

**MODELOS QUIMIOMÉTRICOS A PARTIR DE CARACTERIZACIONES
ESPECTROSCÓPICAS FT-NIR PARA EL BLENDING DE GASOLINAS EN LA
GRB TENIENDO EN CUENTA NÚMERO DE OCTANO INVESTIGACIÓN,
NÚMERO DE OCTANO MOTOR, CONTENIDO DE AROMÁTICOS Y BENCENO**

HERVIN DELGADO REMOLINA

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍA QUÍMICA
ESCUELA DE INGENIERÍA QUÍMICA
MAESTRÍA EN INGENIERÍA QUÍMICA
BUCARAMANGA**

2019

**MODELOS QUIMIOMÉTRICOS A PARTIR DE CARACTERIZACIONES
ESPECTROSCÓPICAS FT-NIR PARA EL BLENDING DE GASOLINAS EN LA
GRB TENIENDO EN CUENTA NÚMERO DE OCTANO INVESTIGACIÓN,
NÚMERO DE OCTANO MOTOR, CONTENIDO DE AROMÁTICOS Y BENCENO**

HERVIN DELGADO REMOLINA

**Trabajo de aplicación presentado para optar al título de Magíster en
Ingeniería Química – Modalidad Profundización**

Director:

JULIO ANDRÉS PEDRAZA AVELLA

Doctor en Química

UNIVERSIDAD INDUSTRIAL DE SANTANDER

FACULTAD DE INGENIERÍA QUÍMICA

ESCUELA DE INGENIERÍA QUÍMICA

MAESTRÍA EN INGENIERÍA QUÍMICA

BUCARAMANGA

2019

CONTENIDO

Pág.

INTRODUCCIÓN	16
1. MARCO TEÓRICO	18
1.1 COMPOSICIÓN DE LA GASOLINA	18
1.2 PROPIEDADES DE LA GASOLINA	20
1.2.1 RON	20
1.2.2 MON	21
1.2.3 Aromáticos	21
1.2.4 Benceno	22
1.3 ESPECTROSCOPIA FT-NIR	22
1.4 MÉTODOS QUIMIOMÉTRICOS	23
1.4.1 Análisis por componentes principales (PCA)	24
1.4.2 Regresión por mínimos cuadrados parciales (PLS)	24
2. ESTADO DEL ARTE	26
3. OBJETIVOS	30
3.1 OBJETIVO GENERAL	30
3.2 OBJETIVOS ESPECÍFICOS	30
4. METODOLOGÍA	31
4.1 RECOLECCIÓN DE MUESTRAS DE GASOLINAS DE LA GRB.	31
4.2 MEDICIÓN DE ESPECTROS INFRARROJOS	31
4.3 PRETRATAMIENTOS A LOS ESPECTROS FT-NIR	32
4.4 ANÁLISIS POR COMPONENTES PRINCIPALES	33
4.5 PROPOSICIÓN Y VALIDACIÓN DE REGRESIONES	34
4.6 COMPARACIÓN DE LOS MODELOS DE REGRESIÓN OBTENIDOS	35
5. ANÁLISIS DE RESULTADOS	36

5.1 MEDICIÓN DE ESPECTROS INFRARROJOS	36
5.2 PRETRATAMIENTOS A LOS ESPECTROS INFRARROJOS	38
5.3 ANÁLISIS POR COMPONENTES PRINCIPALES	40
5.4 MODELO DE PREDICCIÓN PARA EL RON	43
5.5 MODELO DE PREDICCIÓN PARA EL MON.....	51
5.6 MODELO DE PREDICCIÓN PARA AROM.....	58
5.7 MODELO DE PREDICCIÓN PARA BEN	66
5.8 RESUMEN DE LAS REGRESIONES PLS	72
6. CONCLUSIONES	75
7. RECOMENDACIONES.....	77
BIBLIOGRAFÍA.....	78
ANEXOS.....	84

LISTA DE FIGURAS

Pág.

Figura 1. Espectros FT-NIR obtenidos para las muestras de gasolina de la GRB.	37
Figura 2. Pretratamientos aplicados a los espectros FT-NIR de las muestras de gasolina.	39
Figura 3. Gráfica en los ejes de los <i>scores</i> y de las propiedades.	41
Figura 4. Gráficas de los <i>loadings</i> para los tres primeros componentes principales. PC1, 49%; PC2, 18%, PC3, 6%.....	42
Figura 5. Espectros FT-NIR rango entre 4400–4000 cm^{-1}	43
Figura 6. Gráficas de distribución de los valores de RON para las muestras de gasolinas de la GRB.	44
Figura 7. Gráficas de los coeficientes de regresión para el RON con sus respectivos intervalos al 95% de confianza.	46
Figura 8. Curva de la raíz del error cuadrático medio de predicción (RMSEP) mostrando el número de variables latentes adecuado (2) en la regresión PLS de la propiedad RON.	47
Figura 9. Valores predichos versus medidos para la regresión PLS del RON con 2 variables latentes. Figura de la izquierda corresponde a la etapa de calibración, mientras que la figura de la derecha corresponde a la etapa de validación cruzada “LOO”. R^2 validación 0.932; RMSEP=0.618.	47
Figura 10. Residuales reportados por la regresión PLS del RON con 2 variables latentes.	49
Figura 11. Gráfica <i>QQ plot</i> comparando las distribuciones dadas por las muestras y la regresión PLS para el RON.....	50
Figura 12. Gráficas de distribución del MON para las muestras de gasolinas de la GRB.	52

Figura 13. Coeficientes de regresión con sus intervalos de confianza (95%) para los correspondientes números de onda en el modelo final de regresión para el MON.	54
Figura 14. Curva de la raíz del error cuadrático medio de predicción (RMSEP) mostrando el número de variables latentes adecuado (1) en la regresión PLS de la propiedad MON.....	55
Figura 15. Valores predichos versus medidos para la regresión PLS del MON con una variable latente. Figura de la izquierda corresponde a la etapa de calibración, mientras que la figura de la derecha corresponde a la etapa de validación cruzada “LOO”. R^2 validación 0.914; RMSEP=0.372.	56
Figura 16. Residuales reportados por la regresión PLS del MON con una variable latente.	56
Figura 17. Gráfica <i>QQ plot</i> comparando las distribuciones dadas por las muestras y la regresión PLS para el MON.	57
Figura 18. Gráficas de distribución de los valores de AROM para las muestras de gasolinas de la GRB.	60
Figura 19. Coeficientes de regresión con sus intervalos de confianza (95%) para los correspondientes números de onda en el modelo final de regresión para el AROM.	62
Figura 20. Curva de la raíz del error cuadrático medio de predicción (RMSEP) mostrando el número de variables latentes adecuado (3) en la regresión PLS de la propiedad AROM.	63
Figura 21. Valores predichos versus medidos para la regresión PLS del AROM con tres variables latentes. Figura de la izquierda corresponde a la etapa de calibración, mientras que la figura de la derecha corresponde a la etapa de validación cruzada “LOO”. R^2 validación 0.904; RMSEP=0.71.....	63
Figura 22. Residuales de la regresión PLS del AROM con tres variables latentes....	65
Figura 23. Gráficas de distribución de los valores de BEN para las muestras de gasolinas de la GRB.	67

Figura 24. Coeficientes de regresión con sus intervalos de confianza (95%) para los correspondientes números de onda en el modelo final de regresión para BEN...	68
Figura 25. Curva de la raíz del error cuadrático medio de predicción (RMSEP) mostrando el número de variables latentes adecuado (5) en la regresión PLS de BEN.	69
Figura 26. Valores predichos versus medidos para la regresión PLS de BEN con 5 variables latentes. Figura de la izquierda corresponde a la etapa de calibración, mientras que la figura de la derecha corresponde a la etapa de validación cruzada “LOO”. R^2 validación 0.919; RMSEP=0.031.	70
Figura 27. Residuales reportados por la regresión PLS del BEN con 5 variables latentes.	71

LISTA DE TABLAS

Pág.

Tabla 1. Requerimientos de calidad para el tipo de gasolina comercial Resolución 1180 de 21 de junio de 2006.	19
Tabla 2. Coeficientes de los números de onda para la predicción del RON. Término independiente 96.42.	45
Tabla 3. Coeficientes de los números de onda para la predicción del MON. Término independiente 92.48.	53
Tabla 4. Coeficientes de los números de onda para la predicción de AROM. Término independiente 18.32 %.	61
Tabla 5. Coeficientes de los números de onda para la predicción de BEN. Término independiente -0.9729 %.	68
Tabla 6. Modelos finales de regresión para las propiedades RON, MON, AROM y BEN.	73

LISTA DE ANEXOS

Pág.

ANEXO A. Instrucciones utilizadas en R para el ajuste de los modelos de regresión por PLS.	84
--	----

RESUMEN

TÍTULO: MODELOS QUIMIOMÉTRICOS A PARTIR DE CARACTERIZACIONES ESPECTROSCÓPICAS FT-NIR PARA EL BLENDING DE GASOLINAS EN LA GRB TENIENDO EN CUENTA NÚMERO DE OCTANO INVESTIGACIÓN, NÚMERO DE OCTANO MOTOR, CONTENIDO DE AROMÁTICOS Y BENCENO*

AUTOR: HERVIN DELGADO REMOLINA**

PALABRAS CLAVES: Gasolinas, RON, MON, aromáticos, benceno, PLS, regresión.

DESCRIPCIÓN

Dentro de los parámetros de calidad para la preparación de las gasolinas, según la Resolución 1180 de 2006, se encuentran: Número Octano de Investigación (RON), Número Octano Motor (MON), contenido de Aromáticos (AROM) y contenido de Benceno (BEN). La determinación de estas propiedades en laboratorio puede tardar hasta 8 horas, lo cual impide un ajuste inmediato del proceso en caso de gasolinas fuera de especificaciones. Las gasolinas sin especificaciones de calidad son reprocesadas, retrasando la producción y disminuyendo el margen de refinación. En este trabajo se desarrollaron regresiones por quimiometría (PLS) basadas en los espectros NIR para la predicción de las propiedades RON, MON, AROM y BEN de las gasolinas de la Gerencia Refinería Barrancabermeja (GRB). Según los resultados, el análisis de las bandas del espectro infrarrojo condujo a la identificación de la gasolina importada como una muestra con características diferentes a las gasolinas nacionales, especialmente en el rango $4400\text{--}4000\text{ cm}^{-1}$. Respecto a los modelos PLS, el modelo para el RON involucró un total de 18 números de onda entre $4239\text{--}4310\text{ cm}^{-1}$, con un valor del RMSEP comparable con la repetibilidad de la norma ASTM D-2699. Por su parte, el modelo de regresión PLS para el MON envolvió cinco números de onda en el rango $4330\text{--}4264\text{ cm}^{-1}$, con un valor de RMSEP menor que la reproducibilidad según la norma ASTM D-2700. También, los modelos de regresión para AROM y BEN involucraron 8 y 7 números de onda, respectivamente, en los intervalos $4338\text{--}4037\text{ cm}^{-1}$ y $4318\text{--}4046\text{ cm}^{-1}$, respectivamente. Los RMSEP para AROM y BEN resultaron menores que las respectivas reproducibilidades según norma ASTM D-5580. Los anteriores resultados sugieren que las regresiones son de utilidad para su aplicación en las predicciones del *blending* de gasolinas en la GRB.

* Trabajo de grado

** Facultad de Ingeniería Química. Escuela de Ingeniería Química. Maestría en Ingeniería Química. Director: Julio Andrés Pedraza Avella, Doctor en Química

ABSTRACT

TITLE: CHEMOMETRIC MODELS FOR RESEARCH OCTANE NUMBER, MOTOR OCTANE NUMBER, AROMATIC CONTENT AND BENCENE FROM FT-NIR SPECTROSCOPY FOR GASOLINE BLENDING IN GRB*

AUTHOR: HERVIN DELGADO REMOLINA**

KEYWORDS: Gasolines, RON, MON, aromatics, bencene, PLS, regression.

DESCRIPTION

Blending of gasolines must satisfy quality demanded by “Resolución 1180 de 2006” for properties such as: research octane number (RON), motor octane number (MON), aromatic content (AROM) and bencene content (BEN). Experimental measurement of these properties can last up to 8 hours, which delays prompt modifications in blending formulation for nonquality preparations. Nonquality gasolines have to be reprocessed hindering new preparations and lowering gross refining margin. In this work, chemometric regressions for RON, MON, AROM and BEN were proposed based on NIR spectra for gasolines from “Gerencia Refinería Barrancabermeja (GRB)”. According to the results, it follows that IR bands can identify imported gasolines from national gasolines, specially in the range 4400–4000 cm^{-1} . With respect to PLS models, the regression for MON involved a total of 18 wavenumbers in the interval 4239–4310 cm^{-1} , with a RMSEP value comparable with repeatability of ASTM D-2699. For MON, the PLS model required 5 wavenumbers from 4330 cm^{-1} to 4264 cm^{-1} , with a RMSEP value less than reproducibility of ASTM D-2700. Likewise, regression models for AROM and BEN involved 8 and 7 wavenumbers, respectively, in the intervals 4338–4037 cm^{-1} y 4318–4046 cm^{-1} , respectively. RMSEP for AROM and BEN were found to be lower than corresponding values for reproducibilities of ASTM D-5580. The results obtained in this work showed that regressions proposed for the aforementioned properties are suitable for predictions in the blending section of gasolines in GRB.

* Degree work

** Faculty of Chemical Engineering. School of Chemical Engineering. Master's Degree in Chemical Engineering. Director: Julio Andrés Pedraza Avella, Doctor in Chemistry.

ACRÓNIMOS Y ABREVIATURAS

AROM:	Contenido de Aromáticos
ASTM:	<i>American Society for Testing and Materials</i> (Sociedad Americana de Pruebas y Materiales)
BEN:	Contenido de Benceno
FT-NIR:	<i>Fourier Transform Near Infrared</i> (Transformada de Fourier del infrarrojo cercano).
GME:	Gasolina Motor Extra
GMR:	Gasolina Motor Regular
GRB:	Gerencia Refinería de Barrancabermeja
IA:	Índice Antidetonante
LOOCV:	<i>(Leave-One-Out Cross Validation)</i> Método de validación cruzada.
MON:	<i>Motor Octane Number</i> (Número Octano Motor)
PC:	<i>Principal Component</i> (Componentes principales)
PCA:	<i>Principal Component Analysis</i> (Análisis de Componentes principales)
PIANO:	Análisis de parafinas, isoparafinas, aromáticos, naftenos y olefinas.
PLS:	<i>Partial Least Squares</i> (Mínimos Cuadrados Parciales)
R2:	Coeficiente de determinación múltiple.
RMSEC:	<i>Root Mean Square Error of Calibration</i> (Raíz Cuadrada Media del Error de Calibración)
RMSEP:	<i>Root Mean Square Error of Prediction</i> (Raíz Cuadrada Media del Error de Predicción)
RMSEV:	<i>Root Mean Square Error of Validation</i> (Raíz Cuadrada Media del Error de Validación)
RON:	<i>Research Octane Number</i> (Número Octano de Investigación)
RTC:	<i>Real Time Clock</i> (Tiempo Real)
RVP:	<i>Reid vapor pressure</i> (Presión de vapor Reid)
SNV:	<i>Standard Normal Variate Transformation</i> (Transformación Variable

Estandar)

TBP: *True Boiling Point* (Punto de ebullición verdadero)

INTRODUCCIÓN

Debido a la alta demanda de los combustibles y a los compromisos adquiridos por ECOPETROL S.A. a nivel nacional se requiere de la preparación apropiada de la gasolina motor regular (GMR) y la gasolina motor extra (GME) con el fin de cumplir en cantidad, oportunidad y calidad según la normatividad colombiana. La mayoría de los componentes de GMR y GME: como la nafta virgen, nafta craqueada y nafta de aromáticos no cumplen los parámetros de calidad de una gasolina y sus especificaciones varían de acuerdo al tipo de crudo y condiciones de operación utilizadas en la refinación; por lo tanto, estos deben ser llevados a un proceso de *blending* o mezclado, con el fin de lograr una gasolina que pueda ser comercializada.

Dentro de los parámetros de calidad que influyen en la preparación de la GMR y la GME se encuentran¹: Índice Antidetonante, Índice de Cierre de Vapor, Azufre, Corrosión al cobre, Contenido de Gomas, Estabilidad a la Oxidación, Destilación, Contenido de Aditivos, Presión de Vapor Reid (RVP, *Reid Vapor Pressure*), Número Octano de Investigación (RON, *Research Octane Number*), Número Octano Motor (MON, *Motor Octane Number*), contenido de Aromáticos (AROM) y contenido de Benceno (BEN). Por su importancia, los parámetros considerados para el presente trabajo fueron RON, MON, AROM y BEN. Estos parámetros son medidos tanto en los componentes como en las gasolinas y su determinación puede tardar hasta 8 horas, lo cual involucra reprocesamientos que retrasan la producción, conllevan incumplimientos de entrega e impactan el margen de refinación de la Gerencia Refinería Barrancabermeja (GRB). Estos parámetros son medidos en la GRB siguiendo los protocolos descritos en las normas ASTM D2699, D2700, ASTM

¹ Resolución 1180 de 21 de junio de 2006.

D5580 o D1319 o D6729 y ASTM D5580 o D3606 – D6729, en orden según las propiedades mencionadas anteriormente.

En este trabajo se propone el desarrollo de una herramienta que permita especificar la GMR y la GME en tiempos más cortos, lo cual incrementaría la oferta de gasolinas y mejoraría la dinámica de la GRB. Esta herramienta mitigaría los inconvenientes que producen los reprocesamientos de productos fuera de especificaciones. En este sentido, un primer paso para la implementación de esta herramienta corresponde a la generación de regresiones a partir de espectros en el infrarrojo cercano por transformada de Fourier (FT-NIR, *Fourier Transform Near Infrared*) de muestras con propiedades representativas para las GMR y GME. Los espectros FT-NIR de estas muestras pueden ser analizados y procesados por medio de la quimiometría.

1. MARCO TEÓRICO

1.1 COMPOSICIÓN DE LA GASOLINA

La gasolina está constituida por una mezcla de hidrocarburos tipo parafinas, olefinas y aromáticos, principalmente. El rango de ebullición de la gasolina se encuentra entre 30 y 220°C, dependiendo del crudo de origen². En la GRB, las naftas se generan en las torres de destilación fraccionada de crudo y en las unidades de craqueo catalítico; en estas últimas, las gasolinas producidas se caracterizan por exhibir un elevado contenido de olefinas y aromáticos, lo cual conlleva a una elevada capacidad antidetonante. Por otro lado, las gasolinas que poseen una cantidad apreciable de azufre son tratadas en las unidades de hidrotratamiento; pues son corrosivos y generan ensuciamiento y desgaste en los motores de combustión interna, contribuyendo con la contaminación ambiental.

Por otra parte, una nafta liviana se encuentra constituida por compuestos desde pentanos hasta aquellos con puntos de ebullición inferior a 80°C. Esta nafta es sometida a isomerización para alcanzar niveles elevados de número de octano; el incremento en el número de octano por isomerización puede alcanzar valores entre 13 y 20 puntos. Asimismo, la gasolina de alquilación es el resultado de la reacción de *i*-buteno con propileno, butileno o pentileno, lo cual conduce a hidrocarburos ramificados elevando el número de octano del producto final.

Las naftas disponibles son mezcladas para generar gasolinas con las especificaciones comerciales. Cada gasolina tiene sus características, por lo cual la elaboración de una gasolina comercial depende de las proporciones utilizadas y de

² PRADELLE, F., BRAGA, S. L., MARTINS, A. R. F. A., TURKOVICS, F., PRADELLE, R. N. C. Gum Formation in Gasoline and Its Blends: A review. *Energy Fuels*, 29, 7753 – 7770. 2015.

la capacidad predictiva de las propiedades de mezcla. Complementariamente, una gasolina comercial puede contener componentes adicionales que mejoran las características antidetonantes, acelera la ignición, evita la tendencia a la evaporación y los niveles de ensuciamientos en el motor³. En Colombia se tienen las dos gasolinas, regular y extra, que se muestran en la siguiente tabla.

Tabla 1. Requerimientos de calidad para el tipo de gasolina comercial Resolución 1180 de 21 de junio de 2006.

PROPIEDAD	GASOLINA MOTOR REGULAR	GASOLINA MOTOR EXTRA
Índice Antidetonante, mínimo	81	87
Índice de Cierre de Vapor, máximo (kPa)	98	98
Aromáticos, máximo (%vol)	28	35
Benceno, máximo (%vol)	1,0	2,0
Azufre, máximo (%peso)	0,03	0,03
Corrosión al Cobre, 3h a 50 °C, máximo	1	1
Contenido de gomas, máximo (mg/100 mL)	5	5
Estabilidad a la oxidación, mínimo (minutos)	240	240
Destilación, 10% vol., máximo (°C)	70	70
Destilación, 50% vol., máximo (°C)	121	121
Destilación, 90% vol., máximo (°C)	190	190
Punto Final Ebullición, máximo (°C)	225	225
RVP, máximo (psia)	8,0	8,0

³ GARY, G.E.; HANDWERK, M.; KAISER, J. Petroleum Refining technology and economics, fifth edition, CRC press. 2007

También, el reformado de gasolinas es utilizado para incrementar el octanaje por medio de la conversión catalítica de parafinas a aromáticos. Algunos aromáticos tienen afinidad con el ozono generando productos contaminantes; adicionalmente, otros aromáticos poseen actividad carcinogénica. Por lo anterior, el AROM y el BEN son propiedades restrictivas para la venta de gasolinas.⁴

1.2 PROPIEDADES DE LA GASOLINA

En Colombia la Resolución 447 de 2003 del Ministerio de Ambiente regula los criterios ambientales de calidad de los combustibles líquidos utilizados en motores de combustión interna. Dentro de las propiedades más importantes para las gasolinas se encuentran:

1.2.1 RON. El nombre de esta propiedad corresponde al acrónimo del Inglés de *Research Octane Number*. El RON o número de octano de investigación mide la resistencia de la gasolina a la autoignición según las condiciones de operación del motor; un valor elevado de RON eleva el desempeño térmico de los motores⁵. El valor del RON para una gasolina se obtiene según la norma ASTM D-2699⁶, utilizando un motor especial con un costo cercano a los \$200 MUSD; el motor posee un cilindro de 611 cm³, el cual recibe la mezcla aire-combustible a una temperatura de 52°C. El motor es puesto en funcionamiento a 600 rpm, mientras que la ignición de la gasolina es medida en comparación con los combustibles *n*-heptano e *iso*-octano, a los cuales se les asignaron los valores de 0 y 100 en la escala de RON, respectivamente. De esta forma, un valor de 90 de RON indica que la gasolina

⁴ Ibíd.

⁵ NAKATA, K.; UCHIDA, D.; OTA, A.; UTSUMI, S.; KAWATAKE, K. The impact of RON on SI engine thermal efficiency. SAE Technical Paper. 2007

⁶ ASTM D2699-15. Standard test method for research octane number of spark-ignition engine fuel. ASTM International, West Conshohocken, PA; 2015.

utilizada posee un comportamiento de autoignición equivalente a una mezcla en volumen de 90% de *iso*-octano y 10% de *n*-heptano⁷. La prueba de medición del RON toma un poco más de 20 min cuando el técnico se encuentra capacitado en el tema.⁸

1.2.2 MON. El nombre de esta propiedad corresponde al acrónimo del Inglés *Motor Octane Number*. El MON o número de octano motor, así como el RON, mide la característica antidetonante o de autoignición de las gasolinas; el MON es determinado por medio de la norma ASTM D-2700⁹, utilizando el mismo motor que para la prueba del RON; en el ensayo para el MON, la mezcla aire-combustible ingresa a una temperatura de 149°C, mientras el motor es puesto en funcionamiento a 900 rpm. El valor del MON también es determinado por comparación con los combustibles *n*-heptano e *iso*-octano. la prueba del MON tiene tiempos de alrededor de 20 min de duración. El valor del MON es tomado como un índice secundario del poder antidetonante¹⁰ y su valor se encuentra por lo general entre 8 y 12 puntos por debajo del respectivo RON¹¹. Por otra parte, para considerar los valores de RON y MON, la resolución establece un valor mínimo de índice antidetonante¹² de 81 y 87 para GMR y GME, respectivamente (Tabla 1).

1.2.3 Aromáticos. El contenido de aromáticos (AROM) es una propiedad regulada, debido a que la ignición de estos compuestos genera emisiones tóxicas¹³. Sin embargo, en las refinerías se adiciona AROM a las preparaciones con el fin de

⁷ SINGH, E.; BADRA, J.; MEHL, M., & SARATHY, S. M. Chemical Kinetic Insights into the Octane Number and Octane Sensitivity of Gasoline Surrogate Mixtures. *Energy & Fuels*, 31(2), 1945–1960. 2017

⁸ DALY, S.R.; NIEMEYER, K.E.; CANNELLA, W.J.; HAGEN, C.L. Predicting fuel research octane number using Fourier-transform infrared absorption spectra of neat hydrocarbons. *Fuel*, 183, 359-365. 2016

⁹ ASTM D2700.-15 Standard test method for motor octane number of spark-ignition engine fuel. ASTM International, West Conshohocken, PA; 2015.

¹⁰ WANG, C.; PRAKASH, A.; ARADI, A.; CRACKNELL, R.; Xu, H. Significance of RON and MON to a modern DISI engine. *Fuel*, 209, 172-183. 2017

¹¹ GARY y colaboradores. Óp. Cit.

¹² El índice antidetonante se define como el valor promedio entre el MON y el RON en una gasolina.

¹³ MICYUS, N. J.; McCURRY, J. D., & SEELEY, J. V. Analysis of aromatic compounds in gasoline with flow-switching comprehensive two-dimensional gas chromatography. *Journal of Chromatography A*, 1086(1-2), 115–121. 2005

elevar el octanaje de la gasolina final. El contenido máximo de AROM de una gasolina se especifica en 28 %vol y 35 %vol, para GMR y GME, respectivamente. El AROM puede ser determinado a través de los métodos ASTM D-5580 o D-1319, los cuales utilizan la técnica de cromatografía de gases, con múltiples columnas capilares y detector por ionización de llama, para la cuantificación de benceno, tolueno y aromáticos de mayor peso. La prueba tiene una duración de aproximadamente 15 minutos por muestra.

1.2.4 Benceno. Aunque el contenido de benceno (BEN) se encuentra incluido en el valor de AROM, el BEN es de especial importancia debido a los efectos tóxicos que genera en los vapores de la gasolina; la exposición a estos vapores se presenta especialmente en las estaciones de servicio¹⁴. Por lo anterior, los límites especificados en la resolución para el BEN (Tabla 1) son exigentes, llegando máximo al 2% en la gasolina. Los métodos utilizados para determinar el BEN son los mismos que para el AROM; es decir, las pruebas ASTM D-5580 o ASTM D-3606 reportan BEN y AROM para las muestras analizadas.

1.3 ESPECTROSCOPIA FT-NIR

La espectroscopia en el FT-NIR es una técnica analítica no destructiva, rápida y económica; esta ha sido una técnica cualitativa y cuantitativa ampliamente utilizada en muchos campos como herramienta de análisis en la determinación de la calidad de los productos. Esta tecnología se basa en la medición de frecuencias de vibración de los enlaces químicos en los grupos funcionales como C-C, C-H, O-H, C=O y N-H. La información contenida en los espectros FT-NIR, es poco selectiva e influenciada por un gran número de factores fisicoquímicos y estructurales. Estas

¹⁴ PAGE, N. P., & MEHLMAN, M. Health Effects of Gasoline Refueling Vapors and Measured Exposures At Service Stations. *Toxicology and Industrial Health*, 5(5), 869–890. 1989

dificultades hacen de la interpretación de los datos del FT-NIR una tarea complicada, por lo cual el método de espectroscopia debe ser combinado con la quimiometría¹⁵. La quimiometría corresponde a una herramienta para analizar sistemas con gran cantidad de variables; en este sentido, la quimiometría aplica protocolos propios de análisis multivariado debido a la dificultad de describir y modelar sistemas químicos complejos. En los casos específicos de la espectroscopia FT-NIR, raramente se puede emplear una única longitud de onda (análisis univariado) para fines cuantitativos y cualitativos. La mayoría de los casos requieren emplear varias variables espectrales para obtener información suficiente para la solución de un procedimiento analítico dado (análisis multivariado)¹⁶.

1.4 MÉTODOS QUIMIOMÉTRICOS

Los modelos de regresión por quimiometría a partir de espectros del FT-NIR han sido aplicados a nivel industrial en las últimas décadas, lo cual muestra un interés por brindar soluciones fáciles y ágiles a los problemas más recurrentes en la industria del petróleo y gas. La construcción de un modelo matemático de predicción, basado en mediciones espectrales de un conjunto de muestras de interés puede ser guiada por la norma ASTM E-1655¹⁷. Según esta norma las muestras utilizadas para el desarrollo del modelo de calibración deben ser representativas de la variabilidad que pueda darse durante el proceso para obtener un modelo robusto. Esta norma recomienda aplicar una serie de pretratamientos a

¹⁵ BURNS, D.A. and CIURCZAK, E.W. Handbook of Near Infrared Analysis, 3e. Boca Raton: CRC Press. 2008

¹⁶ PALOU, A. Desarrollo de nuevas metodologías espectrales para el control analítico de productos y procesos petroquímicos y farmacéuticos. Bellaterra. Tesis doctoral (Programa de doctorado de química). Universidad Autónoma de Barcelona. Facultad de Ciencias. Departamento de Química. 2014

¹⁷ ASTM International. ASTM Standard E1655, Standard Practices for Infrared Multivariate Quantitative Analysis; ASTM International: West Conshohocken, PA, 2017.

los espectros, con lo cual se puede minimizar errores de medición, ruidos y contribuciones espectrales no deseadas.

Dentro de los tratamientos recomendados se tienen suavizado espectral, el cual busca filtrar matemáticamente el ruido aleatorio, y las derivadas, las cuales reducen el solapamiento de las señales y los desplazamientos de la línea base. Asimismo, la aplicación de la variable normal estándar (SNV, *standard normal variate transformation*) elimina los efectos de dispersión (*scattering*) del haz de luz, que se pueden confundir con los cambios de concentración.^{18, 19}

1.4.1 Análisis por componentes principales (PCA). Otra herramienta matemática corresponde a la reducción en la dimensionalidad del espectro utilizando el análisis de componentes principales (PCA); con PCA la variabilidad del espectro inherente a los compuestos es concentrada en un número reducido de nuevas variables constituidas por combinaciones lineales de números de onda de ciertas regiones del espectro. Los análisis de semejanzas entre diferentes muestras pueden ser realizados por medio de la aplicación del PCA. El análisis de las tendencias de estos PC conlleva a la determinación de las fuentes de variabilidad en el conjunto de datos iniciales; este análisis se basa principalmente en la inspección visual de las gráficas de los *scores* y los *loadings*.^{20, 21}

1.4.2 Regresión por mínimos cuadrados parciales (PLS). De los métodos más conocidos para construir modelos estadísticos se encuentra la regresión por mínimos cuadrados parciales (PLS, *Partial Least Squares*). PLS encuentra

¹⁸ BI, Y.M., CHU, G.H., WU, J.Z., YUAN, K.L., WU, J., LIAO, F., XIA, J., ZHANG, G.X., ZHOU, G.J. Ensemble Partial Least Squares Algorithm Based on Variable Clustering for Quantitative Infrared Spectrometric Analysis. *Chin J Anal Chem*, 43, 1086–1091. 2015.

¹⁹ BARNES, R.J., DHANOA, M.S., LISTER, S.J. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, 43, 772-777. 1989.

²⁰ BRUDZEWSKI, K., y colaboradores. Gasoline quality prediction using gas chromatography and FTIR spectroscopy: An artificial intelligence approach. *Fuel*, 85, 553-558. 2006.

²¹ KIM, M., LEE, Y.H., HAN, Ch. Real-time classification of petroleum products using near-infrared spectra. *Computers & Chemical Engineering*, 24, 513-517. 2000.

combinaciones lineales de las variables predictoras (variables latentes) con variables de respuesta. Las combinaciones pueden ser definidas asignando un peso a las variables independientes que manifiestan una correlación alta con las variables respuesta para mejorar la predicción.^{22,23} La regresión se aplica sobre las nuevas variables, eliminando así los problemas de colinealidad que puedan tener los datos originales. En los modelos PLS, se establece una representación de la matriz X en término de estas variables latentes, según:

$$X = TP^T + E$$

Donde, T representa *scores* y P^T *loadings*. La matriz X queda descompuesta en un número de variables latentes, cada una caracterizada por un vector T y un vector P^T . De esta forma, es posible representar una matriz X (dentro de un cierto error) por una matriz T con un número menor de columnas. Si se incluyen todas las variables latentes, el error es cero ($E=0$). El número de variables latentes necesarias para explicar la matriz X (dentro de un error máximo permitido) es una medida de la complejidad del modelo.²⁴

²² BRUDZEWSKI, K., y colaboradores. 2006. Op. Cit.

²³ KIM, LEE, HAN. Óp. Cit.

²⁴ BLANCO, M.; MASPOCH, S.; VILLARROYA, I.; PERALTA, X.; GONZÁLEZ, J. M.; TORRES, J. Determination of physico-chemical parameters for bitumens using near infrared spectroscopy. Anal. Chim. Acta, 434, 133–141. 2001

2. ESTADO DEL ARTE

Kelly y colaboradores²⁵, mostraron la viabilidad de predecir los parámetros de calidad de la gasolina a través de los espectros de absorción dentro de un rango de longitud de onda de 660 a 1215 nm, en esa región del espectro se observaron los grupos metil, metileno, aromáticos y olefinas. Los espectros fueron correlacionados con ayuda de estadística multivariable para parámetros de calidad como el octano. Los resultados del análisis multivariado (PLS) de los espectros de 43 muestras de gasolina dieron como resultado una ecuación que permitió obtener excelentes correlaciones con los datos obtenidos a través de métodos estándar.

Kim y colaboradores²⁶, desarrollaron una metodología para clasificar en tiempo real (RTC, *Real Time Clock*) 6 corrientes de proceso: diésel, gasolina, keroseno, gasóleo liviano, gasóleo pesado y nafta. El RTC propuesto por los autores está basado en un análisis de combinación de PCA para extraer las características más esenciales de cada corriente. El análisis y metodología de la investigación permitió construir un modelo de medición en tiempo real muy preciso donde se presentaron errores menores al 6 %.

Brudzewski y colaboradores,²⁷ presentaron un trabajo enfocado en inteligencia artificial fue realizado para correlacionar espectros infrarrojos de 45 muestras de gasolina de diferentes características, composición química y RON. Las mediciones de los datos inicialmente fueron hechas por cromatografía de gases, seguidamente de la toma de espectros. Los espectros fueron utilizados para predecir y clasificar la calidad de las gasolinas; y la matriz de datos fue procesada utilizando combinación de componentes principales y algoritmos de lógica difusa (*fuzzy*). Los datos luego

²⁵ KELLY, J.J., y colaboradores. Prediction of gasoline octane numbers from near-infrared spectral features in the range 660-1215 nm. *Analytical Chemistry*, 61, 313-320. 1989.

²⁶ KIM, LEE, HAN. Óp. Cit.

²⁷ BRUDZEWSKI, K., y colaboradores. 2006. Op. Cit.

de ser procesados fueron analizados y correlacionados utilizando redes neuronales híbridas y métodos de regresión multilíneal para predecir el RON.

Lee y colaboradores,²⁸ presentaron la importancia de la optimización del rango del Infrarrojo cercano (NIR, *Near Infrared*) para el análisis de fracciones de crudo con composiciones altamente complejas; la investigación se enfocó en el análisis del RON, RVP y TBP (*True Boiling Point*) en la gasolina. La selección del rango espectral utilizado en la regresión por PLS afecta el desempeño estadístico del modelo. La calibración de los espectros de las muestras mejora a medida que se encuentra la longitud de onda óptima. Los resultados mostraron que para el RON la longitud de onda óptima se encuentra entre 4410 – 4010 cm^{-1} para la RVP entre 4290 – 4150 cm^{-1} y la TBP entre 4460 – 4020 cm^{-1} . Encontrar el mejor rango de longitud de onda, facilitó de manera general el desarrollo de los modelos de predicción por mínimos cuadrados parciales.

Martínez y colaboradores,²⁹ aplicaron análisis comparativos de las técnicas quimiométricas para relacionar los espectros del FT-NIR; estos autores compararon resultados obtenidos por diferentes métodos quimiométricos: PLS, redes neuronales artificiales (ANN, *Artificial Neural Networks*), entre otros para predecir las propiedades críticas de la gasolina. Los resultados están enfocados en el entrenamiento y evaluación de redes neuronales.

Balabin y colaboradores,³⁰ enfocaron su estudio en la comparación de nueve métodos quimiométricos de clasificación multivariable: Análisis discriminante lineal (LDA, *Linear discriminant analysis*), análisis discriminante cuadrático (QDA,

²⁸ LEE, Y., CHUNG, H., KIM, N. Spectral range optimization for the near-infrared quantitative analysis of petrochemical and petroleum products: naphtha and gasoline. *Applied spectroscopy*, 60, 892-897. 2006.

²⁹ MARTÍNEZ, E., HUERTAS, S., MÉNEZ, H., PEÑA, J.L., ALCALDE, A. Comparison of chemometric techniques applied to near infrared spectra for a gasoline blending control. *J. Near Infrared Spectrosc.* 16, 297–303. 2008.

³⁰ BALABIN, R. M., SAFIEVA, R. Z., LOMAKINA, E. I. Comparison of linear and nonlinear calibration models based on near infrared (NIR) spectroscopy data for gasoline properties prediction. *Chemometrics and intelligent laboratory systems*, 88, 183-188. 2007.

Quadratic discriminant analysis), análisis discriminante regularizado (RDA, *Regularized discriminant analysis*), independiente suave Modelado de la analogía de clase (SIMCA, *Soft Independent Modeling of Class Analogy*), clasificación de PLS, K-vecino más cercano (KNN, *K-nearest neighbors*), máquinas vectoriales de apoyo (SVM, *Support vector machines*), redes neuronales probabilísticas (PNN, *Probabilistic neural networks*) y perceptron multicapa (ANN-MLP, *Artificial Neural Networks Multilayer perceptron*) para la clasificación de la gasolina. Todos los métodos resultaron eficaces, entre los altamente efectivos KNN, SVM y PNN seguidos de los medianamente efectivos como MLP y PLS.

He y colaboradores,³¹ desarrollaron un modelo adaptativo y de auto aprendizaje. El método está basado en dos aspectos: actualización de datos locales y actualización de pesos. El algoritmo desarrollado mejora cuando se predice en línea y es de gran utilidad en el control del *blending*.

A nivel local, los trabajos de aplicación desarrollados por Echeverría³², y Sandoval,³³ fueron dirigidos al análisis y la regresión de diferentes propiedades del diésel y de la gasolina producidas en la GRB. Echeverría analizó por PCA los parámetros contenido de biocombustible, punto chispa y temperatura de ebullición al 95% de evaporados en el diésel basado en FT-NIR; asimismo, este autor desarrolló regresiones PLS reportando desempeños elevados de $R^2=0,99$ y desviaciones estándar de 0.07 comparables con las pruebas ASTM respectivas. Por su parte, Sandoval, 2017, aplicó el análisis PCA y la regresión PLS en las propiedades contenido de gomas y estabilidad a la oxidación de las gasolinas; los desempeños

³¹ HE, Kaixun; QIAN, Feng; CHENG, Hui; DU, Wenli. A novel adaptive algorithm with near-infrared spectroscopy and its application in online gasoline blending processes. *Chemometrics and Intelligent Laboratory Systems*, 140, 117-125. 2015.

³² ECHEVERRÍA, O. Análisis y Correlación del Contenido de Biocombustibles, Punto de Chispa y T95 para el Diésel Utilizando Espectroscopía Infrarroja y Métodos Quimiométricos. Trabajo de Aplicación Final. Maestría en Ingeniería Química, UIS. 2016

³³ SANDOVAL, D. J. Modelo Predictivo para la Estabilidad a la Oxidación de la Gasolina de la Gerencia Refinería Barrancabermeja (GRB). Trabajo de Aplicación Final. Maestría en Ingeniería Química, UIS. 2017

en las predicciones realizadas por Sandoval reportan valores de $R^2=0.94$ y desviación estándar de 40 min cercana a la reproducibilidad de la norma.

La anterior revisión del estado del arte muestra que los resultados de estos métodos quimiométricos utilizados en espectros FT-NIR han conducido a reproducir los resultados de laboratorio para diferentes propiedades de las gasolinas. Lo anterior soporta el desarrollo de regresiones para los parámetros de la gasolina comercial, a partir de espectros FT-NIR y la aplicación del análisis PCA y del método PLS.

3. OBJETIVOS

3.1 OBJETIVO GENERAL

Desarrollar modelos quimiométricos útiles para la aplicación en el *blending* de GMR y GME, a partir de espectros FT-NIR, teniendo en cuenta que las mezclas cumplan con los parámetros de calidad de RON, MON, AROM y BEN.

3.2 OBJETIVOS ESPECÍFICOS

1. Conformar una base de datos de espectros FT-NIR representativa de las gasolinas producidas en la GRB, con sus respectivos valores de propiedades RON, MON, AROM y BEN.
2. Analizar la relación y la variabilidad de los espectros FT-NIR de las muestras de gasolinas producidas en la GRB por medio del análisis PCA.
3. Proponer regresiones para las propiedades RON, MON, AROM y BEN a partir de la base de datos de espectros FT-NIR y el método PLS.

4. METODOLOGÍA

4.1 RECOLECCIÓN DE MUESTRAS DE GASOLINAS DE LA GRB.

Un total de 80 muestras fueron tomadas de los tanques de gasolinas intermedias (virgen, craqueada, hidrotratada y aromática) y gasolinas comerciales producidas en la GRB durante un periodo entre enero y junio de 2018; los tanques y los días de recolección de muestras fueron seleccionados de forma aleatoria. Estas muestras fueron llevadas al Laboratorio de Calidad de la GRB para la medición de las propiedades RON, MON, AROM y BEN, según los protocolos respectivos. Los valores obtenidos fueron analizados con estadística univariada (mínimo, máximo, promedio, desviación estándar y rango intercuartil) para identificar y eliminar valores atípicos.

4.2 MEDICIÓN DE ESPECTROS INFRARROJOS

Los espectros FT-NIR de las muestras de gasolinas recolectadas fueron medidos en el rango $10000 - 4000 \text{ cm}^{-1}$ en un equipo ATR-FTIR MB3000 de ABB, localizado en el Laboratorio de Calidad de la GRB. Los datos de los respectivos espectros fueron almacenados en archivos con extensión “.spc” (*Thermo Galactic files*).

Los archivos de los espectros fueron leídos por medio de la función *read.spc* incluida en el paquete *hyperSpec* versión 0.99-20180627, codificado para su aplicación en el programa de uso libre R versión 3.4.3.³⁴

4.3 PRETRATAMIENTOS A LOS ESPECTROS FT-NIR

Los espectros FT-NIR fueron pretratados para eliminar la variabilidad en las mediciones del equipo e incrementar la relación señal/ruido³⁵ y compensar las variaciones en la intensidad de la fuente de emisión IR^{36,37}. Los pretratamientos aplicados fueron suavizados según el procedimiento de Savitzky-Golay, primera y segunda derivadas por medio de la aproximación de Savitzky-Golay, y normalización por Transformación Variable Estándar (SNV, *Standard Normal Variate Transformation*). Los pretratamientos fueron aplicados según las funciones *savitzkyGolay* y normalización SNV del paquete *prospectr* versión 0.1.3 codificado para su aplicación en R versión 3.4.3. Las tendencias reportadas por los espectros y los pretratamientos fueron comparadas visualmente por medio del gráfico en las coordenadas intensidad vs número de onda; los resultados fueron graficados

³⁴ El programa R es un ambiente para cálculos y análisis estadísticos, que se encuentra codificado sobre el sistema operativo GNU (<http://www.gnu.org>). R fue desarrollado en los laboratorios Bell (anteriormente AT&T), bajo la filosofía de desarrollo de librerías y funciones abiertas. Diferentes procedimientos de procesamiento y análisis de datos pueden ser ejecutados en R; entre estos procedimientos se tienen, análisis estadísticos univariados, análisis estadísticos multivariados, regresiones lineales múltiples, regresiones logísticas, regresiones PLS, regresiones PCR, ajustes por RNA y análisis de agrupaciones por kmean, entre otros. Existen extensiones o paquetes desarrollados por programadores de R a nivel mundial, los cuales pueden ser descargados e instalados bajo la filosofía del software libre. Asimismo, los desarrollos y las aplicaciones realizadas con base en R pueden ser publicadas en la página web The R Journal. También, los usuarios de R cuentan con libros y sitios web en donde pueden consultar y descargar rutinas, desarrollos y aplicaciones para este programa; los interesados en profundizar y aplicar el programa R cuentan con cursos libres en diferentes páginas de universidades e institutos de educación. Tomado de <https://www.r-project.org>

³⁵ RINNAN, A. Pre-processing in vibrational spectroscopy – when, why and how. *Anal. Methods*, 6, 7124–7129. 2014.

³⁶ BI, CHU, WU, YUAN, WU, LIAO, XIA, ZHANG, ZHOU, Op. Cit.

³⁷ BARNES, DHANOA, LISTER, Óp. Cit.

(intensidad vs número de onda) empleando la función *plotspc* del paquete *hyperSpec*.

4.4 ANÁLISIS POR COMPONENTES PRINCIPALES

Con los espectros FT-NIR pretratados se conformó una matriz en donde las filas o *scores* representaban las muestras de gasolina y las columnas o *loadings* los aportes en cada número de onda, a esta matriz se le aplicaron el análisis por componentes principales (PCA, *principal component analysis*)³⁸. Con los *scores* y los *loadings* de los componentes principales con mayor varianza se procedió a la detección de muestras atípicas y a la identificación de números de onda con contribuciones importantes en los espectros. La información contenida en los *scores* y los *loadings* fue obtenida principalmente por inspección visual de las respectivas gráficas de dispersión de los componentes principales (PC1 vs PC2 vs PC3), tanto en sus valores *scores* como en sus valores *loadings*. El método PCA fue aplicado según la función *prcomp* codificada para R; las gráficas de dispersión referentes a los *scores* y los *loading* resultantes de la aplicación de la función *prcomp* fueron creadas utilizando la función *scatterplot3d* del paquete *scatterplot3d* versión 0.3-41 codificado para R.

³⁸ En términos resumidos, PCA concentra la varianza de los espectros FT-NIR en unas pocas variables denominadas componentes principales; estas variables son obtenidas por medio de la descomposición de los espectros en *scores* y *loadings*, los primeros representando a las muestras, mientras que los segundos representando a los valores de intensidad pretratada en los correspondientes números de onda (ver Wold y colaboradores, 1987).

4.5 PROPOSICIÓN Y VALIDACIÓN DE REGRESIONES

La matriz de espectros FT-NIR pretratados fue complementada con las columnas correspondientes a las propiedades RON, MON, AROM y BEN. Esta matriz fue utilizada en la proposición de regresiones para cada una de las anteriores propiedades por medio de la aplicación del método PLS.³⁹ Los resultados de la aplicación de las regresiones PLS condujeron a la obtención de los respectivos coeficientes para cada número de onda con contribuciones importantes en los espectros pretratados. Los valores para estos coeficientes resultaron dependientes del número de variables latentes seleccionados, según el valor del error durante las etapas de validación; las validaciones fueron aplicadas considerando la matriz inicial con el procedimiento de descarte de una muestra a la vez; dejar uno afuera en validación cruzada (LOOCV, *leave-one-out cross validation*).⁴⁰ Esta depuración fue aplicada repetidamente con el fin de obtener modelos finales de regresión de las propiedades sin sobreajustes y con elevada capacidad predictiva.^{41, 42} Por otra parte, la depuración de las regresiones involucró la supresión de números de onda con baja contribución en el ajuste de las respectivas propiedades. Tras la supresión de números de onda con baja contribución, los procedimientos de ajuste de regresiones fueron repetidos según el método PLS anteriormente mencionado. Las variables estadísticas de desempeño de los modelos de regresión final fueron el coeficiente de determinación múltiple (R^2), la raíz del error cuadrático medio (RMSE) y confiabilidad del RMSE (RMSE%), definidos como,

³⁹ El procedimiento circunscrito en PLS descompone la variación contenida en los espectros FT-NIR y en la propiedad en turno siguiendo el camino del máximo cambio en la covarianza. Esta descomposición concentra la varianza de los espectros en pocas variables independientes denominadas variables latentes, las cuales son representadas como el producto de los respectivos scores y loadings generados por este procedimiento (ver Helland, 2001; Wold y colaboradores, 2001).

⁴⁰ LI, Baibing; MORRIS, Julián; MARTIN, Elaine B. Model selection for partial least squares regression. Chem. Int. Lab. Syst., 64, 79-89. 2002.

⁴¹ Ibid.

⁴² RINNAN, Op. Cit.

$$R^2 = 1 - \frac{\sum (y_i - \tilde{y}_i)^2}{\sum (y_i - \bar{y})^2}, \quad RMSE = \sqrt{\frac{\sum (y_i - \tilde{y}_i)^2}{n}}, \quad RMSE\% = \frac{RMSE}{\bar{y}} * 100$$

Donde, y_i , $\tilde{y}_i$, $\bar{y}$, n corresponden, respectivamente, al valor de la propiedad experimental para la muestra i -ésima, el valor obtenido con el modelo de regresión para la propiedad, el promedio muestral de la propiedad y el número de muestras utilizadas.

El método PLS fue aplicado según la función *pls* del paquete *pls* versión 2.7-0 codificado para su ejecución en R. También, los valores de los coeficientes y los valores de los residuales determinados por PLS fueron extraídos por medio de las funciones *coef* y *residuals* del paquete *pls*; la gráfica de los coeficientes fue elaborada utilizando la función *coefplot* del mismo paquete.

4.6 COMPARACIÓN DE LOS MODELOS DE REGRESIÓN OBTENIDOS

Los valores de los RMSE de las etapas de validación en cada regresión fueron utilizados en la comparación con la repetitividad y reproducibilidad establecida en las normas ASTM de cada propiedad. Asimismo, las regresiones fueron comparadas con reportes encontrados en la literatura abierta.

5. ANÁLISIS DE RESULTADOS

5.1 MEDICIÓN DE ESPECTROS INFRARROJOS

Los espectros FT-NIR de diferentes muestras de gasolinas fueron medidos para conformar la base de datos de espectros. Los 80 espectros fueron utilizados para la matriz de cálculos con 3112 columnas, correspondientes a los números de onda, y 80 filas o muestras con las absorbancias respectivas. La Figura 1 ilustra los IR medidos a las diferentes muestras de gasolinas en diferentes rangos de números de onda. Según la Figura 1a, la cual presenta el espectro en el rango de números de onda entre 10000 y 4000 cm^{-1} , la contribución espectroscópica de los constituyentes de la gasolina presenta tres zonas características para los resultados FT-NIR. La zona 1 se encuentra conformada por números de onda mayores a 6000 cm^{-1} , mientras que las zonas 2 y 3 se limitan en los números de onda entre 6000–5500 cm^{-1} y 4500–4000 cm^{-1} , respectivamente. La Figura 1b muestra en detalle la zona 1 del espectro; la zona 1 exhibe valores inferiores de 5% de la intensidad de la señal, correspondiendo a segundos sobretonos de las bandas de estiramiento del grupo C-H.^{43, 44}

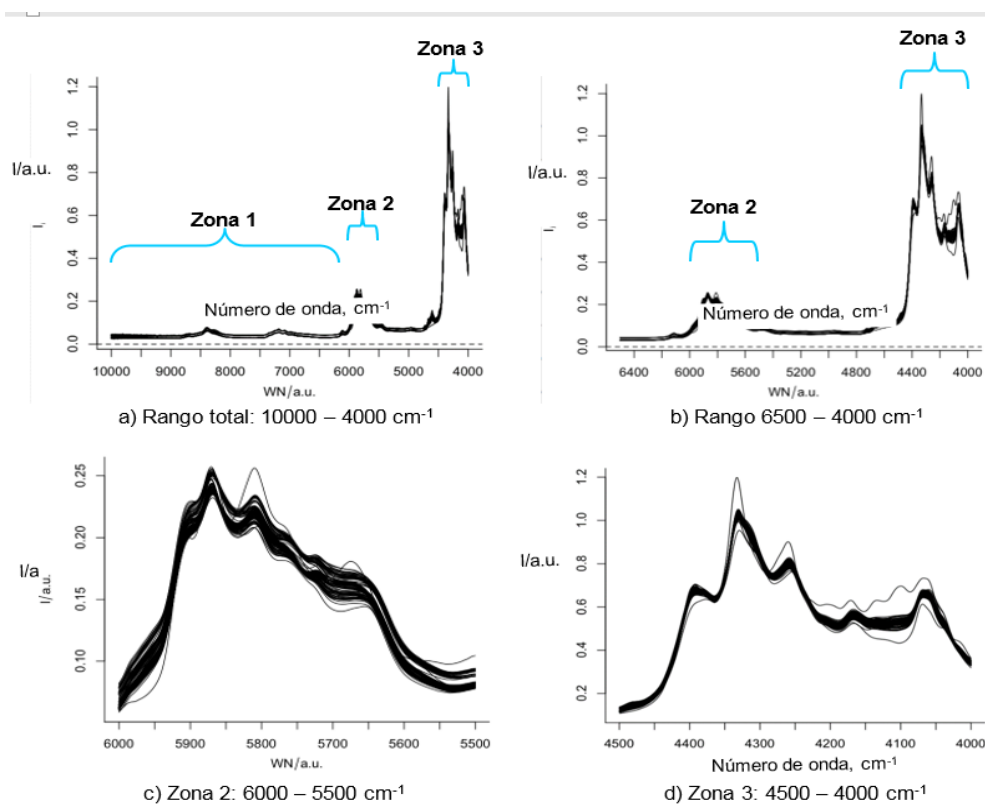
Por otra parte, las Figuras 1c-d ilustra en detalle las zonas 2 y 3, respectivamente las zonas 2 y 3 poseen las mayores contribuciones a la intensidad de la señal espectroscópica; la zona 2, 6000–5500 cm^{-1} , (Figura 1c) expone la contribución de los primeros sobretonos de las bandas de estiramiento del grupo C-H; en la zona 3, 4500–4000 cm^{-1} , (Figura 1d) se presentan combinaciones de bandas de estiramiento para el grupo funcional C-H. Tanto, las regiones de bajas

⁴³ REBOUCAS, Marcio; DE BARROS NETO, Benicio. Near infrared spectroscopic prediction of physical properties of aromatics-rich hydrocarbon mixtures, *J. Near Infrared Spectrosc.*, 9, 263–273. 2001.

⁴⁴ KARDAMAKIS, A.A. & PASADAKIS, N. Autoregressive modeling of near-IR spectra and MLR to predict RON values of gasolines. *Fuel*, 89, 158-161. 2010.

contribuciones como las regiones espectrales con intensidad importante encontradas en las muestras de gasolinas de la GRB coinciden con los aportes espectroscópicos reportados en la literatura para muestras de gasolina de diferentes refinerías a nivel mundial.^{45, 46, 47, 48}

Figura 1. Espectros FT-NIR obtenidos para las muestras de gasolina de la GRB.



⁴⁵ REBOUCAS, NETO, 2001. Óp. Cit

⁴⁶ KARDAMAKIS, A. & PASADAKIS, Óp. Cit.

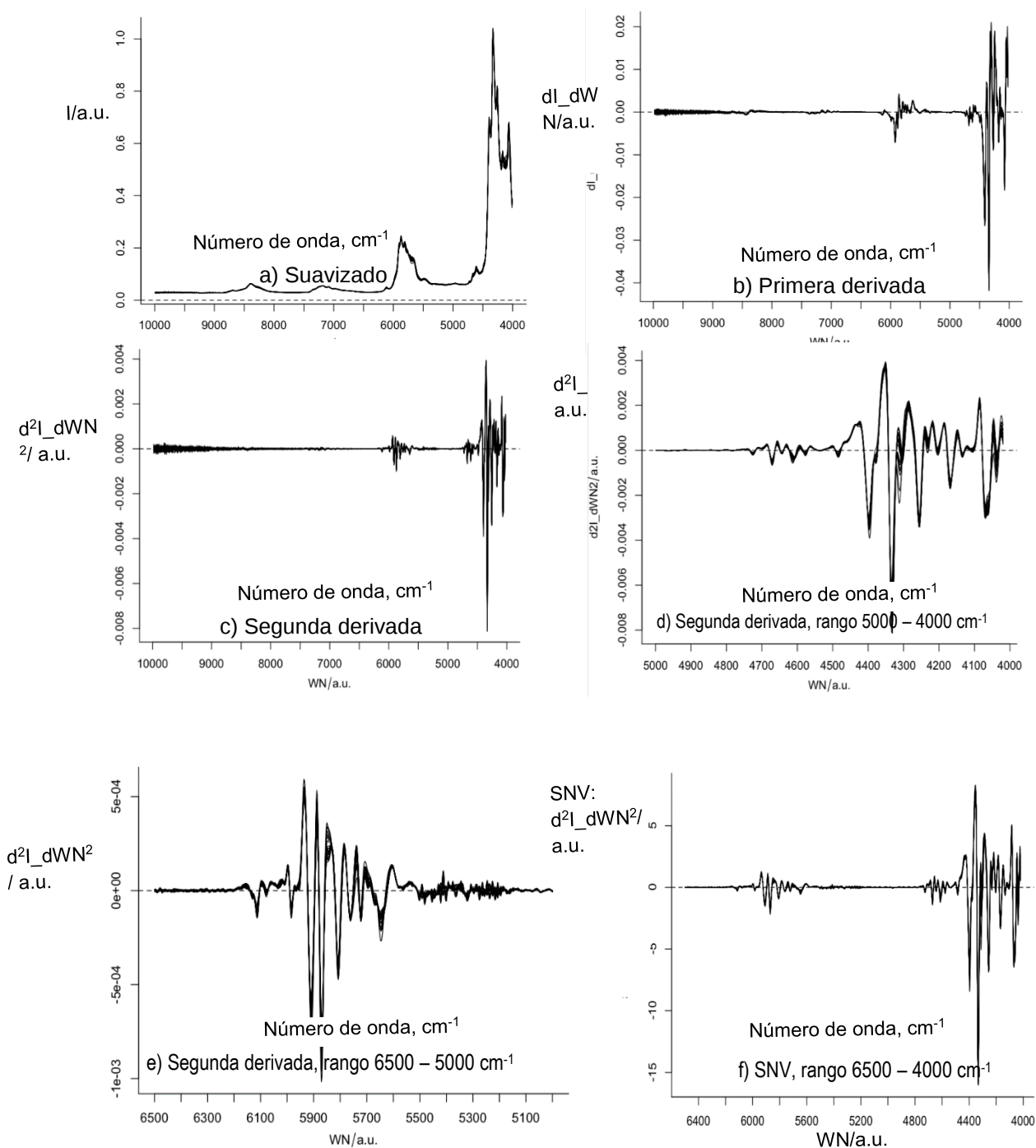
⁴⁷ REBOUCAS, NETO, 2001. Óp. Cit

⁴⁸ REBOUCAS, SANTOS, PIMENTEL, TEIXEIRA, 2011. 185–193. Op. Cit.

5.2 PRETRATAMIENTOS A LOS ESPECTROS INFRARROJOS

La Figura 2 presenta los cambios de los espectros FT-NIR con la aplicación de las diferentes etapas del pretratamiento; las Figuras 2a,b,c,d,e,f presentan los espectros con los pretratamientos de suavizado, primera derivada, segunda derivada, segunda derivada en el rango 5000–4000 cm^{-1} , segunda derivada en el rango 6500–5000 cm^{-1} y pretratamiento SNV, respectivamente. Según la Figura 2a, el suavizado aplicado a la señal representa el comportamiento espectral, disminuyendo las variaciones ocasionadas por el muestreo discontinuo. Asimismo, la Figura 2b exhibe la primera derivada aplicada al espectro suavizado; según esta figura, las contribuciones de las zonas 2 y 3 (Figura 1b-c-d) son acentuadas por medio del uso de la pendiente. Con la segunda derivada (Figura 2c,d,e), que considera las concavidades del espectro, se obtienen las contribuciones más prominentes del espectro en las regiones entre 6000 y 5500 cm^{-1} y 4500 y 4000 cm^{-1} , correspondientes a las zonas 2 y 3, (Figura 1d-e). La dispersión en los valores de las segundas derivadas de los espectros, son reducidas por medio del pretratamiento SNV (Figura 2f).

Figura 2. Pretratamientos aplicados a los espectros FT-NIR de las muestras de gasolina.



5.3 ANÁLISIS POR COMPONENTES PRINCIPALES

Los espectros pretratados fueron tomados para conformar la matriz de entrada del análisis por componentes principales (PCA); la matriz transformada en *scores* y *loadings* en PCA se encuentra constituida por las muestras, en las filas, y la absorbancia pretratada en el rango FT-NIR, en las columnas. Los resultados del análisis PCA son presentados visualmente en las Figuras 3a-b y 4, con las gráficas de la varianza explicada, de los *scores* y los *loadings*, respectivamente. La Figura 3a muestra que las contribuciones a la varianza de los tres primeros componentes principales corresponden a PC1 49%, PC2 18% y PC3 6%; con esto, los tres primeros PC constituyen el 73% de la varianza contenida en los espectros FT-NIR de las muestras de gasolina. La explicación del 73% de la varianza soporta el análisis de las tendencias de los *scores* y los *loadings* de estos tres PCs. La Figura 5.3b localiza las muestras en los ejes de los *scores* de los tres primeros PC; en esta figura, la muestra #78 exhibe valores de los tres primeros componentes principales que la apartan de las muestras restantes; la posición relativa de las muestras en la figura de los *scores* revela agrupaciones entre las mismas. Asimismo, las Figuras 3c y 3d, que presentan las muestras en el espacio de las propiedades RON vs AROM vs MON y RON vs BEN vs MON, respectivamente, confirman que la muestra de gasolina #78 contiene los menores valores de contenido de aromáticos y benceno, por lo cual se diferencia químicamente de las restantes gasolinas.

Figura 3. Gráfica en los ejes de los scores y de las propiedades.

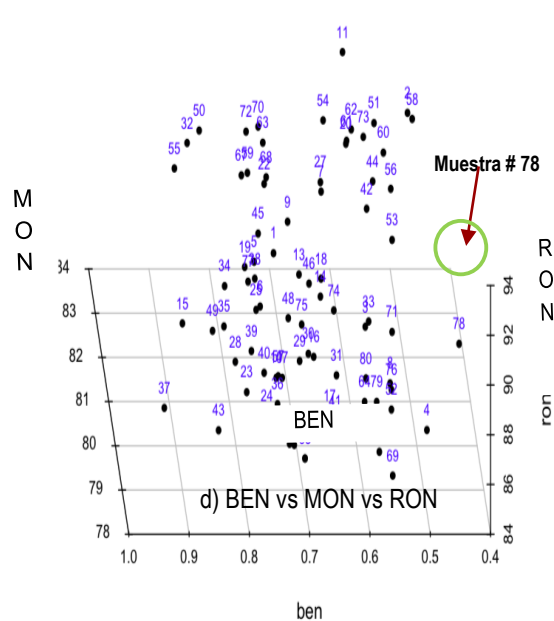
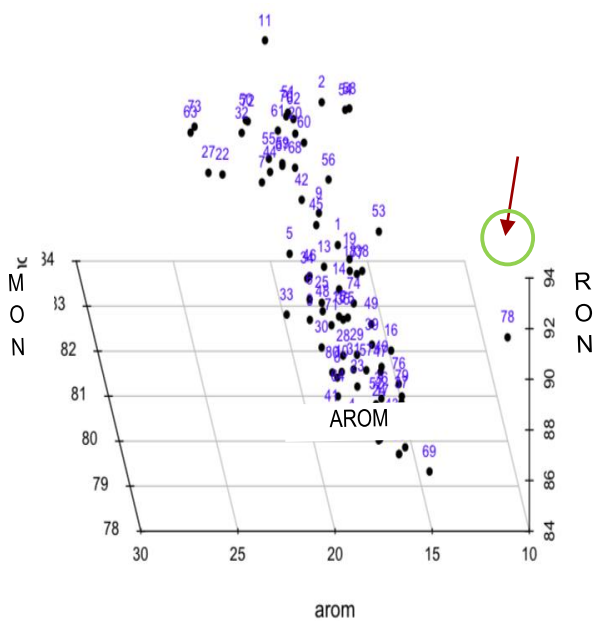
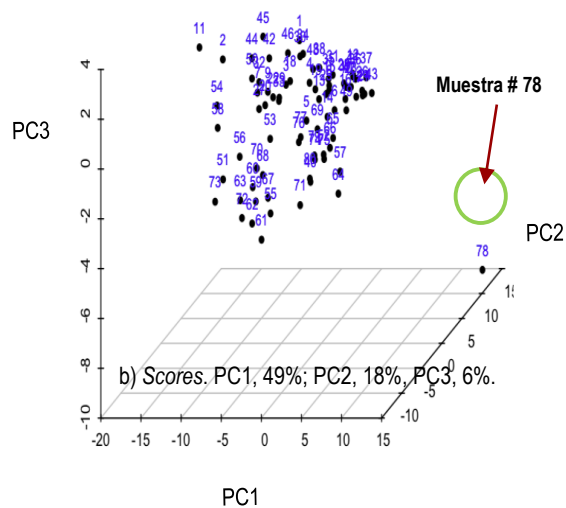
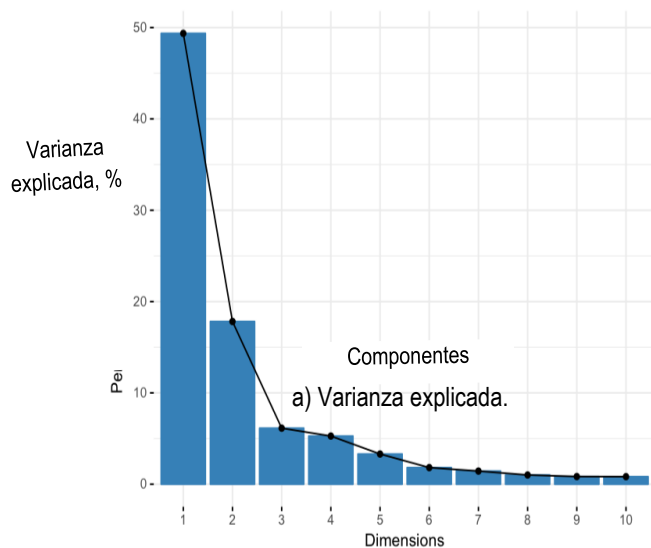
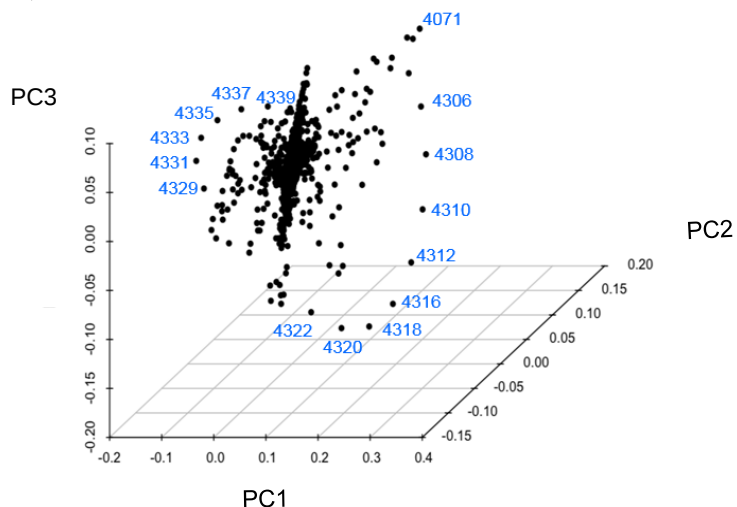


Figura 4. Gráficas de los loadings para los tres primeros componentes principales. PC1, 49%; PC2, 18%, PC3, 6%.

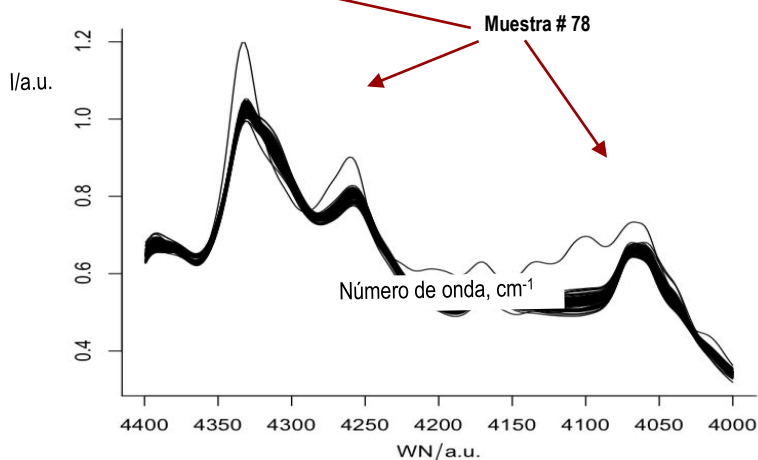


Desde la parte estadística, la muestra #78 es diferente de las otras debido al valor de los *loadings* entre 4300 y 4320 cm^{-1} (Figura 4). En la gráfica de los *loading*, la mayoría de los números de onda se agrupan en el centro de coordenadas (0,0,0) exhibiendo baja influencia en las agrupaciones mostradas en los *scores*. Los números de onda con valores importantes en los *loading* corresponden a la región entre 4300 y 4000 cm^{-1} (Figura 4). La transposición (*biplot*) entre la Figuras 3b (*scores*) y la Figura 4 (*loadings*) ubica en la misma región a la muestra #78 y a las señales espectrales entre 4300 y 4320 cm^{-1} .

La Figura 5 presenta los espectros de las 80 muestras en el intervalo entre 4400 y 4000 cm^{-1} . En la Figura 5 se puede observar que la línea de la muestra #78 sobrepasa las respectivas señales para las otras muestras en diferentes bandas. Una revisión del muestreo revela que la muestra #78 corresponde a una nafta importada, lo cual explica su espectro atípico respecto a las muestras de nafta derivada de crudos nacionales. Con lo anterior, la muestra #78 podría ser

descartada del conjunto base para el ajuste de algunas propiedades y obtener regresiones con desempeños cuantitativos.

Figura 5. Espectros FT-NIR rango entre 4400–4000 cm^{-1} .

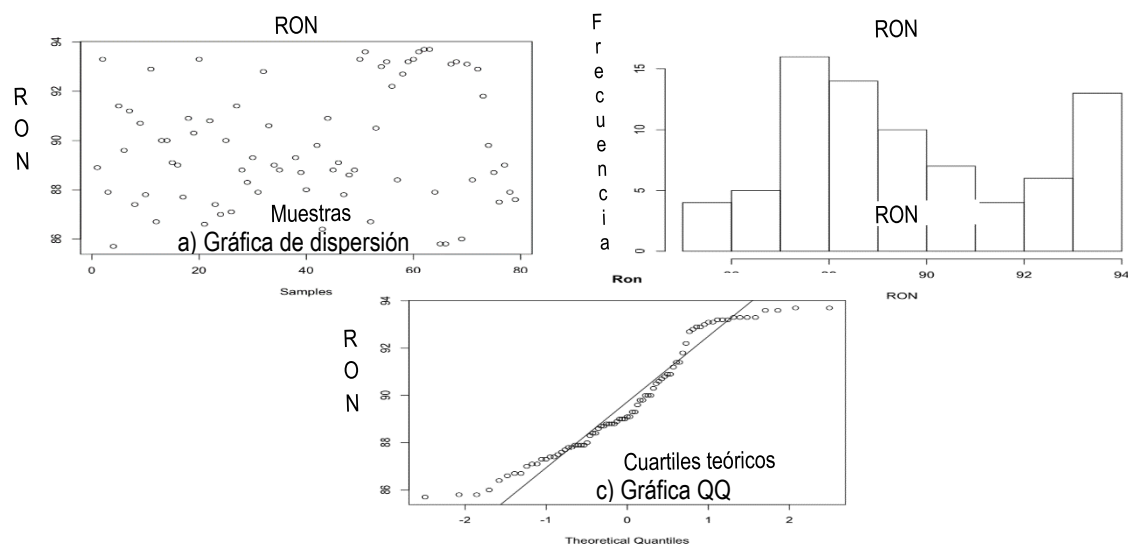


5.4 MODELO DE PREDICCIÓN PARA EL RON

La Figura 6 ilustra los valores de RON medidos a las muestras de gasolina de la GRB. La Figura 6a presenta la gráfica de dispersión de los valores de RON según las muestras, mientras que la Figura 5.6b presenta el histograma para los valores obtenidos. Las muestras exhibieron valores de RON entre 85.8 y 93.8 (Figuras 6a-b), rango característico en donde se tienen GMR y GME. Del histograma presentado en la Figura 5.6b es posible inferir que se tienen dos grupos de muestras de gasolinas, el primero en el intervalo de RON 85.8–91.5 con mayor frecuencia en 87.5 y, el segundo en el intervalo de RON 91.5–94 con mayor frecuencia en 93.5. Lo anterior sugiere la clasificación de las muestras del primer grupo como GMR, mientras que las muestras del segundo grupo corresponden a GME; las muestras pertenecen a gasolinas de destilación y de craqueo catalítico.

Por su parte, la Figura 6c presenta la gráfica *quantil-quantil* (QQ) para la propiedad. Según esta figura, las muestras exhiben una distribución de los RON cercana a la normal. El periodo de muestreo refleja las variaciones operacionales y los cambios en las dietas de los crudos tratados en la refinería; también, las variaciones en las mediciones experimentales y en las prácticas de laboratorio se encuentran circunscritas en las respectivas normas de la industria petrolera. El ajuste del RON a partir de las 79 muestras de gasolina fue aplicado por medio de la regresión PLS considerando la matriz de espectros pretratada; los números de onda que contribuyen a la predicción del RON se encuentran en la región de mayor intensidad espectroscópica de las muestras de gasolina (Figuras 1 y 2). El código desarrollado en R que utiliza la función *pls* del paquete *PLS* para la regresión del RON es mostrado en el anexo 1a.

Figura 6. Gráficas de distribución de los valores de RON para las muestras de gasolinas de la GRB.



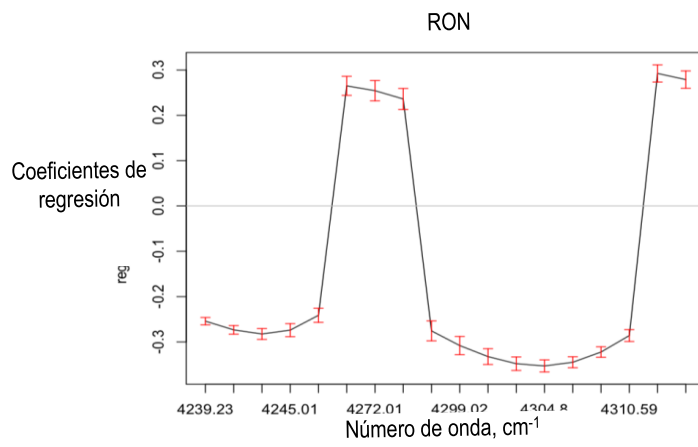
Durante el ajuste a través del método PLS, los números de onda que generaron valores de coeficientes con baja contribución a la regresión fueron descartados del conjunto. El conjunto final de números de onda con contribuciones importantes en la regresión del RON de las muestras de gasolina de la GRB es mostrado en la Tabla 2. Asimismo, esta tabla presenta para cada uno de los 18 números de onda, los respectivos valores del coeficiente y el intervalo considerando un 95% de confianza.

Tabla 2. Coeficientes de los números de onda para la predicción del RON. Término independiente 96.42.

Número de onda, cm^{-1}	Coeficiente	Incertidumbre, 95%
4239	-0.254	± 0.008
4241	-0.273	± 0.009
4243	-0.282	± 0.012
4245	-0.273	± 0.014
4246	-0.241	± 0.016
4270	0.265	± 0.021
4272	0.254	± 0.022
4273	0.236	± 0.023
4297	-0.275	± 0.022
4299	-0.308	± 0.020
4300	-0.332	± 0.017
4302	-0.348	± 0.015
4304	-0.352	± 0.013
4306	-0.344	± 0.012
4308	-0.322	± 0.012
4310	-0.285	± 0.013
4329	0.293	± 0.019
4331	0.279	± 0.019

Una comparación visual entre los diferentes coeficientes con sus respectivos intervalos de confianza, en los correspondientes números de onda aportantes de la regresión del RON puede observarse en la Figura 7. Según esta figura los números de onda aportantes se encuentran entre 4239 y 4331 cm^{-1} , exhibiendo un aporte similar al cálculo de la propiedad.

Figura 7. Gráficas de los coeficientes de regresión para el RON con sus respectivos intervalos al 95% de confianza.



La Figura 8 muestra la variación de la raíz del error cuadrático medio de predicción (RMSEP) con el número de componentes o variables latentes adicionadas al modelo PLS; las variables latentes representan el número de combinaciones lineales de las variables independientes requeridas para el modelo. En esta figura es posible observar que el número adecuado de componentes o variables latentes tomadas para el modelo corresponde a 2, con lo cual se presenta el menor error de RMSEP. La Figura 9 ilustra los resultados de la predicción basada en 2 componentes por medio de gráficos de paridad para calibración (Figura 9a) y validación cruzada (Figura 9b).

Figura 8 Curva de la raíz del error cuadrático medio de predicción (RMSEP) mostrando el número de variables latentes adecuado (2) en la regresión PLS de la propiedad RON.

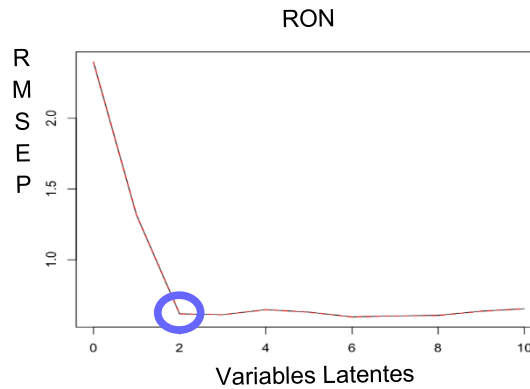
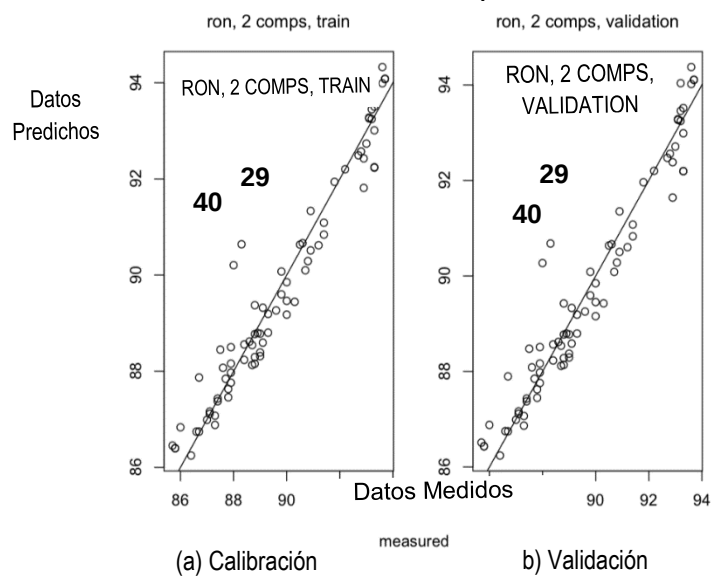


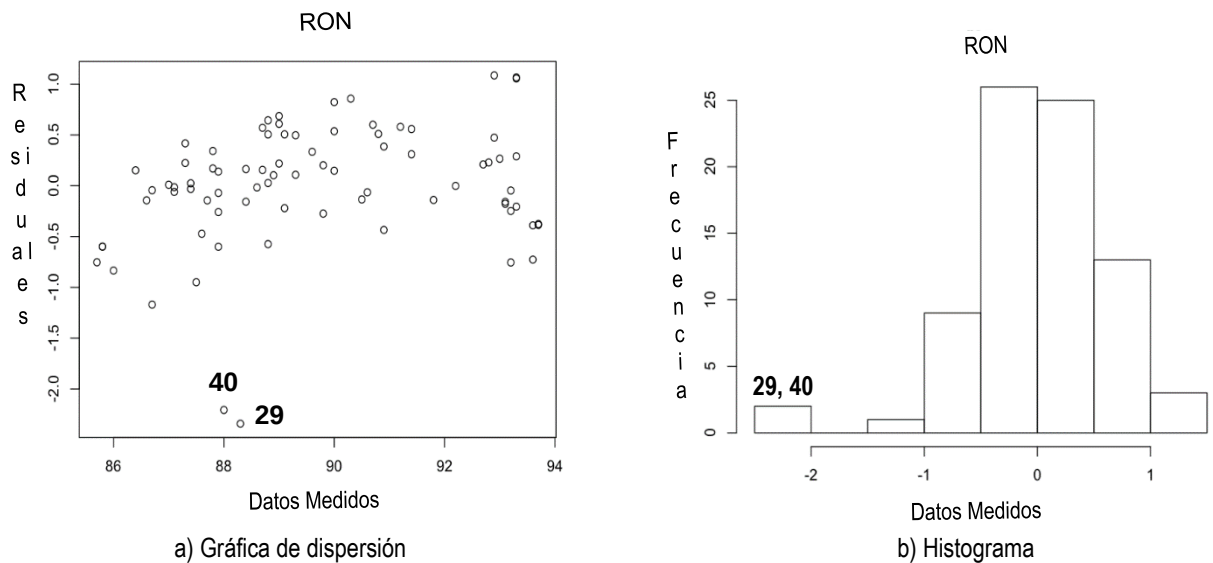
Figura 9. Valores predichos versus medidos para la regresión PLS del RON con 2 variables latentes. Figura de la izquierda corresponde a la etapa de calibración, mientras que la figura de la derecha corresponde a la etapa de validación cruzada “LOO”. R^2 validación 0.932; RMSEP=0.618.



Esta figura sugiere un desempeño adecuado de la regresión desarrollada, a nivel cuantitativo en las etapas de calibración y validación cruzada para las muestras de gasolinas; las muestras 29 y 40 que corresponden a gasolina motor regular presentan las mayores desviaciones respecto a los valores medidos.

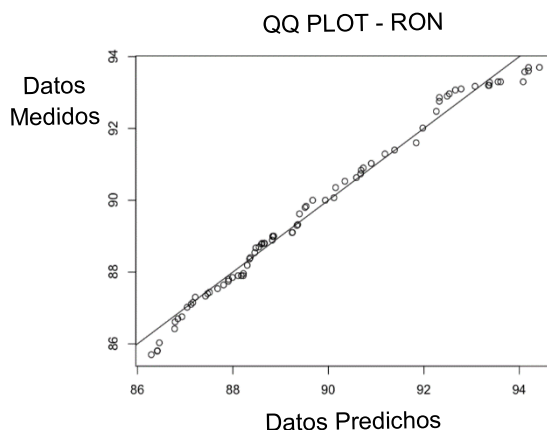
La Figura 10a-b muestra las gráficas de dispersión y el histograma para los residuales obtenidos con la regresión; no se visualizan tendencias en los residuales, lo cual valida el modelo para el RON. Según esta figura, las muestras 29 y 40 presentan valores cercanos a -2. A pesar de los valores de residuales para estas dos muestras la regresión exhibe un desempeño suficiente para su aplicación cuantitativa, soportado por su coeficiente de regresión de 0.932 y un RMSEP de 0.618 en las etapas de calibración y validación cruzada. Una nueva regresión fue desarrollada sin las muestras 29 y 40, obteniendo valores de 0.95 y 0.52 para R^2 y RMSEP, respectivamente; esta mejora en la predicción de la varianza de sólo 2%, conlleva a una disminución en la generalidad de la regresión, por lo cual, las muestras son conservadas y el modelo propuesto para el RON corresponde al expuesto en la Tabla 2. El valor de $RMSEP=0.618$ exhibido por el modelo de predicción del RON (Tabla 2) iguala a la repetibilidad de 0.7, reportada por la norma ASTM D-2699.

Figura 10. Residuales reportados por la regresión PLS del RON con 2 variables latentes.



Complementariamente, el histograma de la Figura 5.10b reporta que cerca de un 90% de los residuales se encuentran en el intervalo entre -1 y 1. La Figura 11 muestra la gráfica *QQ plot* de comparación de los valores de RON de las muestras y las predicciones del modelo. Según esta gráfica *QQ plot*, la mayoría de las predicciones se encuentran sobre la línea recta de pendiente la unidad (Figura 11). Solo un reducido número de muestras se desvían de la recta; estas muestras se localizan al inicio y al final del intervalo de medición, mostrando congruencia con una distribución de tipo normal. Lo anterior sugiere que la regresión desarrollada para el RON (Tabla 2) genera valores según lo esperado para la población de gasolinas motor regular y motor extra de la GRB.

Figura 11. Gráfica *QQ plot* comparando las distribuciones dadas por las muestras y la regresión PLS para el RON.



En cuanto a su comparación, el desempeño del modelo desarrollado (RMSEP=0.618 y $R^2=0.932$) en el presente trabajo supera a varios modelos reportados en la literatura abierta. Farquharson y colaboradores,⁴⁹ presentan una regresión PLS con 8 variables latentes para RON entre 91 y 100; los desempeños de la etapa de validación $R^2=0.86$ y RMSE=1.15 obtenidos por Farquharson y colaboradores, 2016, son característicos de un modelo cualitativo. Asimismo, Martínez y colaboradores, 2008, reportan un RMSEP=0.95 utilizando 243 muestras de gasolina y como pretratamiento la primera derivada.

Por otra parte, Fodor y colaboradores, 1996, reportan valores de predicción para el RON de gasolinas de $R^2=0.961$ y RMSE=0.489 utilizando el rango MIR (4000 – 650 cm^{-1}) de 288 muestras de gasolina de diferentes regiones de Estados Unidos; en el modelo de Fodor y colaboradores, 14% de las muestras reportaron valores por encima de la reproducibilidad según la norma ASTM D-2699. El modelo desarrollado en el presente trabajo reportó el 18% de muestras por fuera de la reproducibilidad

⁴⁹ FARQUHARSON, S., BROUILLETTE, C., SMITH, W., SHENDE, Ch., GLADDING, Z., MORRIS, R.E., CRAMER, J.A., SCHMITIGAL, J. Analysis of Twenty-Two Performance Properties of Diesel, Gasoline, and Jet Fuels Using a Field-Portable Near-Infrared (NIR) Analyzer. *Applied Spectroscopy*, 70, 746–755. 2016.

según la norma ASTM. Este valor ligeramente superior al reportado por Fodor y colaboradores corresponde a una aproximación satisfactoria considerando las fluctuaciones operacionales y la incertidumbre en las composiciones de los crudos procesados en la GRB.

Honorato y colaboradores, 2008, reportan una regresión PLS para 160 muestras de gasolina con contenido de 25% vol de etanol, con un desempeño RMSEP=0.28 para el ajuste del RON. Este desempeño elevado fue obtenido por selección de los números de onda en las regiones FT-MIR y FT-NIR por medio de algoritmos genéticos. Igualmente, Kardamakis & Pasadakis, 2010, desarrollaron un modelo de regresión lineal múltiple para el RON con 384 muestras recolectadas en un periodo de tiempo de 2 años. La región del espectro infrarrojo involucró las regiones FT-MIR (3520 – 4000 cm^{-1}) y FT-NIR (4000 – 4800 cm^{-1}); el desempeño reportado para la regresión fue RMSEP=0.3 y $R^2=0.987$.

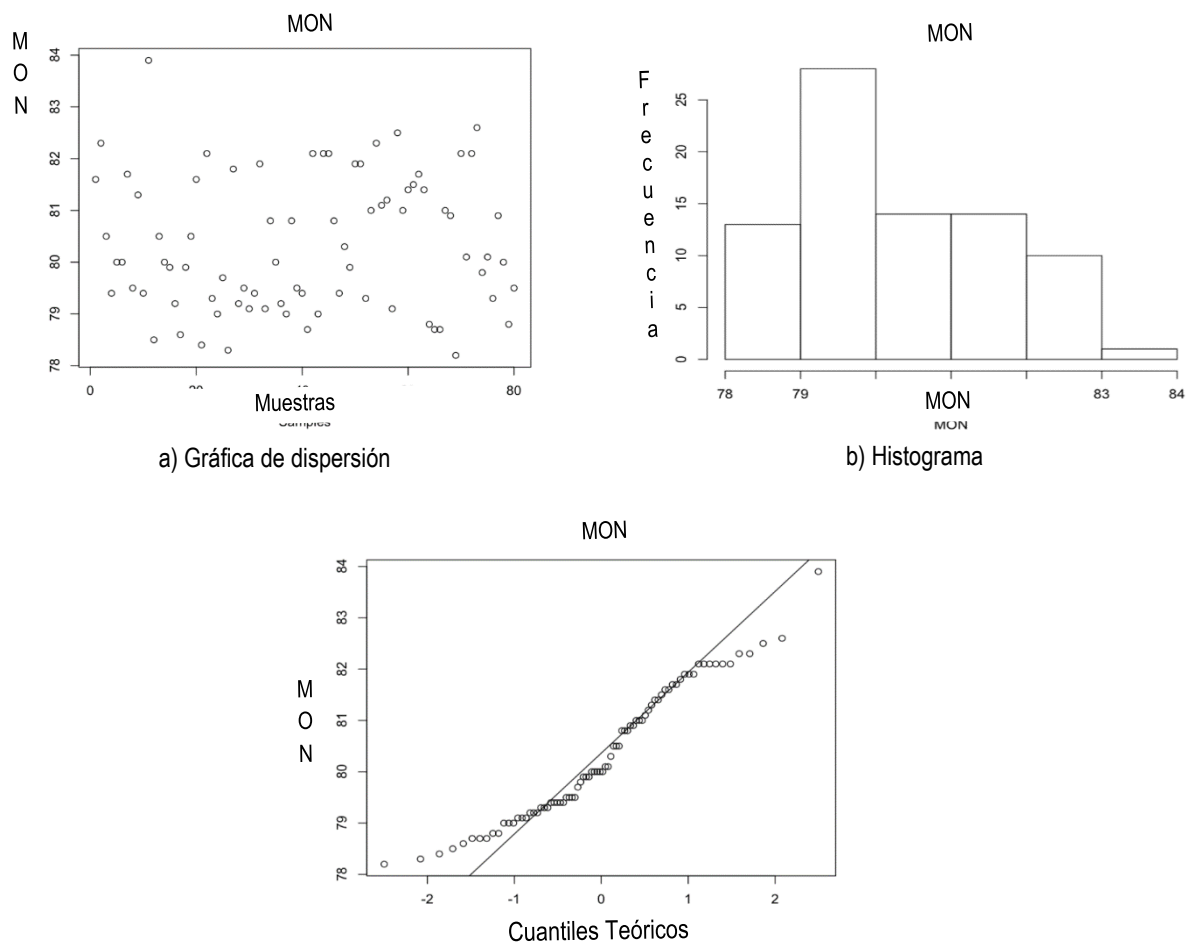
Los anteriores párrafos sugieren que el desempeño de la regresión para el RON desarrollada en el presente documento es comparable con desempeños reportados en trabajos de literatura abierta, lo cual soporta su aplicación en la mezcla de gasolinas. Por otra parte, si se desea una mejora de este modelo es posible recomendar la inclusión de un mayor número de muestras, una combinación entre regiones espectrales FT-MIR y FT-NIR, y la selección de variables por medio de algoritmos genéticos.

5.5 MODELO DE PREDICCIÓN PARA EL MON

La Figura 12 presenta en forma gráfica los valores de MON reportado por las muestras de gasolina. Estos valores se presentan mediante el gráfico de dispersión (Figura 12a), el histograma (Figura 12b) y el gráfico *QQ plot* (Figura 12c). La

propiedad MON medida para las muestras de gasolina de la GRB reporta valores en el rango 78 – 84 (Figura 12a). Por su parte, el histograma para esta propiedad presentado en la Figura 5.12b exhibe una distribución con un solo pico en el valor promedio de 79.5.

Figura 12. Gráficas de distribución del MON para las muestras de gasolinas de la GRB.



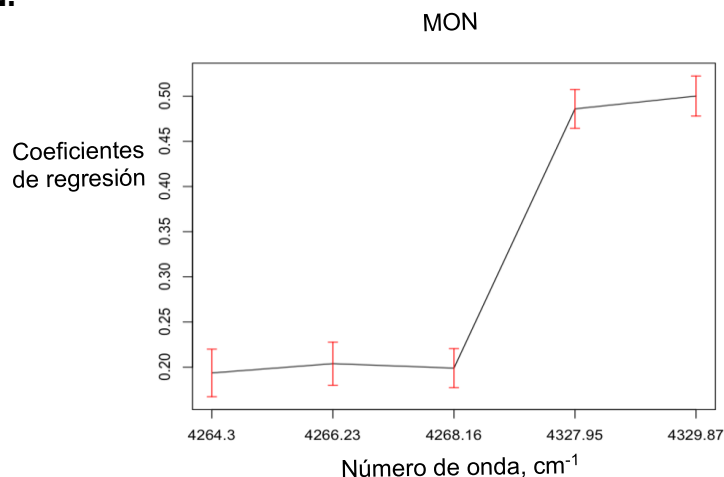
La distribución del MON difiere de la respectiva para el RON, ya que esta última propiedad diferencia los dos tipos de gasolinas comerciales generadas en la GRB. El gráfico QQ (Figura 12c) muestra desviaciones de la normalidad especialmente

en los valores límites de MON. Estas desviaciones son tolerables debido a los valores de repetitividad y reproducibilidad aceptados por la respectiva norma; el método ASTM-D2700 para la determinación experimental del MON reporta valores de 0.2 y 1.1 para la repetitividad y la reproducibilidad, respectivamente. La regresión PLS para el ajuste del MON a partir de las 79 muestras de gasolina fue aplicada según lo definido en la metodología, considerando la matriz de espectros con pretratamiento; el código desarrollado en R que utiliza la función *p/sr* del paquete *PLS* para la regresión del MON es mostrado en el anexo 1b. Diferentes números de onda, especialmente los mayores a 4500 cm^{-1} , fueron descartados durante la aplicación del PLS debido a su bajo aporte explicativo. Los números de onda que conforman el modelo de regresión final para el MON son presentados en la Tabla 5.2; los números de onda del modelo final de regresión del MON se encuentran en la región de mayor intensidad espectroscópica de las muestras de gasolina (Figuras 1d y 5). En la Tabla 3 también se presentan para cada número de onda, los respectivos valores del coeficiente y el intervalo considerando el 95% de confianza. Una comparación visual entre los coeficientes de regresión con sus respectivos intervalos puede obtenerse de la Figura 13.

Tabla 3. Coeficientes de los números de onda para la predicción del MON. Término independiente 92.48.

Número de onda, cm^{-1}	Coeficiente	Incertidumbre, 95%
4264	0.194	± 0.026
4266	0.204	± 0.024
4268	0.199	± 0.022
4328	0.486	± 0.021
4330	0.500	± 0.022

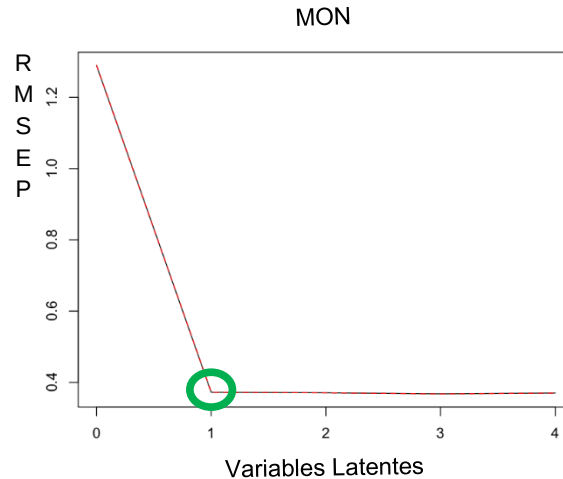
Figura 13. Coeficientes de regresión con sus intervalos de confianza (95%) para los correspondientes números de onda en el modelo final de regresión para el MON.



Según la Figura 13, los aportes de los números de onda son diferentes; los números de onda 4327 y 4329 cm^{-1} presentan la mayor contribución al cálculo del MON. Entre estos números de onda se localiza la banda de mayor contribución de los espectros FT-NIR de las muestras de gasolina (Figuras 1 y 5).

El modelo de regresión final incluye un término independiente con valor 92.48, el cual reproduce el 22% de la varianza del MON (Tabla 3), justificando los demás términos incluidos en la regresión. La Figura 5.14 presenta la variación del error RMSEP con el número de variables latentes que conforman el modelo. El valor de RMSEP de la regresión final corresponde a 0.372 utilizando una variable latente (Figura 14); con esto, la cantidad total de variables independientes fueron combinadas linealmente una sola vez.

Figura 14. Curva de la raíz del error cuadrático medio de predicción (RMSEP) mostrando el número de variables latentes adecuado (1) en la regresión PLS de la propiedad MON.



Por su parte, la Figura 15 presenta los gráficos de paridad para las etapas de calibración y validación. Esta figura sugiere un desempeño adecuado de la regresión desarrollada, a nivel cuantitativo en las etapas de calibración y validación cruzada para las muestras de gasolina. Las Figuras 16a y 16b muestran los residuales obtenidos con el modelo de regresión, utilizando las gráficas de dispersión y el histograma, respectivamente. Estas figuras muestran que los residuales obtenidos se encuentran dispersos alrededor del valor cero, en el rango entre -0.7 y 0.7. Además, no se visualiza una tendencia en los residuales, lo cual valida los supuestos del modelo.

Figura 15. Valores predichos versus medidos para la regresión PLS del MON con una variable latente. Figura de la izquierda corresponde a la etapa de calibración, mientras que la figura de la derecha corresponde a la etapa de validación cruzada “LOO”. R^2 validación 0.914; RMSEP=0.372.

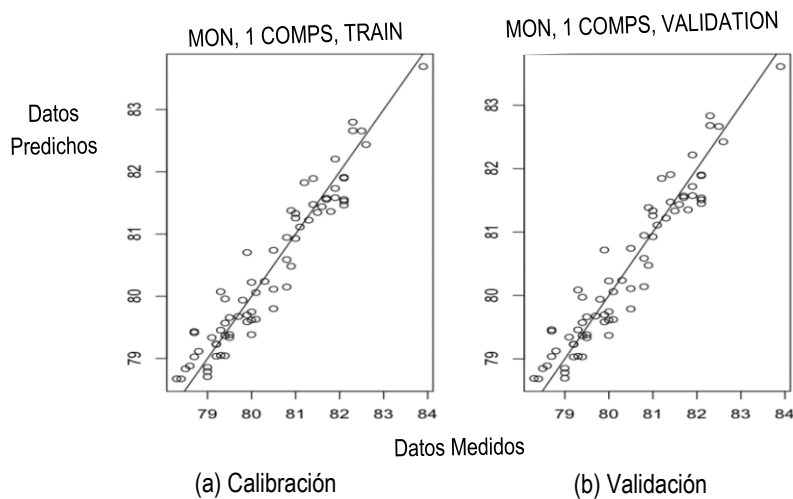
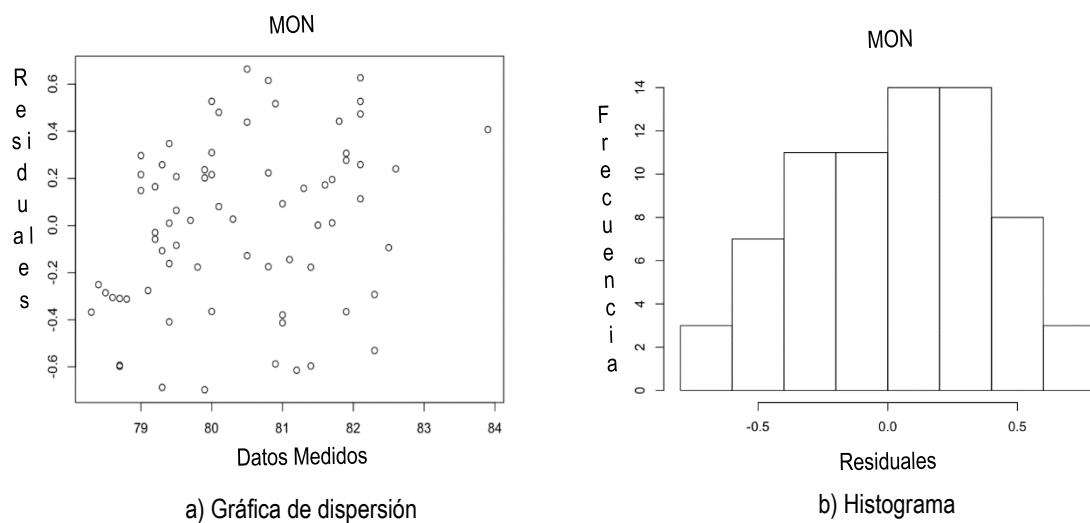
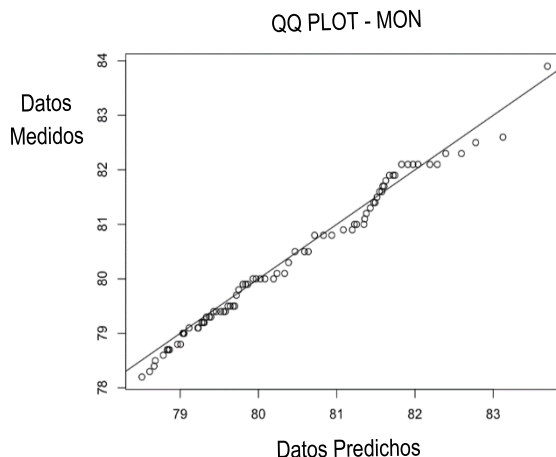


Figura 16. Residuales reportados por la regresión PLS del MON con una variable latente.



El coeficiente de determinación múltiple obtenido con el modelo corresponde a 0.914, con un valor de RMSEP de 0.372 para las etapas de calibración y validación cruzada. El valor de RMSEP=0.372 exhibido por el modelo de predicción del MON (Tabla 3), es menor que el valor de reproducibilidad promedio de 1.1 reportado por la norma ASTM D-2700. También, la gráfica *QQ plot*, mostrada en la Figura 17, presenta una importante cantidad de puntos sobre la línea recta de pendiente la unidad, con una leve desviación en los extremos, cumpliendo el supuesto de normalidad para la regresión aplicada; lo anterior sugiere que la regresión final para el MON (Tabla 3) genera valores representativos de la población de gasolinas producidas por la GRB.

Figura 17. Gráfica *QQ plot* comparando las distribuciones dadas por las muestras y la regresión PLS para el MON.



El RMSE obtenido con la regresión desarrollada en el presente documento es menor que el reportado en algunos trabajos de la literatura abierta. Farquharson y colaboradores, 2016, presentan un modelo para la predicción del MON para valores entre 83 y 91, con $R^2=0.95$ y $RMSE=0.45$, utilizando la calibración en el rango 4700

y 4000 cm^{-1} . Asimismo, Martínez y colaboradores⁵⁰, reportan RMSE de 0.71 utilizando 243 muestras de gasolina y como pretratamiento la primera derivada. También, Fodor y colaboradores⁵¹, reportan valores de predicción para el MON de gasolinas de $R^2=0.962$ y $RMSE=0.41$ utilizando el rango MIR ($4000 - 650\text{ cm}^{-1}$) de 293 muestras de gasolina de diferentes regiones de Estados Unidos; estos autores afirman que la calibración de su modelo no reporta valores mayores que la reproducibilidad según la norma. El modelo desarrollado en el presente documento tampoco reporta durante su etapa de validación ni calibración, valores por encima de la repetibilidad promedio según la norma ASTM.

Por otra parte, el RMSE del modelo final para el MON iguala al modelo propuesto por Honorato y colaboradores, 2008, basado en 160 muestras con contenido de 25% vol de etanol. Este desempeño elevado reportado por estos autores fue obtenido por selección de los números de onda en las regiones FT-MIR y FT-NIR aplicando algoritmos genéticos.

5.6 MODELO DE PREDICCIÓN PARA AROM

La Figura 18 repasa los valores obtenidos para la propiedad AROM en las muestras de gasolinas, en términos de la gráfica de dispersión (Figura 18a), el histograma (Figura 18b) y el gráfico QQ (Figura 18c). Según la Figura 18a, los valores de AROM se encuentran en el rango 10–26 %vol. El histograma para esta propiedad exhibe una distribución con un solo pico en el valor promedio de 19 %vol; al igual que para la propiedad MON, la distribución del contenido de Aromáticos exhibe un sólo valor promedio (Figura 18b). El gráfico QQ, Figura 5.18c, presenta un comportamiento

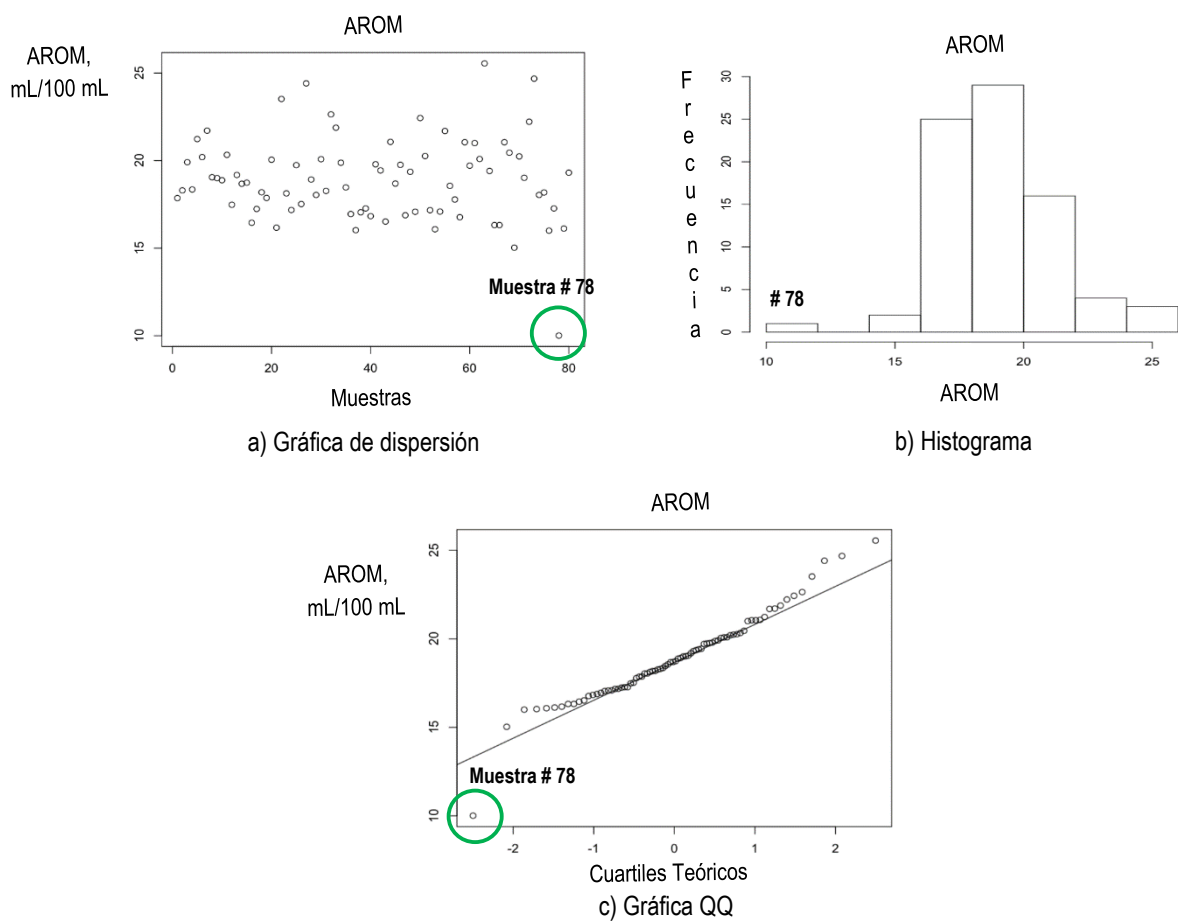
⁵⁰ MARTÍNEZ, HUERTAS, MÉNEZ, PEÑA, ALCALDE. Óp. Cit.

⁵¹ FODOR, G.; KOHL, K.; MASON, R. Analysis of gasoline by FT-IR spectroscopy. Anal Chem; 68:23–30. 1996

cercano a la normalidad, especialmente en la región central. Entre las muestras de regresión se incluyó la #78 descartada según el análisis PCA; la consideración de esta muestra en el modelo de AROM se debe a que esta propiedad es de carácter extensivo, dependiendo de la presencia de los aromáticos, la cual se encuentra relacionada con la intensidad de la señal⁵²; con esto, las desviaciones de la señal espectroscópica de la muestra #78 se encuentran incluidas en el valor de su respectivo AROM. Las desviaciones de la normalidad son tolerables debido a los valores de repetibilidad y reproducibilidad aceptados por la respectiva norma; el método ASTM-D5580 para la determinación experimental del AROM reporta valores de 0.2 %vol y 2.3 %vol para la repetibilidad y la reproducibilidad, respectivamente.

⁵² REBOUCAS, SANTOS, PIMENTEL, TEIXEIRA. 185–193. 2011. Óp. Cit.

Figura 18. Gráficas de distribución de los valores de AROM para las muestras de gasolinas de la GRB.



La regresión PLS para el AROM consideró las 80 muestras iniciales de gasolina, según la matriz de espectros con pretratamiento; el código desarrollado en R que utiliza la función *pls* del paquete *PLS* para la regresión del AROM es mostrado en el anexo 1c. Los ocho números de onda que constituyen el modelo de regresión final para el AROM pueden ser analizados en la Tabla 4.

Tabla 4. Coeficientes de los números de onda para la predicción de AROM. Término independiente 18.32 %.

Número de onda, cm ⁻¹	Coeficiente	Incertidumbre, 95%
4037	-0.786	±0.226
4073	-2.375	±0.232
4135	-0.888	±0.176
4137	-0.974	±0.194
4338	1.022	±0.308
9516	-1.000	±0.212
9881	1.253	±0.224
9942	0.906	±0.248

Parte de los números de onda del modelo final se encuentran en la región de mayor intensidad espectroscópica de las muestras de gasolina (Figuras 1 y 5). Tres números de onda del modelo se encuentran en la región entre 10000 y 9500 cm⁻¹, lo cual puede deberse a la cantidad de compuestos diaromáticos, triaromáticos y otros poliaromáticos presentes en las gasolinas. La contribución de los poliaromáticos a la señal espectroscópica se presenta a partir de los 6500 cm⁻¹; en los crudos estos aromáticos incluidos en los asfaltenos son los responsables de un incremento monotónico no lineal de la línea base en la señal espectroscópica^{53,54}. La contribución de los poliaromáticos en números de onda elevados procede de transiciones electrónicas $\pi \rightarrow \pi^*$ / $\Phi \pi \rightarrow \pi^*$ / residuales del espectro UV-vis⁵⁵. Por otra parte, los otros números de onda del modelo desarrollado en el presente trabajo coinciden con los reportados por Reboucas & Neto⁵⁶, y Reboucas y colaboradores⁵⁷; en los espectros de los monoaromáticos se presentan las máximas contribuciones en la región 4500–4000 cm⁻¹⁵⁸. La Tabla 5.3 incluye para cada número de onda, los respectivos valores del coeficiente y el intervalo considerando 95% de confianza.

⁵³ PASQUINI, C.; BUENO, A.F. Characterization of petroleum using near-infrared spectroscopy: quantitative modeling for the true boiling point curve and specific gravity. *Fuel*. 86:1927–34. 2007

⁵⁴ LAXALDE, J., CAILLOL, N., Wahl, F., RUCKEBUSCH, C., DUPONCHEL, L. Combining near and mid infrared spectroscopy for heavy oil characterisation. *Fuel*, 133, 310-316. 2014.

⁵⁵ MULLINS, 1990. Óp. Cit.

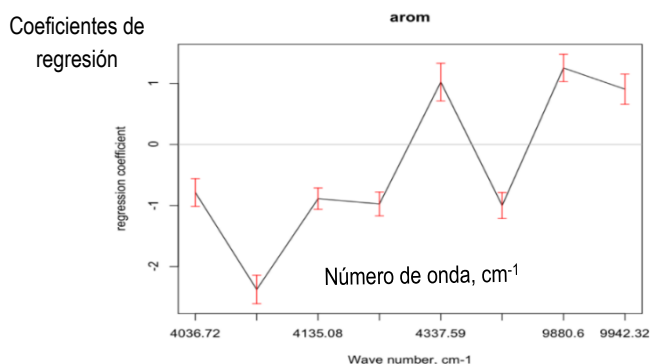
⁵⁶ REBOUCAS, NETO, 263–273. 2001. Óp. Cit.

⁵⁷ REBOUCAS, SANTOS, PIMENTEL, TEIXEIRA, 185–193. 2011. Óp. Cit.

⁵⁸ REBOUCAS, NETO, 263–273. 2001. Óp. Cit.

La Figura 19 presenta un referente de comparación visual entre los coeficientes de regresión con sus respectivos intervalos de confianza.

Figura 19. Coeficientes de regresión con sus intervalos de confianza (95%) para los correspondientes números de onda en el modelo final de regresión para el AROM.



Según la Tabla 3 y la Figura 19, el número de onda en 4073 cm⁻¹ presenta el mayor aporte a la regresión; el término independiente en el modelo final reporta el valor de 18.32 %vol, con una reproducción en la varianza del AROM 15% (RMSEP=2.42, Figura 20). El resto de la varianza se distribuye de forma equitativa entre los restantes siete números de onda (Tabla 4). La Figura 20 presenta la variación del error RMSEP con la inclusión de variables latentes en el modelo de regresión. Según esta figura, el valor de RMSEP de la regresión final corresponde a 0.71 utilizando tres variables latentes; es decir, tres combinaciones lineales entre las variables independientes fueron realizadas para llegar al menor error posible.

Por otra parte, la Figura 21 muestra el desempeño del modelo con las gráficas de paridad para las etapas de validación y calibración. Esta figura sugiere un desempeño a nivel cuantitativo adecuado para el modelo final, en las etapas de calibración (21a) y validación (21b) cruzada para las muestras de gasolina.

Figura 20. Curva de la raíz del error cuadrático medio de predicción (RMSEP) mostrando el número de variables latentes adecuado (3) en la regresión PLS de la propiedad AROM.

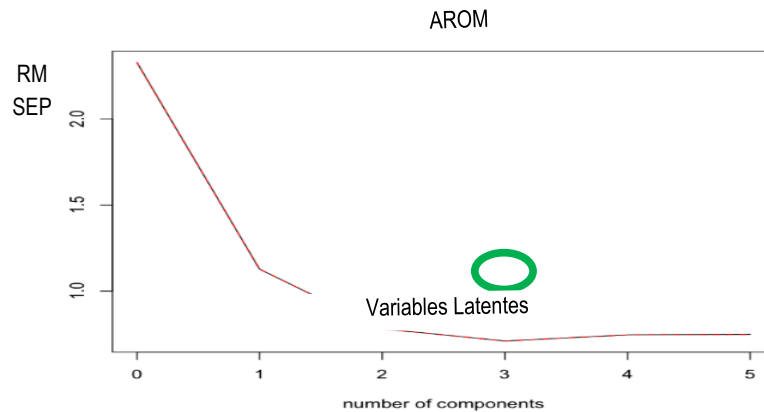
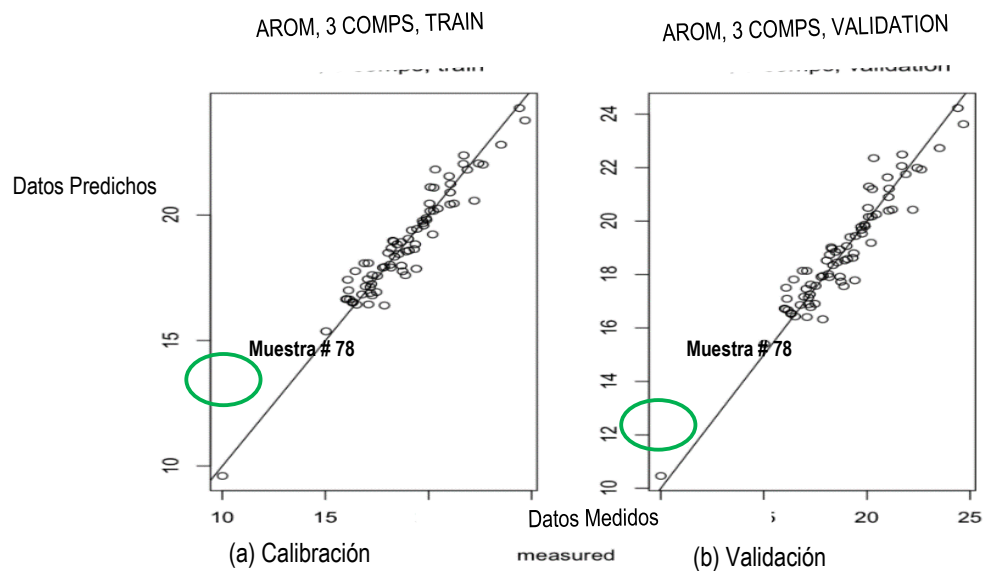
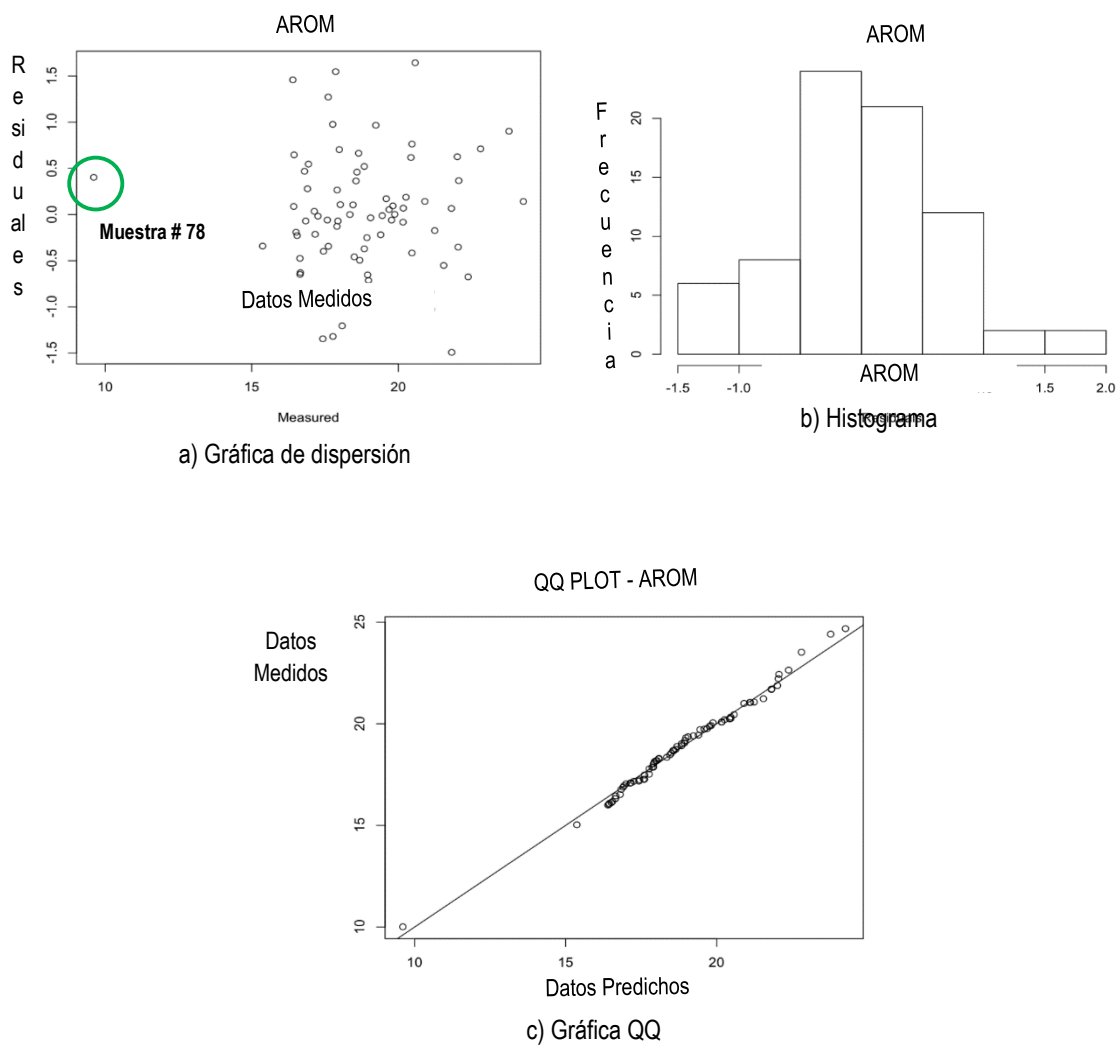


Figura 21. Valores predichos versus medidos para la regresión PLS del AROM con tres variables latentes. Figura de la izquierda corresponde a la etapa de calibración, mientras que la figura de la derecha corresponde a la etapa de validación cruzada “LOO”. R^2 validación 0.904; RMSEP=0.71.



Este desempeño incluye a la muestra 78, para la cual se obtiene un residual de 0.47 %vol (Figura 22a). Las Figuras 22a-b (dispersión e histograma) muestran que los residuales obtenidos al aplicar la regresión para el AROM no presentan ninguna tendencia; estos residuales se encuentran en el rango entre -1.5 y 2.0. El coeficiente de determinación múltiple obtenido corresponde a 0.904, con un valor de RMSEP de 0.71 para las etapas de calibración y validación cruzada. El valor de RMSEP=0.71 exhibido por el modelo final de regresión (Tabla 4), es menor que el valor de reproducibilidad promedio de 2.3 % reportado por la norma ASTM D-5580. También, la gráfica *QQ plot* presenta una importante cantidad de puntos sobre la línea recta de pendiente la unidad (Figura 22c), incluyendo a la muestra #78, con lo cual se cumple la condición de normalidad del ajuste. Esto sugiere que la regresión final para el AROM (Tabla 4) genera valores representativos de la población de gasolinas producidas por la GRB y la gasolina importada.

Figura 5.22 Residuales de la regresión PLS del AROM con tres variables latentes.



El modelo final de regresión para el AROM mejora o iguala los valores de RMSE reportados por varios autores en la literatura abierta. Reboucas y colaboradores⁵⁹, reportan un modelo PLS con 8 variables latentes, el cual posee un desempeño RMSE=0.61% y R²=0.992 para el AROM %w entre 20.20 y 55.18, calibrados en el

⁵⁹ REBOUCAS, M.V., SANTOS, J.B., PIMENTEL, M.F., TEIXEIRA, L.S.G. 185–193. 2011. Óp. Cit.

rango 4700–4000 cm^{-1} . Este valor se encuentra entre la repetitividad y la reproducibilidad reportadas por el método ASTM. Asimismo, Martínez y colaboradores, 2008, reportan un RMSEP de 1.26 % utilizando 243 muestras de gasolina y como pretratamiento la primera derivada. También, Fodor y colaboradores, 1996, reportan valores de predicción para el contenido de aromáticos de gasolinas de $R^2=0.95$ y $\text{RMSE}=1.8$ % utilizando el rango MIR (4000 – 650 cm^{-1}) en 186 muestras de gasolina; según Fodor y colaboradores, 4% de las muestras reportaron valores sin el cumplimiento de la reproducibilidad de la respectiva norma. El modelo final obtenido en el presente documento para el AROM no reporta durante las etapas de calibración y validación diferencias superiores a la reproducibilidad definidas en la norma ASTM.

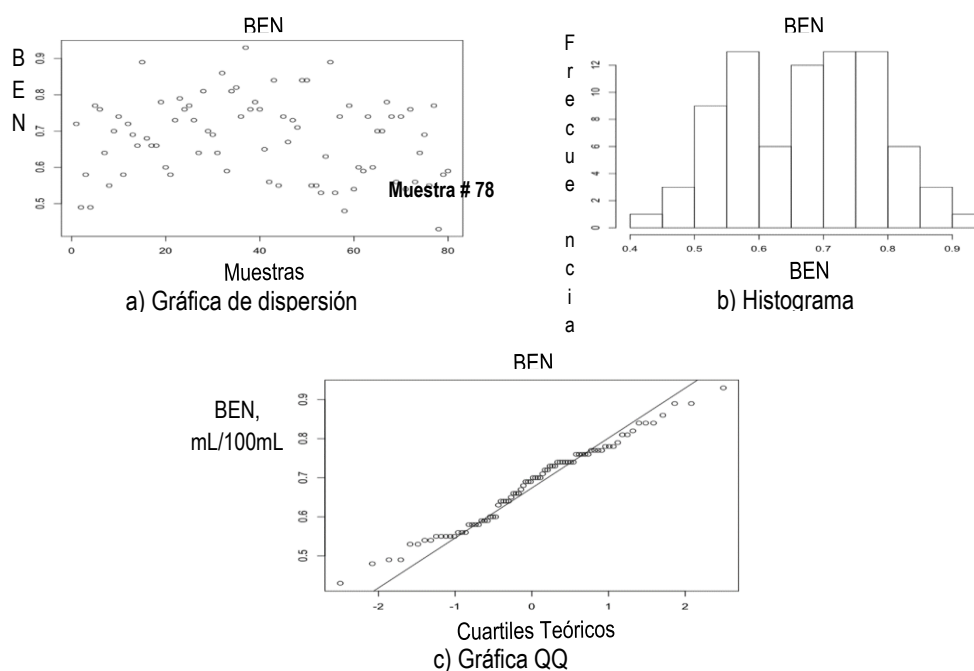
Por otra parte, Honorato y colaboradores, 2008, reportan una regresión PLS para 160 muestras con contenido de 25% vol de etanol con un desempeño $\text{RMSEP}=0.67\%$ para Aromáticos %vol. Este desempeño fue obtenido por selección de los números de onda en las regiones FT-MIR y FT-NIR por medio de algoritmos genéticos. Según la comparación realizada anteriormente, el desempeño del modelo final obtenido en el presente documento es igual o superior a los modelos reportados en la literatura; esto probablemente sea consecuencia de la inclusión de tres bandas en la región entre 10000–9000 cm^{-1} en el modelo final de regresión (Tabla 4 y Figura 19). Por otra parte, si se desea una mejora del modelo AROM es posible recomendar la aplicación de algoritmos genéticos para la selección de variables predictoras.

5.7 MODELO DE PREDICCIÓN PARA BEN

La Figura 23 exhibe el rango de valores para BEN de las 80 muestras de gasolina que conforman la base de datos, utilizando el gráfico de dispersión (Figura 23a), el

histograma (Figura 23b) y el gráfico QQ (Figura 23c). Según la Figura 23, el BEN se encuentra entre los valores (0.4 – 1.0) %vol, exhibiendo una zona de distribución de frecuencias para el máximo entre (0.55 – 0.8) %vol (Figura 23b). De igual manera, la Figura 23c muestra que la distribución de BEN presenta una ligera desviación de la normalidad, especialmente en los valores cercanos a los extremos. La desviación de la normalidad se encuentra considerada en la repetitividad <0.01 y la reproducibilidad promedio de 0.2, según la norma ASTM-D5580.

Figura 23. Gráficas de distribución de los valores de BEN para las muestras de gasolinas de la GRB.



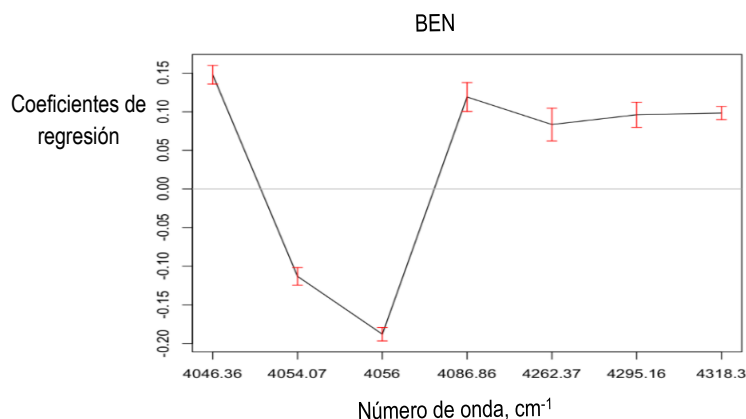
La regresión PLS de BEN consideró las 80 muestras iniciales de gasolina que constituyen la matriz de espectros FT-NIR con pretratamiento; al igual que para la propiedad AROM, la muestra #78 fue incluida en la regresión para la propiedad BEN debido a la característica extensiva de estas propiedades. El código desarrollado en

R que utiliza la función *pls*r del paquete *PLS* para la regresión del BEN es mostrado en el anexo 1d. Según la Tabla 5, el modelo de regresión final para BEN se encuentra formado por un término independiente (-0.9729 %) y siete números de onda en el rango de mayor contribución de los espectros (4500–4000 cm^{-1}). Esta tabla incluye para cada número de onda, los respectivos valores del coeficiente y el intervalo considerando 95% de confianza. Asimismo, la Figura 24 presenta un referente de comparación visual entre los coeficientes de regresión y sus intervalos de confianza.

Tabla 5. Coeficientes de los números de onda para la predicción de BEN. Término independiente -0.9729 %.

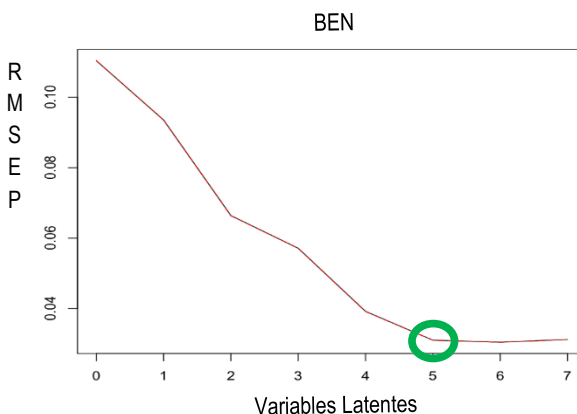
Número de onda, cm^{-1}	Coeficiente	Incertidumbre, 95%
4046	0.1479	± 0.0120
4054	-0.1130	± 0.0114
4056	-0.1879	± 0.0087
4087	0.1191	± 0.0187
4262	0.0834	± 0.0212
4295	0.0961	± 0.0163
4318	0.0983	± 0.0084

Figura 5.24 Coeficientes de regresión con sus intervalos de confianza (95%) para los correspondientes números de onda en el modelo final de regresión para BEN.



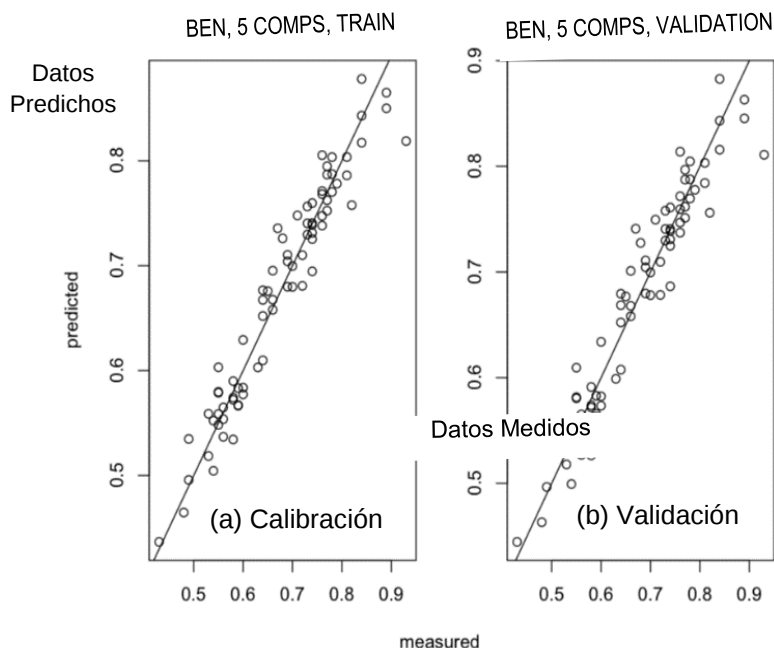
Según la Tabla 5 y la Figura 24, el número de onda en 4056 cm^{-1} presenta el mayor aporte a la regresión; el término independiente reporta una reproducción en la varianza del BEN del 18% ($\text{RMSEP}=0.11$, Figura 26). El resto de la varianza se distribuye de forma aproximadamente equitativa entre los restantes seis números de onda (Tabla 5). La Figura 25 muestra la variación del error RMSEP con la inclusión de variables latentes en el modelo de regresión; según esta figura, el valor de RMSEP de la regresión final corresponde a 0.031 utilizando cinco variables latentes (cinco combinaciones lineales entre las variables independientes). En el modelo final, los números de onda que contribuyen a la predicción del BEN se encuentran en la región de mayor intensidad espectroscópica de las muestras de gasolina (Figuras 1 y 5). Lo anterior se diferencia de lo obtenido para el AROM.

Figura 25 Curva de la raíz del error cuadrático medio de predicción (RMSEP) mostrando el número de variables latentes adecuado (5) en la regresión PLS de BEN.



Por otra parte, la Figura 26 muestra un desempeño cuantitativo para el modelo final, en las etapas de calibración y validación cruzada. Las gráficas de paridad en calibración y validación exhiben a la mayoría de muestras del modelo en una posición cercana a la línea de pendiente la unidad.

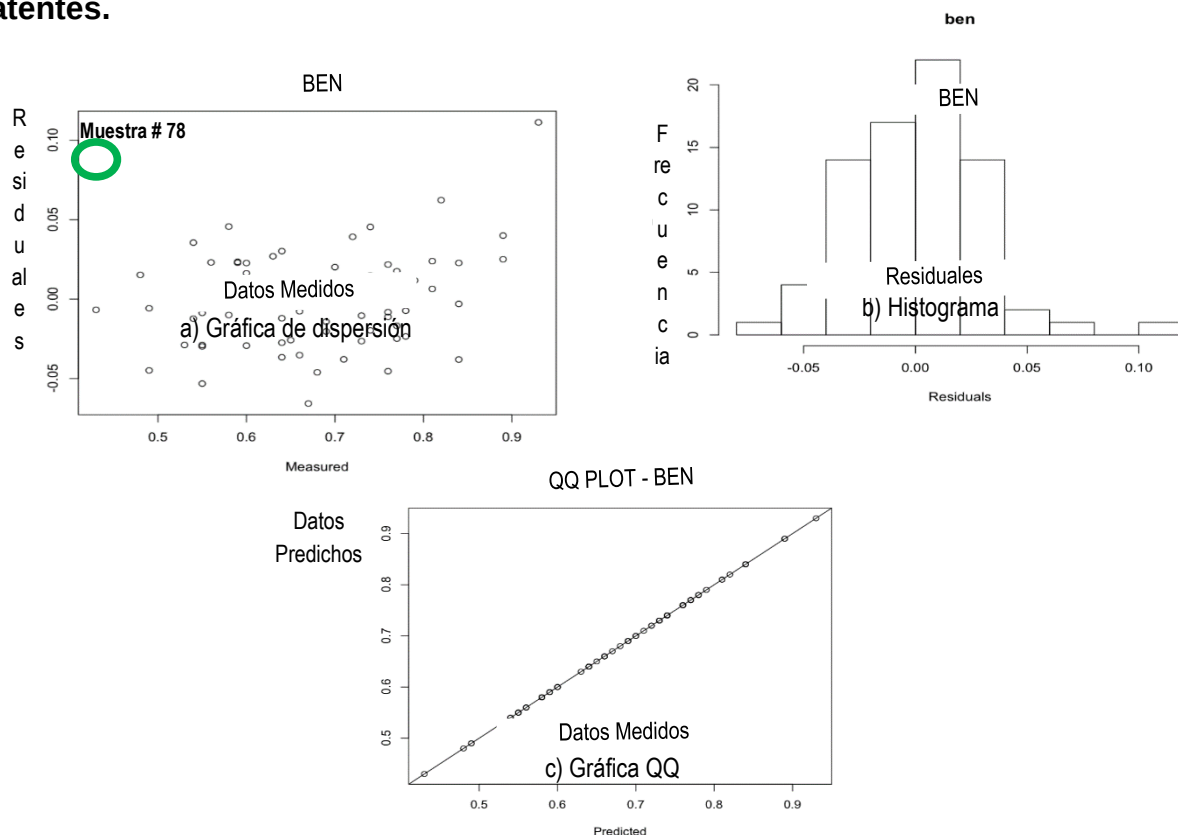
Figura 26. Valores predichos versus medidos para la regresión PLS de BEN con 5 variables latentes. Figura de la izquierda corresponde a la etapa de calibración, mientras que la figura de la derecha corresponde a la etapa de validación cruzada “LOO”. R^2 validación 0.919; RMSEP=0.031.



La Figura 27 muestra los residuales en los diagramas de dispersión (Figura 27a) y en el histograma (Figura 27b). Según las Figuras 27a y 27b, los residuales para BEN se encuentran en el rango entre -0.065 y 0.11 %vol. Este rango incluye a la muestra #78, para la cual se obtiene un residual de 0.005 %vol (Figura 27a); los residuales validan el modelo obtenido al no presentar una tendencia visible. El coeficiente de determinación múltiple obtenido corresponde a 0.919, con un valor de RMSEP de 0.031 para las etapas de calibración y validación cruzada. El valor de RMSEP=0.031 %vol exhibido por el modelo final de regresión (Tabla 5), es menor que el valor de reproducibilidad promedio de 1.2%vol reportado por la norma ASTM D-5580. También, la gráfica *QQ plot* presentado en la Figura 5.27c exhibe a los puntos muestrales sobre la línea recta de pendiente la unidad, incluyendo a la

muestra #78. Esto sugiere que la regresión final para BEN (Tabla 5) representa a la población de gasolinas producidas por la GRB, incluyendo a la gasolina importada.

Figura 27 Residuales reportados por la regresión PLS del BEN con 5 variables latentes.



La literatura reporta desempeños RMSE comparables con los obtenidos en el presente trabajo de aplicación final. Reboucas y colaboradores, 2011, reportan un desempeño RMSE=0.04 y $R^2=0.942$ para BEN entre 0 y 1.4, calibrados en el rango 4700 – 4000 cm^{-1} . Este valor se encuentra ligeramente por encima de la reproducibilidad reportada por el método ASTM. Asimismo, Martínez y colaboradores, 2008, reportan un RMSEP de 0.11 utilizando 243 muestras de gasolina y como pretratamiento la primera derivada. También, Fodor y

colaboradores, 1996, reportan valores de predicción para BEN de gasolinas de $R^2=0.902$ y $RMSE=0.167$ utilizando el rango MIR (4000–650 cm^{-1}); el modelo propuesto por Fodor y colaboradores, reporta que un 2% de las muestras exhibe valores superiores a la reproducibilidad promedio según la norma ASTM. El modelo desarrollado en el presente documento no reporta durante sus etapas de calibración y validación, diferencias superiores a la estipulada en la norma ASTM.

5.8 RESUMEN DE LAS REGRESIONES PLS

Las regresiones desarrolladas en el presente trabajo para las propiedades RON, MON, AROM y BEN exhibieron desempeños comparables con los reportes encontrados en la literatura abierta. La Tabla 6 resume los desempeños de los modelos finales obtenidos; esta tabla presenta la ecuación PLS en términos de las absorbancias pretratadas en los respectivos números de onda (A_{xxx}), el RMSEP y el R^2 ; asimismo, la Tabla 6 incluye en la columna de desempeños, los valores obtenidos por reporte encontrados en la literatura, así como el valor de la reproducibilidad con la correspondiente norma ASTM. Según esta columna, los modelos finales obtenidos superan los desempeños en términos de RMSE de varios reportes encontrados en la literatura abierta. Asimismo, los modelos presentan desempeños de RMSE% muy cercanos a los reportes de literatura.

La comparación de desempeños permite afirmar que los modelos finales propuestos en el presente documento para las propiedades RON, MON, AROM y BEN pueden ser utilizados en la predicción a nivel cuantitativo empleando naftas colombianas, con resultados cercanos a los esperados experimentalmente. Desde luego, verificaciones adicionales de los desempeños de los modelos propuestos son requeridas para validar su aplicación en la predicción en línea de estas propiedades.

Tabla 6. Modelos finales de regresión para las propiedades RON, MON, AROM y BEN.

MODELO FINAL DE REGRESIÓN	DESEMPEÑOS		
$RON = 96.42 - 0.254A_{4239} - 0.273A_{4241} - 0.282A_{4243} - 0.273A_{4245} - 0.241A_{4246} + 0.265A_{4270} + 0.254A_{4272} + 0.236A_{4273} - 0.275A_{4297} - 0.308A_{4299} - 0.332A_{4300} - 0.348A_{4302} - 0.352A_{4304} - 0.344A_{4306} - 0.322A_{4308} - 0.285A_{4310} + 0.293A_{4329} + 0.279A_{4331}$	RMSE	RMSE%	R2
	0.618	0.691	0.932 (*)
	1.150	NA	0.860 (1)
	0.950	NA	NA (2)
	0.489	0.514	0.961 (3)
	0.300	0.315	0.987 (4)
	0.280	NA	NA (5)
$MON = 92.48 + 0.194A_{4264} + 0.204A_{4266} + 0.199A_{4268} + 0.486A_{4328} + 0.500A_{4330}$	RMSE	RMSE%	R2
	0.372	0.462	0.914 (*)
	0.371	NA	NA (5)
	0.410	0.480	0.962 (3)
	0.450	NA	0.950 (1)
	0.710	NA	NA (2)
$AROM = 18.32 - 0.786A_{4037} - 2.375A_{4073} - 0.888A_{4135} - 0.974A_{4137} + 1.022A_{4338} - A_{9516} + 1.253A_{9881} + 0.906A_{9942}$	RMSE	RMSE%	R2
	0.710	3.800	0.904 (*)
	1.800	5.630	0.950 (3)
	1.260	NA	NA (2)
	0.670	NA	NA (5)
	0.610	1.918	0.992 (6)
$BEN = -0.9729 + 0.1479A_{4046} - 0.1130A_{4054} - 0.1879A_{4056} + 0.1191A_{4087} + 0.0834A_{4262} + 0.0961A_{4295} + 0.0983A_{4318}$	RMSE	RMSE%	R2
	0.031	4.100	0.919 (*)
	0.167	15.180	0.902 (3)
	0.110	NA	NA (2)
	0.004	4.000	0.942 (6)
$BEN = -0.9729 + 0.1479A_{4046} - 0.1130A_{4054} - 0.1879A_{4056} + 0.1191A_{4087} + 0.0834A_{4262} + 0.0961A_{4295} + 0.0983A_{4318}$	0.200	NA	NA (10)

(*) Desempeños del presente trabajo. (1) Farquharson y colaboradores, 2016. (2) Martínez y colaboradores, 2008. (3) Fodor y colaboradores, 1996. (4) Kardamakis & Pasadakis, 2010. (5) Honorato y colaboradores, 2008. (6) Reboucas y colaboradores, 2011. (7) Reproducibilidad promedio norma ASTM D-2699. (8) Reproducibilidad promedio norma ASTM D-2700. (9) Reproducibilidad promedio norma ASTM D-1319. (10) Reproducibilidad promedio norma ASTM D-5580.

Por otro lado, los reportes de literatura que exhiben regresiones con desempeños superiores a los modelos obtenidos en el presente trabajo utilizan diferentes estrategias como inclusión de regiones FT-MIR y FT-NIR y selección de variables con algoritmos genéticos^{60, 61}, para gasolinas y Laxalde y colaboradores⁶², para crudos pesados). La aplicación de estas estrategias utilizando la base de datos de espectros FT-NIR es sugerida para verificar un posible incremento en el desempeño de los modelos de regresión propuestos en el presente documento.

⁶⁰ HONORATO, F.A., NETO, F.B., PIMENTEL, M.F., STRAGEVITCH, L., GALVAO, R.K.H. Using principal component analysis to find the best calibration settings for simultaneous spectroscopic determination of several gasoline properties. *Fuel*, 87, 3706-3709. 2008.

⁶¹ KARDAMAKIS, & PASADAKIS, Op. Cit.

⁶² LAXALDE, CAILLOL, Wahl, RUCKEBUSCH, DUPONCHEL, Op. Cit.

6. CONCLUSIONES

- El análisis por componentes principales condujo a la identificación de la gasolina importada como una muestra con características diferentes a las gasolinas nacionales. En efecto, las bandas del espectro infrarrojo entre 4000 y 4400 cm^{-1} en la gasolina importada poseen mayor intensidad en la señal que las exhibidas por las gasolinas nacionales. Esta muestra de gasolina presenta los menores valores de contenido de aromáticos y benceno, por lo cual se diferencia químicamente de las restantes gasolinas. A pesar de lo anterior, la gasolina importada fue incluida en los modelos de regresión PLS para AROM y BEN ya que estas propiedades son de carácter extensivo, lo cual se encuentra relacionado con la intensidad de la señal.
- El modelo de regresión PLS para el RON se encuentra conformado por un total de 18 números de onda entre 4239 y 4310 cm^{-1} , más un término constante. Según el valor de los coeficientes, cada uno de los números de onda del modelo presentan un aporte similar al cálculo de la propiedad. El coeficiente de regresión y el RMSEP del modelo corresponden a 0.932 y 0.618, respectivamente, utilizando sólo dos variables latentes; el valor del RMSEP es comparable con la repetibilidad reportada en la norma ASTM D-2699. Por su parte, el modelo de regresión PLS para el MON depende de cinco números de onda en el rango 4330–4264 cm^{-1} ; los números de onda 4327 y 4329 cm^{-1} presentan la mayor contribución al cálculo del MON. El coeficiente de determinación múltiple y el RMSEP obtenidos con el modelo para el MON corresponden a 0.914 y 0.372, respectivamente, utilizando sólo una variable latente. Este valor de RMSEP es menor que el valor de reproducibilidad promedio de 1.1 reportado por la norma ASTM D-2700. Los desempeños obtenidos con las regresiones del RON y el MON desarrolladas en el presente trabajo, justifican la utilidad de estos modelos en las predicciones del *blending* de gasolinas en la GRB.

- Los modelos de regresión para AROM y BEN dependen de 8 y 7 números de onda, respectivamente. Los números de onda del modelo para AROM incluyen 5 números de onda en el intervalo 4338–4037 cm^{-1} y tres en el intervalo 9942–9516 cm^{-1} , mientras que el modelo para BEN incluye números de onda exclusivamente en el rango 4318–4046 cm^{-1} . Los números de onda con mayor aporte en los modelos para AROM y BEN corresponden a 4073 cm^{-1} y 4056 cm^{-1} , respectivamente. Los coeficientes de determinación múltiple de los modelos para AROM y BEN corresponden a 0.904 y 0.919, respectivamente, mientras que sus errores RMSEP corresponden a 0.71 %vol y 0.031 %vol, respectivamente. Las desviaciones estándar de los modelos para AROM y BEN son menores que los respectivos valores de reproducibilidad reportados en la norma ASTM D-5580, lo cual sugiere que estas regresiones son de utilidad para su aplicación en las predicciones del *blending* de gasolinas en la GRB.

7. RECOMENDACIONES

- Para futuros trabajos en el desarrollo de modelos de predicción de propiedades para las gasolinas se recomienda la inclusión de las regiones FT-MIR y FT-NIR, así como la aplicación de algoritmos de selección de variables, como los algoritmos genéticos.
- Los modelos deberían ser validados con muestras externas antes de su aplicación en algún código de predicción en línea.
- La validación de los modelos con otros programas (por ejemplo, *Unscrambler*) es recomendable para corroborar los códigos utilizados en R.

BIBLIOGRAFÍA

BALABIN, R. M., SAFIEVA, R. Z., LOMAKINA, E. I. Comparison of linear and nonlinear calibration models based on near infrared (NIR) spectroscopy data for gasoline properties prediction. *Chemometrics and intelligent laboratory systems*, 88. 183-188. 2007

BARNES, R.J., DHANOA, M.S., LISTER, S.J. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, 43. 772-777. 1989

BI, Y.M., CHU, G.H., WU, J.Z., YUAN, K.L., WU, J., LIAO, F., XIA, J., ZHANG, G.X., ZHOU, G.J. Ensemble Partial Least Squares Algorithm Based on Variable Clustering for Quantitative Infrared Spectrometric Analysis. *Chin J Anal Chem*, 43, 1086–1091. 2015

BI, Y.M., XIE, Q., PENG, S.L., TANG, L., HU, Y., TAN, J., ZHAO, Y.H., LI, C.W. Dual stacked partial least squares for analysis of near-infrared spectra. *Anal. Chim. Acta*, 792. 19–27. 2013

BLANCO, M.; MASPOCH, S.; VILLARROYA, I.; PERALTA, X.; GONZÁLEZ, J.M.; TORRES, J. Determination of physico-chemical parameters for bitumens using near infrared spectroscopy. *Anal. Chim. Acta*, 434, 133–141. 2001

BRUDZEWSKI, K., y colaboradores. Gasoline quality prediction using gas chromatography and FTIR spectroscopy: An artificial intelligence approach. *Fuel*, 85. 553-558. 2006

BURNS, D.A. and CIURCZAK, E.W. Handbook of Near Infrared Analysis, 3e. Boca Raton: CRC Press. 2008

DALY, S.R.; NIEMEYER, K.E.; CANNELLA, W.J.; HAGEN, C.L. Predicting fuel research octane number using Fourier-transform infrared absorption spectra of neat hydrocarbons. *Fuel*, 183, 359-365. 2016

DE BARROS NETO, Reboucas, B. Near infrared spectroscopic prediction of physical properties of aromatics-rich hydrocarbon mixtures, *J. Near Infrared Spectrosc.* 9. 263–273. 2001

ECHEVERRÍA, O. Análisis y Correlación del Contenido de Biocombustibles, Punto de Chispa y T95 para el Diésel Utilizando Espectroscopía Infrarroja y Métodos Quimiométricos. Trabajo de Aplicación Final. Maestría en Ingeniería Química, UIS. 2016

FARQUHARSON, S., BROUILLETTE, C., SMITH, W., SHENDE, Ch., GLADDING, Z., MORRIS, R.E., CRAMER, J.A., SCHMITIGAL, J. Analysis of Twenty-Two Performance Properties of Diesel, Gasoline, and Jet Fuels Using a Field-Portable Near-Infrared (NIR) Analyzer. *Applied Spectroscopy*, 70. 746–755. 2016

FODOR, G.; KOHL, K.; MASON, R. Analysis of gasoline by FT-IR spectroscopy. *Anal Chem*; 68:23–30. 1996

GARY, G.E.; HANDWERK, M.; KAISER, J. Petroleum Refining technology and economics, fifth edition, CRC press. 2007

HE, K. y colaboradores. A novel adaptive algorithm with near-infrared spectroscopy and its application in online gasoline blending processes. *Chemometrics and Intelligent Laboratory Systems*, 140. 117-125. 2015

HELLAND, I.S. Some theoretical aspects of partial least squares regression. *Chem. Int. Lab. Syst.*, 58. 97-107. 2001

HONORATO, F.A., NETO, F.B., PIMENTEL, M.F., STRAGEVITCH, L., GALVAO, R.K.H. Using principal component analysis to find the best calibration settings for simultaneous spectroscopic determination of several gasoline properties. *Fuel*, 87. 3706-3709. 2008

KARDAMAKIS, A.A. & PASADAKIS, N. Autoregressive modeling of near-IR spectra and MLR to predict RON values of gasolines. *Fuel*, 89. 158-161. 2010

KELLY, J.J., y colaboradores. Prediction of gasoline octane numbers from near-infrared spectral features in the range 660-1215 nm. *Analytical Chemistry*, 61. 313-320. 1989

KIM, M., LEE, Y.H., HAN, Ch. Real-time classification of petroleum products using near-infrared spectra. *Computers & Chemical Engineering*, 24. 513-517. 2000

LAXALDE, J., CAILLOL, N., Wahl, F., RUCKEBUSCH, C., DUPONCHEL, L. Combining near and mid infrared spectroscopy for heavy oil characterisation. *Fuel*, 133. 310-316. 2014

LEE, Y., CHUNG, H., KIM, N. Spectral range optimization for the near-infrared quantitative analysis of petrochemical and petroleum products: naphtha and gasoline. *Applied spectroscopy*, 60. 892-897. 2006

LI, B., MORRIS, J., MARTIN, E.B. Model selection for partial least squares regression. *Chem. Int. Lab. Syst.*, 64. 79-89. 2002

MARTÍNEZ, E., HUERTAS, S., MÉNEZ, H., PEÑA, J.L., ALCALDE, A. Comparison of chemometric techniques applied to near infrared spectra for a gasoline blending control. *J. Near Infrared Spectrosc.* 16. 297–303. 2008

MICYUS, N. J.; MCCURRY, J. D., & SEELEY, J. V. Analysis of aromatic compounds in gasoline with flow-switching comprehensive two-dimensional gas chromatography. *Journal of Chromatography A*, 1086(1-2), 115–121. 2005

NAKATA, K.; UCHIDA, D.; OTA, A.; UTSUMI, S.; KAWATAKE, K. The impact of RON on SI engine thermal efficiency. SAE Technical Paper. 2007

PAGE, N. P., & MEHLMAN, M. Health Effects of Gasoline Refueling Vapors and Measured Exposures At Service Stations. *Toxicology and Industrial Health*, 5(5), 869–890. 1989

PALOU, A. Desarrollo de nuevas metodologías espectrales para el control analítico de productos y procesos petroquímicos y farmacéuticos. Bellaterra. Tesis doctoral (Programa de doctorado de química). Universidad Autónoma de Barcelona. Facultad de Ciencias. Departamento de Química. 2014

PASQUINI, C.; BUENO, A.F. Characterization of petroleum using near-infrared spectroscopy: quantitative modeling for the true boiling point curve and specific gravity. *Fuel*; 86:1927–34. 2007

PRADELLE, F., BRAGA, S. L., MARTINS, A. R. F. A., TURKOVICS, F., PRADELLE, R. N. C. Gum Formation in Gasoline and Its Blends: A review. *Energy Fuels*, 29. 7753 – 7770. 2015

REBOUCAS, M.V., NETO, B. de B. Near infrared spectroscopic prediction of physical properties of aromatics-rich hydrocarbon mixtures, *J. Near Infrared Spectrosc.*, 9. 263–273. 2001

REBOUCAS, M.V., SANTOS, J.B., PIMENTEL, M.F., TEIXEIRA, L.S.G. A novel approach for development of a multivariate calibration model using a Doehlert experimental design: Application for prediction of key gasoline properties by Near-infrared Spectroscopy. *Chemometrics and Intelligent Laboratory Systems*, 107. 185–193. 2011

RINNAN, A. Pre-processing in vibrational spectroscopy – when, why and how. *Anal. Methods*, 6. 7124–7129. 2014

SANDOVAL, D. J. Modelo Predictivo para la Estabilidad a la Oxidación de la Gasolina de la Gerencia Refinería Barrancabermeja (GRB). Trabajo de Aplicación Final. Maestría en Ingeniería Química, UIS. 2017

SASTRY, M.I.S., CHOPRA, A., SARPAL, A.S., JAIN, S.K., SRIVASTAVA, S.P., BHATNAGAR, A.K. Determination of Physicochemical Properties and Carbon-Type Analysis of Base Oils Using Mid-IR Spectroscopy and Partial Least-Squares Regression Analysis. *Energy & Fuels*, 12. 304-311. 1998

SINGH, E.; BADRA, J.; MEHL, M.; & SARATHY, S. M. Chemical Kinetic Insights into the Octane Number and Octane Sensitivity of Gasoline Surrogate Mixtures. *Energy & Fuels*, 31(2), 1945–1960. 2017

VAINIOTALO, S.; PELTONEN, Y.; RUONAKANGAS, A.; PFÄFFLI, P. Customer exposure to MTBE, TAME, C6 alkyl methyl ethers, and benzene during gasoline refueling. *Environmental Health Perspectives*, 107, 133-140. 1999

WANG, C.; PRAKASH, A.; ARADI, A.; CRACKNELL, R.; XU, H. Significance of RON and MON to a modern DISI engine. *Fuel*, 209, 172-183. 2017

WOLD, S., ESBENSEN, K., GELADI, P. Principal Component Analysis. *Chem. Int. Lab. Syst.*, 2. 37-52. 1987

WOLD, S., SJOSTROM, M., ERIKSSON, L. PLS-regression: a basic tool for chemometrics. *Chem. Int. Lab. Syst.*, 58. 109-130. 2001

ANEXOS

ANEXO A. Instrucciones utilizadas en R para el ajuste de los modelos de regresión por PLS.

a) RON:

```
> rNIR.ron<-plsr(ron ~NIR, ncomp=10, data = rNIR, validation = "LOO",jackknife =TRUE)
> plot(rNIR.ron, "val")
> coefplot(rNIR.ron, ncomp = 5,se.whiskers=TRUE)
> predplot(rNIR.ron, ncomp = 3, which = c("train","validation"),line=TRUE)
> re<-residuals(rNIR.ron)
> re<-re[,1,3]
> plot(re)
```

b) MON:

```
> rNIR.mon<-plsr(mon ~NIR, ncomp=10, data = rNIR, validation = "LOO",jackknife
=TRUE)
> plot(rNIR.mon, "val")
> coefplot(rNIR.mon, ncomp = 5,se.whiskers=TRUE)
> predplot(rNIR.mon, ncomp = 3, which = c("train","validation"),line=TRUE)
> re<-residuals(rNIR.mon)
> re<-re[,1,3]
> plot(re)
```

c) Contenido de Aromáticos:

```
> rNIR.arom<-plsr(arom ~NIR, ncomp=10, data = rNIR, validation = "LOO",jackknife
=TRUE)
> plot(rNIR.arom, "val")
```

```

> coefplot(rNIR.arom, ncomp = 5, se.whiskers=TRUE)
> predplot(rNIR.arom, ncomp = 3, which = c("train", "validation"), line=TRUE)
> re<-residuals(rNIR.arom)
> re<-re[,1,3]
> plot(re)

```

d) Contenido de Benceno:

```

> rNIR.ben<-plsr(ben ~NIR, ncomp=10, data = rNIR, validation = "LOO", jackknife
=TRUE)
> plot(rNIR.ben, "val")
> coefplot(rNIR.ben, ncomp = 5, se.whiskers=TRUE)
> predplot(rNIR.ben, ncomp = 5, which = c("train", "validation"), line=TRUE)
> y<-rNIR$ben
> re<-residuals(rNIR.ben)
> re<-re[,1,5]
> plot(re/y*100)
> plot(re)

```