

**DESARROLLO DE MODELOS PREDICTIVOS DE LAS PROPIEDADES
FISICOQUÍMICAS PRESENTES EN LOS FONDOS DE VACÍO A TRAVÉS DEL
MÉTODO DE MÍNIMOS CUADRADOS PARCIALES (PLS) Y
ESPECTROSCOPIA INFRARROJA**

OSCAR MAURICIO VALENCIA LOZANO

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOQUÍMICAS
ESCUELA DE INGENIERÍA QUÍMICA
BUCARAMANGA**

2018

**DESARROLLO DE MODELOS PREDICTIVOS DE LAS PROPIEDADES
FISICOQUÍMICAS PRESENTES EN LOS FONDOS DE VACÍO A TRAVÉS DEL
MÉTODO DE MÍNIMOS CUADRADOS PARCIALES (PLS) Y
ESPECTROSCOPIA INFRARROJA**

OSCAR MAURICIO VALENCIA LOZANO

**Plan de trabajo de grado presentado como requisito parcial para optar al
título de Ingeniero Químico**

Director

ENRIQUE MEJÍA OSPINO, QUÍMICO, PHD

Codirector

LEIDY XIMENA CALDERON CASCAVITA

Química

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOQUÍMICAS
ESCUELA DE INGENIERÍA QUÍMICA
BUCARAMANGA**

2018

CONTENIDO

	pág.
INTRODUCCIÓN	12
1. FORMULACIÓN DEL PROBLEMA Y JUSTIFICACIÓN	18
1.1. FORMULACIÓN DEL PROBLEMA	18
1.2. DELIMITACIÓN DEL PROBLEMA	18
1.2.1. Espacial	18
1.2.2. Conceptual	18
1.2.3. Cronológica	18
1.3 ALCANCE	18
1.4 JUSTIFICACIÓN	19
2. OBJETIVOS	21
2.1 OBJETIVO GENERAL	21
2.2 OBJETIVOS ESPECÍFICOS	21
3. MARCO TEÓRICO	22
3.1 GENERALIDADES DE LOS FONDOS DE VACÍO	22
3.2 COMPUESTOS DE AZUFRE	22
3.3 CARBÓN CONRADSON O RCC	23
4. DESCRIPCIÓN METODOLÓGICA	24
4.1 DISEÑO DE EXPERIMENTOS	24
4.2 MUESTRAS	25
4.3 INSTRUMENTACIÓN	26
4.4 MONTAJE EXPERIMENTAL	26
4.5 DETERMINACIÓN DE PARÁMETROS	27

4.6 REPETIBILIDAD	27
4.7 PROCESAMIENTO DE DATOS	27
4.8 REGIONES ESPECTRALES	29
4.9 ANÁLISIS POR COMPONENTES PRINCIPALES (PCA)	29
4.10 MÍNIMOS CUADRADOS PARCIALES (PLS)	30
4.11 FACTORES A INCLUIR EN EL MODELO	31
4.11.1 Detección de outliers	31
4.11.2 Leverage	31
4.11.3 Estadístico T^2 de Hotelling	31
4.12 ERRORES DE CALIBRACIÓN Y PREDICCIÓN	32
4.12.1 Raíz cuadrada del error cuadrado medio de calibración	32
4.12.2	32
5. RESULTADOS: PRESENTACIÓN, ANÁLISIS Y DISCUSIÓN	33
5.1 ANÁLISIS POR COMPONENTES (PCA)	33
5.2 DESARROLLO DE MODELOS PLS	37
5.2.1 Modelo PLS para la predicción de %w de azufre	37
5.3 DISCUSIÓN	44
6. CONCLUSIONES	46
BIBLIOGRAFÍA	47

LISTA DE FIGURAS

Figura 1. Esquema metodológico	pág. 24
--------------------------------	------------

LISTA DE GRÁFICOS

	pág.
Gráfico 1. Espectroscopia MIR-ATR	26
Gráfico 2. Espectro original de las 49 muestras de crudo	28
Gráfico 3. Espectro normalizado suavizado y derivado	28
Gráfico 4. Varianza explicada en el modelo PCA	34
Gráfico 5. Gráfica de scores en dos dimensiones para los 49 fondos de vacío	34
Gráfico 6. Estadístico Hotelling's T^2 elipse aplicado a la gráfica de score de los 49 fondos de vacío	35
Gráfico 7. Grafica de influencia para los 49 fondos de vacío	36
Gráfico 8. Grafica de los errores RMSEC y RMSECV para el modelo PLS del azufre	38
Gráfico 9. Coeficiente de regresión para el modelo PLS del azufre	39
Gráfico 10. Curva de calibración y validación del modelo de predicción de %w de azufre	40
Gráfico 11. Curva de calibración y validación del modelo de predicción de %w de RCC	42
Gráfico 12. Gráfica del modelo PLS del %w de RCC	43

LISTA DE TABLAS

	pág.
Tabla 1. %w azufre y %w RCC de los fondos de vacío	25
Tabla 2. Varianza explicada por los componentes principales de los 49 fondos de vacío.	33
Tabla 3. Parámetros estadísticos del modelo desarrollado para la predicción %w de azufre	37
Tabla 4. Parámetros estadísticos del modelo generado para la predicción de %w de RCC en muestras de fondos de vacío	41
Tabla 5. Comparativo de los modelos predictivos analizados con otros investigadores	44

RESUMEN

Título: DESARROLLO DE MODELOS PREDICTIVOS DE LAS PROPIEDADES FISICOQUÍMICAS PRESENTES EN LOS FONDOS DE VACÍO A TRAVÉS DEL MÉTODO DE MÍNIMOS CUADRADOS PARCIALES (PLS) Y ESPECTROSCOPIA INFRARROJA*

Autor: OSCAR MAURICIO VALENCIA LOZANO**.

Palabras Clave: Fondos de Vacío, Espectroscopia Infrarroja, PCA, PLS

Descripción

El propósito principal de este proyecto es saber cómo influye los modelos predictivos de propiedades fisicoquímicas de fondos de vacío, utilizando como técnica el método de mínimos cuadrados parciales (PLS) y espectroscopia infrarroja para determinar la calidad y ciertas propiedades de los crudos; por tal razón, este estudio al desarrollar dichos modelos predictivos de propiedades fisicoquímicas busca encontrar las causas de las condiciones proporcionadas a través del método, donde se ordena un conjunto de datos de acuerdo a su importancia. Además, realiza un análisis de componentes principales (PCA) de los espectros correspondientes a las muestras de fondo de vacío y definir el número de componentes principales que mejora la incertidumbre en los datos, determinando las causas de la variabilidad de un conjunto de datos y ordenarlas por importancia, para proyectar los datos en términos de mínimos cuadrados. Los análisis de componentes principales (PCA) permitieron identificar de manera cualitativa el parentesco de los 49 espectros correspondientes a las 49 muestras de fondos de vacío y definir el número de componentes principales que mejora la incertidumbre en los datos; así como el análisis por regresión parcial por mínimos cuadrados (PLS), permitieron identificar de manera cuantitativa el parentesco de los 49 espectros correspondientes a las 49 muestras de fondos de vacío para determinar la composición de azufre y RCC. En conclusión, el modelo mostró un desempeño satisfactorio para la predicción composicional de las ya mencionadas propiedades.

* Trabajo de grado

** Facultad de ingenierías fisicoquímicas. Escuela de ingeniería química. Director: Enrique Mejía Ospino. Co director: Leidy Ximena Calderon Cascavita.

ABSTRACT

Title: DEVELOPMENT OF PREDICTIVE MODELS OF PHYSICOCHEMICAL PROPERTIES PRESENT IN VACUUM FUNDS THROUGH THE METHOD OF PARTIAL MINIMUM SQUARES (PLS) AND INFRARED SPECTROSCOPY*

Author: Oscar Mauricio Valencia Lozano**

Keywords: Vacuum Funds, Infrared Spectroscopy, PCA, PLS

Description

The main purpose of this project is to know how it influences the predictive models of physicochemical properties of vacuum bottoms, using as a technique the method of partial least squares (PLS) and infrared spectroscopy to determine the quality and certain properties of the crude; for this reason, this study in developing such predictive models of physicochemical properties seeks to find the causes of the conditions provided through the method, where a set of data is ordered according to its importance. In addition, it performs a principal components analysis (PCA) of the spectra corresponding to the vacuum bottom samples and defines the number of main components that improves the uncertainty in the data, determining the causes of the variability of a data set and ordering them by importance, to project the data in terms of least squares. The principal components analysis (PCA) allowed qualitatively identify the kinship of the 49 spectra corresponding to the 49 samples of vacuum funds and define the number of main components that improves the uncertainty in the data; as well as the analysis by partial regression by least squares (PLS), allowed to identify in a quantitative way the kinship of the 49 spectra corresponding to the 49 samples of vacuum funds to determine the composition of sulfur and RCC. In conclusion, the model showed a satisfactory performance for the compositional prediction of the aforementioned properties.

* Bachelor Thesis

** Faculty of physicochemical engineering. School of chemical engineering. Directress: Enrique Mejia Ospino. Codirectress: Leidy Ximena Calderon Cascavita.

INTRODUCCIÓN

Actualmente el mundo está sujeto a grandes cambios, mayor desarrollo, aumento de la calidad en los procesos y mejora en las condiciones de vida, producto de la creciente globalización; exigiendo a las empresas lograr una ventaja competitiva, a la vanguardia e innovando permanentemente los productos y servicios que ofrecen. Con respecto a las industrias, el desarrollo está relacionado con la capacidad energética y los métodos empleados para su transformación.

Por consiguiente, Paredes y Ramírez,¹ establecen la importancia de la energía renovable, mediante recursos o medios naturales, a través del sol, el agua o el aire, de acuerdo al crecimiento económico; sin embargo, la importancia del uso de las energías no renovables como lo es el petróleo, son una fuente de energía potente para el movimiento de grandes motores, pero con efecto o impacto ambiental negativo. Así mismo Kharrat, Zacharia, Cherian y Anyatonwu,² señalan que el petróleo como una fuente de energía no renovable estudiada desde hace 100 años, su apariencia varía desde breas sólidas a líquidos claros, su composición, e incluso en su mismo campo de un pozo a otro y su profundidad. Por ende, es difícil determinar la composición de un crudo, por su gran número de componentes, que pueden llegar a decenas de miles, haciendo difícil una caracterización molecular completa. No obstante, la caracterización molecular de un crudo es importante debido a que sus propiedades fisicoquímicas están relacionadas con su composición.³

¹ PAREDES, Juan Roberto y RAMÍREZ, John. Energías renovables variables y su contribución a la seguridad energética: complementariedad en Colombia [en línea]. BID, 2017. Disponible en Internet: <URL: <https://publications.iadb.org/bitstream/handle/11319/8146/Energias-renovables-variables-y-su-contribucion-a-la-seguridad-energetica-Complementariedad-en-Colombia.PDF?sequence=5>>

² KHARRAT, A. M., ZACHARIA, J., CHERIAN, V. J. y ANYATONWU, A. Issues with comparing sara methodologies [en línea]. *En: Energy and fuels* 21, 2007, p. 3618–3621. Disponible en Internet: <URL: <https://bibliotecavirtual.uis.edu.co:2277/doi/pdf/10.1021/ef700393a>>

³ HAMADOU, R., KHODJA, M., KARTOUT, M. y JADA, A. Permeability reduction by asphaltenes and resins deposition in porous media [en línea]. *En: Fuel* 87, 2008, p. 2178–2185. Disponible en Internet: <URL: <https://bibliotecavirtual.uis.edu.co:2191/science/article/pii/S001623610700542X>>

En consecuencia, se señala la importancia de los análisis de porcentaje del peso de azufre utilizado para determinar la corrosividad potencial de los fondos de vacío, ciertos compuestos de este que son corrosivos por su acidez como SH_2 y mercaptanos, se eliminan con tratamientos químicos; la norma ASTM D-4294 es crucial para determinar la composición del mismo. Por ello, el análisis de porcentaje del peso RCC es utilizado para proporcionar una indicación de las tendencias de formación de coque de un aceite, en un análisis cuantitativo, la prueba logra determinar la cantidad de residuo carbonoso que queda después de la evaporación y la pirolisis del aceite.

Por esta razón, a nivel mundial, el uso del petróleo juega un papel muy importante que influye en la población y vastos sectores económicos como elemento esencial para la movilización de pequeños y grandes vehículos; su composición media puede ser de un 85% de carbono, 12% de hidrógeno, 3% de la suma de elementos azufre, oxígeno y nitrógeno y varios componentes metálicos, imposibilitando así a la energía natural para remplazar el efecto que estos elementos producen en determinados aparatos, motores u otros medios que requiere mayor combustión para su movilización.

Otro aspecto se vincula a los métodos de predicción para establecer la calidad de los procesos de extracción, producción, refinación y transporte de los crudos; para ello existen técnicas analíticas, como la espectrometría de masas (MS) y espectroscopias de fluorescencia, Raman,⁴ las cuales han sido ampliamente aplicadas en el análisis de hidrocarburos y sus derivados arrojando consistentes resultados; siendo algunas de ellas muy costosas.⁵ De igual forma se utilizan

⁴ FREDERICKS, Peter; RINTOUL, Llewellyn and COATES, John. Chapter 7. Vibrational Spectroscopy: Instrumentation for Infrared and Raman Spectroscopy. En: CAZES, Jack. Analytical Instrumentation Handbook. 2008, p.76.

⁵ BORGES, B. y ACEVEDO, S. Caracterización estructural de distintas fracciones aisladas de crudos extrapesado Carabobo [en línea]. En: Revista latinoamericana de metalurgia y materiales.

técnicas de análisis multivalente llamadas quimiométricas, aplicadas a datos espectroscópicos que son obtenidos por el análisis de componentes principales (PCA) y la regresión por mínimos cuadrados parciales (PLS), permitiendo avanzar en la identificación y determinación de las propiedades fisicoquímicas de las muestras.⁶ Por todo lo anteriormente mencionado, se tiene como propósito desarrollar un modelo predictivo de propiedades fisicoquímicas de fondos de vacío, utilizando mínimos cuadrados parciales (PLS) y espectroscopia infrarroja.

Desde esta perspectiva, el proyecto investigativo pretende dar respuesta al siguiente interrogante: ¿Cómo influyen los modelos predictivos de propiedades fisicoquímicas de fondos de vacío, utilizando mínimos cuadrados parciales (PLS), análisis por componentes principales (PCA), y espectroscopia infrarroja para determinar la calidad y ciertas propiedades de los crudos?

Ajustándose a la realidad del contexto colombiano y su necesidad imperante de analizar de manera rápida los crudos y sus fracciones como son los fondos de vacío, relevantes por cuanto se tienen crudos pesados con un alto porcentaje de fondos de vacío y estos hay que tratarlos catalíticamente para poder seguir extrayendo compuestos como la gasolina o productos de valor agregado.

En concordancia, las propiedades fisicoquímicas juegan un papel importante al permitir determinar la cantidad de alcanos, el punto de ebullición, el punto de fusión, la densidad, los alquenos, los alquinos, la solubilidad de todos los hidrocarburos y que junto al Instituto Colombiano del Petróleo (ICP) se está trabajando para obtener métodos de análisis rápidos como los espectros de infrarrojos, reemplazando el análisis de esas propiedades a través de los métodos

2007. Vol. 2 p 83-94. Disponible en Internet: <URL: http://www.scielo.org.ve/scielo.php?pid=S0255-69522007000200003&script=sci_arttext>

⁶ KALLEVIK, H. Characterization of crude oil and model oil emulsions by means of near infrared spectroscopy and multivariable analysis [en línea]. Noruega, 1999. Disponible en Internet: <URL: <https://www.osti.gov/etdeweb/biblio/20119612>>

convencionales que causan demoras, elevan los costos y consumen mucho dinero.

Los diferentes estudios encontrados abordan la problemática con modelos predictivos matemático-estadísticos con aplicación de espectroscopia fotoacústica como el desarrollado por Leguizamo,⁷ el cual analiza 20 muestras de fondo de vacío con equipos que difieren del actual proyecto, lo que permite comparar el desempeño del modelo presentado.

Un segundo trabajo lo realiza Parra,⁸ quien igualmente utiliza la espectroscopia fotoacústica con transformada de Fourier, la cual hace uso de referencias primarias (resultados de laboratorio) y secundarias (sistema de información), el trabajo principalmente representa un aporte para el presente proyecto por la metodología empleada.

Por otro lado, Rodríguez⁹ emplea la espectroscopia de fluorescencia y con aplicación de regresión por mínimos cuadrados (PLS), relevante para el presente estudio, investigación que permite comparar los modelos de predicción del azufre.

En la misma línea de investigación se encuentra Ballesteros,¹⁰ la cual utiliza un espectrómetro de fluorescencia inducido por láser, para predecir la composición

⁷ LEGUÍZAMO CASTELLANOS, Cristian Ferney. Predicción del análisis elemental por espectroscopia fotoacústica en la región del infrarrojo medio y métodos quimiométricos a las fracciones de los fondos de vacío [en línea]. Trabajo de Grado de Químico, UIS. Bucaramanga, 2012. Disponible en Internet: <URL: <http://tangara.uis.edu.co/biblioweb/tesis/2012/143262.pdf>>

⁸ PARRA PICÓN, Sara Inés. Predicción de la curva de destilación de fondos de vacío de crudos colombianos usando espectroscopia fotoacústica con transformada de Fourier en el infrarrojo medio y quimiometría [en línea]. Trabajo de Grado de Químico, UIS. Bucaramanga, 2012. Disponible en Internet: <URL: <http://tangara.uis.edu.co/biblioweb/tesis/2012/146837.pdf>>

⁹ RODRÍGUEZ CAMACHO, Sidne Stefany. Predicción de la composición elemental de fracciones del proceso de Delayed Coking de fondos de vacío de crudos colombianos empleando espectroscopia de fluorescencia y métodos quimiométricos [en línea]. Trabajo de Grado de Químico, UIS. Bucaramanga, 2014. Disponible en Internet: <URL: <http://tangara.uis.edu.co/biblioweb/tesis/2014/152138.pdf>>

de resinas, asfaltenos, aromáticos y saturados (SARA) en fondos de vacío, metodología que sirvió de guía para el presente proyecto.

En consecuencia, se puede observar que los datos obtenidos por espectroscopia se introducen en un programa estadístico, que en el presente caso se utiliza the Unscrambler, obteniendo un modelo predictivo y una curva de calibración, que señala ciertas propiedades de los crudos.

Por esta razón, se hace necesario llevar a cabo un estudio que desarrolle modelos predictivos de propiedades fisicoquímicas a través del método de mínimos cuadrados parciales y espectroscopia infrarroja, hallando las causas de las condiciones proporcionadas a través del método, donde se ordena un conjunto de datos de acuerdo a su importancia.

El presente trabajo establece como objetivo general: desarrollar un modelo predictivo de propiedades fisicoquímicas de fondos de vacío mediante el análisis por regresión parcial por mínimos cuadrados (PLS) análisis por componentes principales (PCA), y espectroscopia infrarroja para determinar ciertas propiedades de los crudos.

El desarrollo del trabajo se fundamenta en los siguientes referentes teóricos, Spiecker, Gawrys, Trail, y Kil Patrick,¹¹ quienes definen el petróleo como una mezcla de gran cantidad de compuestos orgánicos, formados a partir de residuos animales y vegetales, se encuentran sometidos a determinadas condiciones de temperatura y presión, que junto a una acción bacteriana anaeróbica da lugar a los

¹⁰ BALLESTEROS LANCHEROS, Sandra Lucía. Predicción de las fracciones SARA de fondos de vacío de crudos colombianos, por métodos quimiométricos utilizando espectroscopia de fluorescencia inducida por láser (LIF) [en línea]. Trabajo de Grado de Químico, UIS. Bucaramanga, 2010. Disponible en Internet: <URL: <http://tangara.uis.edu.co/biblioweb/tesis/2010/133470.pdf>>

¹¹ SPIECKER, P. M., GAWRYS, K. L., TRAIL, C. B. y KILPATRICK, P. K. Effects of petroleum resins on asphaltene aggregation and water-in-oil emulsion formation. *Colloids surfaces a physicochem* [en línea]. Eng. Asp. 220, 2003, p. 9–27. Disponible en Internet: <URL: <https://www.sciencedirect.com/science/article/pii/S0927775703000797?via%3Dihub>>

hidrocarburos. Trejo y Rana,¹² definen el petróleo como un fluido complejo que contiene componentes de acuerdo a su peso molecular; el número de átomos de carbono de sus componentes proporciona al petróleo diferentes propiedades físicas y químicas, así, para hidrocarburos compuestos por uno a cuatro átomos de carbono son gaseosos, los que contienen de 5 a 20 son líquidos, y los de más de 20 son sólidos a la temperatura ambiente.

Conaway,¹³ manifiestan que los crudos tienen características físicas y químicas muy variables, dependiendo de un campo a otro e incluso dentro de un mismo yacimiento, por lo tanto, el petróleo se clasifica en crudos pesados y livianos y esto va de acuerdo a la densidad que la conforma. De acuerdo al contenido de compuestos con azufre, el crudo puede clasificarse como agrio, si presenta cantidades apreciables de estos compuestos o como dulce, si presenta cantidades muy pequeñas.

¹² TREJO, F. y RANA, M. S.; Ancheyta J. Thermogravimetric determination of coke from asphaltenes, resins and sediments and coking kinetics of heavy crude asphaltenes [en línea]. *Catalysis today* 150, 2010, p. 272-278. Disponible en Internet: <URL: <https://bibliotecavirtual.uis.edu.co:2150/record/display.uri?eid=2-s2.0-77949658688&origin=inward&txGid=c228bef6087f5f2750986e738f2c12d1>>

¹³ CONAWAY, Charles. *The petroleum industry: a nontechnical al guide*. Tulsa; pennwell publishing co. Oklahoma: PennWell, 1999.

1. FORMULACIÓN DEL PROBLEMA Y JUSTIFICACIÓN

1.1. FORMULACIÓN DEL PROBLEMA

¿Cómo influyen los modelos predictivos de propiedades fisicoquímicas de fondos de vacío, utilizando mínimos cuadrados parciales (PLS), análisis por componentes principales (PCA), y espectroscopia infrarroja para determinar la calidad y ciertas propiedades de los crudos?

1.2. DELIMITACIÓN DEL PROBLEMA

1.2.1. Espacial. el estudio será realizado en la ciudad de Bucaramanga en el laboratorio leam de Guatiguará de la universidad industrial de Santander (UIS).

1.2.2. Conceptual. todos los conceptos relacionados con el estudio, serán realizados dentro del referente teórico.

1.2.3. Cronológica. el desarrollo del estudio se realizará durante el primer semestre del 2017. Terminando el primer semestre del 2018 se desarrollará proyecto y además se estarán realizando sus ajustes y correcciones para sustentación.

1.3 ALCANCE

Con el presente proyecto se pretende desarrollar un modelo predictivo de las propiedades fisicoquímicas, presentes en los fondos de vacío a través del método de mínimos cuadrados parciales (PLS), análisis por componentes principales (PCA), y espectroscopia infrarroja.

1.4 JUSTIFICACIÓN

En Colombia hay una necesidad importante de analizar de manera rápida los crudos y sus fracciones, dentro de las fracciones la de fondos de vacío es muy importante, porque en Colombia, la mayoría de los crudos que se encuentran son crudos pesados, los crudos pesados tienen un alto porcentaje de fondo de vacíos y estos hay que tratarlos catalíticamente para poder seguir extrayendo de esos fondos de vacíos cosas como la gasolina o productos de valor agregado.

Las propiedades fisicoquímicas, permiten determinar la cantidad de alcanos, el punto de ebullición, el punto de fusión, la densidad, los alquenos, los alquinos, la solubilidad de todos los hidrocarburos. Con el ICP se está trabajando para obtener métodos de análisis rápidos en lugar de analizar esas propiedades a través de los métodos convencionales que son muy demorados, y consumen mucho dinero; entonces lo que se hace es tomar espectros infrarrojos o Raman a los fondos de vacío. Tomar estos datos que se obtienen por espectroscopia, se introducen en unos modelos estadísticos, utilizando the Unscrambler, o software determinados. Con esa estadística obtenida de esos datos, se obtiene un modelo, una curva de calibración, un modelo predictivo que permite determinar ciertas propiedades de los crudos.

Por tal razón, realizar un estudio para desarrollar modelos predictivos de propiedades fisicoquímicas a través del método de mínimos cuadrados parciales y espectroscopia infrarroja, con el fin de hallar las causas de las condiciones proporcionadas a través del método, donde se ordena un conjunto de datos de acuerdo a su importancia.

Desde el punto de vista personal, la realización del estudio es importante, primero porque es una iniciativa para establecer un modelo predictivo eficiente, utilizando como técnica mínimos cuadrados y segundo porque da la oportunidad de

desarrollar un proyecto de investigación real, que ayuda a desarrollar modelos predictivos, mediante el uso de datos estadísticos, a través de una revisión sistemática a un problema real y concreto.

2. OBJETIVOS

2.1 OBJETIVO GENERAL

Desarrollar un modelo predictivo de propiedades fisicoquímicas de fondos de vacío mediante el análisis por regresión parcial por mínimos cuadrados (PLS) análisis por componentes principales (PCA), y espectroscopia infrarroja para poder determinar ciertas propiedades de los crudos.

2.2 OBJETIVOS ESPECÍFICOS

- Observar el fenómeno de la absorción infrarroja para muestras de fondos de vacío y comprender su importancia para el análisis de crudos.
- Desarrollar nuevas metodologías aplicadas a la predicción de propiedades fisicoquímicas, mediante la espectroscopia MIR-ATR en combinación con técnicas de calibración multivariable para la determinación del contenido de azufre y RCC en muestras de fondos de vacío.
- Analizar parámetros estadísticos e interpretar los comportamientos de los modelos generados para la predicción del contenido de azufre y RCC.
- Establecer los aportes del desarrollo del modelo a la investigación en la ingeniería química.

3. MARCO TEÓRICO

3.1 GENERALIDADES DE LOS FONDOS DE VACÍO

Altgelt y Boduszynski,¹⁴ definen los fondos de vacío como fracciones pesadas que provienen de la destilación al vacío del petróleo y pueden ser utilizados para alimentar las unidades de craqueo térmico, donde se producen productos más ligeros, o para alimentar la unidad de producción de coque. El uso de fondos de vacío es muy limitado, ya que considerablemente afectan el medio ambiente porque presentan en su estructura azufre y nitrógeno.¹⁵ Su composición química, según Morawski y Mosio-Mosiewski,¹⁶ específicamente los hidrocarburos presentes en esta fracción del petróleo poseen elevados puntos de ebullición (generalmente más de 500 °C) y moléculas complejas, ya que presenta una diversidad de estructuras químicas cuyos pesos moleculares varían en un amplio rango.

3.2 COMPUESTOS DE AZUFRE

El petróleo tiene compuestos de azufre, desde 0,03% capaz de determinar la corrosividad potencial del crudo, el cual contiene SH₂ y se denomina "crudo ácido" y su compuesto son complejos y generalmente térmicamente inestables, los cíclicos como el tiofeno son más estables. La determinación de azufre se efectúa según la norma ASTM D-4294. Actualmente se utiliza la técnica de quimioluminiscencia. La radiación característica derivada a partir de una fuente de

¹⁴ ALTGELT, Klaus H. y BODUSZYNSKI, Mieczslaw. Composition and analysis of heavy petroleum fractions. Segunda edición. New York: CRC Press, 1994. 495 p.

¹⁵ MORAWSKI, I. y MOSIO-MOSIEWSKI, J effects of parameters in ni-mo catalyzed hydrocracking of vacuum residue on composition and quality of obtained products [en línea]. *En: Fuel Processing Technology* 87 (2006) 659-669. Disponible en Internet: <URL: <https://www.sciencedirect.com/science/article/pii/S037838200600021X?via%3Dihub>>

¹⁶ *Ibid.*, p.659

rayos x, es comparada con la de muestras patrón, cuyo contenido de azufre (% en peso) es conocido.

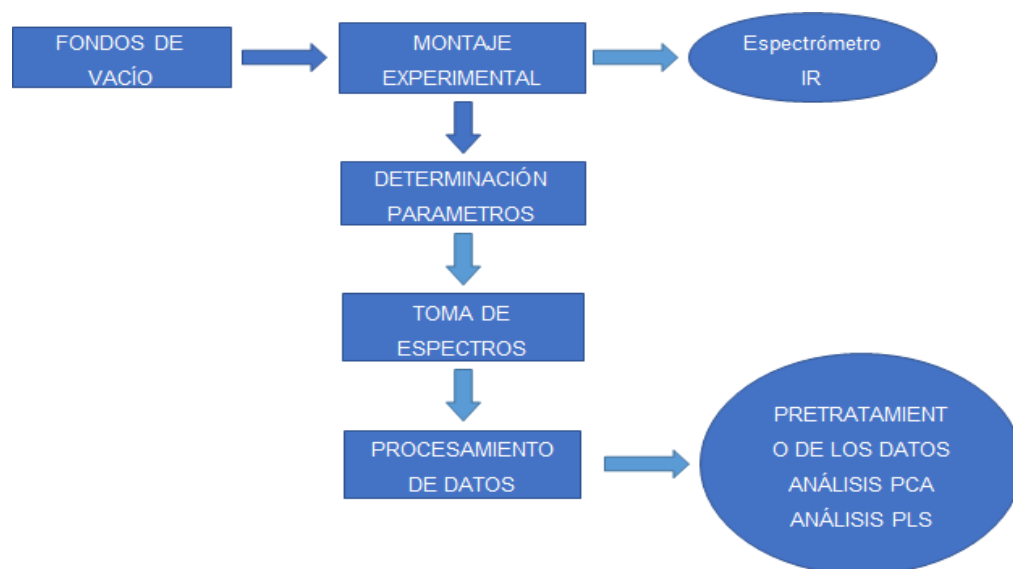
3.3 CARBÓN CONRADSON O RCC

Es un método utilizado para proporcionar una indicación de las tendencias de formación de coque de un aceite; cuantitativamente, la prueba mide la cantidad de residuo carbonoso que queda después de la evaporación y la pirolisis del aceite. En general, el ensayo es aplicable a productos derivados del petróleo que son relativamente no volátiles, y que se descomponen en la destilación a presión atmosférica. La frase "conradson carbono de residuos" y sus nombres comunes pueden hacer referencia a la prueba o el valor numérico obtenido de ella.

4. DESCRIPCIÓN METODOLÓGICA

En la siguiente figura se observa el esquema metodológico de los pasos que se realizaron para el desarrollo del proyecto. Para el desarrollo del proyecto se tuvo en cuenta las pruebas que realiza el ICP en sus laboratorios, acerca del análisis de muestras de los diferentes pozos que existen en Colombia.

Figura 1. Esquema metodológico



4.1 DISEÑO DE EXPERIMENTOS

Teniendo en cuenta la diversidad en los procesos de extracción del crudo, se tomaron muestras aleatorias en los diferentes pozos de Colombia, de las cuales llegaron al laboratorio del ICP para su posterior análisis; una vez analizadas se entregaron las muestras ya analizadas con el respectivo % de peso. Se procedió a hacer el montaje del experimento y a determinar los parámetros necesarios para los modelos predictivos de cada uno de los porcentajes (% de azufre y % RCC) como se detalla a continuación.

4.2 MUESTRAS

Se analizaron 49 muestras de fondos de vacío con diferentes características que fueron provistos por el Instituto Colombiano de Petróleo (ICP), proporcionando el %w azufre y el %w RCC (carbono conradson) como se observa en el anexo (tabla 1); a través de las normas ASTM D-4124.

Tabla 1. %w azufre y %w RCC de los fondos de vacío

Propiedad	(% w) azufre	(%w) RCC	Propiedad	(% w) azufre	(%w) RCC
Fv1	2.14	20.10	Fv49	2.01	39.3
Fv2	1.14	28.80	Fv50	2.09	23.9
Fv3	1.89	21.9	Fv51	3.74	32.1
Fv4	1.87	37.6	Fv52	1.80	21.6
Fv6	1.56	24.9	Fv53	1.70	21.8
Fv7	2.35	36.53	Fv54	3.06	20.4
Fv8	2.96	19.30	Fv55	4.26	32.50
Fv9	2.74	35.50	Fv57	0.91	20.60
Fv29	2.94	12.85	Fv58	4.33	32.6
Fv30	3.81	32.20	Fv60	1.07	16.3
Fv31	3.54	35.4	Fv61	3.31	37.0
Fv32	4.09	32.4	Fv62	3.33	33.4
Fv33	2.48	19.8	Fv63	1.46	19.4
Fv34	1.72	24.8	Fv64	2.9	23.6
Fv35	1.14	18.20	Fv65	1.59	19.6
Fv36	1.53	13.5	Fv66	4.10	32.0
Fv37	2.16	16.3	Fv67	2.28	20.5
Fv38	2.84	18.30	Fv68	1.95	21.5
Fv39	2.84	21.0	Fv70	3.98	30
Fv40	3.27	39.6	Fv72	3.79	35.10
Fv42	3.48	34.0	Fv73	2.32	39.2
Fv44	0.96	26.10	Fv74	3.58	31.90
Fv46	2.49	32.1	Fv75	1.17	25.7
Fv47	3.15	24.20	Fv76	2.77	35.30
Fv48	2.54	33.5			

4.3 INSTRUMENTACIÓN

Se empleó el espectrómetro MIR Thermo Scientific Nicolet IS50 espectrómetro FT-IR con un diamante incorporado reflexión total atenuada (ATR). Los detectores en el banco principal son MCT-A (11700-600 cm^{-1}) y DLATGS (12500-350 cm^{-1}). ATR PIKE Miracle con cristales de diamante de reflexión simple de índice de refracción 2.4 y un detector con control de temperatura (DLATGS) que permite cubrir las regiones de infrarrojo medio en el rango de 4000 a 650 cm^{-1} , ángulo de incidencia del haz IR 45°, velocidad del espejo interferómetro de 2.8mm/s.

Gráfico 1. Espectroscopia MIR-ATR



Fuente. UV/VIS SPECTROSCOPY.

4.4 MONTAJE EXPERIMENTAL

El ICP proporcionó 49 muestras de fondos de vacío, al abrir las latas que contenían dichos fondos de vacío estos se encontraban completamente secos y duros, por lo que se tuvo que golpear el fondo de vacío con un destornillador de pala para extraer pequeños trozos y posteriormente triturados en un mortero para convertirlos en polvo; luego con una espátula de laboratorio se recogieron las partículas que habían quedado después de la trituración; para luego depositarlas en el espectrómetro y proceder a tomar el espectro de dicha muestra.

4.5 DETERMINACIÓN DE PARÁMETROS

Los espectros MIR fueron tomados en el laboratorio de espectroscopia atómica molecular (LEAM). El espectrómetro IS50 se trabajó con un número de escaneos igual a 32, se trabajó con una longitud de onda de 400 cm^{-1} a 4000 cm^{-1} con una resolución de 4 cm^{-1} con una velocidad de escaneo de $0,4747\frac{\text{cm}}{\text{s}}$, a una temperatura de $25\text{ }^{\circ}\text{C}$, estas condiciones mejoran la relación señal / ruido de los espectros.

4.6 REPETIBILIDAD

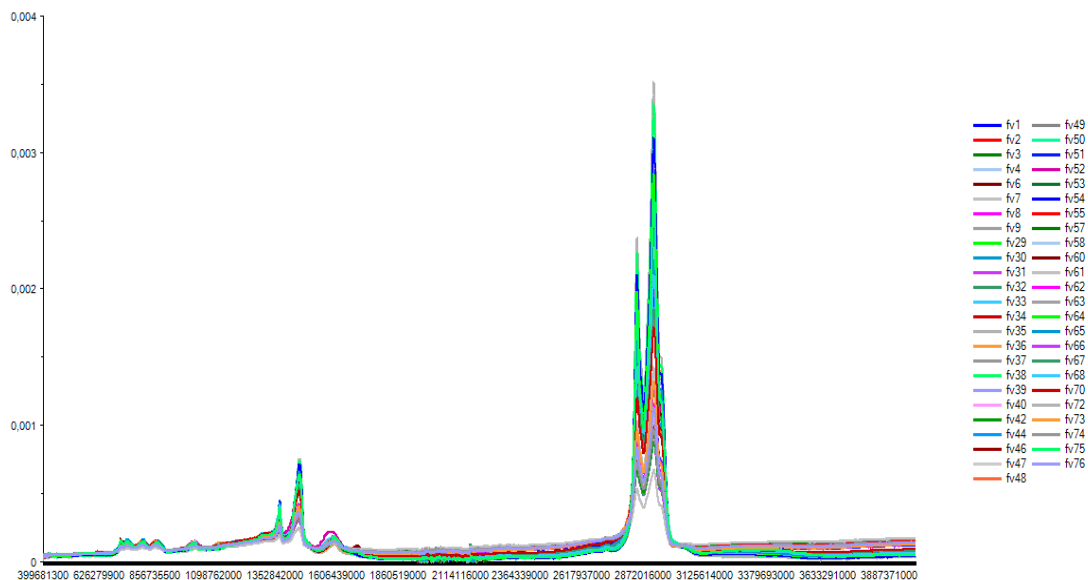
Las 49 muestras se distribuyen en grupos de 10 muestras, cada grupo tiene un patrón, que es la muestra número 5 (según el listado de muestras proporcionadas por el ICP), cada vez que se tomó el espectro en el espectrómetro de 2 muestras se toma de nuevo el espectro del patrón en el espectrómetro, una vez se llega a la muestra número 10, se revuelve al azar y se toman nuevamente los espectros de las 10 muestras como se hizo anteriormente, teniendo en cuenta el patrón, procedimiento que se repite tres veces en forma secuencial.

4.7 PROCESAMIENTO DE DATOS

Los datos espectroscópicos obtenidos se analizaron en el software Unscrambler¹⁷ versión 10.3.

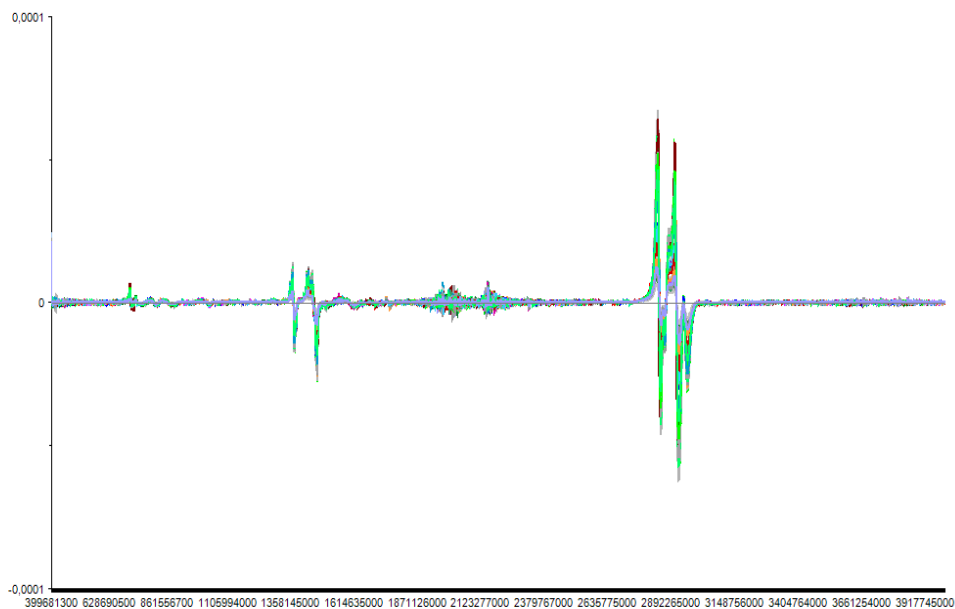
¹⁷ THE UNSCRAMBLER 10.3. Methods. Software para diseño de experimentos y análisis multivariado.

Gráfico 2. Espectro original de las 49 muestras de crudo



Fuente. Programa de análisis multivariado THE UNSCRAMBLER versión 10.3.

Gráfico 3. Espectro normalizado suavizado y derivado



Fuente. Programa de análisis multivariado THE UNSCRAMBLER versión 10.3.

Las técnicas utilizadas para los modelos de predicción fueron las siguientes:

4.8 REGIONES ESPECTRALES

Aunque el espectro de infrarrojo se extiende desde 10 a 14300 cm^{-1} desde un punto de vista funcional se divide en tres zonas: IR lejano, el IR medio (MIR o simplemente IR), y el IR cercano (NIR), donde se producen absorciones debidas a sobre tonos y combinaciones de las bandas fundamentales, en nuestro caso se trabajó el IR medio.

4.9 ANÁLISIS POR COMPONENTES PRINCIPALES (PCA)

El análisis por componentes principales (PCA: Principal Component Analysis) es una técnica de descomposición multivariante, la cual reduce el número de variables perdiendo la menor cantidad de información posible y reduciendo la contribución del ruido. El PCA desarrolla una modificación lineal que selecciona un nuevo sistema de coordenadas para el grupo original de datos denominados componentes principales (PC) el cual explica la máxima dispersión de las muestras; el segundo PC se selecciona de manera que sea ortogonal al primero presentando la máxima dispersión de las muestras una vez restada la explicada por el primer PC y se define a través de la norma de la ASTM E131, para resolver conjuntos de datos hallando componentes o vectores ortogonales entre sí, llamados componentes principales (PC), cuya combinación lineal se aproxima a los datos originales de acuerdo al grado de exactitud deseado. La idea de emplear esta técnica es la de hallar componentes principales $Z_1, Z_2... Z_n$ Que sean acoplamientos lineales de las variables originales X_1, X_2, X_n que detallan cada muestra de acuerdo a la ecuación 1.

$$Z_1 = a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + a_{1n}X_n \text{ (Ecuación 1)}$$

Los coeficientes a_{11} , a_{12} , etc., se escogen de forma que las nuevas variables, a diferencia de las variables originales, no se encuentran correlacionadas unas con otras. Cuando solo hay dos variables y, por lo tanto, solo dos componentes principales, el PCA es equivalente a una rotación de los ejes originales; de tal manera que PC1 se encuentra en la dirección de la máxima variación, pero manteniendo el ángulo entre los ejes.

4.10 MÍNIMOS CUADRADOS PARCIALES (PLS)

La regresión por mínimos cuadrados parciales (PLS) es una técnica matemática que modeló sincronizadamente las matrices x y para hallar un conjunto de variables latentes (VL) en x que mejor predicen las variables latentes en y . Como en el proceso del análisis de componente principal, estas nuevas variables en x y se pueden manifestar como un producto de matrices según se muestra en las ecuaciones (2) y (3).

$$X = TPT + E = \sum t_0 p_0^r + E \text{ (ecuación 2)}$$

$$Y = UQT + F = \sum u_0 q_0^r + E \text{ (ecuación 3)}$$

En el cual:

T y u son las matrices de puntuación (scores) de x y respectivamente;

P y q son las matrices de carga de x y respectivamente;

E y f son los residuos. Las variables originales x y se pueden asociar mediante los scores de cada una de las nuevas variables latentes como lo muestra la (ecuación 4).

$$U_0 = b_0 t_0 \text{ (ecuación 4)}$$

Donde b_0 es el coeficiente de regresión para cada variable latente obtenido a través de la (ecuación 5).

$$b_0 = \frac{U_0^r t_0}{t_0^r t_0} \text{ (ecuación 5)}$$

Los coeficientes b_0 encontrados para cada componente son reunidos en una matriz diagonal b , que contiene los coeficientes de regresión de los scores u y t de las matrices y y x respectivamente. De tal manera, la matriz y puede ser estimada por medio de la (ecuación 6).

$$Y = TBQ^r + F \text{ (ecuación 6)}$$

Los PLS son utilizados para analizar la estructura de covarianza entre los espacios correspondientes a dos matrices x y y , para predecir un grupo de variables dependientes a partir de un grupo enorme de variables independientes.

4.11 FACTORES A INCLUIR EN EL MODELO

4.11.1 Detección de outliers. En la fase de instauración del modelo se puede emplear información de la influencia de los objetos en el grupo de calibración (leverage) y los residuales, tanto en la propiedad de interés como en la respuesta instrumental.

4.11.2 Leverage. Es un valor de la posición (o influencia) de una muestra en relación al modelo. Las muestras con un alto valor de leverage están muy separadas del centro del modelo, por lo que tendrán una influencia muy elevada sobre el mismo. Este valor se calcula como:

$$h = \frac{1}{I} + t_i^r (T^r T)^{-1} t_i \text{ (ecuación 7)}$$

Donde t_i representa el vector de score de la muestra i , t , la matriz de score del modelo y I el número de muestras de calibración.

4.11.3 Estadístico T^2 de Hotelling. Fue planteado originalmente por Hotelling y calcula la variación de cada muestra dentro del modelo PCA, su medición es a través de la suma de los cuadrados de los scores. El grafico T^2 controla la distancia de una nueva medida al valor de referencia en el espacio reducido de los factores PCA y encuentra si la variación añadida en los componentes principales

analizados es mayor que la que le correspondería si solo añadieran variaciones aleatorias.

4.12 ERRORES DE CALIBRACIÓN Y PREDICCIÓN

4.12.1 Raíz cuadrada del error cuadrado medio de calibración. (RMSEC por sus siglas en inglés) se ocupa de detallar el nivel de correspondencia entre las propiedades estimadas por el modelo de calibración para las muestras de calibración y los valores admitidos como verdaderos para las muestras de calibración empleados para obtener los parámetros del modelo.

$$\text{RMSEC} = \sqrt{\left(\frac{\sum_{i=1}^n (y_i - Y_i)^2}{n-m-1}\right)} \quad (\text{Ecuación 8})$$

4.12.2 Raíz cuadrada del error cuadrado medio de calibración cruzada. (RMSECV por sus siglas en inglés) trata en el empleo de los datos del grupo de calibración para validar el modelo. El grupo de calibración inicial se divide en diferentes grupos de cancelación. Cada vez que se calcula el modelo se deja un grupo de cancelación fuera y se calcula la ecuación de la regresión con los demás datos. Una vez elaborada la calibración se ensaya el modelo con los datos del grupo de cancelación excluido, conforme a la ecuación

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_{cvi})^2}{n_c}} \quad (\text{Ecuación 9})$$

4.12.3 Residuales en la respuesta instrumental. Los residuales en la respuesta (o residuales espectrales) revela la falta de ajustes entre las respuestas experimentales utilizadas en la calibración, r y las respuestas restauradas por el modelo con A factores ($R = TP^T$).

$$E = R - TP^T \quad (\text{Ecuación 10})$$

5. RESULTADOS: PRESENTACIÓN, ANÁLISIS Y DISCUSIÓN

A continuación, se presentan los resultados correspondientes al análisis por componentes principales y regresión por mínimos cuadrados realizados, a espectros IR de fondos de vacío.

5.1 ANÁLISIS POR COMPONENTES (PCA)

En primera medida se debe definir el número de componentes principales, mejorando la incertidumbre según la norma ASTM E1655, como alternativa para reducir costes y optimizar tiempo se emplea un espectrómetro. Para comprobar su efectividad se mide el espectro de absorbancia de 7470 longitudes de onda en 49 muestras de azufre y carbono. Esta técnica estadística de síntesis de la información o reducción de la dimensión tiene como objetivo reducirlo a un menor número partiendo de la menor cantidad posible.

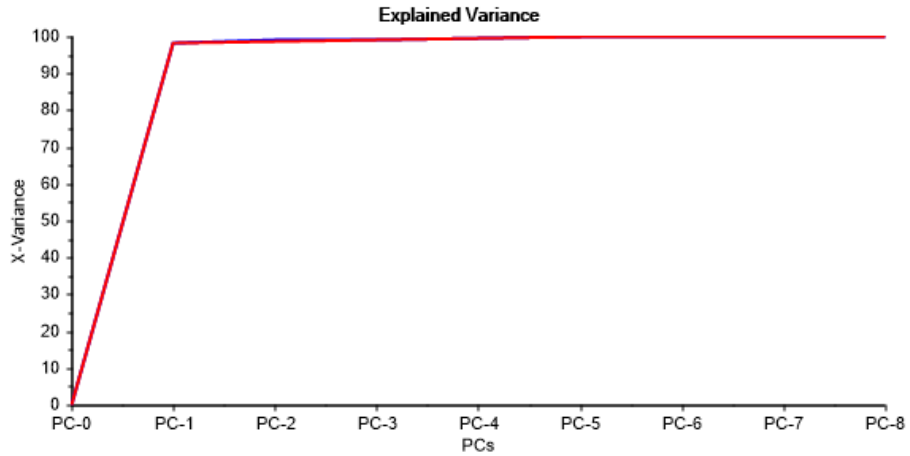
Tabla 2. Varianza explicada por los componentes principales de los 49 fondos de vacío.

(PCA)	Varianza explicada (%)	Varianza explicada acumulada (%)
PC- 1	98,6561	98,6561
PC- 2	0,6322	99,2883
PC- 3	0,398	99,6863
PC- 4	0,1522	99,8385
PC- 5	0,0369	99,8754
PC- 6	0,027	99,9024
PC- 7	0,0202	99,9226
PC-8	0,0132	99,9358

Fuente. Programa de análisis multivariado THE UNSCRAMBLER versión 10.3.

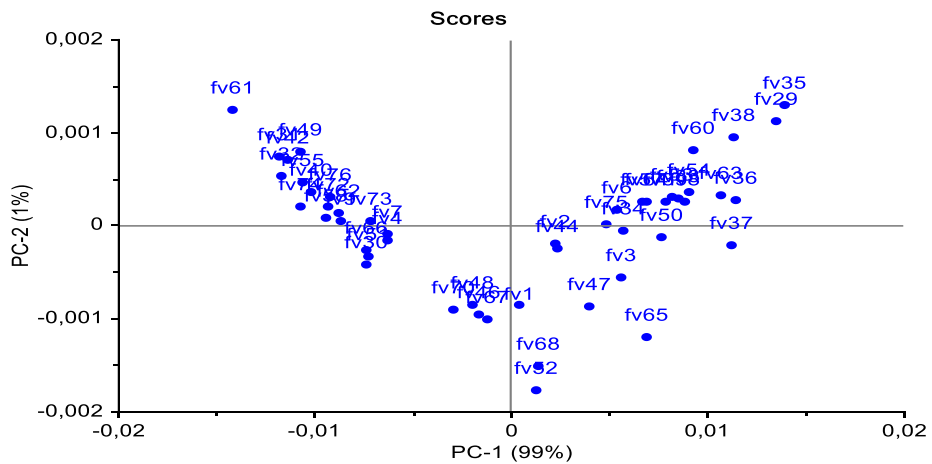
Observando la Tabla 2 se determinó que 5 es un número óptimo de componentes a utilizar en los modelos de predicción ya que a partir de PC-6 el número de la varianza explicada acumulada es menor a 0.027%. Teniendo en cuenta que añadiendo uno más, se tiene el 0.027% y quitando uno se pierde el 0,0369%

Gráfico 4. Varianza explicada en el modelo PCA



Fuente. Programa de análisis multivariado THE UNSCRAMBLER versión 10.3.

Gráfico 5. Gráfica de scores en dos dimensiones para los 49 fondos de vacío

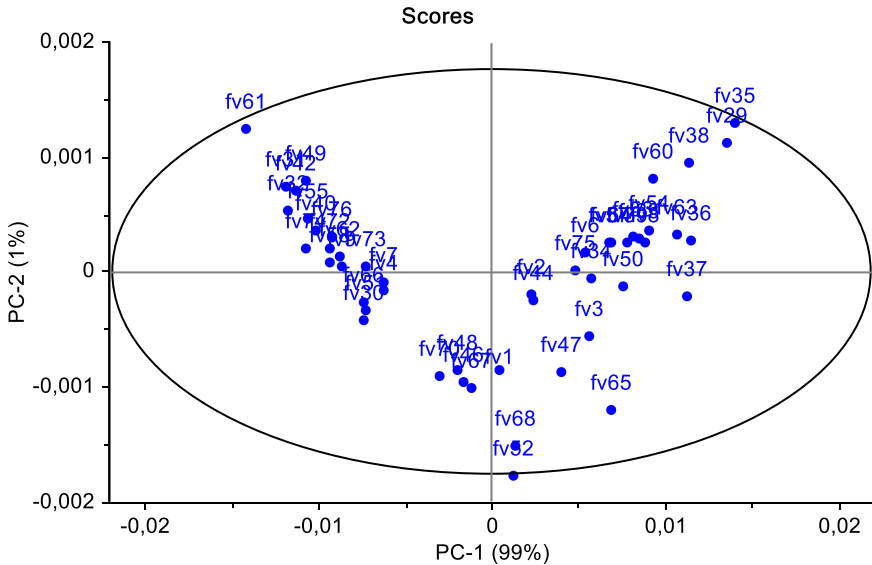


Fuente. Programa de análisis multivariado THE UNSCRAMBLER versión 10.3.

La gráfica de scores, es un gráfico de variabilidad en dos dimensiones de los componentes principales PC1 y PC2 en los que se agrupa el 98,6561% y el 0,6322% de varianza explicada respectivamente, lo que señala que estos dos componentes abarcan más del 99% de la varianza total de las intensidades espectrales, igualmente permite establecer las interacciones entre las muestras según su proximidad al plano PC-1 y PC-2 debido a que entre más cerca están las muestras al eje de coordenadas del gráfico, más semejantes son con respecto a los dos componentes PC-1 y PC-2.

La aplicación de Hotelling T^2 elipse, concluyó si los 49 fondos de vacío eran potenciales outliers (puntos fuera de la elipse o muestras atípicas) (Gráfico 6).

Gráfico 6. Estadístico Hotelling's T^2 elipse aplicado a la gráfica de score de los 49 fondos de vacío

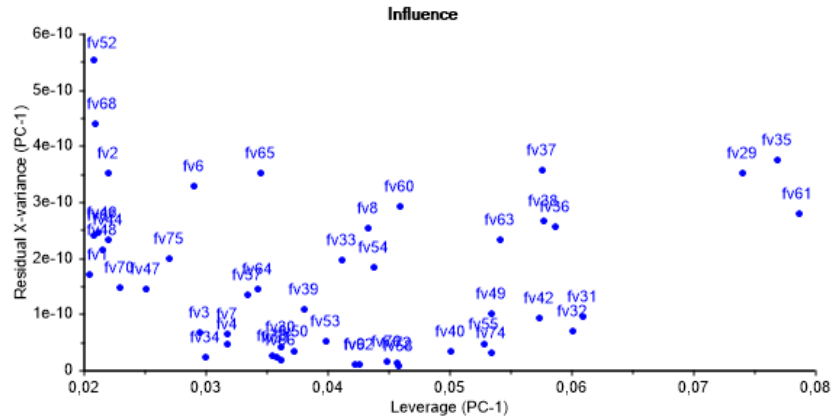


Fuente. Programa de análisis multivariado THE UNSCRAMBLER versión 10.3.

Las muestras 61, 29, 52 68 y 35 quedaron dentro del promedio de la elipse, señalando, que a pesar de estar separadas de las demás, estas presentan semejanza con algunas muestras en su composición. El Gráfico 7 cataloga de

forma más minuciosa los valores atípicos (outliers), influyentes y peligrosos (dangerous outliers) de las muestras analizadas.

Gráfico 7. Grafica de influencia para los 49 fondos de vacío



Fuente. Programa de análisis multivariado THE UNSCRAMBLER versión 10.3.

Objetos con varianza residual elevada se considera outliers, si además tiene un leverage alto son outliers peligrosos para el modelo, debido a que tienen mucha influencia sobre él. Las muestras con un alto leverage son muestras influyentes y no necesariamente outliers. Dos de las muestras presentan alta varianza (52 y 68), las cuales se definen como probables outliers, las muestras 29, 35 y 61 se definen como dangerous outliers. Las muestras que presentan alto leverage, son las que mejor describen el modelo de predicción; para este caso se observaron las muestras 31, 32 y 42.

El presente análisis en el primer componente recogió la gran cantidad de proporción posible de variabilidad original, en el segundo componente la máxima proporción posible de variabilidad no recogida en el primero, pero muy inferior. Una vez realizado el análisis de componentes principales y definido el número de componentes principales (5) que mejora la incertidumbre en los datos se determinó las causas de la variabilidad de un conjunto de datos y se ordenó por

importancia, para proyectar los datos en términos de mínimos cuadrados, se procede a desarrollar el modelo mediante el algoritmo PLS.

5.2 DESARROLLO DE MODELOS PLS

A través del algoritmo de regresión PLS y el programa de análisis multivariado The Unscrambler versión 10.3, se evaluó el aporte de la espectroscopia MIR-ATR para predecir el porcentaje peso de azufre y RCC. Los modelos se calcularon eliminando las variables x de menor importancia (con coeficiente de regresión cercanos a cero) hasta que el coeficiente de correlación (R^2) y la raíz cuadrada media del error de calibración y validación (RMSEC y RMSECV, respectivamente) permanecieron estables. Las muestras alejadas de la curva de calibración y validación también fueron eliminadas. Para probar que las muestras excluidas eran potencialmente atípicas, se realizó nuevamente el modelo, pero sacando inicialmente las muestras elegidas y seguidamente se eliminaron las variables x.

5.2.1 Modelo PLS para la predicción de %w de azufre. El rango en el que se desarrollaron los análisis fue de 4.33 – 0.91 %w de azufre. La Tabla 3 expone el error estándar de calibración (RMSEC), de validación (RMSECV) y la varianza explicada para el modelo de predicción seleccionado con 6 factores para obtener el mejor rendimiento en la calibración en concordancia con los parámetros estudiados.

Tabla 3. Parámetros estadísticos del modelo desarrollado para la predicción %w de azufre

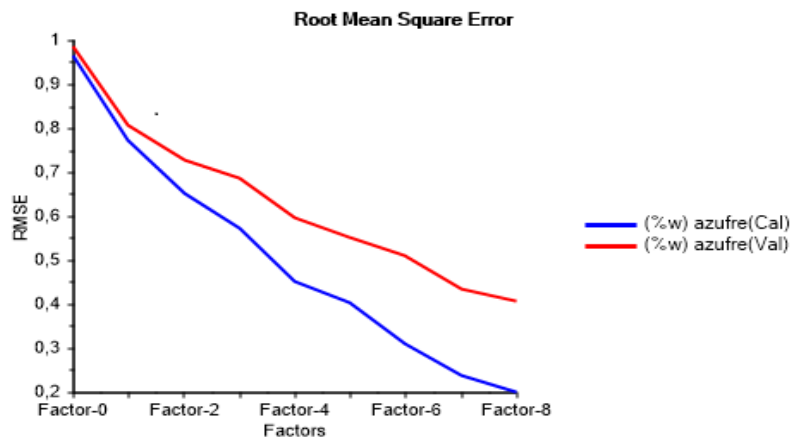
Muestras excluidas	factores	Varianza explicada	RMSEC	RMSECV
52, 61, 36	6	99%	0.32	0.51

Fuente. Programa de análisis multivariado THE UNSCRAMBLER versión 10.3.

El modelo requirió el uso de 6 factores que explica más del 99 % de la variabilidad de los datos. La muestra 61 fue excluida del modelo de predicción por tener comportamiento dangerous outliers (alto varianza residual y alto leverage) y la muestra 52 fueron excluidas del modelo de predicción por ser potenciales outliers (ver Gráfico 7). La muestra 36 se eliminó por prueba y error. Finalmente se determinó que las muestras excluidas mejoraron el modelo

En el Gráfico 8 se visualiza la raíz cuadrada del error de validación (RMSECV) y calibración (RMSEC), ya que, el RMSEC es el error promedio de los residuales calculados por la diferencia entre los valores de referencia y predicción por el modelo para las muestras de calibración; y el RMSECV se basa en un método iterativo que escoge un grupo de muestras dentro del grupo de calibración para desarrollar el modelo de predicción y luego evaluarlo sobre el grupo de muestras restantes.

Gráfico 8. Grafica de los errores RMSEC y RMSECV para el modelo PLS del azufre



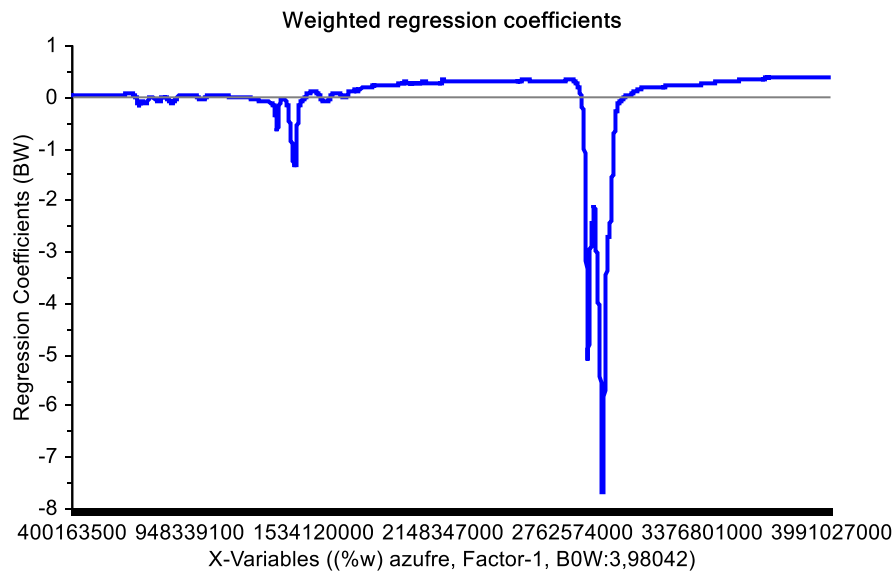
Fuente. Programa de análisis multivariado THE UNSCRAMBLER versión 10.3.

El RMSEC es el error de calibración, es decir, los residuos de los datos de calibración, el cual mide la bondad de ajuste entre sus datos y el modelo de calibración. Dependiendo del tipo de datos, modelo y aplicación, esto puede estar

sujeto a un gran sesgo optimista debido al sobreajuste en comparación con el RMSE observado en casos reales al aplicar la calibración. Si el modelo no es lo suficientemente complejo (insuficiente), el error de calibración se aproxima al error de predicción. Pero no puede indicar exceso de ajuste.

Coefficiente de regresión: cada componente principal es un acoplamiento lineal de las absorbancias medidas en los diferentes números de onda del rango espectral, multiplicadas por un coeficiente de regresión que especifica el peso o influencia de los números de onda sobre el componente, la regresión exhibe los coeficientes encontrados para cada componente. El Gráfico 9, expone los coeficientes encontrados para el primer componente principal ya que este explica la mayor dispersión de los datos.

Gráfico 9. Coeficiente de regresión para el modelo PLS del azufre

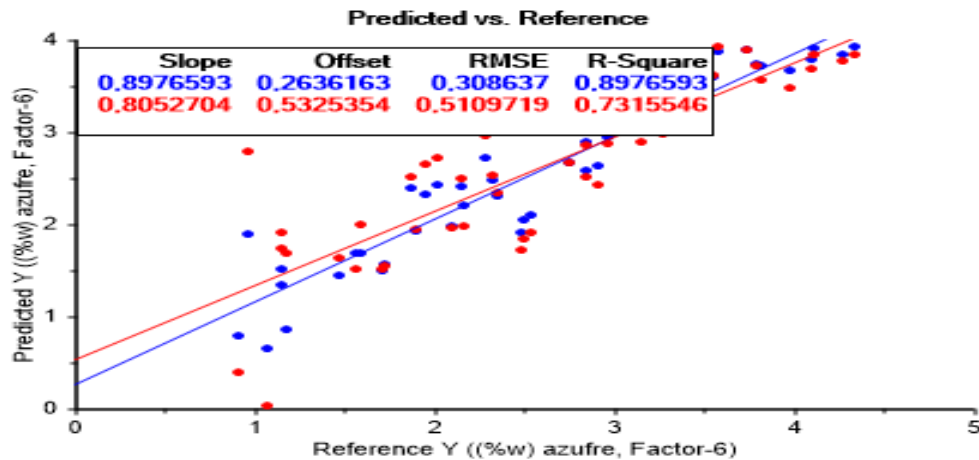


Fuente. Programa de análisis multivariado THE UNSCRAMBLER versión 10.3.

En el gráfico anterior se observan notoriamente dos regiones espectrales como las de mayor influencia y que coinciden con las regiones de mayor absorbancia en los espectros obtenidos para las muestras de fondo de vacío: la región de 1353-

1480 cm^{-1} (región 1) y la región de 2800-3000 cm^{-1} (región 2). El coeficiente de correlación entre los valores de referencia y de predicción fue encontrado superior al 0.89 para la curva de calibración y superior a 0.73 para la curva de validación (Gráfico 10).

Gráfico 10. Curva de calibración y validación del modelo de predicción de %w de azufre



Fuente. Programa de análisis multivariado THE UNSCRAMBLER versión 10.3.

Para un rango de análisis de 4.33 – 0.91 %w de azufre, se tiene un %w promedio de 2,55 %w de azufre, entonces se tiene que:

$$\%error\ promedio\ RMSEC = \frac{0,3 * 100}{2,55}$$

$$\%error\ promedio\ RMSEC = 11,76\%$$

$$\%error\ promedio\ RMSECV = \frac{0,51 * 100}{2,55}$$

$$\%error\ promedio\ RMSECV = 20\%$$

Este %error RMSEC=11,76% es un valor un poco alto del error aceptado por la comunidad científica (RMSEC<10%). La dispersión del modelo mostró un desempeño adecuado para la predicción de %w azufre, ya que el coeficiente de correlación entre los valores de referencia y de predicción fue hallada superior al

0.89 para la curva de calibración y superior al 0.73 para la curva de validación. (Ver Gráfico 10)

Todos los parámetros estadísticos anteriormente mencionados demuestran que el modelo desarrollado presenta un desempeño medianamente satisfactorio en la predicción de la composición %w de azufre. Este modelo se puede mejorar volviéndolo más robusto, es decir, realizando el modelo con más muestras.

5.2.2 Modelos PLS para predicción de %w de RCC. El rango de calibración en el que desarrollaron los análisis fue de 39.6- 12.85 %w de RCC.

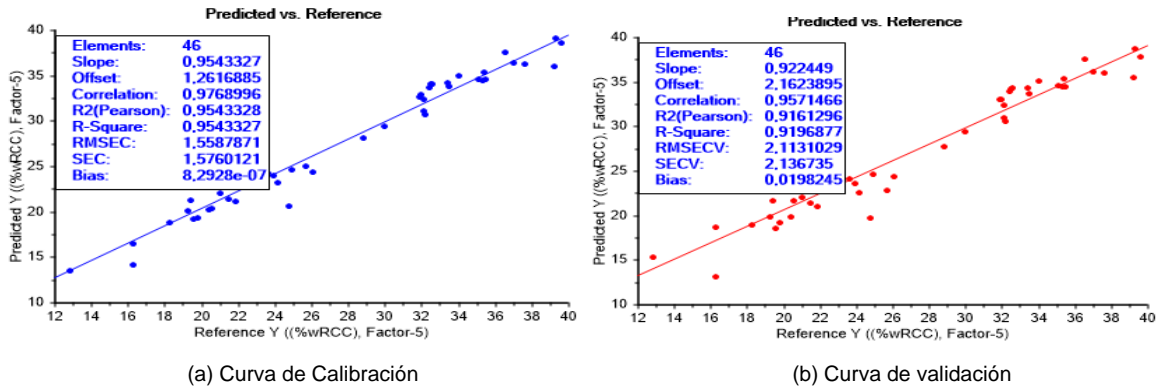
Tabla 4. Parámetros estadísticos del modelo generado para la predicción de %w de RCC en muestras de fondos de vacío

Muestras excluidas	factores	Varianza explicada	RMSEC	RMSECV
52,35,36	5	99%	1.55	2.11

Fuente. Programa de análisis multivariado THE UNSCRAMBLER versión 10.3.

La muestra 36 fue excluida del modelo de predicción seleccionado, por presentar baja variabilidad en su composición de RCC entre ellas y con otras muestras, siendo este valor de 13.5. Además, se determinó que, para el modelo seleccionado, modelo de predicción de %w de RCC, la muestra 35 se comporta como dangerous outliers y la 52 se comporta como potencial outliers aumenta el porcentaje de error de predicción del modelo (ver Gráfico 7).

Gráfico 11. Curva de calibración y validación del modelo de predicción de %w de RCC

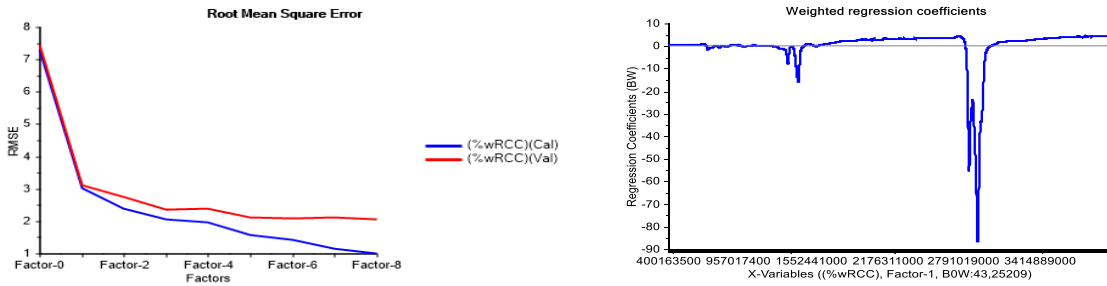


Fuente. Programa de análisis multivariado THE UNSCRAMBLER versión 10.3.

El coeficiente de correlación entre los valores de referencia y de predicción fue encontrado superior al 0.95 para la curva de calibración (Gráfico 11, curva de calibración) y superior a 0.91 para la curva de validación (Gráfico 11, curva de validación) para el modelo de RCC. El error promedio en las curvas de calibración y validación disminuye considerablemente hasta el quinto componente, donde obtiene 1.55% y 2.11% respectivamente (Gráfico 12 (c)).

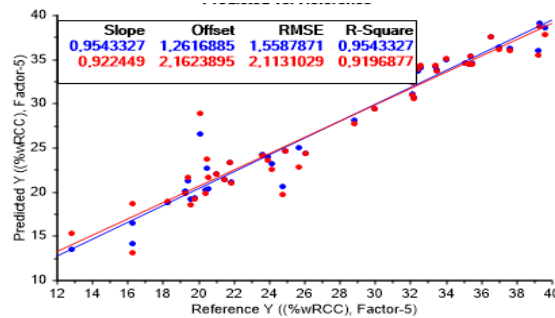
En el Gráfico 12 (b), los coeficientes de regresión calculados para el PC1 explican la mayor variabilidad en los datos (99%), se observan notoriamente dos regiones espectrales como las de mayor influencia y que coinciden con las regiones de mayor absorbancia en los espectros obtenidos para las muestras de fondo de vacío: la región de $1353\text{-}1480\text{ cm}^{-1}$ (región 1) y la región de $2800\text{-}3000\text{ cm}^{-1}$ (región 2).

Gráfico 12. Gráfica del modelo PLS del %w de RCC



(a) Errores RMSEC y RMSECV

(b) Coeficiente de regresión



(c) Curva de calibración y validación

Fuente. Programa de análisis multivariado THE UNSCRAMBLER versión 10.3.

Para un rango de análisis de 39.6 - 12.85 %w de RCC, se tiene un %w promedio de 27,02 %w de RCC, entonces se tiene que:

$$\%error\ promedio\ RMSEC = \frac{1,55 * 100}{27,02}$$

$$\%error\ promedio\ RMSEC = 5,73\%$$

$$\%error\ promedio\ RMSECV = \frac{2,11 * 100}{27,02}$$

$$\%error\ promedio\ RMSECV = 7,8\%$$

Este %error RMSEC=5,73% es un valor de error aceptado por la comunidad científica (RMSEC<10%). La dispersión del modelo mostró un desempeño adecuado para la predicción de %w RCC, ya que el coeficiente de correlación entre los valores de referencia y de predicción fue hallada superior al 0.95 para la

curva de calibración y superior al 0.91 para la curva de validación. (Ver Gráfico 12, (c))

Todos los parámetros estadísticos anteriormente mencionados demuestran que el modelo desarrollado presenta un desempeño satisfactorio en la predicción de la composición %w de RCC. Este modelo se puede mejorar volviéndolo más robusto, es decir, realizando el modelo con más muestras.

5.3 DISCUSIÓN

El modelo predictivo desarrollado en el presente trabajo presenta diferencias significativas con los realizados por Leguizamó¹⁸ y Rodríguez,¹⁹ quienes utilizaron espectroscopia fotoacústica y fluorescencia respectivamente.

Tabla 5. Comparativo de los modelos predictivos analizados con otros investigadores

Criterios Propiedad	Investigador 1		Investigador 2		Modelo propio	
	Azufre	Hidrógeno	Azufre	Nitrógeno	Azufre	RCC
Número de componentes	3	3	3	3	6	5
RMSEC	0,044	0,106	4,74 E ⁻²	6,08 E ⁻⁴	0,32	1,55
RMSECV	0,085	0,293	9,43 E ⁻²	2,63 E ⁻³	0,51	2,11
R square de calibración	0,99	0,99	0,96	0,99	0,898	0,954
R square de validación	0,99	0,95	0,89	0,99	0,731	0,9197
Muestras excluidas	6/20	4/20	6/16	5/16	3/49	3/49

Igualmente se puede observar en la tabla 5, que la diferencia principal en los modelos se relaciona con el tamaño de las muestras analizadas y excluidas, esto en parte se debe al número de pozos objeto de estudio, lo que conllevó a diferencias en el número de componentes encontrados, errores RMSEC Y RMSECV, y los R square de calibración y validación. Se puede evidenciar el

¹⁸ LEGUÍZAMO CASTELLANOS, Cristian Ferney, Op. cit.

¹⁹ RODRÍGUEZ CAMACHO, Sidne Stefany, Op. cit.

Azufre es un elemento esencial para el tratamiento de datos, y que su calibración se diferencia en poca medida de acuerdo a la validación de los datos.

De lo anterior, se deduce que el modelo propuesto en el presente trabajo tiene un mayor número de componentes que mejora la incertidumbre en los datos.

6. CONCLUSIONES

Realizar un análisis de componentes principales (PCA) de los espectros correspondientes a las muestras de fondo de vacío y definir el número de componentes principales que mejora la incertidumbre en los datos, no fue una tarea fácil, los análisis de componentes principales (PCA) concedieron determinar de forma cualitativa la semejanza de los 49 espectros de las 49 muestras de fondos de vacío y determinar el número de componentes principales que mejora la precisión en los datos.

Se logró determinar las causas de la incertidumbre de un conjunto de datos y ordenarlas por importancia, para proyectar los datos en términos de mínimos cuadrados, mediante el análisis por regresión parcial por mínimos cuadrados (PLS) permitiendo determinar de forma cuantitativa la semejanza de los 49 espectros de las 49 muestras de fondo de vacío para poder hallar el porcentaje peso de azufre y RCC.

El uso de la espectroscopia MIR-ATR junto con métodos de calibración multivariable es otra opción, a la cromatografía líquida de alta resolución, resonancia magnética nuclear, entre otras; para la determinación del porcentaje peso de azufre y RCC en muestras de fondos de vacío, evitándose así, el pretratamiento de los fondos de vacío, no usar solventes perjudiciales para el medio ambiente y aumentar la replicación en los resultados. Por medio de esta técnica se logró desarrollar un Modelo Predictivo de Propiedades Físicoquímicas de Fondos de Vacío mediante el análisis por regresión parcial por mínimos cuadrados (PLS). Se logró además evaluar el aporte de la espectroscopia MIR-ATR para predecir la composición de porcentaje peso azufre y porcentaje de peso de RCC y establecer los aportes del desarrollo del modelo a la investigación en la Ingeniería Química.

BIBLIOGRAFÍA

ALTGELT, Klaus H. y BODUSZYNSKI, Mieczslaw. Composition and analysis of heavy petroleum fractions. Segunda edición. New York: CRC Press, 1994. 495 p.

BALLESTEROS LANCHEROS, Sandra Lucía. Predicción de las fracciones SARA de fondos de vacío de crudos colombianos, por métodos quimiométricos utilizando espectroscopia de fluorescencia inducida por láser (LIF) [en línea]. Trabajo de Grado de Químico, UIS. Bucaramanga, 2010. Disponible en Internet: <URL: <http://tangara.uis.edu.co/biblioweb/tesis/2010/133470.pdf>>

BORGES, B. y ACEVEDO, S. Caracterización estructural de distintas fracciones aisladas de crudos extrapesado Carabobo [en línea]. En: Revista latinoamericana de metalurgia y materiales. 2007. Vol. 2 p 83-94. Disponible en Internet: <URL: http://www.scielo.org.ve/scielo.php?pid=S0255-69522007000200003&script=sci_arttext>

CONAWAY, Charles. The petroleum industry: a nontechnical al guide. Tulsa; pennwell publishing co. Oklahoma: PennWell, 1999.

FREDERICKS, Peter; RINTOUL, Llewellyn and COATES, John. Chapter 7. Vibrational Spectroscopy: Instrumentation for Infrared and Raman Spectroscopy. En: CAZES, Jack. Analytical Instrumentation Handbook. 2008, p.76.

HAMADOU, R., KHODJA, M., KARTOUT, M. y JADA, A. Permeability reduction by asphaltenes and resins deposition in porous media [en línea]. En: Fuel 87, 2008, p. 2178–2185. Disponible en Internet: <URL: <https://bibliotecavirtual.uis.edu.co:2191/science/article/pii/S001623610700542X>>

KALLEVIK, H. Characterization of crude oil and model oil emulsions by means of near infrared spectroscopy and multivariable analysis [en línea]. Noruega, 1999. Disponible en Internet: <URL: <https://www.osti.gov/etdeweb/biblio/20119612>>

KHARRAT, A. M., ZACHARIA, J., CHERIAN, V. J. y ANYATONWU, A. Issues with comparing several methodologies [en línea]. En: Energy and fuels 21, 2007, p. 3618–3621. Disponible en Internet: <URL: <https://bibliotecavirtual.uis.edu.co:2277/doi/pdf/10.1021/ef700393a>>

LEGUÍZAMO CASTELLANOS, Cristian Ferney. Predicción del análisis elemental por espectroscopia fotoacústica en la región del infrarrojo medio y métodos quimiométricos a las fracciones de los fondos de vacío [en línea]. Trabajo de Grado de Químico, UIS. Bucaramanga, 2012. Disponible en Internet: <URL: <http://tangara.uis.edu.co/biblioweb/tesis/2012/143262.pdf>>

MELÉNDEZ CORREA, Lesly Viviana. Predicción de propiedades fisicoquímicas de productos destilados del Delayed Coking de fondos de vacío a partir de parámetros estructurales determinados por espectroscopia infrarroja FTIR-ATR y métodos quimiométricos [en línea]. Trabajo de Grado de Magíster en Química, UIS. Bucaramanga, 2014. Disponible en Internet: <URL: <http://tangara.uis.edu.co/biblioweb/tesis/2014/152400.pdf>>

MORAWSKI, I. y MOSIO-MOSIEWSKI, J effects of parameters in ni-mo catalyzed hydrocracking of vacuum residue on composition and quality of obtained products [en línea]. En: Fuel Processing Technology 87 (2006) 659-669. Disponible en Internet: <URL: <https://www.sciencedirect.com/science/article/pii/S037838200600021X?via%3Dihub>>

PAREDES, Juan Roberto y RAMÍREZ, John. Energías renovables variables y su contribución a la seguridad energética: complementariedad en Colombia [en línea]. BID, 2017. Disponible en Internet: <URL: <https://publications.iadb.org/bitstream/handle/11319/8146/Energias-renovables-variables-y-su-contribucion-a-la-seguridad-energetica-Complementariedad-en-Colombia.PDF?sequence=5>>

PARRA PICÓN, Sara Inés. Predicción de la curva de destilación de fondos de vacío de crudos colombianos usando espectroscopia fotoacústica con transformada de Fourier en el infrarrojo medio y quimiometría [en línea]. Trabajo de Grado de Químico, UIS. Bucaramanga, 2012. Disponible en Internet: <URL: <http://tangara.uis.edu.co/biblioweb/tesis/2012/146837.pdf>>

RODRÍGUEZ CAMACHO, Sidne Stefany. Predicción de la composición elemental de fracciones del proceso de Delayed Coking de fondos de vacío de crudos colombianos empleando espectroscopia de fluorescencia y métodos quimiométricos [en línea]. Trabajo de Grado de Químico, UIS. Bucaramanga, 2014. Disponible en Internet: <URL: <http://tangara.uis.edu.co/biblioweb/tesis/2014/152138.pdf>>

SPIECKER, P. M., GAWRYS, K. L., TRAIL, C. B. y KILPATRICK, P. K. Effects of petroleum resins on asphaltene aggregation and water-in-oil emulsion formation. Colloids surfaces a physicochem [en línea]. Eng. Asp. 220, 2003, p. 9–27. Disponible en Internet: <URL: <https://www.sciencedirect.com/science/article/pii/S0927775703000797?via%3Dihub>>

THE UNSCRAMBLER 10.3. Methods. Software para diseño de experimentos y análisis multivariado.

TREJO, F. y RANA, M. S.; Ancheyta J. Thermogravimetric determination of coke from asphaltenes, resins and sediments and coking kinetics of heavy crude asphaltenes [en línea]. Catalysis today 150, 2010, p. 272-278. Disponible en Internet: <URL: <https://bibliotecavirtual.uis.edu.co:2150/record/display.uri?eid=2-s2.0-77949658688&origin=inward&txGid=c228bef6087f5f2750986e738f2c12d1>>