

# **PREDICCIÓN DE PROPIEDADES FISICOQUÍMICAS DE GASOLINAS MEDIANTE CROMATOGRAFÍA GASEOSA Y MÉTODOS DE REGRESIÓN MULTIVARIADOS**

## **PREDICTION OF PHYSICOCHEMICAL PROPERTIES IN GASOLINE BY GAS CHROMATOGRAPHY AND MULTIVARIATE REGRESSION METHODS**

**Reinaldo Fernández<sup>1</sup>, Ángel Dago<sup>1</sup>, Roberto Oropesa<sup>1</sup>, Yumirka Comesaña<sup>1</sup>, Alicia Romero<sup>1</sup>**

(1) Centro de Investigación del Petróleo, Churrucá No. 481 esquina Washington, Cerro, La Habana - Cuba  
(Email: reyn@ceinpet.cupet.cu)

*Recibido: 14/01/2019- Evaluado: 22/02/2019 - Aceptado: 12/06/2019*

### **RESUMEN**

En este trabajo se desarrolla un procedimiento alternativo para la estimación del número de octano experimental y de las temperaturas de destilación atmosférica en gasolinas mediante la aplicación de métodos de regresión multivariados a partir de datos cromatográficos, que resulta de gran utilidad en el ensayo de micro actividad para la evaluación de catalizadores de craqueo catalítico fluidizado. En el análisis detallado de las gasolinas se utilizó el método PIANO. Las matrices de datos se conformaron tomando como variables el por ciento de área de los picos más representativos de los cromatogramas. En el proceso de calibración se aplicó el método de regresión de mínimos cuadrados parciales. Los errores residuales medios de los modelos en las etapas de calibración y validación cruzada resultaron adecuados. Los modelos fueron validados empleando muestras externas y los errores de predicción obtenidos estuvieron acordes a la reproducibilidad de los métodos estandarizados.

### **ABSTRACT**

In this work an alternative procedure by means of the application of multivariate regression methods for the estimation of the research octane number (RON) and the atmospheric distillation temperatures in gasoline from the chromatographic data, was developed, which is very useful in the microactivity test (MAT) for the evaluation of fluid catalytic cracking (FCC) catalysts. In the detailed analysis of the gasolines PIANO method was used. The data matrixes were conformed to taking as variables the percent of area of the more representative peaks of the chromatograms. In the calibration process the method of regression of partial least squares (PLS) was used. The mean residual errors of the models in the calibration and cross validation stages were appropriate. The models were validated using external samples and the obtained prediction errors were in agreement to the reproducibility of the standardized methods.

Palabras clave: temperaturas de destilación, número de octano experimental, modelos de regresión, calibración multivariada

Keywords: distillation temperatures, research octane number, regression models, multivariate calibration

## INTRODUCCIÓN

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, como 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 (Tanaka *et al.*, 2011; Mendes *et al.*, 2012). Según lo indicado en la norma ASTM D86-16a (2016), entre los análisis más importantes que se le realizan se encuentra la destilación atmosférica, el que nos brinda valiosa información acerca de la composición, las propiedades y el comportamiento de las gasolinas durante su uso y almacenamiento, pero este método utiliza equipos que consumen un volumen de 100 mL de muestra. Por otro lado, el índice de octano experimental (Research Octane Number, RON), que permite la evaluación de las propiedades antidetonantes de la gasolina, se determina mediante la norma ASTM D2699 (2016); este ensayo consume un volumen considerable de muestra y de reactivos (aproximadamente 1 litro de muestra y 1 litro de una mezcla de isooctano y n-heptano).

En la actualidad se han desarrollado diferentes procedimientos alternativos que emplean métodos multivariados para la predicción de propiedades físico químicas de gasolinas a partir de técnicas analíticas que utilizan pequeñas cantidades de muestras y reactivos, como la espectroscopía infrarroja con transformada de Fourier de rango medio y cercano (Balabin & Safieva, 2008; Balabin *et al.*, 2010; Dadson *et al.*, 2018; De Souza *et al.*, 2014) y la cromatografía gaseosa (Carvalho *et al.*, 2014; Cocco *et al.*, 2005; Flumignan *et al.*, 2007; Flumignan *et al.*, 2008).

El desarrollo de una metodología para la predicción de determinados parámetros físico químicos a partir de los datos cromatográficos resulta de gran utilidad en el análisis de micro actividad (microactivity test, MAT) que se realiza con el propósito de evaluar la actividad de los catalizadores de craqueo catalítico fluidizado (fluid catalytic cracking, FCC) en equilibrio o desactivados. Los productos de la reacción de craqueo obtenidos del MAT son gases no condensables, una fracción líquida (gasolina craqueada) y el carbón depositado sobre el catalizador. Las cantidades de los productos de reacción son pequeñas; en el caso de la fracción líquida es de alrededor de 2 mL, y se analiza usualmente por cromatografía gaseosa para la cuantificación de parafinas, isoparafinas, aromáticos, naftenos y olefinas (ASTM D6729, 2014). La extracción de mayor información de estos datos experimentales es de interés para los especialistas del área de catálisis y constituye el propósito de este estudio.

El objetivo del trabajo es desarrollar un procedimiento para la predicción del número de octano experimental y de las temperaturas de destilación de la fracción líquida de los ensayos de microactividad de catalizadores de craqueo catalítico mediante el empleo de los datos de cromatografía gaseosa y el método de regresión de mínimos cuadrados parciales.

## MATERIALES Y MÉTODOS

### Muestras de gasolinas objeto de análisis

Se tomaron 68 muestras de gasolinas procedentes de una refinería cubana, en el período comprendido entre los años 2015 y 2017. Teniendo en cuenta la elevada volatilidad y posible degradación de las muestras, su traslado y manipulación se llevó a cabo según establece la norma ASTM D4057 (2012). Se utilizaron viales cromatográficos de 5 ml sellados, que fueron embadurnados con varias alícuotas antes de depositar la muestra, y se almacenaron a una temperatura entre 0-5 °C.

### Análisis cromatográfico: Matriz X

Las muestras fueron analizadas mediante cromatografía de gases de alta resolución en un cromatógrafo de la firma Agilent modelo 7890A equipado con detector de ionización a la llama, automuestreador G4513A y columna capilar DB-Petro de 100 metros de longitud, 0.25 milímetros de diámetro interno y 0.5 micrómetros de espesor de película.

En el análisis detallado de las gasolinas se estableció lo descrito en la norma ASTM D6729 (2014). La estabilidad de la respuesta instrumental se verificó mediante el análisis del patrón Alphasgaz PIANO Calibration Standards Supelco 4-4594-U y del material de referencia Sigma WSPA RM-1 de gasolina California fase II. Los componentes fueron separados en la columna utilizando hidrógeno como gas portador, identificados por su tiempo de retención y cuantificados a partir del área normalizada de los picos y su factor de respuesta. Para la identificación y cuantificación de los picos cromatográficos se utilizó el programa GC ChemStation versión B-04-02, 2010. La matriz X de datos experimentales se conformó tomando como variables el por ciento de área de 70 picos representativos, los cuales aparecen representados en la Figura 1 y relacionados por su denominación en la Tabla 1.

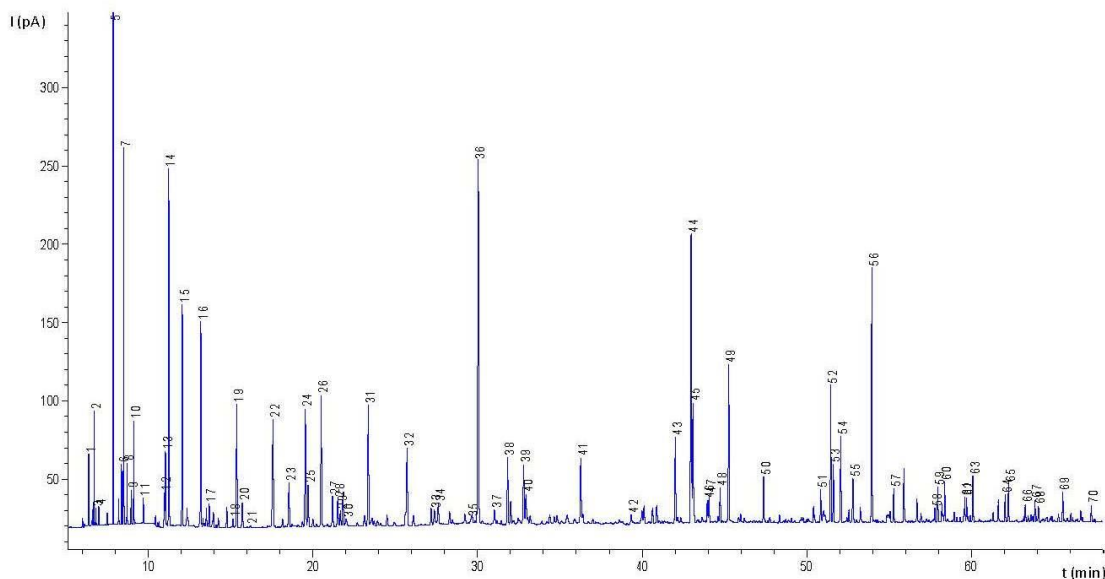


Fig. 1: Cromatograma típico de una gasolina con los picos seleccionados.

#### Determinación de los parámetros de destilación y el RON: Matriz Y

Los parámetros de destilación atmosférica utilizados como valores de referencia fueron medidos en destiladores Stanhope-Seta modelo 11860-2, según la norma ASTM D86 (2016). Los valores de temperatura fueron determinados para el punto inicial de ebullición (PIE) y el punto final de ebullición (PFE), para los que se establecen valores de reproducibilidad de 4.7 y 7.1 °C, respectivamente.

Los valores de índice de octano experimental adoptados como referencia se determinaron en un equipo Wakesha de la serie G-45765, de acuerdo con lo indicado en la norma ASTM D2699 (2016); la reproducibilidad estimada en estudios interlaboratorios es de 1.0 en el valor del RON.

#### Métodos multivariados

Se empleó el análisis por componentes principales (Principal Component Analysis, PCA) como método de reconocimiento de patrones para realizar el análisis exploratorio de los datos (Doble *et al.*, 2003; Ballabio, 2015), y el método de regresión por mínimos cuadrados parciales (Partial Least Squares, PLS) para calcular los modelos (Wentzell & Montoto, 2003; Zayas *et al.*, 2014).

Se utilizaron la normalización (área=1) y el autoescalado de las variables como métodos de preprocesamiento previo al cálculo del modelo PCA. El número óptimo de componentes principales se seleccionó teniendo en cuenta el por ciento de varianza que cada componente principal logra capturar. Se usaron los descriptores residuales Q y T<sup>2</sup> de Hotelling para definir la presencia de muestras o variables discrepantes (Brereton, 2009).

Tabla 1: Relación de los 70 picos seleccionados del análisis PIANO.

No.	Compuesto	No.	Compuesto
1	Isobutano (I)	36	Toluene (A)
2	N-Butano (P)	37	2,3-dimetilhexano (I)
3	Trans-2-buteno (O)	38	2-metilheptano (I)
4	Cis-2-buteno (O)	39	3-metilheptano (I)
5	Isopentano (I)	40	3-etilhexano (I)
6	2-metil-1-buteno (O)	41	n-Octano (P)
7	n-pentano (P)	42	2,4-DMheptano (I)
8	Trans-2-penteno (O)	43	Etilbenceno (A)
9	Cis-2-penteno (O)	44	m-Xileno (A)
10	2-metil-2-buteno (O)	45	p-Xileno (A)
11	2,2-dimetilbutano (I)	46	4-MC8+C9-olefina (I)
12	Cypentano (N)	47	2-metiloctano (I)
13	2,3-dimetilbutano (I)	48	3-metiloctano (I)
14	2-metilpentano (I)	49	o-Xileno (A)
15	3-metilpentano (I)	50	n-Nonano (P)
16	n-hexano (P)	51	n-Propilbenceno (A)
17	2-metil-2-penteno (O)	52	1-etil-3-metilbenceno (A)
18	2,2diMePentano (I)	53	1-etil-4-Mbz (A)
19	Metilcypentano (N)	54	1,3,5-triMbz (A)
20	2,4-dimetilpentano (I)	55	1-etil-2-Mbenceno (A)
21	2,2,3-trimetilbutano (I)	56	1,2,4-triMbz (A)
22	Benceno (A)	57	n-Decano (P)
23	Cyhexano (N)	58	1,3-dietilbenceno (A)
24	2-metilhexano+C7-olefina (I)	59	1-M-3-n-propilbz (A)
25	2,3-dimetilpentano (I)	60	1,3-Dimetil-5-Etilbenceno (A)
26	3-metilhexano (I)	61	1,4-Dimetil-2-Etilbz (A)
27	cis-1-3-dimetilcypentano (N)	62	1,3-DM-4-Ebz (A)
28	Trans-1-3-dimetilcypentano (N)	63	1,2-Dimetil-4-Etbz+C-1-Indano (A)
29	3-etilpentano (I)	64	1,2,4,5-tetraMbenceno (A)
30	2,2,4-trimetilpentano (I)	65	1,2,3,5-tetraMbz (A)
31	n-Heptano (P)	66	1-etil-2-propilbz (A)
32	Metilcyhexano (N)	67	1-metil-3-butilbz (A)
33	2,5-dimetilhexano+C8-olefina (I)	68	1,2,3,4,-tetraMbz+C11aromático (A)
34	2,4-dimetilhexano (I)	69	Naftaleno (A)
35	2,3,4-trimetilpentano (I)	70	n-Dodecano (P)

( ): P=parafina; I=isoparafina; A=aromático; N=nafteno; O=olefina.

Para la calibración por el método de PLS los valores de referencia se sometieron a un proceso de centrado en la media. El número óptimo de variables latentes y la detección de las muestras discrepantes se determinaron mediante el procedimiento de validación cruzada con el criterio "leave one out"; el empleo del análisis de residuales y el examen estadístico de influencia (leverage) permitieron evaluar la precisión de los valores estimados, según lo descrito en la norma ASTM E2617-09a (2009). Se fijó en un 5 % el nivel de significación de los diferentes estadígrafos utilizados en la evaluación de todos los modelos.

Los modelos de calibración obtenidos para el RON y los rangos de ebullición estudiados fueron validados empleando muestras ajenas a los mismos con valores de referencia conocidos. Como criterio general de aceptación de los modelos, se consideró que la raíz cuadrada del error medio de predicción (RMSEP) debe estar en el entorno de la reproducibilidad del método referencia utilizado, de acuerdo con lo indicado en la norma ASTM E1655 (2012). Para los cálculos se empleó el sistema de programas PLS Toolbox 6.7 (2012) soportado en Matlab 7.7 (2008).

## RESULTADOS Y DISCUSIÓN

### Análisis exploratorio de datos: Modelo PCA

La dimensión de la matriz de trabajo al iniciar el proceso de cálculo y refinamiento del modelo exploratorio fue de 68 x 70 (68 muestras y 70 variables). En el modelo preliminar se apreció que tres muestras presentaron valores altos de los estadígrafos residuales Q y T<sup>2</sup> de Hotelling, lo que revela un comportamiento diferente al del resto debido a relaciones anómalas entre sus variables. Al eliminar estos datos la dimensión de la matriz de trabajo final fue de 65 x 70.

El modelo PCA se calculó con seis componentes principales que explican el 90 % de la varianza de los datos. De acuerdo con lo citado por Brereton (2009), el apropiado nivel de agrupación que se observa en la Figura 2 nos indicó que el criterio de eliminar las muestras discrepantes fue adecuado.

Este modelo PCA nos permitió determinar la asignación de las muestras para la confección de los conjuntos de calibración y validación; también puede utilizarse para excluir muestras o datos experimentales anómalos en el proceso de actualización de los modelos de regresión.

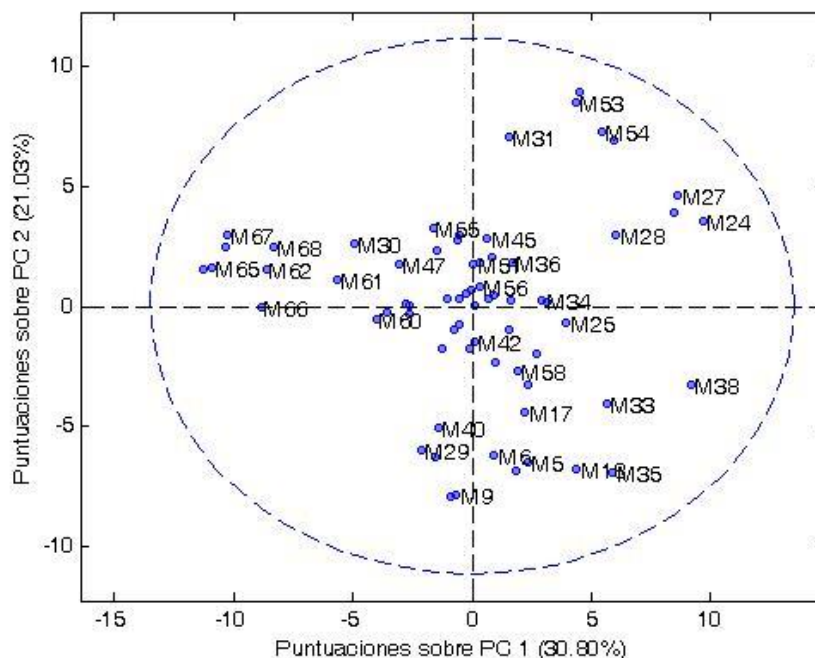


Fig. 2: Modelo PCA. Puntuaciones sobre los dos primeros componentes principales (con líneas discontinuas se destacan los límites de las puntuaciones para un nivel de significación del 5%).

### Modelos de regresión PLS

De las 65 muestras de gasolinas definidas en el análisis exploratorio, se utilizaron 45 para conformar el conjunto de calibración (matriz de 45 x 70), seleccionadas a partir de su distribución en la Figura 2, y con el criterio de que ambos conjuntos en su conformación contengan muestras de los 3 años del muestreo. Las restantes 20 muestras van a conformar la matriz de validación (20 x 70) que nos permitirá evaluar la capacidad de predicción de los modelos.

### Modelo PLS para el Punto Inicial de Ebullición (PIE).

El número óptimo de variables latentes se seleccionó acorde al criterio que la raíz media cuadrática del error en validación cruzada (RMSECV) sea un mínimo (Andrade-Garda, 2009), escogiéndose 8 variables latentes que explican un 90.6 por ciento de la varianza capturada en la matriz X y un 82.8 en la Y, lo cual indica la buena calidad del ajuste en el modelo. Se logró una adecuada concordancia entre los valores de PIE determinados por el método de referencia y los estimados por el modelo (Figura 3).

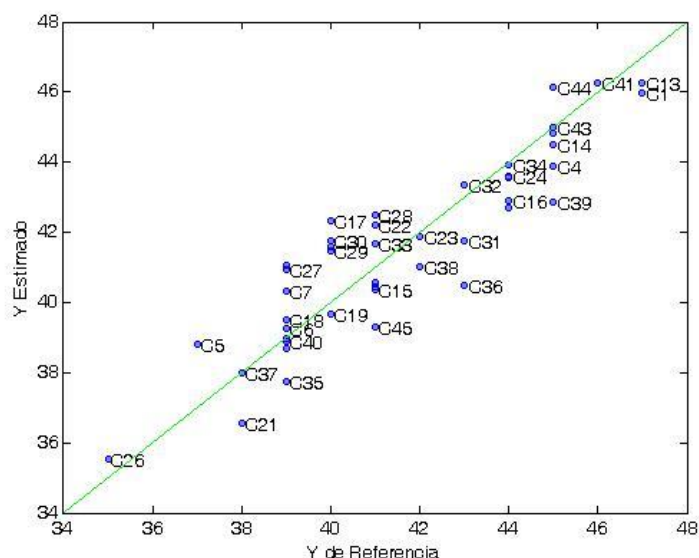


Fig. 3: Modelo PLS. Valores de referencia y estimados para el PIE.

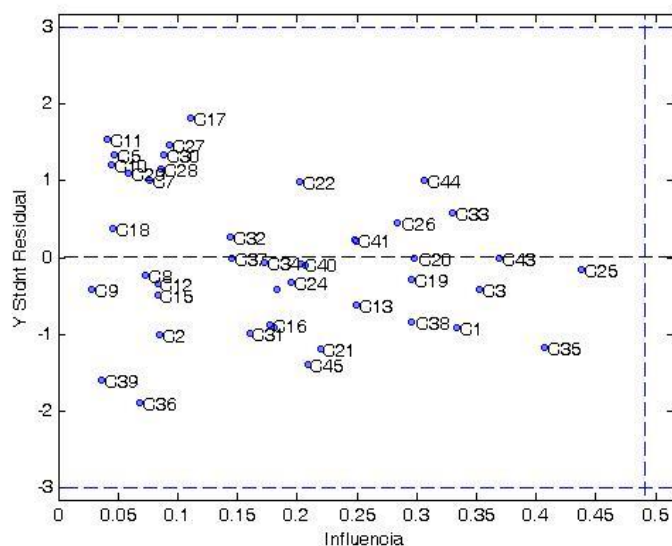


Fig. 4: Modelo PLS. Influencia vs Y Student residual (con líneas discontinuas se destacan los límites de cada uno de los estadígrafos para un nivel de significación del 5%).

En la Figura 4 se examina la influencia (leverage) que ejerce cada una de las muestras en el modelo respecto a los residuales de Student, observándose una apropiada distribución de las mismas.

Los errores residuales medios del modelo en calibración y validación cruzada fueron de 1.2 y 2.7 °C respectivamente, resultado conveniente que nos permite validar el modelo, según lo indicado por Andrade-Garda (2009).

#### Validación del modelo PLS

En la Tabla 2 se presentan los valores de temperatura determinados por el método de referencia para las 20 muestras del conjunto de validación y los estimados por el modelo PLS en predicción, así como la diferencia entre ellos. Puede observarse que esta desigualdad es menor que el valor de reproducibilidad por la norma ASTM D86 (2016) (4.7 °C) en el 100% de los casos.

Tabla 2: Modelo PLS en predicción. Valores de PIE de referencia, PIE de predicción y su diferencia.

Muestra	Valor de referencia (°C)	Valor de predicción (°C)	Diferencia (°C)
V1	41	38.0	-3.0
V2	43	43.0	0.0
V3	41	40.1	-0.9
V4	38	41.7	3.7
V5	45	41.4	-3.6
V6	42	41.6	-0.4
V7	40	40.7	0.7
V8	40	43.5	3.5
V9	36	36.4	0.4
V10	41	38.9	-2.1
V11	43	40.0	-3.0
V12	40	38.2	-1.8
V13	45	44.0	-1.0
V14	44	42.1	-1.9
V15	37	34.9	-2.1
V16	39	37.0	-2.0
V17	39	38.1	-0.9
V18	42	42.1	0.1
V19	42	41.6	-0.4
V20	38	39.3	1.3

El error residual medio de predicción RMSEP obtenido es 2.0, valor muy inferior al de la reproducibilidad del método de referencia, lo que según la ASTM E1655 (2012) es indicativo de la capacidad de predicción del modelo.

#### Modelo PLS para el Punto Final de Ebullición (PFE)

El modelo de calibración se calculó con 7 variables latentes que explican un 89.1 por ciento de la varianza capturada en la matriz X y un 87.9 en la Y. En la Figura 5 se observa la adecuada correlación entre los valores de PFE determinados por el método de referencia y los evaluados por el modelo.

En la Figura 6 se muestra la influencia que ejerce cada una de las muestras en el modelo respecto a los residuales de Student, advirtiéndose que existe una apropiada distribución.

Los errores residuales medios del modelo en calibración y validación cruzada fueron de 3.6 y 6.1 °C respectivamente, lo cual resulta satisfactorio y nos admite validar el modelo.

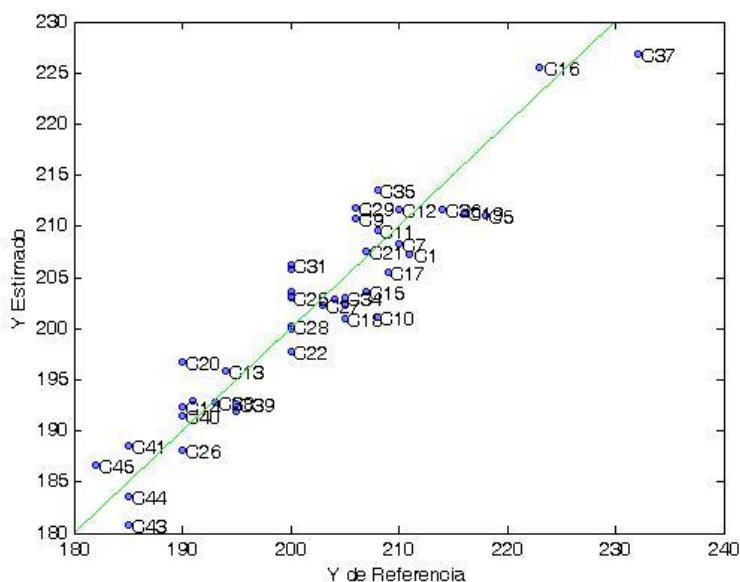


Fig. 5: Modelo PLS. Valores de referencia y estimados para el PFE.

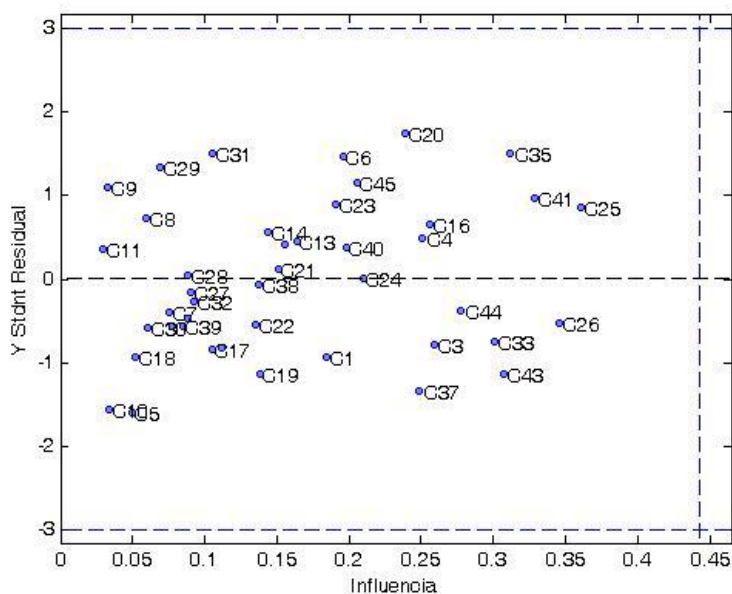


Fig. 6: Modelo PLS. Influencia vs Y Student residual (con líneas discontinuas se destacan los límites de cada uno de los estadígrafos para un nivel de significación del 5%).

### Validación del modelo PLS

Se aplica el modelo PLS en muestras ajenas al mismo, utilizando la matriz de validación de 20x70. Se muestran los valores de temperatura determinados por el método de referencia y los considerados por el modelo PLS en predicción, así como su diferencia (Tabla 3). Puede advertirse que el valor de los residuales es menor que el de la reproducibilidad del método estandarizado (7.1 °C) en el 95% de los casos (excepto para V10).



Tabla 3: Modelo PLS en predicción. Valores de PFE de referencia, PFE de predicción y su diferencia.

Muestra	Valor de referencia (°C)	Valor de predicción (°C)	Diferencia (°C)
V1	188	187.2	-0.8
V2	191	191.7	0.7
V3	202	195.8	-6.2
V4	216	213.1	-2.9
V5	203	205.1	2.1
V6	196	194.9	-1.1
V7	200	206.4	6.4
V8	219	221.7	2.7
V9	220	222.9	2.9
V10	203	194.8	<b>-8.2</b>
V11	197	197.1	0.1
V12	191	187.9	-3.1
V13	185	192.0	7.0
V14	195	195.6	0.6
V15	216	221.5	5.5
V16	215	215.7	0.7
V17	218	219.8	1.8
V18	210	214.6	4.6
V19	205	202.8	-2.2
V20	212	211.7	-0.3

El error residual medio de predicción RMSEP obtenido es 4.4, valor inferior al de la reproducibilidad del método de referencia, lo que demuestra la calidad del modelo en predicción.

#### Modelo PLS para el RON

El modelo de calibración final se calculó con 6 variables latentes que explican un 87.9 por ciento de la varianza capturada en la matriz X y un 98.3 en la Y. Los resultados muestran (Figura 7) la adecuada concordancia entre los valores de RON determinados por el método de referencia y los estimados por el modelo.

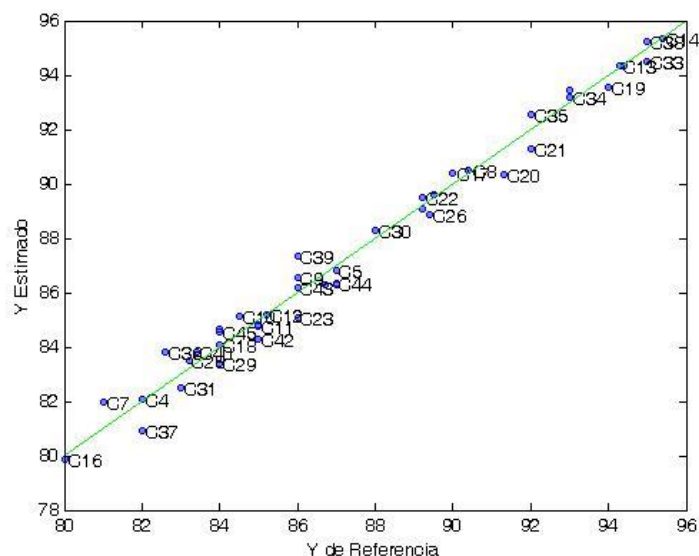


Fig. 7: Modelo PLS. Valores de referencia y estimados para el RON.

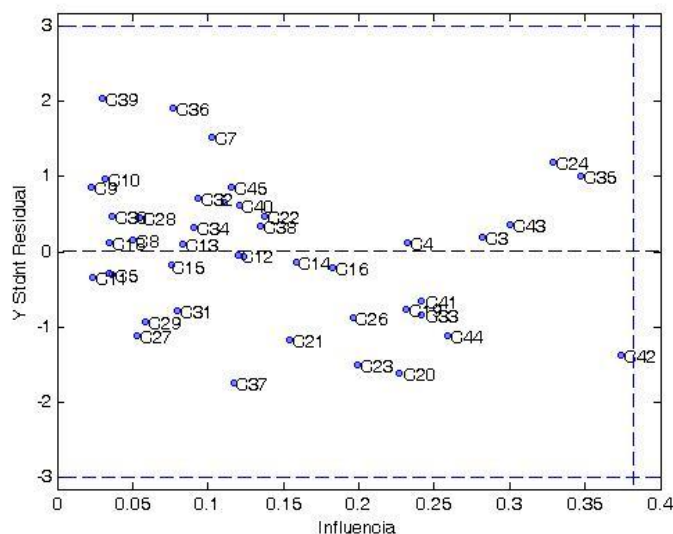


Fig. 8: Modelo PLS. Influencia vs Y Student residual (con líneas discontinuas se destacan los límites de cada uno de los estadígrafos para un nivel de significación del 5%).

En la Figura 8 se analiza la influencia que ejerce cada una de las muestras en el modelo respecto a los residuales de Student, observándose que existe una apropiada disposición de las mismas.

Los errores residuales medios del modelo en calibración y validación cruzada fueron de 0.6 y 0.9 respectivamente, lo cual resulta satisfactorio y permite efectuar la validación del modelo.

Tabla 4: Modelo PLS en predicción. Valores de RON de referencia, RON de predicción y su diferencia

Muestra	Valor de referencia	Valor de predicción	Diferencia
V1	84.5	85.2	0.7
V2	83.0	82.3	-0.7
V3	90.0	90.3	0.3
V4	91.0	91.9	0.9
V5	84.0	84.9	0.9
V6	83.0	82.9	-0.1
V7	83.3	84.3	1.0
V8	92.0	92.4	0.4
V9	95.0	94.9	-0.1
V10	80.0	80.2	0.2
V11	93.4	93.6	0.2
V12	91.0	90.3	-0.7
V13	94.5	95.1	0.6
V14	94.0	93.1	-0.9
V15	85.0	86.2	<b>1.2</b>
V16	91.5	92.1	0.6
V17	84.2	84.1	-0.1
V18	92.0	92.8	0.8
V19	88.0	88.2	0.2
V20	93.0	93.6	0.6

## Validación del modelo PLS

Se emplea el modelo PLS en muestras externas, utilizando la matriz de validación. Se tabulan los valores de RON determinados por el método de referencia y los considerados por el modelo PLS en predicción (Tabla 4); puede observarse que su diferencia es menor que el valor de reproducibilidad por el método referenciado (1.0) en el 95% de los casos (excepto para V15).

El error residual medio de predicción RMSEP obtenido es 0.9, valor en el entorno de la reproducibilidad del método de referencia, lo que revela la calidad del modelo en predicción.

A manera de resumen del trabajo, en la Tabla 5 se muestran los resultados de los modelos para el RON y los rangos de temperaturas de ebullición estudiados.

Los modelos desarrollados para las temperaturas de destilación PIE y PFE y el RON fueron aceptados al tomar como criterio que el error estándar de predicción RMSEP se encuentre en el entorno de la reproducibilidad del método ASTM de referencia.

Tabla 5: Resultados de los modelos de regresión PLS.

Modelo	Número de VL	RMSEC	RMSECV	RMSEP	R <sub>ASTM</sub>	Aceptación
PIE	8	1.2	2.7	2.0	4.7	Si
PFE	7	3.6	6.1	4.4	7.1	Si
RON	6	0.6	0.9	0.9	1.0	Si

## CONCLUSIONES

Se desarrollaron modelos de regresión multivariados para la predicción de propiedades físico químicas de la fracción líquida de los ensayos del MAT de catalizadores de craqueo catalítico a partir de los datos de cromatografía gaseosa y el método PLS.

Los resultados de los modelos de regresión para los rangos de temperaturas de destilación PIE y PFE, así como para el RON, fueron adecuados. El RMSEP del RON fue de 0.9 y para las temperaturas de destilación PIE y PFE fueron de 2.0 y 4.4 °C, respectivamente.

## REFERENCIAS

Andrade-Garda, J.M. (2009). *Basic Chemometric Techniques in Atomic Spectroscopy*. En RSC Analytical Spectroscopy Monographs No. 10 (pp. 181-243): The Royal Society of Chemistry, Cambridge, UK.

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

ASTM D4057-12 (2012). *Standard Practice for Manual Sampling of Petroleum and Petroleum Products*.

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 D86-16a (2016). *Standard Test Method for Distillation of Petroleum Products and Liquid Fuels at Atmospheric Pressure*.

ASTM E1655-12 (2012). *Standard Practices for Infrared Multivariate Quantitative Analysis*.

- ASTM E2617-09a (2009). Standard Practice for Validation of Empirically Derived Multivariate Calibrations.
- Balabin, R. & Safieva, R. (2008). Gasoline classification by source and type based on near infrared (NIR) spectroscopy data. *Fuel*, *87*, 1096–1101.
- Balabin, R., Safieva, R. & Lomakina, E. (2010). Gasoline classification using near infrared (NIR) spectroscopy data: Comparison of multivariate techniques. *Analytica Chimica Acta*, *671*, 27-35.
- Ballabio, D. (2015). A MATLAB toolbox for Principal Component Analysis and unsupervised exploration of data structure. *Chemometrics and Intelligent Laboratory Systems*, *149*, 1–10.
- Brereton, R. (2009). *Chemometrics for Pattern Recognition*, Cap. 3, pp. 47-106. John Wiley & Sons Ltd, UK.
- 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.
- Cocco, L., Yamamoto, C. & Von Meien, O. (2005). Study of correlations for physicochemical properties of Brazilian gasoline. *Chemometrics and Intelligent Laboratory Systems*, *76*, 55-63.
- 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.
- De Souza, L., Mitsutake, H., Caixeta, L. & Borges Neto, W. (2014). Quantification of residual automotive lubricant oil as an adulterant in Brazilian S-10 diesel using MIR spectroscopy and PLS. *Fuel*, *130*, 257-262.
- 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.
- Flumignan, D.L., Ferreira, F., Tininis, A., & De Oliveira, J.E. (2008). Multivariate calibrations in gas chromatographic profiles for prediction of several physicochemical parameters of Brazilian commercial gasoline. *Chemometrics and Intelligent Laboratory Systems*, *92* (1), 53-60.
- 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. *Analytica Chimica Acta*, *595*, 128-135.
- MATLAB: The Language of Technical Computing. (2008). The Math Works, version 7.7.0.471. <http://www.mathworks.com>.
- Mendes, G., Aleme, H.G. & Barbeira, P. (2012). Determination of octane numbers in gasoline by distillation curves and partial least squares regression. *Fuel*, *97*, 131-136.
- PLS Toolbox for use with MATLAB. (2012). Eigenvector Research, version 6.7.1. <http://www.eigenvector.com>.
- Tanaka, G.T., De Oliveira Ferreira, F., Ferreira da Silva, C.E., Flumignan, D.L. & De Oliveira, J.E. (2011). Chemometrics in fuel science: demonstration of the feasibility of chemometrics analyses applied to physicochemical parameters to screen solvent tracers in Brazilian commercial gasoline. *Journal of Chemometrics*, *25*, 487-495.
- Wentzell, P.D. & Montoto, V.L. (2003). Comparison of principal component regression and partial least squares regression through generic simulations of complex mixtures. *Chemometrics and Intelligent Laboratory Systems*, *65*, 257-279.
- Zayas, E., Lorenzo, M., Hernández, D. & Blanco, I. (2014). Clasificación multivariante de roncs añejos cubanos. *Revista Cubana de Ingeniería*, *5* (2), 62-67.