

**PREDICCIÓN DE LA DENSIDAD EN ASFALTENOS COLOMBIANOS
UTILIZANDO ESPECTROSCOPIA FOTOACÚSTICA EN LA REGIÓN DEL
INFRARROJO MEDIO Y MÉTODOS QUIMIOMÉTRICOS**

LAURA NAIR MONROY VÁSQUEZ

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE QUÍMICA
BUCARAMANGA**

2010

**PREDICCIÓN DE LA DENSIDAD EN ASFALTENOS COLOMBIANOS
UTILIZANDO ESPECTROSCOPIA FOTOACÚSTICA EN LA REGIÓN DEL
INFRARROJO MEDIO Y MÉTODOS QUIMIOMÉTRICOS**

LAURA NAIR MONROY VÁSQUEZ

**Trabajo de Grado presentado como Requisito parcial para optar al título de
Química**

Director:

Enrique Mejía Ospino. Químico Ph.D

Codirector:

Jorge Armando Orrego. Químico M.Sc.

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE QUÍMICA
BUCARAMANGA**

2010

A mi familia..... Mi mami, mi Papi y mis hermanas
A mi nueva familia.....Mario y Juan Diego

AGRADECIMIENTOS

A Dios por su presencia en mi vida y por ser mi guía en todo momento.

Al Instituto Colombiano del Petróleo (ICP) por el apoyo suministrado mediante los convenios de cooperación tecnológica que sostiene con la Universidad Industrial de Santander.

Al Profesor Enrique Mejía Ospino (Director) por darme la oportunidad de trabajar en esta investigación, por ser un ejemplo de profesional y persona y por su ayuda incondicional en cada paso de la realización de este trabajo.

Al Msc. Jorge Armando Orrego (Codirector) por sus conocimientos y aportes que ayudaron a direccionar este proyecto.

A mis padres y mis hermanas por ser la columna que sostiene mi vida, por su constante apoyo y amor.

A mi mami por ser mi compañera de siempre y mi punto de referencia para todo.

A mis amiguitas vivi, adri, glori, sandris, andre..... por brindarme una amistad sincera y por permitirme disfrutar de muchos momentos de alegrías. Gracias por su compañía, cariño y ayuda en todo momento.

A mis demás compañeritas y compañeros del Laboratorio de Espectroscopia Atómica y Molecular, por su valiosa colaboración.

Y por último, pero no menos importante, a mi hijo Juan Diego por ser la razón de mi vida, la luz de mis ojos y el motor de mis sueños. A mi cosa hermosa Mario por ser ese compañero, amigo, novio y esposo que supo sobrellevarme y ayudarme con amor, apoyo y mucha comprensión durante toda mi etapa de estudiante y espero que siga siendo así por muchos años más.... Te amooo.

Así mismo, quiero agradecer a todas las personas que de una u otra forma estuvieron involucradas en el desarrollo de este trabajo.

CONTENIDO

| | PÁG. |
|--|-------------|
| INTRODUCCIÓN | 14 |
| | |
| 1. CONSIDERACIONES TEÓRICAS | 16 |
| 1.1 EL PETRÓLEO | 16 |
| 1.1.1 Generalidades | 16 |
| 1.1.2 Clasificación Química del Petróleo | 17 |
| 1.1.3 Caracterización Básica | 18 |
| 1.2 ASFALTENOS | 20 |
| 1.3 ESPECTROSCOPIA FOTOACÚSTICA (PAS) | 23 |
| 1.4 QUIMIOMETRÍA | 28 |
| 1.4.1 Calibración Multivariable | 30 |
| 1.4.1.1 Clasificación de los métodos de calibración multivariable: | 30 |
| 1.4.2 Construcción de modelos de calibración multivariable. | 32 |
| 1.4.3 Pretratamiento de los datos registrados | 35 |
| 1.4.4 Análisis por componentes principales (PCA) | 37 |
| 1.4.5 Regresión por mínimos cuadrados parciales (PLS) | 39 |
| | |
| 2. METODOLOGÍA | 42 |
| 2.1 SELECCIÓN DE MUESTRAS | 42 |
| 2.2 CARACTERIZACIÓN POR ESPECTROSCOPIA FOTOACÚSTICA | 43 |
| 2.2.1 Instrumentación | 43 |
| 2.2.2 Verificación del Espectrómetro | 47 |
| 2.2.3 Adquisición y análisis de Espectros PAS | 48 |
| 2.3 PROCESAMIENTO DE DATOS | 50 |
| | |
| 3. RESULTADOS Y ANÁLISIS | 53 |

| | |
|--|----|
| 3.1 ANÁLISIS DE COMPONENTES PRINCIPALES (PCA) | 53 |
| 3.2 ANALISIS DE REGRESIÓN POR MÍNIMOS CUADRADOS PARCIALES (PLS-REGRESION) | 58 |
| 3.2.1 Modelo PLS para la predicción de la densidad en Asfaltenos. | 58 |
| 3.2.1.1 Calibración del modelo PLS: | 58 |
| 3.2.1.2 Validación Cruzada del modelo PLS: | 62 |
| 4. CONCLUSIONES | 66 |
| 5. RECOMENDACIONES | 68 |
| REFERENCIAS BIBLIOGRAFÍA | 69 |

LISTA DE FIGURAS

| | PAG |
|---|------------|
| Figura 1. Diagrama para separación de Asfaltenos | 21 |
| Figura 2. Estructura de algunos Asfaltenos | 22 |
| Figura 3. Espectro PAS-IR de un asfalteno. | 25 |
| Figura 4. Generación de la señal fotoacústica | 26 |
| Figura 5. Interpretación geométrica de un PCA. | 38 |
| Figura 6. Descripción gráfica del modelo de regresión PLS | 40 |
| Figura 7. Sistema PAS acoplado al equipo IR Tensor 27 marca Bruker | 43 |
| Figura 8. Celda PAS modelo MTEC 200 | 44 |
| Figura 9. Efecto de la relación señal ruido en un espectro de una muestra de Asfalteno | 46 |
| Figura 10. Prueba de repetibilidad del espectrómetro FT-IR | 48 |
| Figura 11. Espectros PAS de todas las muestras de Asfaltenos | 49 |
| Figura 12. Espectro con longitudes de onda de interés | 51 |
| Figura 13. Espectros normalizados | 52 |
| Figura 14. Descomposición por componentes principales de la información espectral. Gráfica de Score | 55 |
| Figura 15. Gráfico de X-Loadings de PC1 y PC2 | 56 |
| Figura 16. Gráfica de Influencia | 57 |
| Figura 17. Varianza explicada en el modelo PLS de Asfaltenos | 60 |
| Figura 18. Error calculado en función del número de componentes para el modelo PLS de Asfaltenos | 61 |
| Figura 19. Coeficiente de regresión para el primer componente principal PLS de Asfaltenos | 62 |
| Figura 20. Curvas de calibración del modelo de predicción de la densidad en | 64 |

LISTA DE TABLAS

| | PÁG. |
|---|-------------|
| Tabla 1. Composición porcentual de Crudos | 16 |
| Tabla 2. Clasificación del crudo por gravedad API | 19 |
| Tabla 3. Análisis próximo de Asfaltenos analizados | 42 |
| Tabla 4. Varianza de los componentes principales de las 14 muestras de Asfaltenos | 54 |
| Tabla 5. Parámetros estadísticos de los modelos de predicción generados para la predicción de la densidad en Asfaltenos | 59 |
| Tabla 6. Validación cruzada del modelo de predicción de la densidad de las 14 muestras de asfaltenos | 63 |

RESUMEN

TÍTULO

PREDICCIÓN DE LA DENSIDAD EN ASFALTENOS COLOMBIANOS UTILIZANDO ESPECTROSCOPIA FOTOACÚSTICA EN LA REGIÓN DEL INFRARROJO MEDIO Y MÉTODOS QUIMIOMÉTRICOS.*

AUTOR

LAURA NAIR MONROY VÁSQUEZ. **

PALABRAS CLAVES

FOTOACÚSTICA, ASFALTENOS, PLS, PCA.

DESCRIPCIÓN

En el presente trabajo se evidencia la potencialidad de la Espectroscopia Fotoacústica en la región del infrarrojo medio (PAS-MIR) como una técnica de caracterización de muestras complejas como los asfaltenos de crudos colombianos. La obtención de los espectros fotoacústicos de catorce muestras de asfaltenos se llevó a cabo utilizando una celda fotoacústica MTEC 200 que se acopló a un espectrómetro IR 27 marca Bruker. Como parámetros del equipo para la toma de los espectros se utilizó una resolución de 8cm^{-1} , ganancia del amplificador de 1, velocidad del interferómetro de 0.0583 cm/s y 256 barridos. La metodología de caracterización propuesta por espectroscopia fotoacústica requiere menos de diez minutos de análisis y una mínima cantidad de muestra para la toma de espectros.

El trabajo consistió en construir un modelo matemáticos a partir de los espectros fotoacústicos de las muestras de asfaltenos para predecir la densidad de los mismos. Se utilizó análisis de componentes principales (PCA) para reducir la dimensionalidad de los datos espectrales y luego regresión de mínimos cuadrados parciales (PLS-R) para generar el modelo predictivo de la propiedad por medio de correlaciones lineales de las matrices de los datos espectroscópicos y las valores teóricos de densidad.

El modelo de predicción fue desarrollado empleando la región espectral comprendida entre $3101,7\text{-}2750,63\text{ cm}^{-1}$. La validación mostró resultados satisfactorios para la predicción de la densidad en los asfaltenos. Se obtuvo un error estándar de predicción (RMSEP) de 0.002, y un coeficiente de correlación (R^2) entre los valores de referencia y los valores predichos de 0.99.

La metodología de caracterización propuesta por espectroscopia fotoacústica requiere menos de diez minutos de análisis.

* Trabajo de Grado

** Facultad de Ciencias. Escuela de Química. Director: Enrique Mejía Ospino. Codirector: Jorge Armando Orrego.

SUMMARY

TITLE

PREDICTION OF THE DENSITY FROM ASPHALTENES COLOMBIANS USING PHOTOACOUSTIC SPECTROSCOPY IN THE REGION OF MIDDLE INFRARED AND CHEMOMETRIC METHODS.*

AUTHOR

LAURA NAIR MONROY VÁSQUEZ.**

KEY WORDS

PHOTOACOUSTIC, ASPHALTENES, PCA, PLS.

DESCRIPTION

In the present study demonstrates the potential of photoacoustic spectroscopy in the mid-infrared region (PAS-MIR) as a technique for characterization of complex samples such as crude oil asphaltenes Colombians. Obtaining photoacoustic spectra of asphaltenes fourteen samples was carried out using a MTEC 200 photoacoustic cell was coupled to an IR spectrometer Bruker 27 mark. As parameters of the equipment for making the spectra we used a resolution of 8cm⁻¹, an amplifier gain, velocity interferometer 0.0583 cm / s and 256 scans. The characterization methodology proposed by photoacoustic spectroscopy requires less than ten minutes of analysis and minimal sample size for taking spectra.

The work consisted of constructing a mathematical model based on the photoacoustic spectra of asphaltene samples to predict the density of them. We used principal component analysis (PCA) to reduce the dimensionality of spectral data and then partial least squares regression (PLS-R) to generate the predictive model of the property by means of linear correlation matrices and spectroscopic data the theoretical values of density.

Prediction model was developed using the spectral region between 3101.7 to 2750.63 cm⁻¹. The validation showed satisfactory results for predicting the density asphaltenes. We obtained a standard error of prediction (RMSEP) of 0.002 and a correlation coefficient (R²) between the baseline and the predicted values of 0.99.

* Trabajo de Grado

** Facultad de Ciencias. Escuela de Química. Director: Enrique Mejía Ospino. Codirector: Jorge Armando Orrego.

INTRODUCCIÓN

Desde el punto de vista industrial existe gran interés en el desarrollo de metodologías analíticas que proporcionen gran cantidad de información, de calidad y que además lo hagan en el menor tiempo posible. Esta idea, que es aplicable a cualquier proceso industrial, cobra especial relevancia en un sector como el de la industria petroquímica, en la que es habitual el monitoreo y control de los procesos basándose en la información que proporciona el análisis de laboratorio.

Dentro de las técnicas de análisis químico instrumental empleadas en la industria, la espectroscopia fotoacústica en la región del infrarrojo medio (PAS-IR) ocupa un lugar importante para el desarrollo de metodologías de caracterización. La espectroscopia fotoacústica, tanto en la zona del infrarrojo medio como el infrarrojo cercano, resulta una técnica atractiva en el ámbito industrial porque⁽¹⁾:

- a) Proporciona una información química muy versátil. La espectroscopia molecular, como consecuencia del tipo de transiciones propias de esta zona, aporta información de una gran cantidad de compuestos químicos.
- b) La obtención del espectro se hace de una forma rápida y no es necesario un pretratamiento de la muestra, lo que puede facilitar el acoplamiento de esta técnica a la línea de producción (análisis en línea).

El incremento en el número de variables que pueden ser obtenidas con los instrumentos actuales conlleva una clara dificultad en su interpretación. Los métodos estadísticos univariantes permiten el análisis de conjuntos de datos discretos de una forma rápida y simple. Sin embargo, la habilidad de la mente

humana para entender una gran cantidad de información distribuida en n dimensiones es restringida. Un ejemplo claro de este tipo de situaciones se encuentra con la Espectroscopia Fotoacústica (PAS). Cada espectro puede estar compuesto por varios millares de variables, y si además se tiene en cuenta la elevada complejidad de la señal, se hace imprescindible la utilización de técnicas quimiométricas multivariantes que permitan interpretar y entender estos grandes conjuntos de datos.

El tratamiento de la señal con técnicas multivariantes se ha empleado con éxito en el desarrollo de aplicaciones en productos derivados del petróleo a nivel cuantitativo, como la predicción de propiedades fisicoquímicas, cuantificación de componentes específicos y a nivel cualitativo, para la identificación y clasificación de productos.

Este trabajo busca predecir la densidad en la fracción de Asfaltenos que además de ser la fracción más pesada de los crudos, se caracteriza por poseer un alto grado de aromaticidad y alto peso molecular promedio. La presencia de asfaltenos es muy importante, ya que a medida que esta fracción aumenta se puede producir un incremento en la formación del coque y generar problemas de precipitación durante los procesos de generación y recuperación.

La técnica PAS-IR junto con la aplicación de técnicas quimiométricas busca jugar un papel esencial para el análisis de productos derivados del petróleo, específicamente de la fracción de Asfaltenos, buscando desarrollar nuevas metodologías, que permitan operar en condiciones menos rigurosas que los procedimientos estándar con un menor costo de análisis.

1. CONSIDERACIONES TEÓRICAS

1.1 EL PETRÓLEO

1.1.1 Generalidades

Es la materia prima de la industria de refinación, es de naturaleza hidrocarbonada y está constituido por una mezcla compleja de diferentes tipos de hidrocarburos, se compone prácticamente de carbono e hidrógeno, con un pequeño porcentaje de otros elementos como azúfre, nitrógeno y metales integrados en hidrocarburos de estructuras más o menos complejas⁽²⁾; su composición aproximada puede verse en la Tabla 1. Las moléculas de hidrocarburos que los componen son más complejas a medida que aumenta el punto de ebullición de las mismas. Desde el punto de vista estructural están presentes, en mayor o menor proporción, hidrocarburos parafínicos, aromáticos y nafténicos.

Tabla 1. Composición porcentual de Crudos

| CRUDOS | % PESO |
|---------------|---------------|
| Carbono | 84-87 |
| Hidrógeno | 11-14 |
| Azúfre | 0-5 |
| Nitrógeno | 0-0,2 |

Entre las diversas teorías existentes sobre su origen, la más aceptada admite su origen orgánico, formado a partir de residuos animales y vegetales sometidos a determinadas condiciones de temperatura y presión, que junto a una acción bacteriana anaeróbica da lugar a los hidrocarburos. Estos fluyen entre varias capas sedimentarias hasta que se concentran en unas formaciones geológicas

características, no permeables y porosas, que constituyen lo que se denomina *yacimiento* dentro de un campo petrolífero ⁽³⁾

1.1.2 Clasificación Química del Petróleo

El petróleo está constituido de carbono (83-87%), hidrogeno (11.5-14%) y heteroátomos (1-5.5%). La composición elemental de un crudo es muy variada, por lo cual se hace necesario hacer una clasificación basada en la clase de hidrocarburos que predominan en el crudo, esto permite seleccionar un método correcto para su procesamiento ⁽²⁾:

Petróleo de base parafínica

Predominan los hidrocarburos saturados o parafínicos, muy fluidos de colores claros y bajo peso específico (aproximadamente 0,85 kg/L). Por destilación producen alrededor del 5% de parafina y poco asfalto.

Petróleo de base asfáltica

Predominan los hidrocarburos etilénicos y diétilínicos, nafténicos, y bencénicos o aromáticos. Son muy viscosos, de coloración oscura y mayor peso específico (aproximadamente 0,950 kg/L), por destilación produce un abundante residuo de asfalto y menos del 2% de parafinas.

Petróleo de base mixta o híbridas

De composición de bases intermedias, formados por toda clase de hidrocarburos: Saturados, no saturados, no cíclicos (parafínicos y olefinas) y cíclicos (ciclánicos o nafténicos y bencénicos o aromáticos). La mayoría de los yacimientos mundiales son de este tipo. De esta clase podemos encontrar: bases parafínicas-nafténicas, parafínicas- nafténicas-aromáticas.

1.1.3 Caracterización Básica

Como consecuencia de su origen, condiciones de evolución y antigüedad, los crudos tienen diferente composición en cuanto al tipo y cantidad de familias químicas hidrocarbonadas de las que están formados. Las características físicas varían en función de esta composición, y las posibilidades de refinado o aprovechamiento para la obtención de sus productos derivados, principalmente combustibles, lubricantes, asfaltos y materia prima para petroquímica, serán diferentes, así como su valor económico.

Una serie mínima de propiedades permiten un conocimiento básico, primario, de la calidad de un crudo: se trata de su densidad, contenido de azufre y curva TBP (True Boiling Point). El conocimiento de estas propiedades permite una diferenciación entre los crudos en términos de ligero o pesado según el valor de densidad °API, rendimiento en destilación mediante su curva TBP, y el contenido de azufre, que permite una primera valoración de sus dificultades de procesamiento.

- **Densidad:**

La densidad en grados API se define a partir de la densidad relativa o mediante la siguiente expresión ⁽⁴⁾:

$$^{\circ}API = \frac{141,5}{\text{Grav. Espec. a } 60^{\circ}F} - 131,5$$

La gravedad API, de sus siglas en inglés *American Petroleum Institute*, es una medida de densidad que describe cuán pesado o liviano es el petróleo comparándolo con el agua. Si los grados API son mayores a 10, es más liviano que el agua, y por lo tanto flotaría en ésta. La gravedad API es también usada para comparar densidades de fracciones extraídas del petróleo. Por ejemplo, si una fracción de petróleo flota en otra, significa que es más liviana, y por lo tanto

su gravedad API es mayor. Matemáticamente la gravedad API no tiene unidades, sin embargo siempre al número se le coloca la denominación grado API.

El Petróleo es clasificado en liviano, mediano, pesado y extrapesado, de acuerdo con su medición de gravedad API (Tabla 2).

Tabla 2. Clasificación del crudo por gravedad API

| | |
|----------------------|-------------------|
| Extrapesados | Menos de 10° |
| Pesados | 10°- 22.3° |
| Medianos | 22.3°- 31.1° |
| Livianos | Mayores que 31.1° |
| Superlivianos | 40° en adelante |

- **Azúfre:**

El contenido de azúfre de una crudo es un factor importante, ya que se trasladaría casi en su totalidad a los productos de refino, que están sometidos a fuertes restricciones de calidad respecto a su contenido. Su presencia en los hidrocarburos implica la necesidad de mayores procesos de refinamiento, y por ende un mayor costo final, razón por la cual la presencia de azufre es también un determinante del valor comercial del petróleo.

El azúfre puede presentarse como azúfre elemental y sulfuro de hidrógeno, que se encuentran disueltos en el crudo; también puede encontrarse formando parte de compuestos hidrocarbonados como mercaptanos, sulfuros, tiofenos y benzotiofenos polisustituídos.

- **Curva de destilación TBP (True Boiling Point):**

Siendo un crudo una mezcla física de hidrocarburos de diferentes puntos de ebullición, la manera más simple de aprovechamiento será su separación por

destilación en cortes o fracciones de diferentes intervalos de ebullición y con diferente aplicación.

Debido a la compleja composición de los crudos, la caracterización molecular individual no es posible, y un análisis elemental no es atractivo, porque da solo información limitada a cerca de la constitución del crudo. En lugar de eso, el análisis del tipo de hidrocarburo presente es el más empleado⁽⁵⁾.

La separación o fracción SARA es un ejemplo de este tipo de análisis, que consiste en la separación del crudo en cuatro grupos principales con base a diferencias en la solubilidad y polaridad. Las cuatro fracciones SARA corresponden a los saturados (S), aromáticos (A), resinas (R) y asfaltenos (A)⁽⁶⁾. La técnica para determinar la composición SARA de los crudos, y específicamente la fracción de asfaltenos usados para realizar los modelos predictivos (técnica de referencia), fue la cromatografía de columna abierta, esta se realiza siguiendo la norma ASTM D2007⁽⁷⁾.

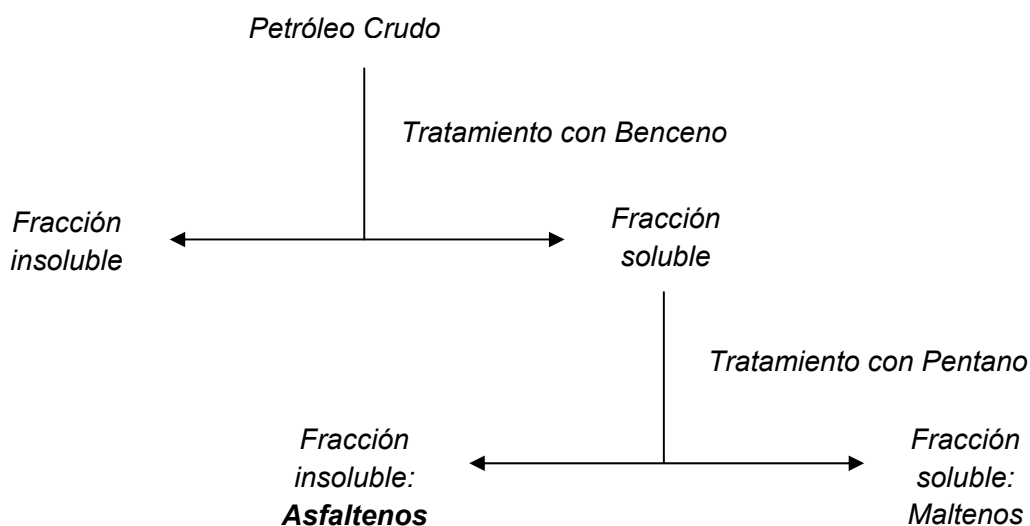
La fracción de saturados está compuesta por hidrocarburos saturados no polares lineales, ramificados y cíclicos. Los aromáticos, que contienen uno o más anillos aromáticos, son más polarizables. Las otras dos fracciones, resinas y asfaltenos, tienen sustituyentes polares. La distinción entre los dos, es que los asfaltenos son insolubles en un exceso de heptano (o pentano), mientras que las resinas son miscibles en estos solventes.

1.2 ASFALTENOS

El estudio de la fracción pesada del crudo (asfaltenos) se ha incrementado en los últimos años debido a los problemas que estos representan en los procesos de producción y conversión.

A pesar de la diversidad de criterios que se adoptan al definir los asfaltenos se ha llegado a un consenso al catalogarlos como la fracción de crudo soluble en tolueno (ó benceno) e insoluble en un exceso de n–alcano (pentano o heptano) (figura 1). Los asfaltenos están constituidos principalmente por anillos aromáticos ligados con cadenas alquílicas y cicloalcanos, además de compuestos heterocíclicos que poseen N, S y O ⁽⁸⁾.

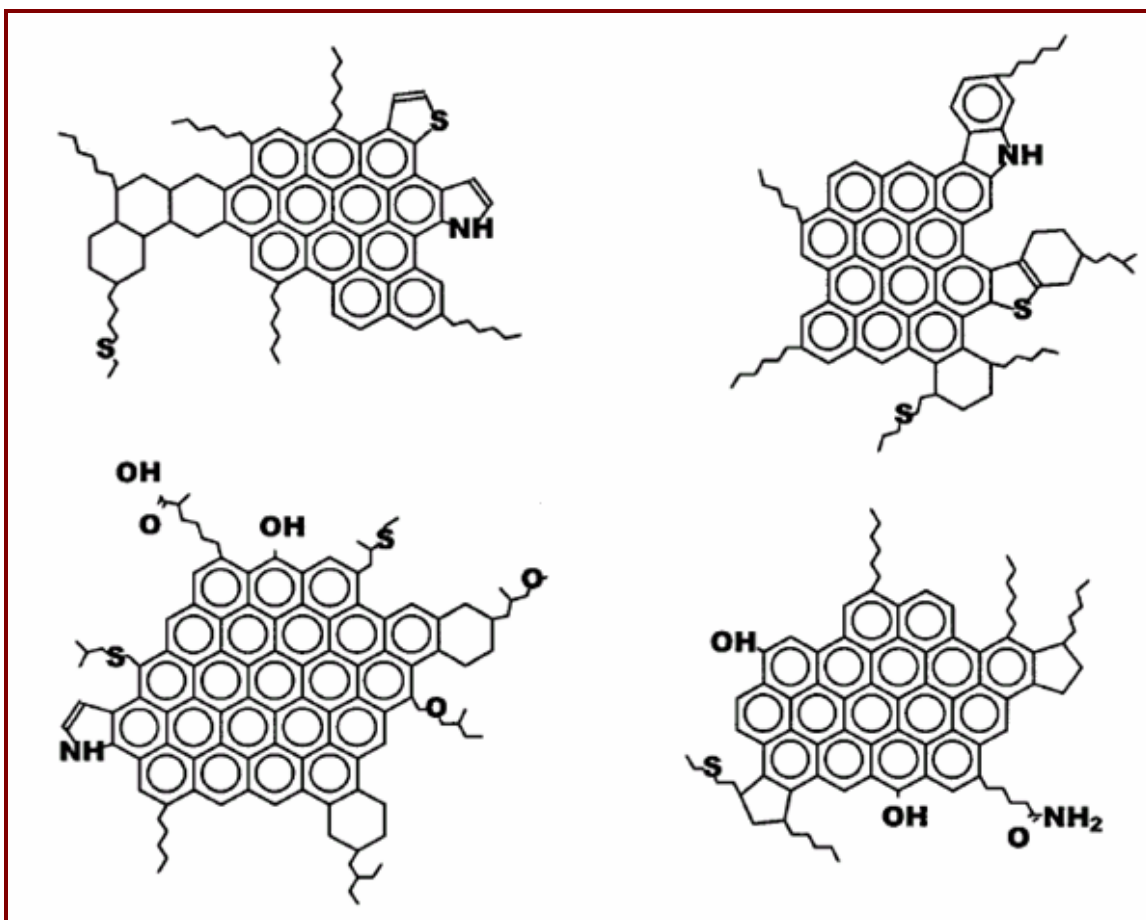
Figura 1. Diagrama para separación de Asfaltenos



Los asfaltenos son moléculas planas, poli aromáticas y poli cíclicas que contienen heteroátomos y metales, que existen en un estado de agregación en suspensión y están rodeados y estabilizados por resinas (agentes peptizantes); no son puros, ni son moléculas idénticas y están poli dispersos.

El contenido de compuestos aromáticos en los asfaltenos está entre 40-60 %, con una relación atómica H/C de 1-1.2. Un alto porcentaje de los anillos aromáticos están conectados en la estructura intermolecular, y por esta razón las moléculas de asfaltenos presentan formas aplanadas. (figura 2).

Figura 2. Estructura de algunos Asfaltenos



Hace poco tiempo, se consideraba que los asfaltenos eran macromoléculas de peso molecular de varias decenas de miles y que poseían una estructura formada por una matriz aromática muy condensada, rodeada por anillos nafténicos y cadenas alifáticas, con incorporación de grupos multifuncionales, tales como: alcoholes, aminas primarias y secundarias, ácidos, éteres, esterres, cetonas y conjuntos más complejos como las porfirinas.

Los últimos estudios ^(8,9), donde se presentan técnicas de análisis, tienden a presentar a los asfaltenos con un peso molecular mucho menor, entre 800 y 3000 y una matriz menos condensada, limitada a 4 ó 5 anillos bencénicos ligados por naftenos y prolongados por grupos alquil.

Los heteroátomos mayormente encontrados en asfaltenos de diferentes crudos son azufre (S), nitrógeno (N) y oxígeno (O). El nitrógeno se encuentra mas como parte de los conjuntos aromáticos, mientras que el oxígeno y el azufre forman puentes entre ellos en una estructura cíclica o lineal. El azufre existe predominantemente como heterociclos tiofénicos (típicamente de 65-85%) con el resto en grupos sulfídicos. El nitrógeno se relaciona con los grupos pirrólicos, piridinicos y quinólicos, siendo el grupo dominante el pirrolico. Los grupos relacionados con el oxígeno son hidroxilico, carbonilo, carboxílico, y éter⁽¹⁰⁾.

La espectroscopia infrarroja, espectroscopia fotoacústica, resonancia magnética nuclear y la difracción de rayos X, se han utilizado para determinar los detalles químicos de los heteroátomos, determinación de grupos funcionales, y tipos de carbono e hidrogeno en los asfaltenos.

1.3 ESPECTROSCOPIA FOTOACÚSTICA (PAS)

Un área muy activa de investigación en la física y química aplicada en estos días viene bajo el título general de fotoacústica (PA). Este término implica la generación de ondas acústicas u otros efectos termoelásticos por cualquier tipo de flujo energético incidente que va de la radiación electromagnética por rayos x, electrones, protones, ultrasonido, etc. El efecto PA fue descubierto por Alexander Graham Bell en 1880 quien observó que se producían sonidos audibles en muestras ópticamente absorbentes, cuando estas se exponían a la luz solar modulada⁽¹¹⁾.

La generación de la señal fotoacústica se debe al efecto causado por el calentamiento de la muestra después de haber absorbido energía óptica -llamado efecto fototérmico- y que produce una expansión termoelástica en la muestra. La detección de la señal puede ser directa, cuando la onda acústica se produce y

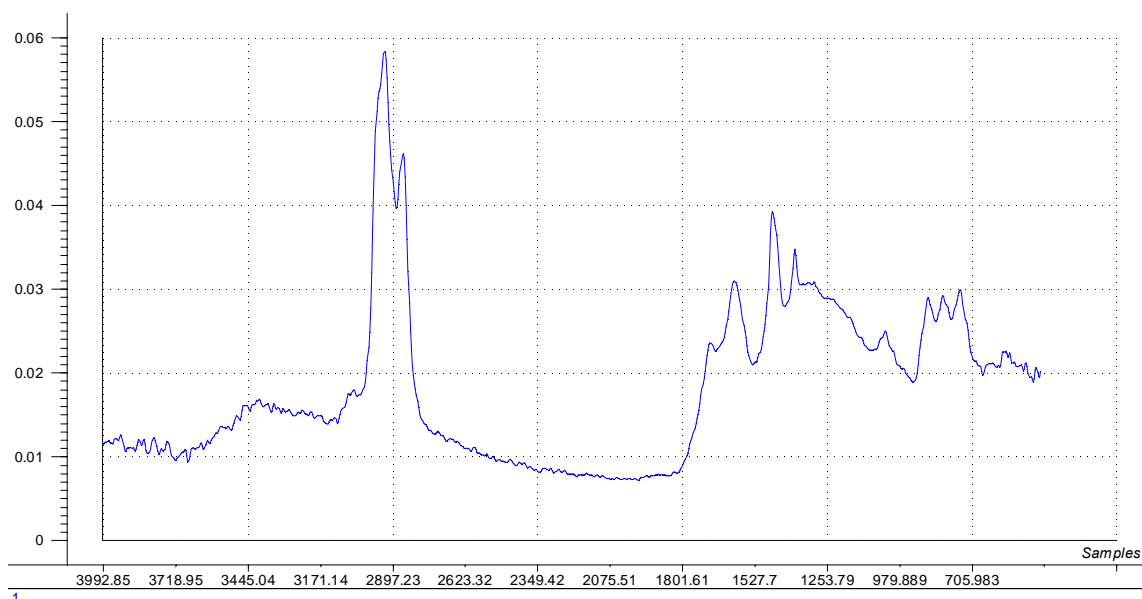
detecta en la muestra; o indirecta cuando la onda acústica se genera en un medio de acoplamiento (que puede ser un gas o líquido) adyacente a la muestra, y se debe a la transferencia de calor entre la muestra y el medio de acoplamiento, y a la transmisión acústica desde la muestra.

El efecto fotoacústico es muy débil; solamente una pequeña porción de la energía óptica absorbida se puede convertir en energía acústica: la relación entre la energía acústica generada y la energía óptica absorbida es del orden de 10^{-12} a 10^{-8} . No obstante, en los últimos años resurgió como técnica espectral debido a varios factores ⁽¹²⁾:

- a) Se han fabricado fuentes de radiación láser de gran potencia.
- b) Se han desarrollado detectores acústicos de alta sensibilidad.
- c) Ha demostrado ser una de las técnicas más versátiles y sensibles tanto para muestras gaseosas como para muestras en fase condensada.
- d) La técnica permite aplicaciones que son únicas en su género:
 - Estudios espectroscópicos no destructivos de materiales opacos y materiales pulverizados.
 - Estudios de perfilamiento superficial de subcapas .

La espectroscopia fotoacústica se origina en la fracción no radiativa (producción de calor) de la energía liberada por el decaimiento desde un estado excitado al estado fundamental; ella complementa a la espectroscopia de absorción y fluorescencia, proporcionando información sobre los procesos no radiativos ocurridos entre los procesos de absorción y emisión. Proporciona espectros similares a los de absorción óptica en cualquier tipo de material líquido, sólido, semisólido, gaseoso, cristalino, pulverizado o amorfo (Figura 3).

Figura 3. Espectro PAS-IR de un asfalteno.



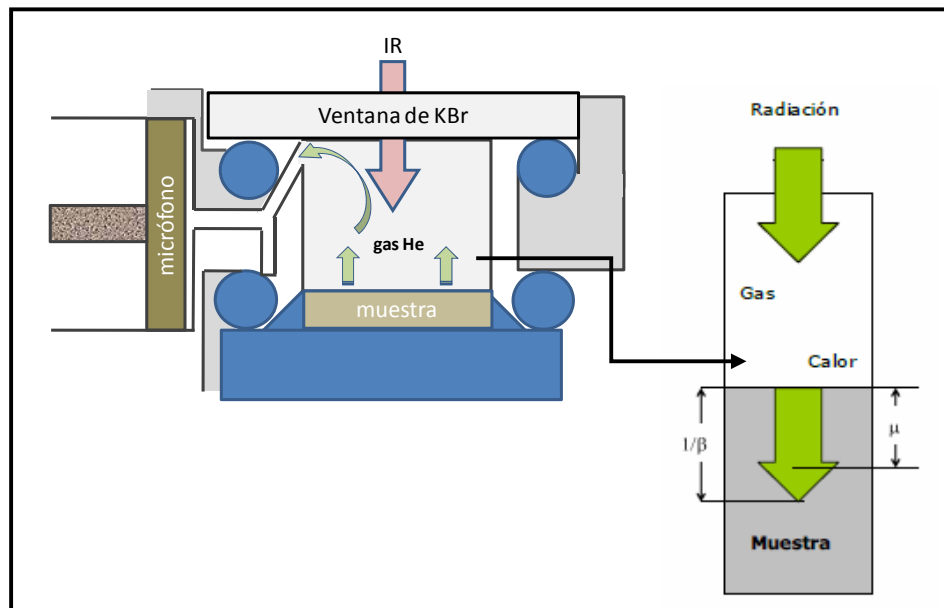
La diferencia fundamental entre la espectroscopia fotoacústica y la espectroscopia de absorción convencional, es la forma en que se detecta la absorción. En la espectroscopia convencional esta medición es indirecta, en realidad se detecta la radiación transmitida y posteriormente se realiza la conversión a unidades de absorbancia. Las muestras con bajos coeficiente de absorción, β (ver figura 4), difícilmente pueden ser analizadas por esta espectroscopia porque la magnitud de la señal transmitida es cercana a la señal de referencia, y es difícil discriminar entre las dos señales ⁽¹³⁾.

Lo contrario ocurre con muestras con altos valores de β , porque la magnitud de la señal transmitida es muy pequeña comparada con la señal de referencia y fácilmente se pierde en el ruido de fondo. El rango de detección lineal de éste tipo de experimentos, está limitado a las muestras con un rango medio de β .

En contraste, la absorción de la radiación electromagnética en espectroscopia fotoacústica se detecta con un transductor acústico (micrófono o cristal piezoeléctrico), que produce una señal eléctrica cuya amplitud pico a pico, es

directamente proporcional a la magnitud de los cambios de presión en la muestra. Por lo tanto, el rango de detección lineal de este tipo de experimento es muy bueno para las muestras con valores altos de β , debido a que el cambio inducido de presión es directamente proporcional a la energía absorbida.

Figura 4. Generación de la señal fotoacústica



El modelo normal que describe la señal PA de una muestra sólida, se propuso por Rosencwaig y Gersho ⁽¹⁴⁾. Estos autores, mostraron que el flujo de calor periódico entre la muestra y el gas de la cámara cerrada tiene un mecanismo básico responsable de la generación de la señal acústica. Se basaron en la opacidad, la difusividad térmica y el grosor total de la muestra. Para muestras ópticamente opacas, la magnitud de la señal fotoacústica, q , y la fase ϕ , se pueden expresar como se muestra en las ecuaciones 1 a 3.

$$q = C \left(\frac{1}{\rho_s c_{p,s} K_s} \right)^{1/2} \left[\frac{\beta \mu^2}{(\beta \mu + 1)^2 + 1} \right]^{1/2} \quad (\text{Ecuación 1})$$

$$\phi = \tan^{-1} \left(1 + \frac{2}{\beta \mu_s} \right) \quad (\text{Ecuación 2})$$

$$C = \frac{I_0 \gamma P_0 k_g^{1/2}}{4 \lambda T_0 \pi f (\rho_g C_{p,g})^{1/2}} \quad (\text{Ecuación 3})$$

Donde ρ es la densidad, $C_{p,g}$ la capacidad calorífica del gas circundante, k la conductividad térmica, β el coeficiente de absorción óptica, μ la profundidad de muestreo térmico, I_0 la intensidad de la radiación incidente, γ el cociente de las capacidades caloríficas específicas, P_0 la presión del gas, λ la longitud de la celda, T_0 la temperatura del gas y f la frecuencia de modulación de la radiación.

La señal fotoacústica es proporcional a la intensidad de la radiación, inversamente proporcional a la temperatura y volumen de la celda, y dependiente de las propiedades termofísicas de la muestra y del gas circundante. El coeficiente de absorción óptica, β , propio de cada material, se define como el inverso de la profundidad en donde el 63% de la radiación incidente es absorbida. Mientras que, μ , es la profundidad a la cual la energía absorbida constituye un 63% de la señal fotoacústica. Esto está dado por la ecuación 4

$$\mu = \left(\frac{\alpha}{\pi f} \right)^{1/2} \quad (\text{Ecuación 4})$$

Donde α ($\alpha = \beta^{-1}$) es la difusividad térmica y está dada por la ecuación 5

$$\alpha = \left(\frac{k}{\rho C_p} \right) \quad (\text{Ecuación 5})$$

La señal fotoacústica se conduce desde μ hasta la superficie donde calienta una capa delgada del gas sobre la superficie de la muestra. La cantidad de calor transportada de este modo es dependiente en las características termofísicas de la muestra y del gas. La capa calentada del gas actúa como pistón acústico que se amplía y contrae contra el gas restante en la celda. Este proceso causa una oscilación periódica de la presión (onda acústica), que se propaga a través del volumen del gas. Un micrófono situado en el espacio del gas detecta esta señal y la retransmite a un amplificador. Este toma la señal y la amplifica rechazando las señales de altas frecuencias que normalmente corresponden al ruido electrónico. Un sistema de adquisición de datos recoge y registra la señal de salida del amplificador⁽¹⁵⁾.

En PAS-IR se utilizan celdas que se acoplan a equipos FTIR comerciales en los cuales la celda fotoacústica sustituye al detector del instrumento. La celda fotoacústica contiene la muestra y el micrófono donde se detectan los cambios de presión del gas circundante. Después de una filtración adecuada para remover las frecuencias de la señal de salida inherentes al rango espectral de trabajo, la señal es procesada por el equipo IR y obtenido el interferograma, es transformada para producir un espectro de absorción.

1.4 QUIMIOMETRÍA

El análisis de muestras que contienen mezclas de compuestos químicos constituye uno de los retos más importantes de la química analítica. La dificultad aumenta cuando se pretende determinar simultáneamente estos compuestos a bajas concentraciones o en presencia de otras sustancias que tienen estructuras y propiedades similares, las cuales pueden interferir en su determinación. El análisis de muestras complejas requiere utilizar un conjunto de muestras de referencia que tenga las mismas características que las muestras a analizar y de algún método

instrumental que mida alguna propiedad característica de los analitos. En la utilización de un sistema de detección para las muestras a analizar nos podemos encontrar con problemas debido a interferencias, a efectos de matriz, ruido de fondo, etc., que precisan una especial atención.

El método tradicionalmente utilizado para resolver la falta de selectividad en muestras complejas, consiste en la separación previa de los compuestos químicos puros antes de analizarlos. Si la separación ha sido completa (ej. mediante cromatografía) tan solo es necesaria una medida univariable para tener la información cuantitativa deseada. No obstante, ésta presenta un costo elevado, ya que se necesita disponer de una instrumentación analítica sofisticada e invertir bastante tiempo para optimizar las condiciones de separación. Como alternativa, se han desarrollado los métodos quimiométricos, los cuales permiten extraer la información requerida sobre los componentes de interés de las muestras a partir del tratamiento de datos, sin necesidad de su separación previa.

La quimiometría es una disciplina ampliamente utilizada en el área de la química, nacida aproximadamente al final de los años 70, y que, actualmente, está siendo ampliamente utilizada en diversos campos, especialmente en química analítica. El desarrollo rápido de la quimiometría en química analítica se debe fundamentalmente a la utilización masiva de los ordenadores acoplados a la moderna instrumentación. Se ha pasado de hacer medidas puntuales a una longitud de onda (univariables), a medidas espectrales (multivariables) con gran contenido de información donde los métodos tradicionales de análisis y procesado de datos resultan totalmente insuficientes.

Por tanto, uno de los campos de trabajo fundamental de la quimiometría es el desarrollo y aplicación de los nuevos métodos de *calibración multivariable*, que permiten extraer la máxima información química del análisis de muestras a partir de señales o respuestas instrumentales con muy poca selectividad⁽¹⁶⁾.

1.4.1 Calibración Multivariable

La calibración multivariable se puede definir como la actividad de encontrar relaciones entre una o más variables de respuesta “**y**” y una matriz de variables predictoras “**x**”, de manera que se cumpla la ecuación 6⁽¹⁷⁾ .

$$y = g(x) \quad \text{(Ecuación 6)}$$

La variable “**y**” puede ser un parámetro cuantitativo o cualitativo que representa una propiedad de interés en el sistema, y la matriz “**x**” contiene información relevante de la muestra determinada por un método de análisis químico. La función “**g(x)**” puede ser difícil de determinar por el número de variables en la matriz “**x**” y por la incertidumbre asociada a la determinación de los parámetros “**y**” y “**x**”. La forma de la función “**g(x)**” depende del método de regresión empleado, por lo cual puede existir más de una posibilidad de ajuste de los datos diferenciándose básicamente en la complejidad de la función y en sus parámetros estadísticos.

1.4.1.1 Clasificación de los métodos de calibración multivariable: Los métodos de calibración pueden tener una base totalmente empírica o bien estar soportados por una base teórica que explica el fenómeno, físico o químico, responsable de la señal analítica (p.ej. la ley de Lambert-Beer en espectrometría de absorción). Se han propuesto diversas clasificaciones de los métodos de calibración. Una de ellas es la presentada por Martens & Naes [1989] en su teoría de química analítica; proponen una clasificación de los métodos de calibración basados en conceptos menos matemáticos:

(a) Según la relación entre las variables dependiente e independiente:

- *Calibración lineal* descrita por un modelo lineal,

$$y = b_0 + \sum_{k=1}^K b_k X_k \quad \text{(Ecuación 7)}$$

siendo b_0 , y b_k los parámetros a determinar y, X_k las variables independientes. Es decir, son modelos lineales en los parámetros.

- *Calibración no-lineal*: basada en modelos no lineales en los parámetros:

$$y = \alpha x^\beta \quad (\text{Ecuación 8})$$

donde α y β son los parámetros a determinar.

(b) Según la forma de encontrar la relación entre las variables:

- *Métodos directos*: los parámetros de la calibración se calculan directamente a partir de la señal de cada uno de los analitos de forma individual.

- *Métodos indirectos*: los parámetros de la calibración se calculan a partir de las señales analíticas de mezclas de los componentes.

(c) *Rígidos - Flexibles*: en los primeros es necesario tener información de todas las especies presentes que pueden contribuir a la señal, mientras que en los flexibles únicamente es necesario tener información de los analitos que se desea cuantificar.

(d) Según cuál sea la variable dependiente y cuál la independiente:

- *Calibración clásica*: sigue un criterio directamente relacionado con la ley de Beer-Lambert. La señal analítica actúa como variable dependiente de la concentración, que es la variable independiente.

- *Calibración inversa*: se utiliza la concentración como variable dependiente y la señal analítica como independiente.

1.4.2 Construcción de modelos de calibración multivariable.

El objetivo de los métodos de calibración multivariable es establecer modelos de calibración que sean capaces de predecir el valor de la propiedad estudiada para nuevas muestras, a las que se ha registrado la señal analítica. Para que estas predicciones sean fiables, los modelos obtenidos han de ser robustos y, para ello, se deben seguir las etapas que se describen a continuación:

a) *Definición de un conjunto de calibración o conjunto de entrenamiento (training set)*. De él se conoce la propiedad a determinar, es representativo del sistema en estudio y debe permitir realizar predicciones futuras sobre un conjunto nuevo de muestras. Se deben considerar tanto las fuentes de variación químicas como las físicas (origen de las muestras, tratamientos recibidos, etc), las cuales pueden afectar al espectro registrado.

b) *Registro de la información analítica*. Incluye la selección de condiciones experimentales óptimas de adquisición espectral y tratamientos previos de acondicionamiento de la muestra.

c) *Pretratamiento de los datos*. En esta etapa, se minimizan las contribuciones no deseadas presentes en la señal analítica que disminuyen la reproducibilidad y pueden provocar que el sistema presente comportamientos que darían lugar a estimaciones erróneas de los parámetros deseados. Los pretratamientos utilizados en este trabajo se detallan en el apartado siguiente.

d) *Selección del modelo*. Establece la relación más simple posible entre la propiedad a determinar y la señal analítica. Esto, en muchas situaciones, implica un estudio complejo de la matriz de datos analíticos obtenidos, ya que no siempre es necesaria la utilización de toda la información registrada. Es necesaria la aplicación de técnicas de regresión sobre los datos espectrales.

Dentro de las técnicas de correlación más comunes se tiene la regresión lineal múltiple (MLR), la regresión por componentes principales (PCR), y la regresión por mínimos cuadrados parciales (PLS).

e) *Validación del modelo*. La validación se desarrolla aplicando el modelo generado sobre un grupo de muestras y estos resultados son comparados estadísticamente con los valores de referencia. Si se emplean muestras diferentes a las empleadas en la calibración del modelo se tiene el método de validación externa. Si se emplean muestras incluidas en la calibración del modelo se tiene el procedimiento de validación cruzada.

Después de calcular la ecuación para una determinada calibración es esencial determinar la capacidad de la misma para predecir valores y desconocidos. Uno de los procedimientos más utilizados es el cálculo del error medio en predicción basado en el cuadrado de las diferencias entre los valores de referencia y los predichos por el modelo de calibración.

En todo modelo de calibración se puede definir la raíz cuadrada de la media del error de calibración al cuadrado (RMSEC) como:

$$RMSEC = \sqrt{\frac{\sum_{i=1}^C (y_i - \hat{y}_i)^2}{n_c}} \quad (\text{Ecuación 9})$$

donde y_i es el valor de referencia de la respuesta, \hat{y}_i es el valor de la respuesta predicho por el modelo y n_c es el número total de muestras en el conjunto de calibración.

- Validación mediante cross-validación:

La cross-validación o validación cruzada, consiste en la utilización de los datos del conjunto de calibración para validar el modelo. En la cross-validación, el conjunto 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 realizada la calibración se pone a prueba el modelo con los datos del grupo de cancelación excluido. Se repite el proceso de manera sucesiva dejando cada vez un grupo de cancelación fuera del proceso de cálculo. Si el número de grupos de cancelación es igual al número de objetos en el conjunto de calibración, cada vez quedará un objeto fuera del cálculo del modelo y el proceso se denomina *full-cross-validación* o *leave-one-out*. Cuando se utiliza la cross-validación para validar un modelo se puede definir la raíz cuadrada de la media del error en cross-validación al cuadrado (RMSECV) como:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^c (y_i - \hat{y}_{cv,i})^2}{n_c}} \quad (\text{Ecuación 10})$$

- Validación mediante conjunto de validación externa:

La validación mediante un conjunto de datos externo consiste en dividir inicialmente el conjunto de muestras en dos grupos: calibración y validación. El modelo de regresión se calcula a partir de los datos del conjunto de calibración y, el modelo calculado, se pone a prueba con los datos del conjunto de validación que no han sido utilizados para calcular la ecuación de calibración del modelo. En este caso se puede definir la raíz cuadrada de la media del error de predicción al cuadrado (RMSEP) para el conjunto de validación.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n_p} (y_i - \hat{y}_i)^2}{n_p}} \quad (\text{Ecuación 11})$$

Donde n_p es el número de muestras en el conjunto de validación.

f) *Predicción de nuevas muestras*. Con el modelo construido y validado, se calcula la propiedad estudiada en nuevas muestras a las que, previamente, se ha registrado la señal analítica^(18, 19,20).

1.4.3 Pretratamiento de los datos registrados

Los pretratamientos espectrales son correcciones del espectro basadas en tratamientos matemáticos con la finalidad de aumentar la calidad de la señal instrumental, reduciendo al máximo la información no asociada a la propiedad estudiada, como pueden ser desplazamientos de la línea base, ruido espectral, etc. Se denomina ruido al componente espectral no relacionado con ninguna propiedad física o química de la muestra como puede ser el ruido instrumental o de iluminación. Además, otras fuentes de variación como la temperatura y humedad pueden aportar interferencias a la señal analítica.

Existen bastantes tratamientos matemáticos que se pueden aplicar antes de intentar hallar la relación señal-concentración para mejorar la calidad de los datos, entre las más utilizadas están⁽¹⁹⁾:

a) *Normalización*. Es utilizada para obtener todos los datos en aproximadamente la misma escala, o para obtener una distribución más equitativa de las variaciones y los valores medios. Existen varios tipos de normalización:

- Media de la normalización (Mean Normalization): Es el tipo de normalización más clásica. Consiste en dividir cada fila de una matriz de datos por su media, neutralizando así la influencia de los factores ocultos.

- Máxima normalización (Maximum normalization): Esta es una alternativa a la normalización clásica. Divide cada fila por su valor máximo absoluto en lugar de la media.

Propiedad de un máximo de muestras normalizadas:

- Si todos los valores son positivos: el valor máximo se convierte en 1.
 - Si todos los valores son negativos: el valor mínimo se convierte en -1.
 - Si el signo de los cambios de valores en la curva: o bien el valor máximo se convierte en 1 o el mínimo valor se convierte en -1.
- Normalización por rangos (Range normalization): En esta transformación se normaliza un espectro X_i calculando el área bajo la curva del espectro. Se intenta corregir el espectro de longitud de la trayectoria indeterminada cuando no hay forma de medirla, o aislar a un grupo de un componente constante.

b) *Suavizado espectral*. Mediante el suavizado se intenta reducir el ruido aleatorio existente en la señal instrumental. La técnica más utilizada es la propuesta por Savitzky y Golay⁽²¹⁾. El principio en el que se basa este método es que, para un pequeño intervalo de longitudes de onda, se puede ajustar un polinomio del grado adecuado. Los nuevos valores tras el ajuste son una mejor estima que los valores medidos ya que se ha eliminado parte del ruido que los afectaba.

c) *Derivadas*. La diferenciación o cálculo de derivadas permite acentuar las diferencias existentes en los datos espectrales. Tanto la primera como la segunda derivada se utilizan a menudo para el tratamiento de los datos. La segunda derivada elimina el ruido de fondo lineal y constante. Los dos principales algoritmos de diferenciación son el de Savitzky-Golay y el de Norris⁽²²⁾. El primero, permite calcular derivadas de primer orden o mayor incluyendo un factor de suavizado que determina el número de variables adyacentes que se usarán en la estimación de la aproximación polinómica utilizada en la derivación. El algoritmo

de Norris, a diferencia del anterior, solo permite el cálculo de derivadas de primer orden.

Una desventaja del uso de las derivadas es que disminuyen el valor de la relación señal-ruido, por esta razón, se recomienda realizar un suavizado de la señal antes de la diferenciación de los datos. Otra desventaja es que en ocasiones los modelos de calibración obtenidos mediante datos espectrales tratados con primera o segunda derivada, son menos robustos frente a cambios instrumentales, como por ejemplo derivas de la longitud de onda, que ocurren a lo largo del tiempo, por lo que habría que revisar las calibraciones

1.4.4 Análisis por componentes principales (PCA)

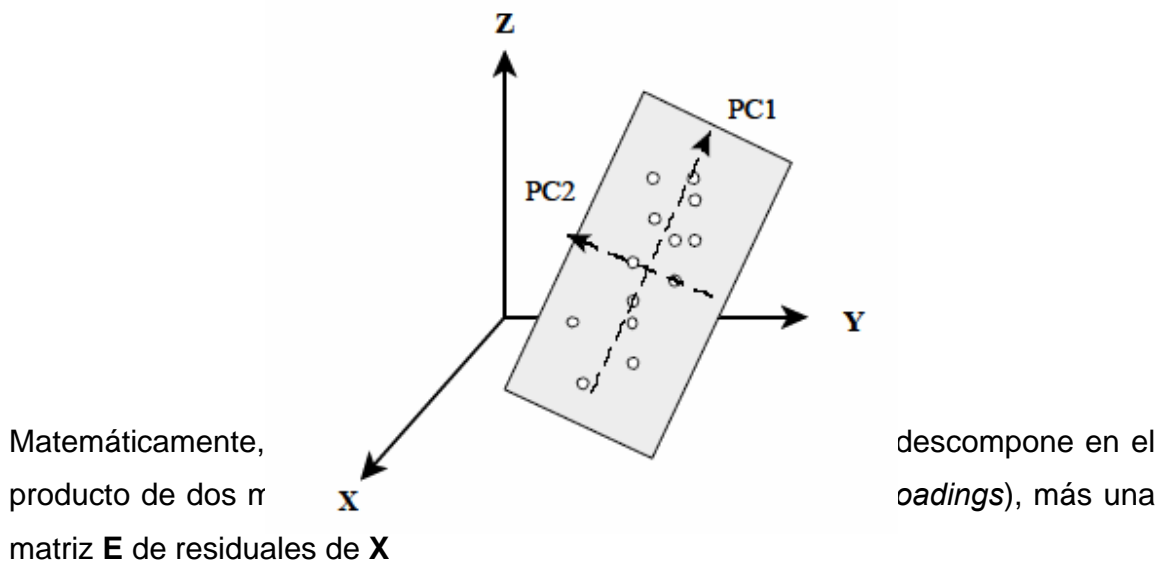
El análisis en componentes principales (PCA) se aplica en química con diferentes finalidades [Jackson 1991; Wold, 1987]. Entre ellas cabe destacar la clasificación de muestras y la reducción en la dimensionalidad de los datos, siendo éste el paso previo de muchos métodos multivariantes de análisis de datos.

En la técnica multivariante PCA, un número elevado de variables, más o menos correlacionadas, son transformadas en un número inferior de variables ortogonales (no correlacionadas) y que contienen la información principal del sistema, eliminando la información redundante o reduciendo la contribución del ruido.

Se parte de una matriz de datos \mathbf{X} , donde cada fila es una muestra u objeto, y cada columna una variable, que puede ser la absorbancia a una longitud de onda y tiempo determinados. La dimensionalidad de esta matriz, si se han registrado m muestras y k variables, será $m \times k$.

Geoméricamente, el PCA es un cambio en la dirección y dimensionalidad de los ejes (figura 5). Busca las direcciones ortogonales que explican la máxima variabilidad de las muestras y las utiliza como nuevos ejes de coordenadas llamados componentes principales (PCs) para representarlas. El primer componente principal es la dirección que explica la máxima variabilidad; el segundo se escoge de tal forma que sea perpendicular al primero y que explique la máxima variabilidad una vez extraída la explicada por el primer componente principal, y así sucesivamente. Para poder definir estos nuevos ejes se utilizan sus *loadings* que son los cosenos de los ángulos que forman con los ejes antiguos. Los *scores* son las coordenadas de las muestras en estos nuevos ejes^(23,25).

Figura 5. Interpretación geométrica de un PCA.



$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (\text{Ecuación 12})$$

Los diferentes componentes principales no contienen la misma información. Los primeros describen la fuente de variación más importante de los datos, que se puede asociar a la información más relevante. Para tener una reproducción

perfecta de la matriz \mathbf{X} se necesitan k vectores de *loadings* y *scores*, sin embargo, es posible representar la matriz original con un número $a < k$ de vectores

$$\mathbf{X} = t_1 p_1^T + t_2 p_2^T + \dots + t_a p_a^T + \mathbf{E} \quad (\text{Ecuación 13})$$

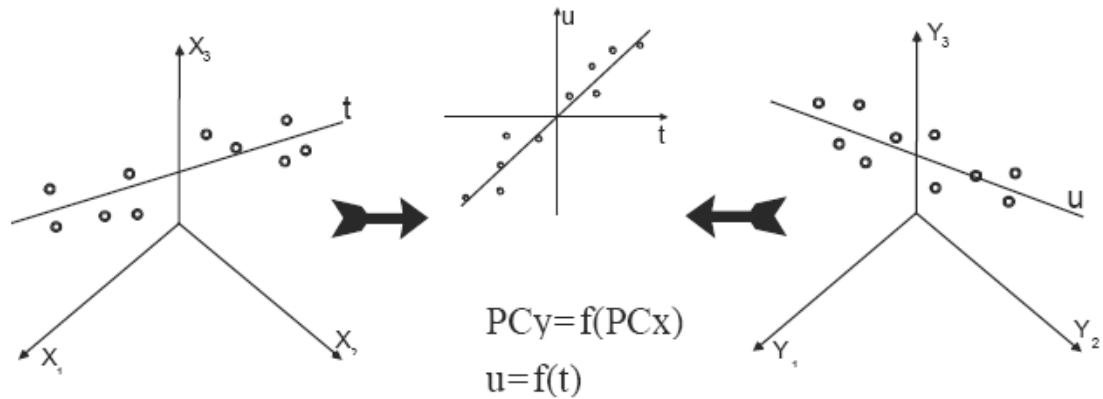
puesto que la información relevante está contenida en los a primeros componentes, mientras que los últimos sólo describen variaciones debidas al ruido. Esto permite una reducción importante del número de variables.

Resumiendo, se puede decir que el conjunto de datos \mathbf{X} , que estaba descrito por variables correlacionadas, en este momento está definido por un nuevo conjunto de variables, llamadas componentes principales, que son variables no correlacionadas entre sí, en un nuevo sistema de ejes ortogonales.

1.4.5 Regresión por mínimos cuadrados parciales (PLS)

El método de Regresión Parcial por Mínimos Cuadrados (PLS, *Partial Least-Squares Regression*) fue desarrollado por H. Wold en 1975 [Wold, H.; 1975]. Este método pretende que los primeros componentes contengan la mayor información para la predicción de la matriz \mathbf{Y} . Para ello, durante la etapa de calibración, el algoritmo PLS utiliza tanto la información contenida en la matriz de datos (matriz \mathbf{X} , p. ej. Datos espectroscópicos) como la información contenida en la matriz de la propiedad a determinar (matriz \mathbf{Y} , p. ej. densidad), obteniéndose unas variables auxiliares llamadas variables latentes. ^(23,24,25) (Figura 6).

Figura 6. Descripción gráfica del modelo de regresión PLS



Como en el caso del análisis por componentes principales, estas nuevas variables en \mathbf{X} y \mathbf{Y} se pueden representar como un producto de matrices según se muestra en las ecuaciones 14 y 15.

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} = \sum t_a p_a^T + \mathbf{E} \quad (\text{Ecuación 14})$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} = \sum u_a q_a^T + \mathbf{F} \quad (\text{Ecuación 15})$$

Donde:

\mathbf{T} y \mathbf{U} son las matrices de puntuación (scores) de \mathbf{X} y \mathbf{Y} respectivamente;

\mathbf{P} y \mathbf{Q} son las matrices de carga (loadings) de \mathbf{X} y \mathbf{Y} respectivamente;

\mathbf{E} y \mathbf{F} son los residuos.

Las variables originales \mathbf{X} y \mathbf{Y} se pueden relacionar mediante los scores de cada una de las nuevas variables latentes como lo muestra la ecuación 16.

$$u_a = b_a t_a \quad (\text{Ecuación 16})$$

Donde b_a es el coeficiente de regresión para cada variable latente obtenido a través de la ecuación 17.

$$b_a = \frac{u_a^T t_a}{t_a^T t_a} \quad (\text{Ecuación 17})$$

Los coeficientes b_a hallados para cada componente son agrupados en una matriz diagonal \mathbf{B} , que contiene los coeficientes de regresión de los scores \mathbf{U} y \mathbf{T} de las

matrices \mathbf{Y} y \mathbf{X} respectivamente. De tal manera, la matriz \mathbf{Y} puede ser calculada por medio de la ecuación 18.

$$\mathbf{Y} = \mathbf{TBQ}^T + \mathbf{F} \quad (\text{Ecuación 18})$$

2. METODOLOGÍA

En el marco del convenio UIS-ICP, el Laboratorio de Espectroscopia Atómica y Molecular, LEAM de la Universidad Industrial de Santander, ha venido desarrollando en los últimos años investigaciones en la caracterización del petróleo Colombiano, generando aportes científicos y tecnológicos en esta área.

2.1 SELECCIÓN DE MUESTRAS

En este trabajo se estudiaron 14 muestras de Asfaltenos de diferentes crudos Colombianos que fueron suministradas por el Instituto Colombiano del Petróleo (ICP), así como los valores de la propiedad a predecir. (Tabla 3).

Tabla 3. Análisis próximo de Asfaltenos analizados

| NOMBRE ASFALTENO | DENSIDAD |
|-----------------------------|-----------------|
| *ASF 140554 | 1.08135 |
| ASF 140599 | 1.07048 |
| ASF 145774 | 1.10851 |
| ASF 147887 | 1.12913 |
| ASF 147888 | 1.11917 |
| ASF 148680 | 1.10942 |
| ASF 152682 | 1.09117 |
| ASF 166391 | 1.14217 |
| ASF 168829 | 1.10286 |
| ASF 168830 | 1.12868 |
| ASF 172067 | 1.08028 |
| ASF 173567 | 1.14223 |
| ASF 174891 | 1.1002 |
| ASF 175347 | 1.08392 |

*Asfalteno

2.2 CARACTERIZACIÓN POR ESPECTROSCOPIA FOTOACÚSTICA

2.2.1 Instrumentación

El laboratorio de espectroscopia atómica y molecular (LEAM) de la UIS cuenta con una celda fotoacústica MTEC 200 que se adaptó a un equipo IR Tensor 27 marca Bruker, como primer paso para la adquisición de espectros PAS-IR.

Figura 7. Sistema PAS acoplado al equipo IR Tensor 27 marca Bruker



El sistema PAS IR que fue acoplado consta de las siguientes partes:

a. *Celda PAS* (figura 7a). Está formada principalmente por un compartimiento para la muestra (portamuestra), por una cámara que mantiene confinado el gas inerte y garantiza que esté en contacto con la muestra mientras se adquiere cada espectro, y por un micrófono que detecta las variaciones de presión del gas. La figura 8 muestra el exterior de la celda PAS. En el interior de la celda la cámara es

sellada para garantizar que el gas no se escape a la atmósfera durante la toma de cada espectro.

Figura 8. Celda PAS modelo MTEC 200



b. *Amplificador* (figura 7b) usado para amplificar la señal y controlar remotamente la ganancia del pre-amplificador localizado dentro del micrófono. La ganancia puede modificarse desde 1 hasta 128. La magnitud de la ganancia se determina de acuerdo a la velocidad del espejo del interferómetro.

c. *Interface* (figura 7c) usada para llevar la señal fotoacústica hasta el procesador de la señal del equipo FTIR.

d. *Flujómetro* (Figura 7d) se encarga de controlar el flujo de gas He (figura 7e) que es introducido a la celda. El gas He además de servir como el portador de la señal, se usa para purgar la celda de los gases atmosféricos (vapor de H₂O y CO₂). El flujo de gas debe ser controlado para minimizar el riesgo de daño del micrófono.

Luego de acoplar la celda se variaron parámetros del sistema PAS y del equipo IR para obtener espectros con relación señal ruido aceptable. La relación señal-ruido

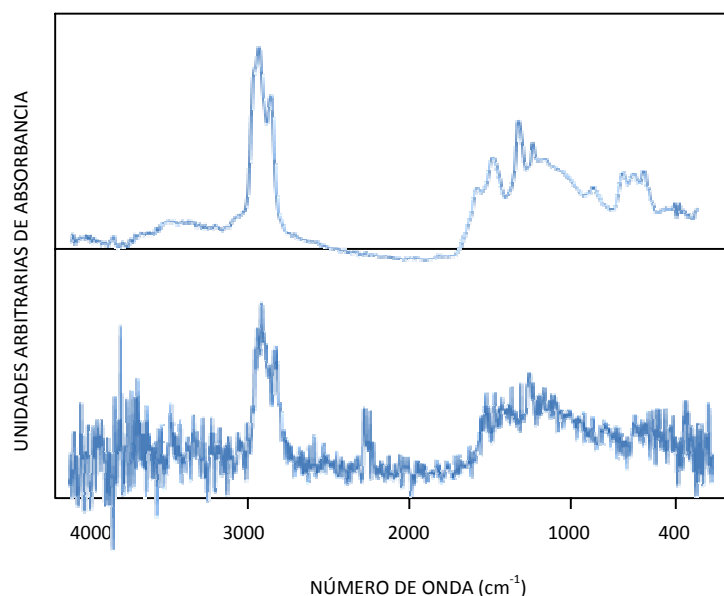
(S/N) es una medida útil para describir la calidad de un método analítico o el funcionamiento de un instrumento. La señal lleva la información relativa al analito, y el ruido la información ajena al analito. En general, el ruido degrada la exactitud y la precisión, y establece un límite inferior en la cantidad de analito que se puede detectar. Es imposible obtener datos sin ruido, pues algunos tipos de ruido tienen su origen en factores termodinámicos y cuánticos.

En la mayoría de medidas el valor promedio del ruido es constante e independiente de la señal. Por esto el efecto del ruido en el error relativo de una medida aumenta con la disminución de la cantidad de analito. La magnitud del ruido puede definirse como la desviación estándar de las medidas de la intensidad de la señal, que a su vez viene dada por la media de las intensidades. De este modo la relación S/N se expresa por el siguiente cociente (Ecuación 19)

$$\frac{S}{N} = \frac{\text{media}}{\text{desviación estándar}} = \frac{\mu}{\sigma} \quad (\text{Ecuación 19})$$

Esta relación se determinó a partir de la ecuación 19. La figura 10 muestra el aspecto de dos espectros de Asfaltenos con diferente relación S/N.

Figura 9. Efecto de la relación señal ruido en un espectro de una muestra de Asfalteno



Los parámetros de trabajo como resolución, ganancia, velocidad del espejo y número de barridos se variaron para establecer cuáles eran las condiciones que mostraban la mejor relación señal ruido en los espectros.

La señal fotoacústica, como se mostró en las ecuaciones 1 y 3, depende drásticamente de la frecuencia de modulación que a su vez es función de la velocidad del espejo del interferómetro ($f = \nu V$ donde V es la velocidad del espejo y ν es el número de onda). El equipo IR Bruker Tensor 27 donde se acopló el sistema PAS, permite trabajar a tres únicas velocidades de espejo del interferómetro: 0.0583, 0.633 y 1.27 cm/s. La relación existente entre la velocidad del espejo y la señal fotoacústica es la siguiente: Una velocidad alta implica que el tiempo de interacción de la radiación con la muestra sea corto y por lo tanto la intensidad de la señal fotoacústica es baja. A medida que la velocidad del espejo disminuye la intensidad de la señal aumenta. Un aumento de la señal podría causar una saturación de la señal.

La optimización de condiciones se realizó a dos velocidades de espejo: 0.0583 y 0.633 cm/s.

La selección de las mejores condiciones para la toma de espectros se basó en las relaciones señal ruido que se calcularon variando cada conjunto de parámetros, encontrando que a partir de una velocidad de espejo de 0.0583 cm/s, una resolución de 8 cm^{-1} , 256 barridos y una ganancia de 1, la relación alcanzaba un valor aproximado y máximo de 253.42, por lo cual se establecieron estas condiciones como los parámetros para la toma de todos los espectros de las muestras de Asfaltenos.

2.2.2 Verificación del Espectrómetro

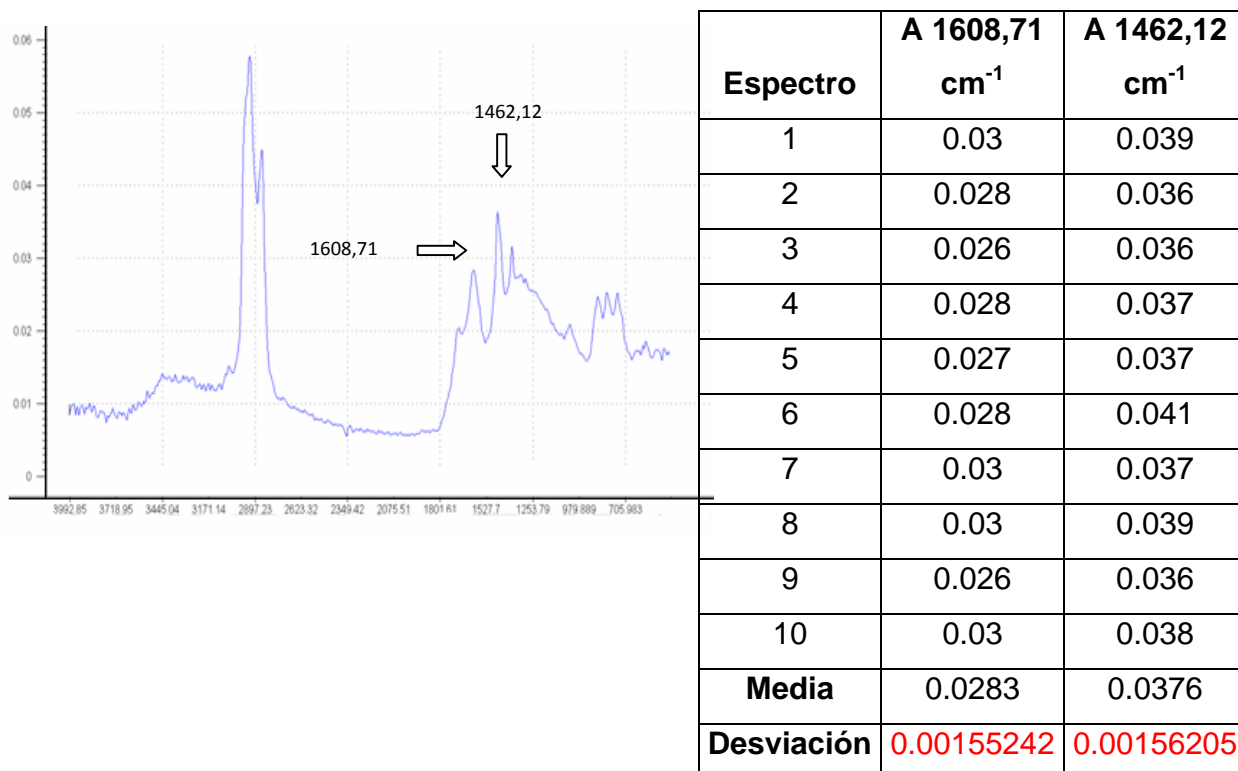
El desempeño del espectrómetro FT-IR fue verificado realizando una prueba establecida dentro de los protocolos de funcionamiento del equipo: repetibilidad en las medidas de absorbancia.

La repetibilidad es una condición necesaria para un análisis más profundo de la realidad de un sistema (espectrómetro). Cuando se hace un experimento por segunda vez, éste se repite en algunos aspectos pero no en todos; puesto que siendo lo suficientemente analítico, se puede considerar que, por ejemplo, las condiciones externas e internas del experimento no son exactamente las mismas y que, por tanto, el experimento es diferente para cada evento.

La calidad de las medidas espectrométricas depende de las propiedades analíticas del instrumento, destacando principalmente la exactitud de la longitud de onda, la linealidad y exactitud de la respuesta. Estas propiedades deben verificarse previamente para comprobar el estado del instrumento.

Para el presente caso, se calcularon las desviaciones estándar para las absorbancias medidas a 1608,71 y 1462,12 cm^{-1} en 10 espectros adquiridos sobre una misma muestra de asfalteno. Las desviaciones encontradas para cada uno de los espectros fueron inferiores a 0.001 (Figura 11), asegurando de esta manera la repetibilidad del espectrómetro.

Figura 10. Prueba de repetibilidad del espectrómetro FT-IR

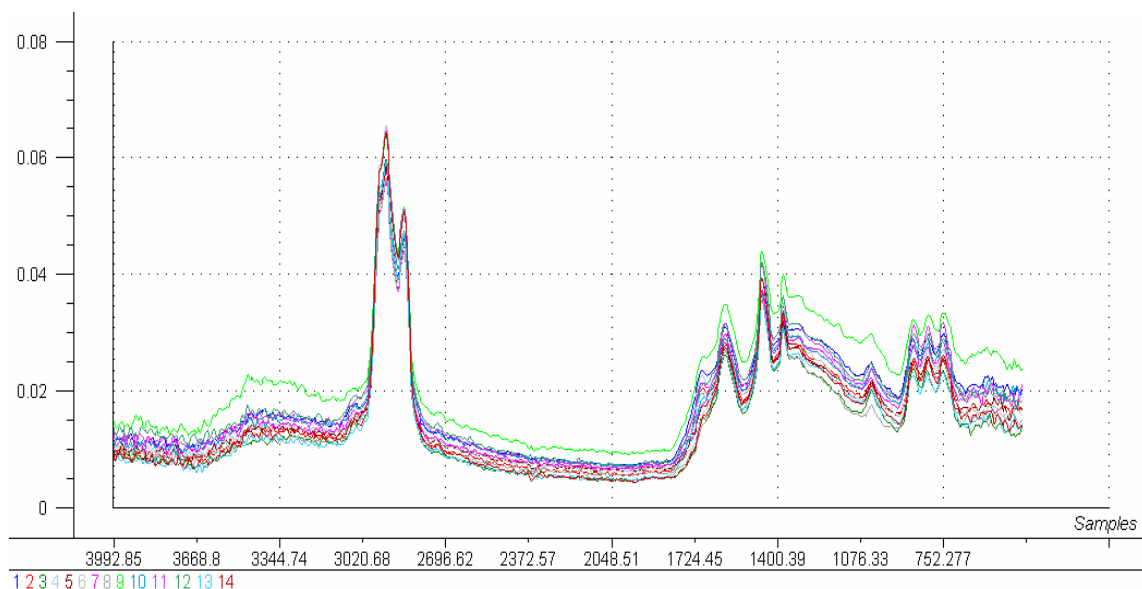


2.2.3 Adquisición y análisis de Espectros PAS

Los espectros PAS-IR de cada muestra, se tomaron en un intervalo espectral comprendido entre 4000 y 400 cm^{-1} , con 256 barridos, una velocidad de espejo de 0.0583 cm/s y una resolución de 8 cm^{-1} . Se usó negro de humo como material de referencia (background) y helio como gas de purga antes de la adquisición de

cada espectro. La figura 12 muestra los espectros PAS para las 14 muestras de asfaltenos.

Figura 11. Espectros PAS de todas las muestras de Asfaltenos



En cuanto al análisis espectral, en general los espectros de los 14 asfaltenos presentan las mismas bandas típicas para hidrocarburos: vibraciones de estiramiento simétrico y asimétrico de enlaces C-H aromáticos, $\nu_{CH_{AR}}(3057 - 3000 \text{ cm}^{-1})$, estiramiento simétrico y asimétrico de enlaces C-H alifáticos, $\nu_{CH_3 - CH_2}(2922 - 2852 \text{ cm}^{-1})$, bandas de deformación de los grupos metilo y metileno, $\delta_{CH_3}(1375 \text{ y } 1365 \text{ cm}^{-1})$ y $\delta_{CH_3 - CH_2}(1460 - 1440 \text{ cm}^{-1})$, las deformaciones de sustitución de los anillos aromáticos, $\gamma_{CH_{AR1}}(870 - 860 \text{ cm}^{-1})$, $\gamma_{CH_{AR2,3}}(800 - 810 \text{ cm}^{-1})$, $\gamma_{CH_{AR4}}(760 - 740 \text{ cm}^{-1})$.

Además de estas bandas características, algunos autores (Speight, 1991) muestran que los asfaltenos presentan bandas de absorción en 1680 y 1760 cm^{-1} y las atribuyen a presencia de cetonas. Sin embargo, Borrego *et al.* (1996) atribuyen esta banda (1600 cm^{-1}) a vibraciones de estiramiento de carbonos aromáticos C=C. Los espectros de los asfaltenos de este estudio, muestran

bandas de vibración entre $1609\text{-}1580\text{ cm}^{-1}$, que son asignadas a las reportadas por Borrego *et al.* (1996).

2.3 PROCESAMIENTO DE DATOS

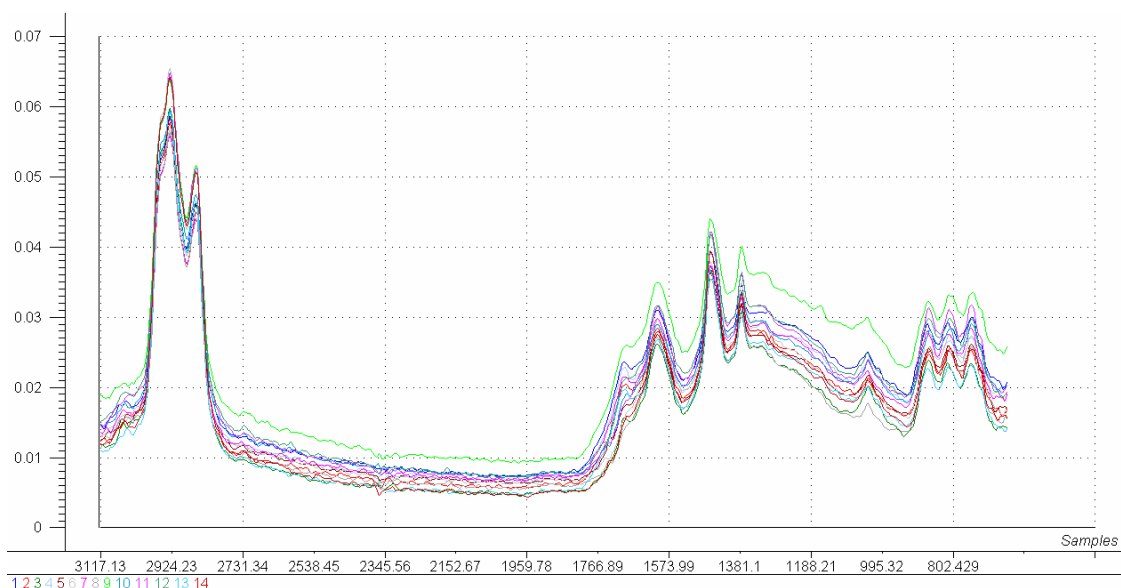
El pretratamiento y análisis de los datos espectroscópicos se realizó utilizando las herramientas del software The Unscrambler® versión 9.7⁽²⁷⁾ licenciado por Ecopetrol S.A.

Para verificar la calidad de la información espectral obtenida se realizó inicialmente una inspección visual de los espectros PAS con el fin de descartar la presencia de muestras atípicas o la influencia de posibles errores instrumentales y humanos durante la toma de muestras y posterior adquisición de espectros.

De esta manera se pudo visualizar también la región del espectro que contenía información valiosa y la región que no aportaba ningún tipo de información debiéndose esta a ruido aleatorio presente en el medio o el equipo.

La región comprendida desde 3120 cm^{-1} hasta 650 cm^{-1} es la que registra la señal correspondiente a las vibraciones existentes en las moléculas que componen las muestras; por lo tanto las regiones restantes en un rango espectral de 4000 cm^{-1} a 400 cm^{-1} simplemente no registran información de utilidad. (Figura 13)

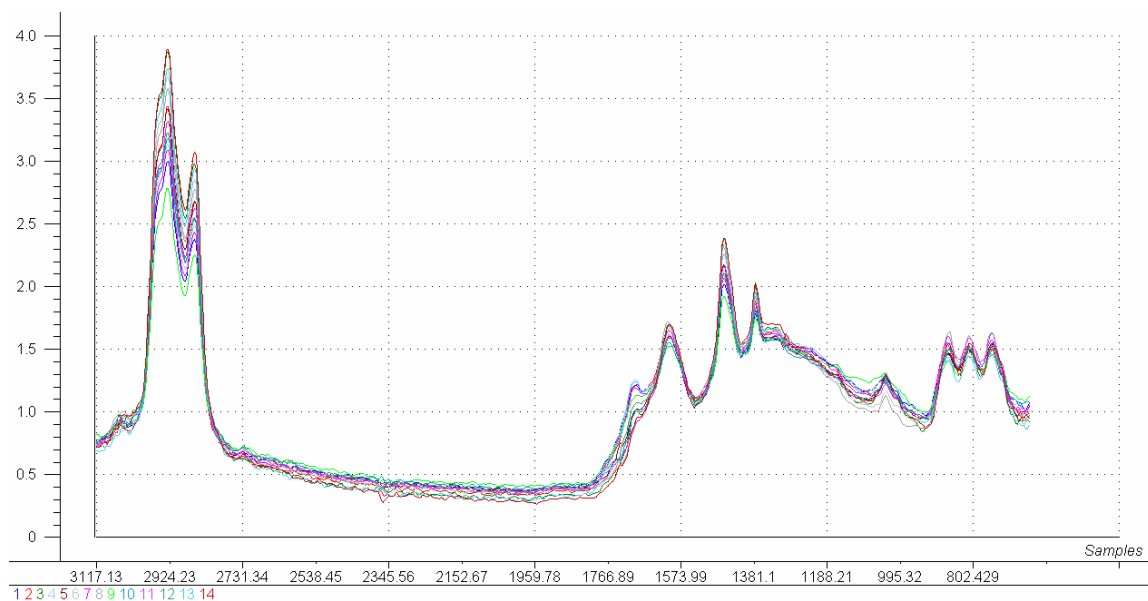
Figura 12. Espectro con longitudes de onda de interés



Como segunda medida se inició el pretratamiento de datos utilizando algunas herramientas disponibles en el software, haciendo uso solamente de aquellas que permitieran de alguna manera obtener la máxima y más útil información de los datos. En la figura 13 se pueden observar los espectros de las 14 muestras de Asfaltenos sin ningún tipo de pretratamiento y con las longitudes de onda de interés.

La herramienta utilizada como pretratamiento de datos fue la Normalización media (Figura 14) que permitió obtener todos los datos en aproximadamente la misma escala, con una distribución mucho más equitativa de las variaciones y los valores medios. Esta herramienta permitió obtener la mayor información de los datos que se evidencia más adelante en el análisis por componentes principales, de manera que no fue necesario realizar otro tipo de pretratamiento.

Figura 13. Espectros normalizados



Luego del análisis PCA para determinar la variabilidad que podían tener el conjunto de datos con las 14 muestras y se aplicó el PLS al conjunto de datos para obtener la predicción de la densidad en las muestras de Asfaltenos.

3. RESULTADOS Y ANÁLISIS

A continuación se presentan los principales resultados obtenidos en los análisis de componentes principales (PCA) y la regresión por mínimos cuadrados parciales (PLS), para los Asfaltenos analizados.

3.1 ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

Inicialmente se realizó una descomposición por componentes principales realizada sobre la información espectral de las 14 muestras de Asfaltenos con los pretratamientos realizados, para encontrar el número de componentes óptimo en las posteriores etapas de calibración y validación del modelo.

Los datos de varianza explicada son muy importantes para saber cuántos componentes principales se van a utilizar en el análisis. Debido a que no existe una regla definida sobre el número que se debe utilizar, se buscó decidir en función del número de variables iniciales (recordando que se trata de reducirlas en la medida de lo posible) y de la proporción de varianza explicada acumulada. (Tabla 4)

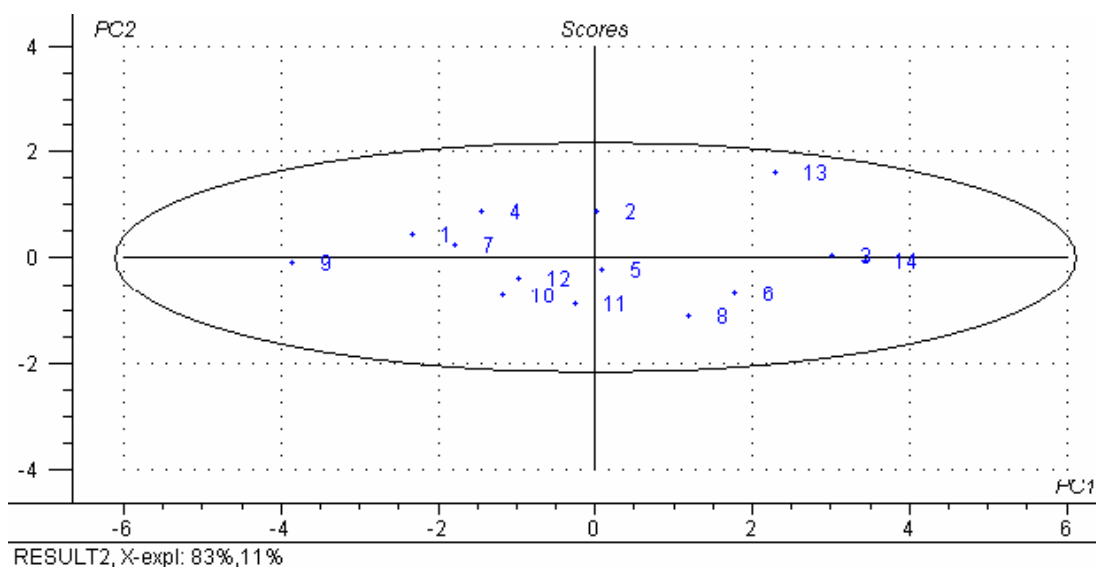
Tabla 4. Varianza de los componentes principales de las 14 muestras de Asfaltenos

| Componente Principal | Varianza Explicada % | Varianza Explicada Acumulada % |
|-----------------------------|-----------------------------|---------------------------------------|
| 1 | 83.27 | 83.27 |
| 2 | 10.91 | 94.18 |
| 3 | 2.26 | 96.44 |
| 4 | 1 | 97.44 |
| 5 | 0.67 | 98.11 |
| 6 | 0.51 | 98.62 |
| 7 | 0.39 | 99.01 |
| 8 | 0.27 | 99.28 |

Observando el porcentaje de varianza de los datos espectrales que explican los primeros 8 componentes principales se determinó que 5 es el número de componentes a utilizar en los modelos de predicción ya que a partir del PC6 el aumento de la varianza explicada acumulada es bajo.

Cada una de estas componentes contiene cierta información que permite la clasificación de las muestras a partir de su espectro PAS (Figura 15).

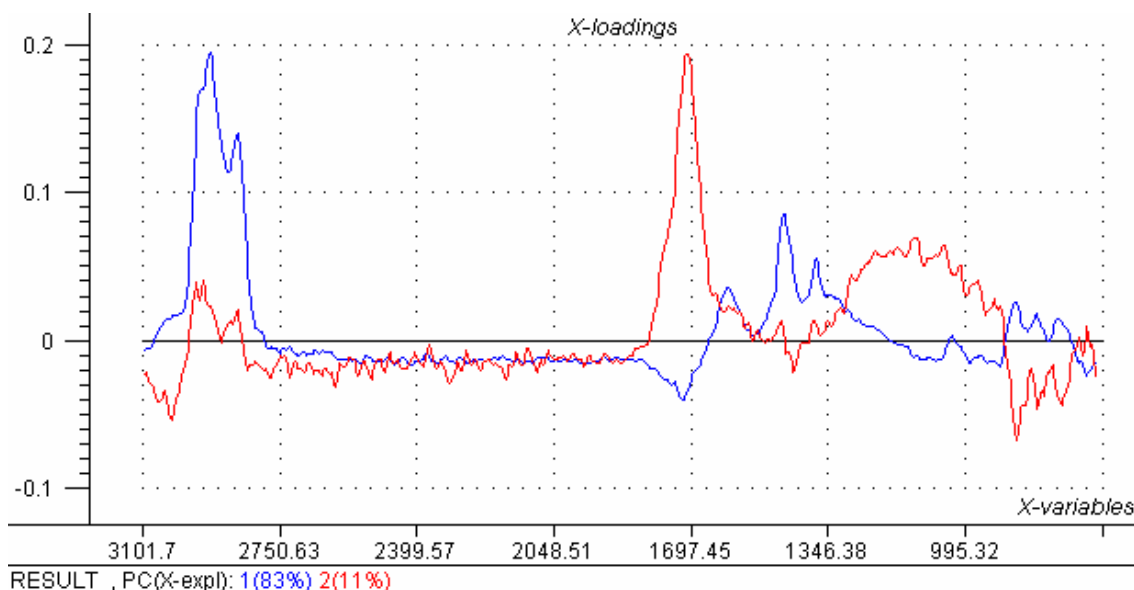
Figura 14. Descomposición por componentes principales de la información espectral. Gráfica de Score



Mediante la descomposición por componentes principales se identificó claramente que la primera y segunda componente permiten identificar y clasificar las muestras en un grupo, además de representar el 94.18% de la varianza total de las intensidades espectrales. Utilizando la aplicación de la elipse Hotelling T^2 , se buscaba encontrar outliers potenciales (puntos alejados que quedan fuera de la elipse) o muestras atípicas, observando que todas las muestras están dentro del promedio y que según este gráfico no se presentan este tipo de muestras.

La gráfica de *X-loadings* (figura 16) se utilizó para determinar los números de onda de mayor importancia los cuales se encuentran aproximadamente en los intervalos de $3100-2750\text{ cm}^{-1}$, $1700-648\text{ cm}^{-1}$ del espectro PAS. Las señales presentes en $2700-1750\text{ cm}^{-1}$ no se tomaron en cuenta en el análisis debido a que estas regiones pertenecen usualmente a vibraciones del agua presente en los crudos y de CO_2 presente.

Figura 15. Gráfico de X-Loadings de PC1 y PC2



En esta figura también se observó la varianza explicada por los dos primeros PC según la información de los espectros, donde el 83% de las variables contribuyen al primer componente, es decir, que este componente describe en su mayoría el perfil del espectro inicial, representado por las regiones correspondientes a vibraciones de alifáticos ($3100-2750\text{ cm}^{-1}$), a la combinación de alifáticos y aromáticos ($1700-900\text{ cm}^{-1}$) y en menor proporción a las vibraciones de grupos aromáticos ($900-680\text{ cm}^{-1}$). El otro 11% de las variables están representadas por el segundo componente haciéndose más evidente su contribución hacia la región de vibraciones de grupos aromáticos.

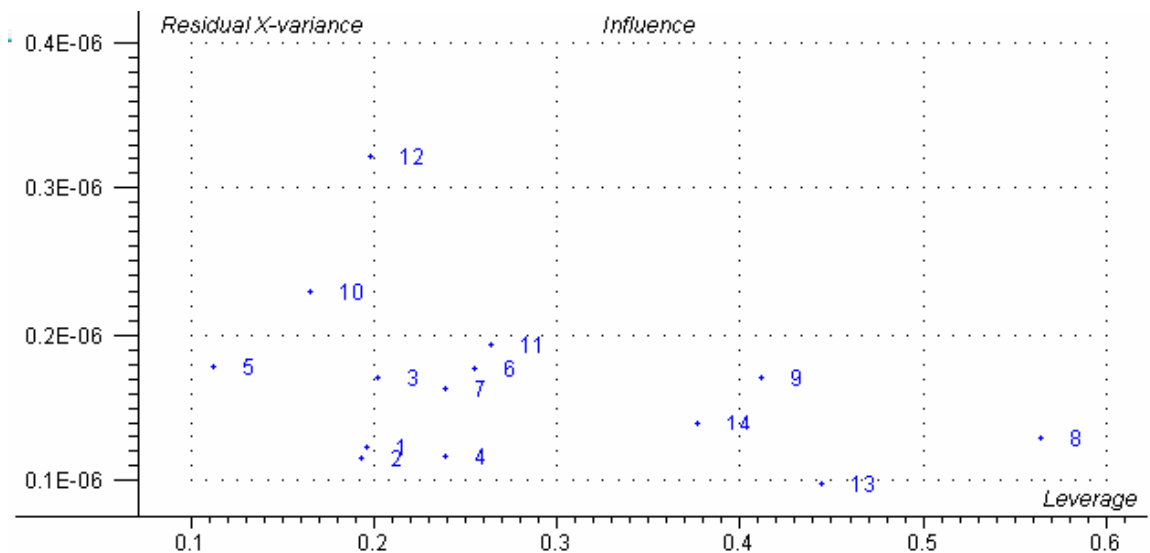
La gráfica de *Influence* (Figura 17) permitió identificar con mucha más certeza aquellos muestras atípicas, de alta influencia y posibles dangerous outliers (valores extremos peligrosos), que podrían de alguna manera distorsionar el modelo. Las muestras con una alta varianza residual, es decir, ubicadas en la parte superior de la gráfica, pueden ser posibles outliers. Las muestras con alto leverage (apalancamiento), ubicadas en la parte derecha de la gráfica, serán

muestras influyentes, lo que significa que tratarán de atraer el modelo para que las describa mejor.

Teniendo en cuenta la información que proporciona la gráfica no existen posibles dangerous outliers, pero la muestra 8 contiene un alto valor de leverage lo que implica que esta muestra atraerá el modelo para que sea descrito de una mejor forma.

Este alto valor de leverage se debe posiblemente a que esta muestra posee el valor de densidad más alto del conjunto de muestras, además ese mismo valor está presente en otra muestra (13), lo que conlleva a que se sitúe en esta parte de la gráfica de esta manera se hace necesario prescindir de esta muestra así observar un mejor comportamiento del modelo.

Figura 16. Gráfica de Influencia



3.2 ANALISIS DE REGRESIÓN POR MÍNIMOS CUADRADOS PARCIALES (PLS-REGRESION)

Los resultados obtenidos en el análisis de regresión por mínimos cuadrados parciales, para la predicción de la densidad de asfaltenos, se muestran a continuación.

3.2.1 Modelo PLS para la predicción de la densidad en Asfaltenos.

Mediante el algoritmo de regresión lineal PLS, se evaluó el potencial de la PAS para predecir la densidad en las muestras de Asfaltenos.

El desarrollo de los modelos de predicción abarcó dos grandes etapas: calibración y validación. A continuación se describe en detalle el procedimiento empleado y los resultados encontrados para la predicción de esta propiedad.

3.2.1.1 Calibración del modelo PLS: Teniendo en cuenta el número óptimo de componentes principales y las regiones espectrales más influyentes, fueron desarrollados cuatro modelos de predicción variando el número de muestras de calibración y las regiones espectrales en función del análisis PCA descrito en la sección 4; además de las muestras influyentes que podrían de alguna manera mejorar o distorsionar el modelo . La selección del modelo de predicción se realizó evaluando parámetros estadísticos como el RMSEP, RMSEC y R^2 . (Tabla 5)

Tabla 5. Parámetros estadísticos de los modelos de predicción generados para la predicción de la densidad en Asfaltenos

| Modelo | Intervalos de números de onda (cm⁻¹) usados | Muestra excluida | R² | Varianza Explicada | RMSEC (%) | RMSEP (%) |
|---------------|---|-------------------------|----------------------|---------------------------|------------------|------------------|
| 1 | 3101,7-682,836 | - | 0.40 | 83 | 0.02 | 0.021 |
| 2 | 1824,75-663.54 | 8 | 0.88 | 85 | 0.035 | 0.0172 |
| 3 | 933,59-663,54 | 8 | 0.93 | 60 | 0.006 | 0.0161 |
| 4 | 3101,7-2750,63 | 8 | 0.99 | 98 | 0.002 | 0.0145 |

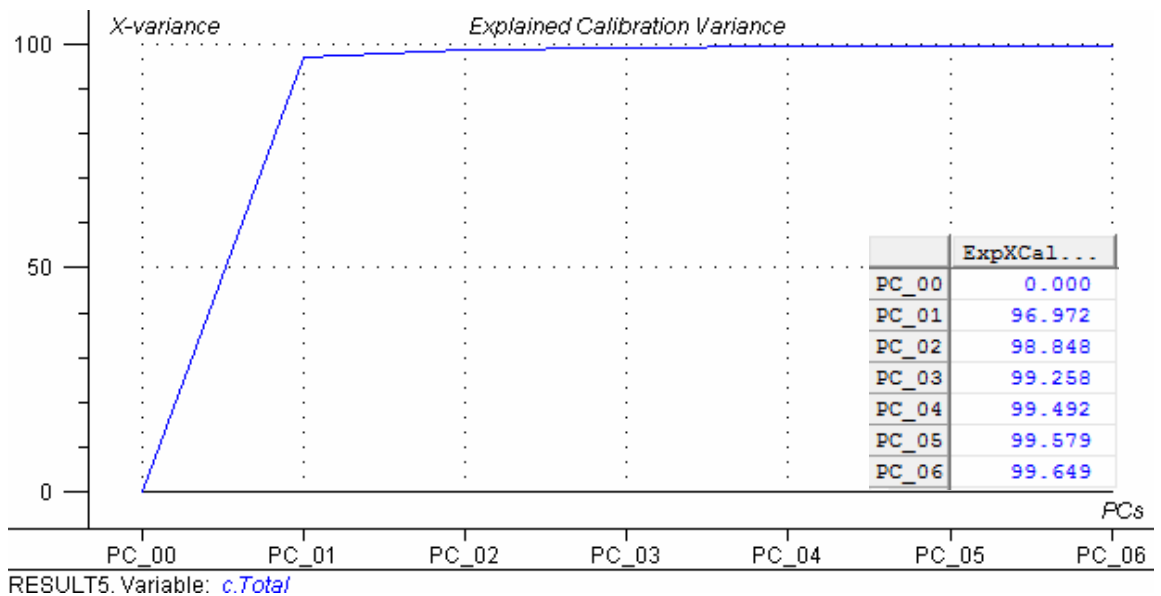
Por presentar el menor % de error tanto de predicción (RMSEP) como de calibración (RMSEC), además de la mejor correlación y la mayor varianza explicada por los componentes (98%), el modelo de predicción seleccionado fue el número 4.

La muestra 8 fue excluida del modelo de predicción seleccionado, por diferentes razones como presentar un valor extremo de densidad, entre otras descritas anteriormente en el análisis por componentes principales (PCA).

El modelo 4 con cinco componentes principales explica, a partir de la señal PAS en la región de 3101,7-2750,63 cm⁻¹, más del 99% en la variabilidad de los datos. (Figura 18). Esta contribución puede deberse posiblemente a que en esta región ocurren las vibraciones asimétricas y simétricas de enlaces C-H aromáticos, además de contribuciones por parte de enlaces C-H alifáticos, además de las mayores interacciones que se puedan presentar entre estos grupos en esta región

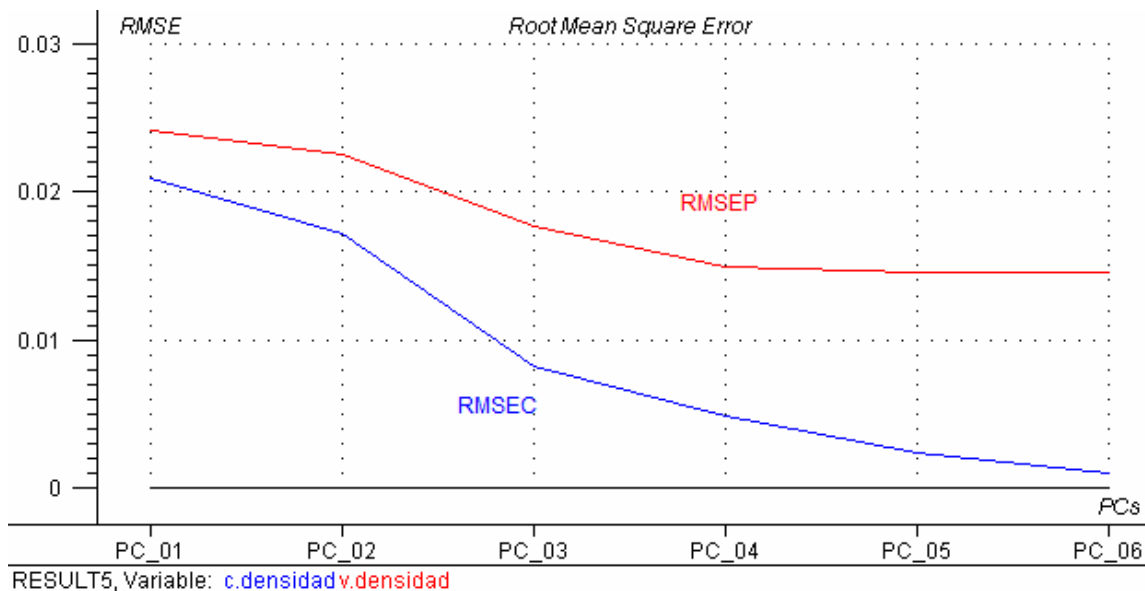
debido a la estructura general de los asfaltenos , aportando así la mayor información de los datos.

Figura 17. Varianza explicada en el modelo PLS de Asfaltenos



La selección del número óptimo de componentes requirió evaluar el comportamiento del error estándar de calibración (RMSEC) y el error estándar de predicción (RMSEP). El primero de estos parámetros es una medida de la desviación estándar de los residuales obtenidos por la diferencia entre los valores observados y predichos por el modelo para las muestras de calibración. El segundo parámetro se basa en un algoritmo iterativo que selecciona muestras dentro del grupo de calibración para desarrollar el modelo de predicción y posteriormente lo evalúa sobre las muestras restantes. (Figura 19)

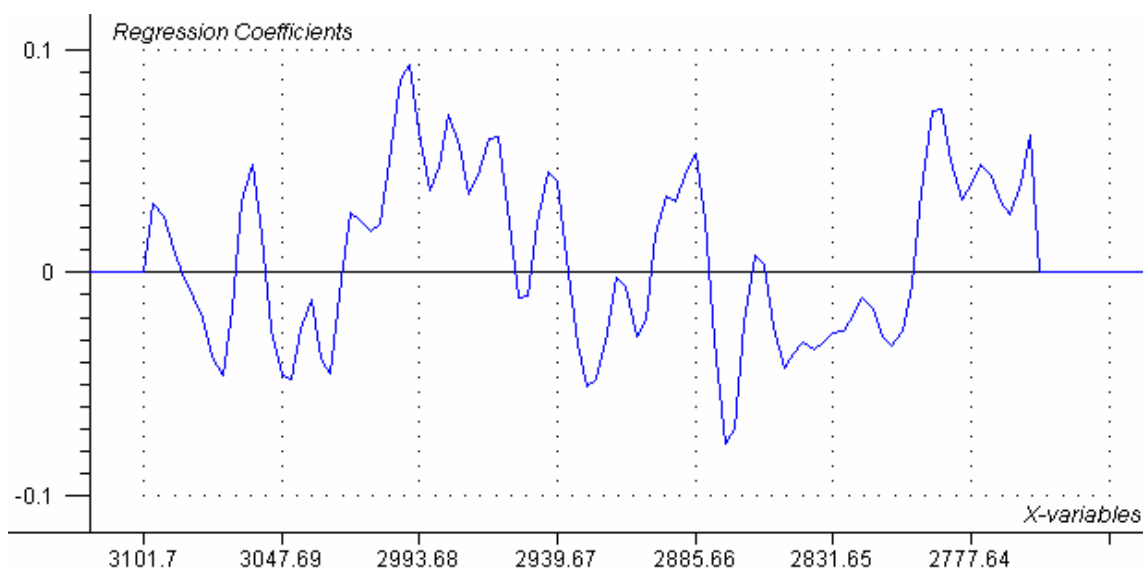
Figura 18. Error calculado en función del número de componentes para el modelo PLS de Asfaltenos



El error estándar en la etapa de calibración y validación disminuye apreciablemente hasta el quinto componente, donde alcanza valores de 0.002% y 0.01% respectivamente.

Debido a que cada componente principal es una combinación lineal de las absorbancias medidas en las diferentes frecuencias del rango espectral, multiplicadas por un coeficiente de regresión que determina el peso o influencia de tales frecuencias sobre el componente; la regresión PLS permite visualizar los coeficientes hallados para cada componente utilizado. La figura 20 muestra los coeficientes de regresión hallados para la primera componente principal ya que este explica la mayor variabilidad en los datos.

Figura 19. Coeficiente de regresión para el primer componente principal PLS de Asfaltenos



de 3101,7-2750,63 cm^{-1} presentaba un efecto positivo en el modelo.

3.2.1.2 Validación Cruzada del modelo PLS: La validación del modelo PLS para la predicción de la densidad en asfaltenos se realizó empleando el grupo de muestras de crudo utilizadas para generar el modelo de calibración. Sobre los datos completos de los espectros PAS pre-tratados y con las regiones seleccionadas, se aplicó el modelo seleccionado (modelo 4).

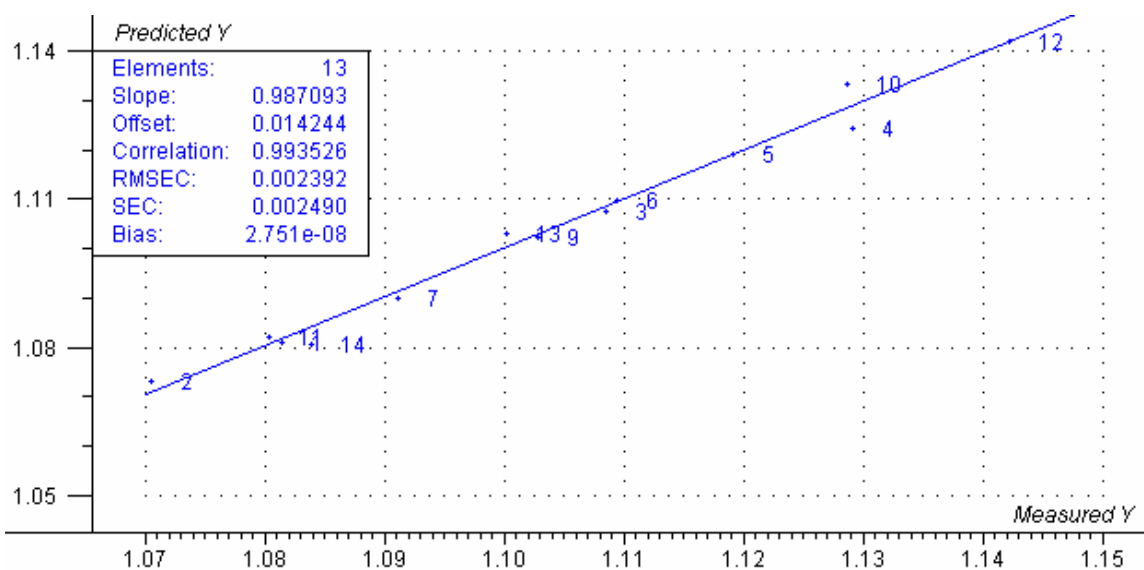
Los valores predichos por el modelo propuesto y los valores de referencia están muy cercanos entre sí, lo que demuestra el buen desempeño del modelo propuesto. Sólo en una de las 14 muestras empleadas (muestra 8) para la etapa de validación, el porcentaje de error calculado entre los valores predichos y de referencia fue superior a la unidad. (Tabla 6).

Tabla 6. Validación cruzada del modelo de predicción de la densidad de las 14 muestras de asfaltenos

| Número de muestra | REFERENCIA (g/cm³) | PREDICHO (g/cm³) | RESIDUAL | % ERROR |
|--------------------------|--------------------------------------|------------------------------------|-----------------|----------------|
| 1 | 1.0813 | 1.081 | -0.0087 | 0.027 |
| 2 | 1.0705 | 1.073 | -0.0025 | 0.23 |
| 3 | 1.1085 | 1.107 | 0.0015 | 0.13 |
| 4 | 1.1291 | 1.124 | 0.0051 | 0.45 |
| 5 | 1.1192 | 1.119 | 0.0002 | 0.017 |
| 6 | 1.1094 | 1.110 | -0.0006 | 0.054 |
| 7 | 1.0912 | 1.090 | 0.0012 | 0.10 |
| 8 | 1.1422 | 1.080 | 0.0622 | 5.4 |
| 9 | 1.1029 | 1.102 | 0.0001 | 0.081 |
| 10 | 1.1287 | 1.133 | -0.0043 | 0.38 |
| 11 | 1.0803 | 1.082 | -0.0017 | 0.15 |
| 12 | 1.1422 | 1.142 | 0.0002 | 0.017 |
| 13 | 1.1002 | 1.103 | -0.0028 | 0.25 |
| 14 | 1.0839 | 1.080 | 0.0039 | 0.36 |

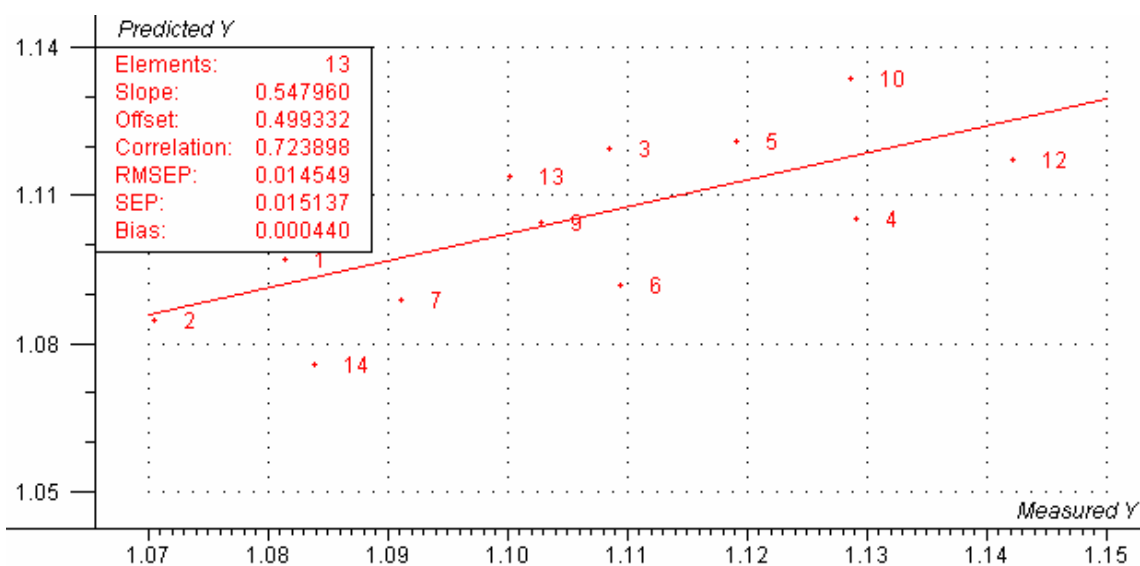
El coeficiente de correlación entre los valores de referencia y de predicción hallado fue superior a 0.99 para la curva de calibración y superior a 0.72 para la curva de validación. (Figura 21)

Figura 20. Curvas de calibración del modelo de predicción de la densidad en



RESULT5, (Y-var, PC): (densidad,5)

asfaltenos



RESULT5, (Y-var, PC): (densidad,5)

El error estándar de predicción calculado para el conjunto de muestras de validación fue de 0.01.

Los resultados hallados en esta etapa dan como satisfactorio el desempeño del modelo para la predicción de la densidad en asfaltenos frente a las muestras empleadas en la etapa de calibración.

4. CONCLUSIONES

El desarrollo del presente trabajo demostró que el empleo de métodos quimiométricos para el tratamiento de la señal de PAS permite el desarrollo de un método alternativo de determinación de la densidad de la fracción de Asfaltenos de crudos colombianos.

La aplicación de técnicas quimiométricas y en concreto el empleo del análisis multivariante, permitió extraer información relevante de los datos iniciales. Así, mediante el Análisis por Componentes Principales (PCA) se pudo reducir el número de variables iniciales a cinco de forma que estos 5 PCs explican por sí solos casi el 100 % de la varianza de los datos. Este análisis además permitió la determinación de las regiones del espectro PAS de mayor importancia localizadas en las regiones comprendidas de $3100\text{-}2750\text{ cm}^{-1}$, $1700\text{-}648\text{ cm}^{-1}$.

Se implementó una técnica alternativa de análisis de fracciones de petróleo por Espectroscopía Fotoacústica, que permite hacer un estudio de muestras sólidas, opacas, a un menor tiempo que las técnicas tradicionales.

El método de regresión PLS empleado sobre el espectro PAS mostró un desempeño satisfactorio para predecir la densidad en la fracción de asfaltenos. El método propuesto de caracterización por espectroscopia PAS proporciona información de la composición química en menos de cinco minutos y con mínima cantidad y preparación de muestra.

Como tendencia general, se observó que la región espectral comprendida de $3101,7\text{-}2750,63\text{ cm}^{-1}$ presentó el mejor desempeño para la generación de modelos de predicción de la densidad. La validación cruzada realizada para el modelo

desarrollado en esta región, mostro buenos resultados como un método alternativo al estándar.

Los modelos desarrollados deben ser únicamente aplicados a muestras de la misma naturaleza. Adicionalmente, las señales espectroscópicas deben ser adquiridas en el espectrómetro empleado bajo las mismas condiciones y parámetros instrumentales. La aplicación de los modelos predictivos a muestras de naturaleza diferente puede llevar a resultados erróneos.

5. RECOMENDACIONES

Conservando de manera adecuada las muestras empleadas en este estudio y aprovechando que se cuenta con su información espectral, es posible desarrollar modelos de predicción de nuevas propiedades de interés (gravedad API, relación C/H, entre otras).

Los resultados del estudio anteriormente presentados muestran el gran potencial de la espectroscopia PAS para la caracterización de asfaltenos de crudos colombianos y muestras opacas. Sin embargo se recomienda que los resultados de calibración y predicción sean continuamente evaluados aunque muestran un buen desempeño en los modelos. Además se recomienda que los modelos sean alimentados con nuevas muestras, para asegurar que estos modelos siempre estarán actualizados.

REFERENCIAS BIBLIOGRÁFICA

- (1) GARCÍA, G. Roberto C. Espectroscopía Fotoacústica: Una teoría antigua con nuevas aplicaciones. Instituto Tecnológico de Tuxtla Gutierrez. Chiapas. 2002.
- (2) SPEIGHT, J. G. *The chemistry and technology of petroleum*; Marcel Dekker: New York, 1999.
- (3) CURTIS, Carl. KOPPER, Robert. y colaboradores. Yacimientos de petróleo pesado. En: *Oilfield Review*, 2003.
- (4) CONAWAY, C. *The petroleum industry: A Nontechnical al guide*. Tulsa: Pennwell Publishing Co. 1999.
- (5) WAUQUIER, J.P. *El Refino del Petróleo*. Editorial Díaz Santos, S.A. Primera edición. México. 2004. Pág, 265-273.
- (6) ASKE, N., KALLEVIK, H., SJÖBLOM, J. Determination of Saturate, Aromatic, Resin and Asphaltenic (SARA) Components in Crude Oils by means of Infrared and Near Infrared Spectroscopy. *Energy & Fuels*, 15(5), 1304-1312, 2001.
- (7) AMERICAN SOCIETY FOR TESTING AND MATERIALS. Standard Test Method for Separation of Asphalt into four fractions. ASTM D4124-09
- (8) BINGER J. W., "*Chemistry of Asphaltenes*," American Chemical Society, pp. 17 – 27, Washington DC, 1984.
- (9) PINEDA-FLORES, Gabriel y MESTA-HOWARD, Ana. Petroleum asphaltenes: generted problematic and posible biodegradation mechanisms. En: *Revista Latinoamericana de Microbiología*. Julio-Septiembre, 2001, vol. 43 N° 3, p. 143-150.
- (10) DELGADO, J.G. *Asfaltenos: Composición, agregación, precipitación*. Versión # 1. Merida-Venezuela.: Universidad de los Andes, 2006. 1-5p
- (11) KLIGER, D. S. *Ultrasensitive laser spectroscopy, Quantum electronics, Principles and applications*, Academic Press, New York 1988.

- (12) ORREGO, J. A. Tesis de Maestría. Estudio de la estructura de cinco carbones Colombianos por Espectroscopía Fotoacústica en el infrarrojo medio. Universidad Industrial de Santander. Bucaramanga. 2008
- (13) McCLELLAND, J.F. BAJIC, S.J. JONES, R.W. and SEAVERTSON, L.M. 'Photoacoustic Spectroscopy' in "Modern Techniques in Applied Molecular Spectroscopy", F.M. Mirabella (Ed.) Wiley, New York, Chapter 6, pp. 221-265 (1998).
- (14) ROSENCWAIG, A. and GERSHO A. Theory of the Photoacoustic effect with solids *J. Appl. Phy.* **1976**, 47(1), 64.
- (15) McCLELLAND, J.F. JONES, R.W. LUO, S. 'A Practical Guide to FT-IR Photoacoustic Spectroscopy' in "Practical Sampling Techniques for Infrared Analysis", P.B. Coleman (Ed.), CRC Press, Boca Raton, FL, Chapter 5, pp. 107-144 (1993).
- (16) RUBINSON, K. A; RUBINSON, J. F. Análisis instrumental, Prentice Hall Hispanoamericana S.A. Madrid.: Editorial Pearson Educación, 2001.
- (17) GELADI, P. Some recent trends in the calibration literature. En: Chemometrics and Intelligent Laboratory Systems. Vol. 60 (2002). p. 211 – 224.
- (18) MILLER, J. N. y MILLER, J. C. Estadística y quimiometría para química analítica. Cuarta edición. Editorial Prentice Hall. España. 2002. 296 p.
- (19) GEMPERLINE, P. Practical guide to chemometrics. Segunda edición. Editorial Taylor y Francis Grupo. 2006. 541 p.
- (20) FERNÁNDEZ, C. M. Quimiometría. Editorial Publicaciones Universitat de València. 2005. 423 p.
- (21) SAVITZKY, A. and GOLAY, M.J.E. (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* 36, 1627-1639.
- (22) NORRIS, K.H. and WILLIAMS, P.C. (1984) Optimization of mathematical treatments of raw near-infrared signal in the measurement of protein in hard red spring wheat. I. Influence of particle size. *Cereal Chemistry*, 61(2), 158-65.

- (23) MACHO, A. S. Metodologías analíticas basadas en espectroscopía de infrarrojo y calibración multivariante. Aplicación a la industria petroquímica. Trabajo de grado Doctor en Química. Tarragona-España. Universitat Rovira I Virgili. Área de química analítica. Departamento de Química analítica y química orgánica. 2002. 19-42 p.
- (24) PORCEL, G. M. Aplicación de técnicas quimiométricas para el desarrollo de nuevos métodos cinético-espectrofotométricos de análisis. Trabajo de grado Doctor en Ciencias Químicas. Bellaterra-España. Universitat Autònoma de Barcelona. Facultad de Ciencias. 2001. 5-26p.
- (25) MARTENS, H. & NAES, T. Multivariate calibration. New York: Jhon Wiley & Sons, 1989. 438 p.
- (26) SKOOG, D., A., HOLLER, F., J. y NIEMAN, T., A. Principios de análisis instrumental. 5ta Edición. Mc Graw Hill, Madrid, pp. 103-118, **2001**.
- (27) THE UNSCRAMBLER 9.7. METHODS. Software para diseño de experimentos y análisis multivariado. (En línea). <http://www.camo.com> [citado el 13 de abril de 2010].