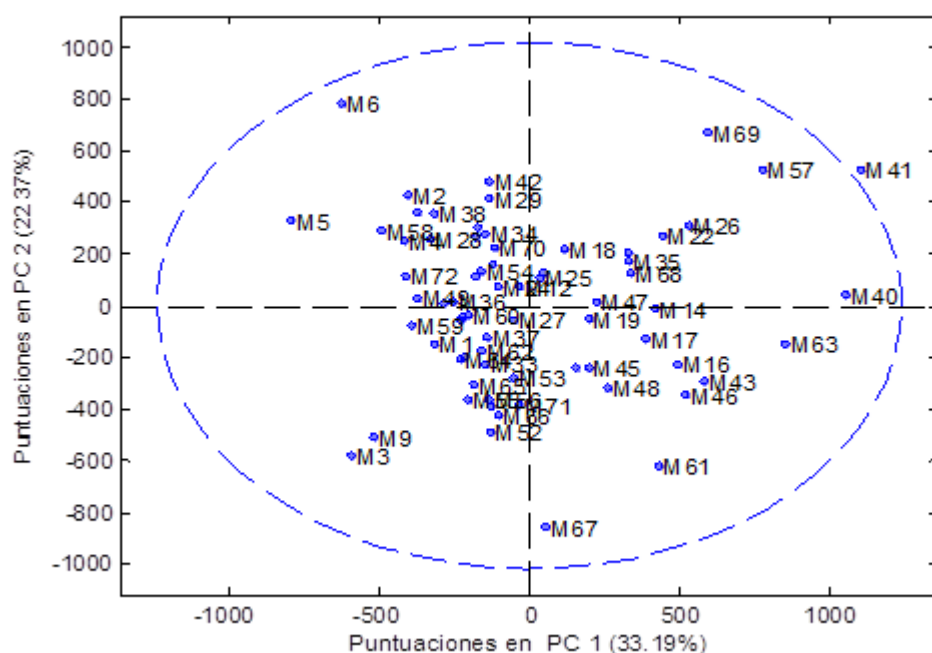


Fig. 3. Modelo PCA (con los perfiles cromatográficos). Puntuaciones sobre los dos primeros componentes principales (la varianza explicada se reporta entre paréntesis): con líneas discontinuas se destacan los límites de las puntuaciones para un nivel de significación del 5 %.

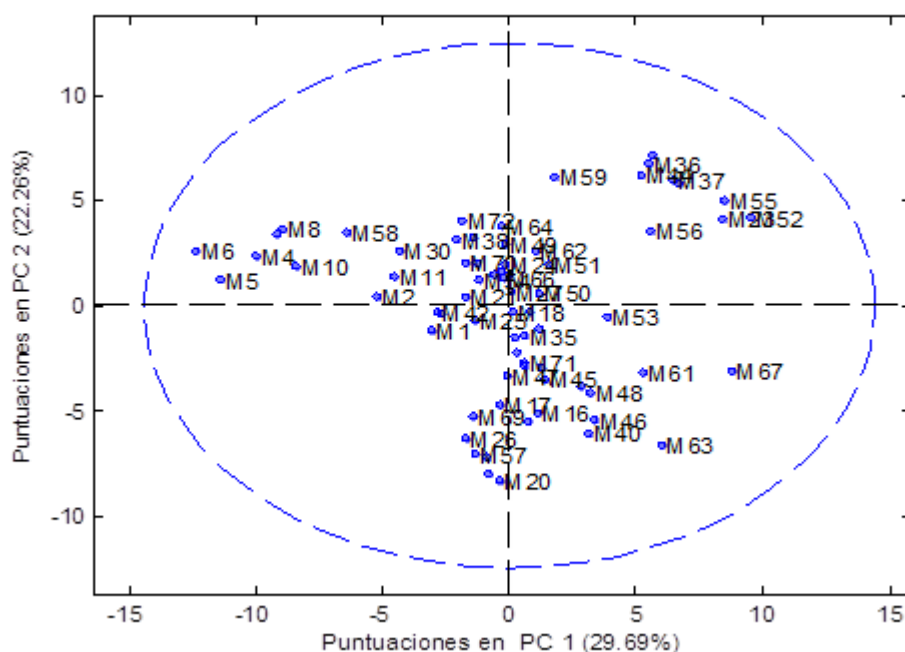


Fig. 4. Modelo PCA (con los 70 picos del cromatograma). Puntuaciones sobre los dos primeros componentes principales (la varianza explicada se reporta entre paréntesis): con líneas discontinuas se destacan los límites de las puntuaciones para un nivel de significación del 5 %.

Tabla 1. Resumen de los resultados de los modelos PCA.

Parámetro	Modelo PCA calculado con los perfiles cromatográficos alineados	Modelo PCA calculado con los porcentos de áreas de 70 picos
Dimensiones matriz X	70 x 6554	70 x 70
Transformaciones	-Filtro de Savitky-Golay -Mínimos cuadrados asimétricos (AsLS) -Alineación optimizada mediante correlación (COW)	Normalización (área 1)
Pre procesamiento	Centrado en la media	Autoescalado
Número de componentes principales	6	7
Porcentaje de varianza explicada	92	91

Tabla 2. Resumen de los resultados de los modelos PLS.

Parámetro	Modelo PLS calculado con los perfiles cromatográficos alineados	Modelo PLS calculado con los porcentos de áreas de 70 picos
Dimensiones matriz X	70 x 6554	70 x 70
Transformaciones en X	-Filtro de Savitky-Golay -Mínimos cuadrados asimétricos (AsLS) -Alineación optimizada mediante correlación (COW)	Normalización (área 1)
Preprocesamiento de X	Centrado en la media	Autoescalado
Dimensiones matriz Y (valores del RON)	70 x 1	70 x 1
Preprocesamiento de Y	Centrado en la media	Centrado en la media
Número de variables latentes	7	6
Porcentaje de varianza explicada para la matriz X	90,5	87,3
Porcentaje de varianza explicada para la matriz Y	95,4	94,7
RMSEC	1,0	1,1
RMSECV	1,7	1,5
R ²	0,95	0,95

ANEXO 1. Picos cromatográficos más significativos de las gasolinas (método PIANO)

No.	Componente	No.	Componente
1	isobutano	36	tolueno
2	<i>n</i> -butano	37	2,3-dimetilhexano
3	<i>trans</i> -2-butenos	38	2-metilheptano
4	<i>cis</i> -2-butenos	39	3-metilheptano
5	isopentano	40	3-etilhexano
6	2-metil-1-butenos	41	<i>n</i> -octano
7	<i>n</i> -pentano	42	2,4-dimetilheptano
8	<i>trans</i> -2-penteno	43	etilbenceno
9	<i>cis</i> -2-penteno	44	<i>m</i> -xileno
10	2-metil-2-butenos	45	<i>p</i> -xileno
11	2,2-dimetilbutano	46	4-metiloctano
12	ciclopentano	47	2-metiloctano
13	2,3-dimetilbutano	48	3-metiloctano
14	2-metilpentano	49	<i>o</i> -xileno
15	3-metilpentano	50	<i>n</i> -nonano
16	<i>n</i> -hexano	51	<i>n</i> -propilbenceno
17	2-metil-2-penteno	52	1-etil-3-metilbenceno
18	2,2-dimetilpentano	53	1-etil-4-metilbenceno
19	metilciclopentano	54	1,3,5-trimetilbenceno
20	2,4-dimetilpentano	55	1-etil-2-metilbenceno
21	2,2,3-trimetilbutano	56	1,2,4-trimetilbenceno
22	benceno	57	<i>n</i> -decano
23	ciclohexano	58	1,3-dietilbenceno
24	2-metilhexano	59	1-metil-3- <i>n</i> -propilbenceno
25	2,3-dimetilpentano	60	1,3-dimetil-5-etilbenceno
26	3-metilhexano	61	1,4-dimetil-2-etilbenceno
27	<i>cis</i> -1,3-dimetilciclopentano	62	1,3-dimetil-4-etilbenzeno
28	<i>trans</i> -1,3-dimetilciclopentano	63	1,2-dimetil-4-etilbenceno
29	3-etilpentano	64	1,2,4,5-tetrametilbenceno
30	2,2,4-trimetilpentano	65	1,2,3,5-tetrametilbenceno
31	<i>n</i> -heptano	66	1-etil-2-propilbenceno
32	metilciclohexano	67	1-metil-3-butilbenceno
33	2,5-dimetilhexano	68	1,2,3,4-tetrametilbenceno
34	2,4-dimetilhexano	69	naftaleno
35	2,3,4-trimetilpentano	70	<i>n</i> -dodecano

CONCLUSIONES

Se desarrolló un procedimiento para la alineación de los perfiles cromatográficos basado en tres transformaciones: minimización del ruido instrumental mediante el filtro de Savitky-Golay; corrección de la línea base mediante mínimos cuadrados asimétricos y ajuste de los perfiles por el método de alineación optimizada mediante correlación. El procedimiento se aplicó a los datos cromatográficos de muestras de gasolinas analizadas por el método PIANO. Se verificó mediante las técnicas de análisis por componentes principales y de regresión por mínimos cuadrados parciales que los resultados no difieren con los reportados por el método que utiliza los porcentajes de áreas de los 70 picos más representativos de los cromatogramas. Los resultados del análisis PCA reflejaron que las varianzas explicadas en ambos casos fueron similares: 92 % cuando se utilizaron los perfiles y 91 % cuando se emplearon los 70 picos; otro aspecto a destacar de este análisis fue la agrupación de las muestras que en ambos casos fue semejante. Los errores residuales medios (RMSEC) de los modelos PLS para la estimación del RON fueron muy similares 1,0 y 1,1; y ambos estuvieron acorde a la reproducibilidad reportada por el método ASTM de referencia. El procedimiento de alineación de los perfiles cromatográficos del análisis de gasolinas por el método PIANO, pudiera utilizarse en el desarrollo de técnicas quimiométricas de reconocimiento de patrones para la solución de problemas de tipo forense.

REFERENCIAS BIBLIOGRÁFICAS

Agilent GC ChemStation Software, version B-04-02, 2010.

ASTM D2699-16 (2016). Standard Test Method for Research Octane Number of Spark-Ignition Engine Fuel.

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.

ASTM E2617-10 (2010). Standard Practice for Validation of Empirically Derived Multivariate Calibrations.

Barakat, A. O., Mostafa, A. R., Qian, Y. R. & Kennicutt, M. C. (2002). Application of petroleum hydrocarbon chemical fingerprinting in oil spill Investigations-Gulf of Suez, Egypt. *Spill Sci. & Technol. Bull.*, 7, 229-239.

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.

Christensen, J. H. & Tomasi, G. (2007). Practical aspects of chemometric for oil spill fingerprinting. *J. Chromatogr. A*, 1169, 1-22.

Coral Martínez, T. I. (2013). Aplicación de métodos quimiométricos para establecer la huella dactilar cromatográfica de *Morinda Panamensis* y determinar su correlación con la

actividad antidiabética. Tesis Doctoral. Universidad Autónoma de Nuevo León, Facultad de Medicina.

- 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.
- Eilers Paul, H. C. (2004). Baseline correction with asymmetric least squares smoothing. *Anal. Chem.*, 76, 404-411.
- Eilers Paul, H.C. (2003). A perfect smooter. *Anal. Chem.*, 75, 3631-3636.
- 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, A., Oropesa, R., Comesaña, Y. & Romero, A. (2019). Predicción de propiedades fisicoquímicas de gasolinas mediante cromatografía gaseosa y métodos de regresión multivariados. *Av. 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.

- Jian, X., Qing-Song, X., Chi-On, C., Kam-Wah Mok, D., Lun-Zhao, Y. & Foo-Tim, C. (2015). Identifying bioactive components in natural products through chromatographic fingerprint. *Anal. Chim. Acta*, 870, 45-55.
- 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.
- Lucio Gutiérrez, J. R. (2012). Aplicación de métodos quimiométricos para la caracterización y control de calidad de plantas medicinales. Tesis Doctoral. Departamento de Química, Facultad de Ciencias, Universidad Autónoma de Barcelona.
- 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-González, D. & Talavera-Bustamante, 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.
- PLS_Toolbox for use with MATLAB. (2012). Eigenvector Research, version 6.7.1.
- Sinkov, N. A., Johnston, B. M., Sandercock, P. M. & Harynuk, J. 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 analyst (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.
- Villamizar, R. W., Mendoza, L-E. & Santafé, P. G. (2013). Implementación de técnicas de reconocimiento de patrones en procesos de selección de parámetros característicos aplicados a sistemas metabolómicos. *Revista Colombiana de Tecnologías de Avanzada*, 21(1), 104-112.
- Wiedemann, L. S. M., d'Avila, L.A. & Azevedo, D.A. (2005). Adulteration detection of Brazilian gasoline samples by statistical analysis. *Fuel*, 84:467-73.