

MODELO PARA LA PREDICCIÓN DE LA COMPOSICIÓN S.A.R.A DE
CRUDOS COLOMBIANOS A PARTIR DE ESPECTROSCOPIA
INFRARROJA (FTIR) EN COMBINACIÓN CON MÉTODOS
MULTIVARIADOS

OSCAR MAURICIO VÁSQUEZ ARDILA

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE QUÍMICA
BUCARAMANGA
2015

MODELO PARA LA PREDICCIÓN DE LA COMPOSICIÓN S.A.R.A DE
CRUDOS COLOMBIANOS A PARTIR DE ESPECTROSCOPIA
INFRARROJA (FTIR) EN COMBINACIÓN CON MÉTODOS
MULTIVARIADOS

Plan de trabajo de grado presentado como requisito parcial para optar
al título de químico

AUTOR

OSCAR MAURICIO VASQUEZ ARDILA

DIRECTOR

ENRIQUE MEJÍA OSPINO, QUÍMICO, PhD

CODIRECTOR

RAFAEL CABANZO, FÍSICO MAGISTER

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE QUÍMICA
BUCARAMANGA
2015

*A Dios por permitir cambiar mi vida y darme valor para
afrentar todos los problemas que se me presentaron durante mi
carrera universitaria.*

*A mi madre Bertha Vásquez Ardila por su amor
incondicional por su apoyo y compañía gracias.*

*A mi hermana Gina Meylan Suescun por su amor
incondicional y por siempre estar a mi lado gracias.*

*A toda mi familia especialmente a mi tía Elsa Mendieta
por su amor y apoyo, a mi abuela Cayetana y a mi nona
María por su amor y apoyo gracias.*

A mi amiga Mercy Angarita por su apoyo gracias

AGRADECIMIENTOS

A la Universidad Industrial de Santander por la oportunidad que me ofreció para mi formación como profesional.

Quiero darle las gracias a mi Director de Tesis PhD. Enrique Mejía Ospino por la oportunidad que me brindo al permitirme trabajar en el Laboratorio de Espectroscopia Atómica y Molecular (LEAM), por su enorme paciencia, conocimientos y en especial por su gran confianza para la realización de este proyecto.

Al Laboratorio de Espectroscopia Atómica y Molecular (LEAM) y a mi Codirector PhD. (c) Rafael Cabanzo por permitirme ser parte de este grupo de investigación.

Quiero darles las gracias a todos aquellos que se alegran y comparten este triunfo conmigo, y a las personas que de una u otra forma hicieron posible la culminación de esta investigación.

Contenido

| | PÁG |
|---|-----|
| INTRODUCCIÓN | 15 |
| 1. CONSIDERACIONES TEORICAS | 17 |
| 1.1 PETRÓLEO..... | 17 |
| 1.1.1.Generalidades..... | 17 |
| 1.1.2. Clasificación del petróleo..... | 17 |
| 1.1.3. Composición química del petróleo..... | 18 |
| 1.1.3.1. Saturados..... | 19 |
| 1.1.3.2. Aromáticos..... | 19 |
| 1.1.3.3. Resinas..... | 20 |
| 1.1.3.4. Asfáltenos..... | 20 |
| 1.1.4. Fraccionamiento SARA..... | 21 |
| 1.2. ESPECTROSCOPIA DE INFRARROJO..... | 22 |
| 1.2.1. Aspectos fundamentales..... | 22 |
| 1.2.2. Regiones espectrales..... | 24 |
| 1.2.3. Tipo de medidas en el infrarrojo..... | 25 |
| 1.2.3.1. Reflectancia total atenuada (ATR) | 26 |
| 1.2.4. Interpretación del espectro..... | 28 |
| 1.3. QUIMIOMETRÍA..... | 29 |
| 1.3.1. Clasificación multivariable..... | 29 |
| 1.3.1.1. Clasificación de los métodos de calibración multivariable..... | 29 |
| 1.3.1.2. Construcción de modelo de calibración multivariable..... | 30 |
| 1.3.2. Técnica de pretratamiento de datos..... | 31 |
| 1.3.2.1. Suavizado espectral..... | 31 |
| 1.3.2.2. Normalización..... | 31 |
| 1.3.2.3. Corrección de la línea base..... | 32 |
| 1.3.2.4. Centralización..... | 32 |
| 1.3.2.5. Derivación..... | 32 |
| 1.3.3. Métodos basados en reducción de variables..... | 32 |
| 1.3.3.1. Análisis por componentes principales (PCA) | 33 |
| 1.3.3.2. Regresión por mínimos cuadrados parciales (PLS) | 35 |
| 1.3.4. Factores a incluir en el modelo..... | 37 |
| 1.3.4.1. Detección de outliers..... | 37 |
| 1.3.4.2. Leverage..... | 37 |

| | |
|---|-----------|
| 1.3.4.3. Estadístico T^2 de Hotelling..... | 38 |
| 1.3.4.4. Errores de calibración y predicción..... | 38 |
| 1.3.4.5. Residuales en la respuesta instrumental..... | 39 |
| 1.3.4.6. Variación del modelo..... | 40 |
| 1.3.4.7. Predicción de muestras desconocidas..... | 41 |
| 1.4. APLICACIÓN DE LA ESPECTROSCOPIA FTIR-ATR EN EL PETRÓLEO.. | 41 |
| 2. PARTE EXPERIMENTAL..... | 42 |
| 2.1 MUESTRA..... | 42 |
| 2.2 ESPECTROSCOPIA MIR-ATR..... | 43 |
| 2.2.1. Instrumentación..... | 43 |
| 2.2.2. Verificación del desempeño del espectro..... | 43 |
| 2.2.3. Adquisición del espectro MIR..... | 44 |
| 2.2.3.1. Características de los espectros obtenidos..... | 45 |
| 2.3. Procesamiento de datos..... | 45 |
| 3. RESULTADOS Y ANÁLISIS..... | 48 |
| 3.1. ANÁLISIS POR COMPONENTES PRINCIPALES (PCA) | 48 |
| 3.1.1. Análisis por componentes principales de muestras livianas..... | 48 |
| 3.2. Desarrollo del modelo del PLS..... | 52 |
| 4. RESULTADOS DE LOS MODELOS DE PREDICCIÓN..... | 53 |
| 4.1. Modelos PLS para la predicción de saturados..... | 53 |
| 4.2. Modelos PLS para la predicción de aromáticos..... | 57 |
| 4.3. Modelos PLS para la predicción de resinas..... | 61 |
| 4.4. Modelos PLS para la predicción de asfáltenos..... | 65 |
| 5. CONCLUSIONES..... | 70 |
| CITAS BIBLIOGRÁFICAS..... | 71 |
| BIBLIOGRAFÍA..... | 76 |

LISTA DE FIGURAS

| | PÁG |
|--|------------|
| Figura 1. Esquema simplificado para la determinación de las fracciones SARA..... | 21 |
| Figura 1.a Procedimiento estándar para el análisis SARA según las normas ASTM..... | 22 |
| Figura 2. Perfil de energía potencial según el modelo de oscilador (a) armónico (b) anarmónico..... | 23 |
| Figura 3. Fenómeno de absorción, transmisión y reflexión de la radiación electromagnética al interactuar con la materia..... | 25 |
| Figura 4. Reflexión total interna y elemental de reflexión interna (IRE) utilizando en el sistema ATR..... | 26 |
| Figura 5. Aclaración del método cuando solo hay dos variables y, por tanto, solo dos componentes principales..... | 33 |
| Figura 6. Notación matricial para la descomposición por componentes principales..... | 34 |
| Figura 7. Notación matricial extendida para la descomposición por componentes principales..... | 35 |
| Figura 8. Descripción gráfica del método de regresión PLS..... | 36 |
| Figura 9. Ejemplo del cálculo del residual de un espectro MIR..... | 39 |
| Figura 10. Gráfico del residual frente al leverage..... | 40 |
| Figura 11. Sistemas de caracterización MIR..... | 43 |
| Figura 12. Prueba de repetibilidad del espectrómetro MIR..... | 44 |
| Figura 13. Espectro original de 20 muestras de crudo..... | 45 |
| Figura 14. Espectro normalizado y derivado..... | 47 |
| Figura 15. Varianza explicada en el modelo PLS de saturados..... | 49 |
| Figura 16. Gráfica de Scores en dos dimensiones para las 20 muestras..... | 49 |

| | |
|---|----|
| Figura 17. Estadístico T^2 aplicado a la gráfica de Score de las 20 muestras..... | 50 |
| Figura 18. Gráfica de influencia para las 20 muestras..... | 51 |
| Figura 19. Gráfico de X-loadings de PC1 y PC2..... | 52 |
| Figura 20. Gráfica de los errores RMSEC y RMSEP calculados para el modelo PLS saturados..... | 54 |
| Figura 21. Coeficiente de regresión para el primer componente principal del modelo PLS saturados..... | 55 |
| Figura 22. Curva de calibración y validación del modelo de predicción de %W de saturados..... | 56 |
| Figura 23. Curva de calibración del modelo de predicción de %W de aromáticos..... | 58 |
| Figura 24. Curva de validación del modelo de predicción de %W de aromáticos..... | 58 |
| Figura 25. Gráfica de los errores RMSEC y RMSEP para el modelo PLS de aromáticos..... | 59 |
| Figura 26. Coeficiente de regresión para el primer componente para el modelo PLS de aromáticos..... | 59 |
| Figura 27. Curva de calibración y validación del modelo de predicción de %W de aromáticos..... | 60 |
| Figura 28. Curva de calibración del modelo de predicción de %W de resinas..... | 62 |
| Figura 29. Curva de validación del modelo de predicción de %W de resinas..... | 62 |
| Figura 30. Gráfica de los errores RMSEC y RMSEP para el modelo PLS de resinas..... | 63 |
| Figura 31. Coeficiente de regresión para el primer componente del modelo PLS de resinas..... | 63 |
| Figura 32. Curva de calibración y validación del modelo de predicción de %W de resinas..... | 64 |

| | |
|---|----|
| Figura 33. Curva de calibración del modelo de predicción de %W de asfaltenos..... | 66 |
| Figura 34. Curva de validación del modelo de predicción de %W de asfaltenos..... | 66 |
| Figura 35. Gráfica de los errores RMSEC y RMSEP para el modelo PLS de asfaltenos..... | 67 |
| Figura 36. Coeficiente de regresión para el primer componente del modelo PLS de asfaltenos..... | 67 |
| Figura 37. Curva de calibración y validación del modelo de predicción de %W de asfaltenos..... | 68 |

LISTA DE TABLAS

| | PÁG |
|---|------------|
| Tabla 1. Especificaciones generales de clasificaciones de crudos..... | 18 |
| Tabla 2. Regiones de absorción en el infrarrojo..... | 25 |
| Tabla 3. Frecuencia de absorción de algunos grupos funcionales en la región MIR..... | 28 |
| Tabla 4. Análisis S.A.R.A de los crudos..... | 42 |
| Tabla 5. Cálculo de los RSD para tres números de onda característicos aplicando diferentes procedimientos..... | 46 |
| Tabla 6. Varianza explicada por los componentes principales de los 20 muestras de crudos..... | 48 |
| Tabla 7. Parámetros estadísticos del modelo seleccionado para la predicción de %W de saturados..... | 53 |
| Tabla 8. Prueba de repetibilidad del modelo de predicción de %W de saturados..... | 56 |
| Tabla 9. Parámetros estadísticos del modelo generado para la predicción de %W de aromáticos en muestras livianas..... | 57 |
| Tabla 10. Prueba de repetibilidad del modelo de predicción de %W de aromáticos..... | 60 |
| Tabla 11. Parámetros estadísticos del modelo generado para la predicción de %W de resinas en muestras livianas..... | 61 |
| Tabla 12. Prueba de repetibilidad del modelo de predicción de %W de resinas en muestras livianas..... | 64 |
| Tabla 13. Parámetros estadísticos del modelo parar la predicción de %W de asfáltenos en crudos livianos..... | 65 |
| Tablas 14. Prueba de repetibilidad del modelo de predicción de %W de asfáltenos en muestras livianas..... | 68 |

RESUMEN

TITULO

MODELO PARA LA PREDICIÓN DE LA COMPOSICIÓN S.A.R.A DE GRUDO COLOMBIANOS APARTIR DE ESPECTROSCOPIA INFRARROJA (FTIR) EN COMBINACIÓN CON MÉTODOS MULTIVARIADOS

AUTOR

OSCAR MAURICIO VASQUEZ ARDILA

PALABRAS CLAVES

ESPECTROSCOPIA MIR-ATH, CRUDO, PCA, PLS.

DESCRIPCIÓN

En el presente trabajo se desarrollan cuatro modelos matemáticos – estadísticos que permiten determinar la composición SARA de crudos colombianos mediante la espectroscopía en la región del infrarrojo medio (MIR). Las muestras para este estudio fueron proporcionadas por el instituto colombiano de petróleo. Todas las muestras fueron caracterizadas por espectroscopía MIR-ATR, y su señal fue correlacionada, mediante análisis de componentes principal (PCA) y regresión por mínimos cuadrados parciales (PLS).

Los modelos de predicción fueron desarrollados empleando el rango espectral de 4000- 400 cm^{-1} la validación mostró resultados satisfactorios para la predicción de análisis SARA de crudos. Para cada fracción, se obtuvieron errores estándares de predicción (RMSEP) para muestras livianas de 0.30, 1.70, 0.68 y 1.53. En todos los casos el coeficiente de correlación (R^2) entre los valores de referencia y predichos por el modelo fue superior a 0.98.

La caracterización por espectroscopía MIR-ATR requiere menos de cinco minutos y un menor costo de análisis en comparación con la cromatografía de exclusión, la cual requiere uso de solvente y mayora tiempo de análisis.

* Trabajo de Grado

** Facultad de ciencias. Escuela de química. Director: Enrique Mejía Ospino. Codirector: Rafael Cabanzo.

SUMMARY

TITLE:

MODEL FOR PREDICTING THE S.A.R.A COMPOSITION COLOMBIAN CRUDE FROM INFRARED SPECTROSCOPY (FTIR) IN COMBINATION WITH MULTIVARIATE METHODS.

AUTHORS:

OSCAR MAURICIO VASQUEZ ARDILA

KEYWORDS:

SPECTROSCOPY MIR-ATR, CRUDE, PCA, PLS

DESCRIPTION

In this work was developed eight chemometrics models using Fourier Transform Spectroscopy coupled to attenuate total reflectance (FTIR-ATR). The samples for this study were provided by the Colombian Petroleum Institute (ICP). The spectra of the samples were correlated by similarity using analysis of main components (PCA).

We use partial least squares regression (PLS) to obtaining prediction chemometrics models employing the spectral range of 4000- 400 cm^{-1} . The validation showed satisfactory results for the prediction of the SARA analysis of crude. For each fraction, were obtained standard errors of prediction (RMSEP) for light samples of 0.30, 1.70, 0.68 y 1.53. In all cases the coefficient of correlation (R^2) between the values of reference and predicted by the models was superior to 0.98.

The characterization methodology proposed by the MIR-ATR spectroscopy requires less than five minutes and a lower cost analysis compared to analytical exclusion chromatography, which requires use of solvents and more analysis time.

* Trabajo de Grado

** Facultad de ciencias. Escuela de química. Director: Enrique Mejía Ospino. Codirector: Rafael Cabanzo.

INTRODUCCIÓN

Desde principios del siglo XX dos nuevos recursos geológicos vinieron a sustituir al carbón como principal fuente de energía: el petróleo y el gas natural. Además el petróleo proporcionó compuestos hidrocarburos que sirven para sintetizar un sin fin de compuestos químicos nuevos, los petroquímicos, que constituyen materiales esenciales en la industria y en la vida actual.

Estos recursos geológicos comprenden el grupo de recursos no renovables de la tierra, puesto que los procesos geológicos para su formación requieren millones de años.

La palabra “petróleo” proviene de las voces latinas *petra* y *óleum*, que significan piedra y aceite, no porque sea aceite de piedra sino por estar aprisionado entre piedras. Este término general abarca la gama de productos comprendidos por un lado entre el petróleo bruto y el gas natural, y por el otro, entre el asfalto y otros hidrocarburos saturados semisólidos emparentados con él. Se ha sugerido que el petróleo bruto procede fundamentalmente de proteínas animales y que el gas natural deriva de materia vegetal alterada. Se define como “líquido oleoso más ligero que el agua y de color oscuro y olor fuerte, que se encuentra nativo en el interior de la tierra y a veces forma grandes manantiales. Es una mezcla de carburos de hidrógeno, que arde con facilidad, y después del refinado tiene diversas aplicaciones”.

El petróleo bruto o crudo en estado natural es una mezcla de una gran variedad de compuestos hidrocarburos, que difieren mucho de unos yacimientos a otros. En general, los compuestos parafínicos (lineales, ramificados, ciclados, aromáticos y especies complejas) son los más abundantes, tanto en el petróleo líquido como en el gas natural. La composición media elemental puede ser de un 85% de carbono, 12% de hidrógeno, 3% de la suma de elementos azufre, oxígeno, nitrógeno, y varios elementos metálicos.

El estudio de la estructura molecular de los crudos ha sido muy importante en el campo de la química del petróleo durante los últimos 100 años, debido a que sus propiedades fisicoquímicas están profundamente relacionadas con su composición y su estructura química ^[1,2].

Para caracterizar los crudos y sus derivados existen parámetros como la viscosidad, la gravedad API y el análisis SARA ^[3]. Este último consiste en el fraccionamiento de la muestra en compuestos Saturados, Aromáticos, Resinas y Asfáltenos por la acción de solventes como n-heptano, tolueno o diclorometano y por la interacción entre la muestra y sólidos superficialmente activos. El análisis se inicia con la precipitación de los asfáltenos por la acción de n-alcanos como pentano o heptano. Posteriormente, la fracción, se separa por cromatografía de elusión, con la ayuda de diferentes fases estacionarias y solventes de polaridad

variada. De cada fracción se remueve el solvente por evaporación y el análisis SARA se completa por la determinación de los pesos de las fracciones. Todo el consume una gran cantidad de solventes poco amigables con el medio ambiente [4,5].

El desarrollo de métodos de análisis precisos y rápidos se ha convertido en una necesidad urgente para el control de calidad de los procesos de producción, refinación y transporte de los crudos. Un gran número de técnicas analíticas incluyendo cromatografía líquida de alta resolución (HPLC), resonancia magnética nuclear (NMR), espectrometría de masas (MS) y espectroscopias de fluorescencia, Raman e infrarroja han sido ampliamente aplicadas en el análisis de hidrocarburos y sus derivados [6]. Estas técnicas han arrojado buenos resultados pero algunas de ellas son costosas y usualmente no están disponibles en los laboratorios.

Dentro de las técnicas espectroscópicas empleadas en la industria. El uso de la reflectancia total atenuada en el infrarrojo medio (ATR-MIR) es una alternativa prometedora para reemplazar los métodos tradicionales debido a que la muestra requiere de un tratamiento mínimo, la toma de espectros toma unos pocos minutos, la región media del IR disminuye el solapamiento de bandas en comparación con la región cercana (NIR) y presenta una repetibilidad aceptable en muestras líquidas [7].

El desarrollo de técnicas de análisis multivariante llamadas quimiométricas aplicadas a datos espectroscópicos ha permitido avanzar en la identificación y determinación de las propiedades fisicoquímicas de las muestras. El análisis por componentes principales (PCA) y la regresión por mínimos cuadrados parciales (PLS) permiten obtener modelos para la predicción de diferentes propiedades a partir de datos espectroscópicos [8].

1. CONSIDERACIONES TEÓRICAS

1.1 PETRÓLEO

1.1.1 Generalidades.

El petróleo es una mezcla de gran cantidad de compuestos orgánicos, 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, cuyas moléculas se componen en carbono e hidrógeno, con un pequeño porcentaje de otros elementos conformando hidrocarburos de estructuras complejas como heterocíclicos de nitrógeno, oxígeno y azufre, compuestos órgano-metálicos, compuestos inorgánicos y agua. El crudo puede estar contenido entre capas sedimentarias y migra para ser concentrado en formaciones geológicas características particulares como poca permeabilidad y algún grado de porosidad, que constituyen lo que se denomina yacimiento dentro de un campo petrolífero. Su estabilidad depende de una multitud de factores, entre las que se incluyen la temperatura, la presión, y su composición, especialmente las parafinas y los asfáltenos^[9].

El petróleo es un fluido complejo que contiene muchos componentes que van de menor a mayor peso molecular^[10]. 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. El petróleo varía mucho en su composición dependiendo de la región o ya bien sea de un yacimiento a otro.

Para caracterizar la composición química y los materiales relacionados es importante determinar el tipo de grupo de hidrocarburo. Una forma de caracterizar un crudo o sus fracciones pesadas es con base en el contenido de las fracciones SARA, el crudo es dividido en saturados, aromáticos, resinas y asfáltenos^[3, 11]. Las dos últimas fracciones (resinas y asfáltenos) son de mayor interés para la industria petrolera, debido a los inconvenientes que originan en la producción, refinación y transporte, causados especialmente por un fenómeno de agregación/floculación que causa la precipitación de estas.

1.1.2 Clasificación del petróleo.

Los crudos tienen características físicas y químicas muy variables de un campo de producción a otro e incluso dentro de un mismo yacimiento. La clasificación más sencilla, pero no menos importante en cuanto a los resultados económicos, es la clasificación en crudos pesados y livianos. Al estar formado principalmente por moléculas hidrocarbonadas, la densidad de un crudo será tanto menor cuanto

mayor sea la relación atómica H/C. La densidad de los crudos puede oscilar entre 0.7 y 1, expresándose con mucha frecuencia en grados API (american petroleum institute) cuyo valor varía entre 70 y 5 esta variabilidad de la densidad es consecuencia de composiciones en familias químicas muy diferentes ^[12].

$$API = \frac{141.5}{\text{densidad relativa estandar } \frac{60^{\circ}F}{60^{\circ}}} - 131.5 \quad (\text{ecuación 1})$$

Los crudos están constituidos por mezclas de un número muy elevado de componentes puros, aumentando la dificultad de la descripción de las distintas fracciones. En términos de la densidad API, los crudos livianos o de baja gravedad específica presentan alto dicho valor. Los crudos con bajo contenido de carbono, alto contenido de hidrógeno y alta densidad API son generalmente ricos en hidrocarburos parafínicos y tienden a producir mayores cantidades de gasolina y productos ligeros; aquellos crudos con alto contenido de carbono, bajo contenido de hidrogeno y baja densidad API son ricos en hidrocarburos nafténicos y aromáticos. 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.

Las especificaciones generales de la clasificación del crudo basándose en la densidad API se muestran en la (tabla 1).

Tabla 1. Especificaciones generales de clasificaciones de crudos.

| Aceites crudos | Densidad (g/cm ³) | Gravedad API |
|----------------|-------------------------------|--------------|
| Extra pesado | >10 | 1 |
| Pesado | 1,0-0,92 | 10 |
| Mediano | 0,92-0,87 | 22,3-31,1 |
| Ligero | 0,87-0,83 | 31,1-39 |
| Super ligero | <0,83 | >39 |

Fuente WAUQUIER.J.P. El refinado del petróleo, petróleo crudo, productos petrolíferos, esquemas de fabricación ^[13].

1.1.3 Composición química del petróleo.

El análisis de la composición de los crudos de petróleo es infinitamente complejo, un esquema de análisis simple consiste en dividir un crudo en saturados, aromáticos, resinas y asfáltenos (fracción SARA). La fracción de saturados está compuesta por hidrocarburos saturados no polares lineales, ramificados y cíclicos. Los aromáticos, que contiene uno o más anillos aromáticos, son más polarizables. Las otras dos fracciones, resinas y asfáltenos son insolubles en un exceso de heptano (o pentano), mientras que las resinas son miscibles en estos solventes.

Este sistema de clasificación es útil porque identifica las fracciones del crudo que se refieren a la estabilidad de asfáltenos y por lo tanto permite identificación de los crudos con potencial para generar problemas por asfáltenos ^[13]. A continuación se da una pequeña descripción de la composición de estas fracciones ^[14].

1.1.3.1 Saturados.

Saturados significa que la molécula contiene el número máximo de átomos de hidrógeno posibles. Son aceites blancos no polares constituidos por hidrocarburos alifáticos lineales o con cadenas laterales alifáticas y aromáticas. El rango de peso molecular medio está comprendido entre 300 y 2000.

Hidrocarburos alifáticos saturados - alcanos o parafinas: Están constituidos por una cadena de átomos de carbono enlazados cada uno de 0 a 3 átomos de hidrógeno, excepto en el más sencillo, el metano (CH_4). Cada carbono está ligado siempre a otro cuatro átomos (carbono o hidrogeno); y su fórmula general es C_nH_{2n+2} . Cuando su estructura es de cadena recta se llaman parafinas normales o n-alcanos los átomos de hidrógeno pueden ser sustituidos por carbonos o cadenas hidrocarbonadas, formando isoparafinas o isoalcanos.

Hidrocarburos cíclicos saturados, cíclicos o naftenos: En estos hidrocarburos hay una ciclación total o parcial del esqueleto carbonado. El número de átomos de carbono del anillo formado pueden ser variables. Tiene temperaturas de ebullición y densidades superiores a los de los alcanos del mismo número de átomos de carbono. En los petróleos crudos, los anillos más frecuentes son los de los cinco o seis carbono. En estos anillos, cada átomo de hidrógeno puede ser sustituido por una cadena parafinica recta o ramificada, llamada alquilo.

1.1.3.2 Aromáticos.

Comprenden los compuestos nafteno-aromático de menor peso molecular; generalmente son líquidos viscosos de color marrón anaranjado. El peso molecular promedio de esta fracción es similar a la de los saturados. Consiste en cadenas no polares de carbono en las que dominan los sistemas de anillos insaturados y tienen una gran capacidad para disolver los otros hidrocarburos de alto peso molecular. Los aromáticos incorporan uno o más anillos de seis átomos de carbono y seis átomos de hidrogeno. El aromático más simple es el benceno C_6H_6 incluye mono- aromáticos y aromáticos policíclicos.

Hidrocarburos aromáticos: Son hidrocarburos cíclicos poli-insaturados que están presentes en una gran proporción en los crudos de petróleo. La presencia en su fórmula de uno o más ciclos son tres dobles enlaces conjugados les confieren unas notables propiedades. Así, los primeros compuestos (benceno, tolueno, xileno) son materias primarias fundamentales de la petroquímica además contribuyen igualmente a aumentar el número de octano de la gasolina.

1.1.3.3 Resinas.

Las resinas son sólidos negros, brillantes y quebradizos que constituyen también la fracción de crudo soluble en n-heptano, son compuestos aromáticos similares en estructura a los asfáltenos, tienen mayor relación parafina/aromático (contienen heteroátomos tales como N, S, O y a veces Ni y V), tienen menor tamaño que los asfáltenos y constituyen las partes polares de los crudos. Tanto los asfáltenos como las resinas son los principales constituyentes que controlan el comportamiento del crudo en el depósito y durante el transporte. Por lo tanto, varios trabajos han demostrado que la estabilidad del crudo depende de la naturaleza, del contenido de las resinas y los asfáltenos^[15].

Las resinas se consideran como agentes estabilizantes naturales para los asfáltenos según lo reportado por Koots y Speight y se pueden adsorber en la superficie de los agregados cuando los asfáltenos son depositados^[10].

1.3.3.4 Asfáltenos.

Representan un grupo importante de compuestos y se definen en términos de su solubilidad^[16, 17] y constituyen la fracción más polar^[6] y pesada del crudo, se espera que cada vez sean más utilizados^[18,19]. De acuerdo a la solubilidad están catalogados como la fracción de crudo soluble en tolueno o benceno e insoluble en un exceso de n-alcano (pentano o heptano)^[3, 15, 20, 21].

La determinación de su peso molecular ha sido, y sigue siendo, un problema difícil en la química de los asfáltenos. El origen de las discrepancias en la caracterización y el peso molecular de los asfáltenos se puede atribuir a tres propiedades básicas de los asfáltenos, es decir, variaciones de composición, tamaño de la polidispersidad y lo más importante, la alta capacidad de las moléculas covalentes de los asfáltenos para formar agregados moleculares^[22].

Los asfáltenos están constituidos principalmente por anillos aromáticos ligados con cadenas alquílicas y ciclo-alcanos, además de compuestos heterocíclicos que poseen N, S y O. Son moléculas relativamente pequeñas con pesos moleculares promedios de aproximadamente 750 Dalton (g/mol) con un rango de 500 – 1000 Dalton^[18, 23, 24] y tienen probablemente de cuatro a diez anillos fusionados en sus PAHS según el examen de absorción óptica y perfiles de emisión de fluorescencia^[23, 24,25].

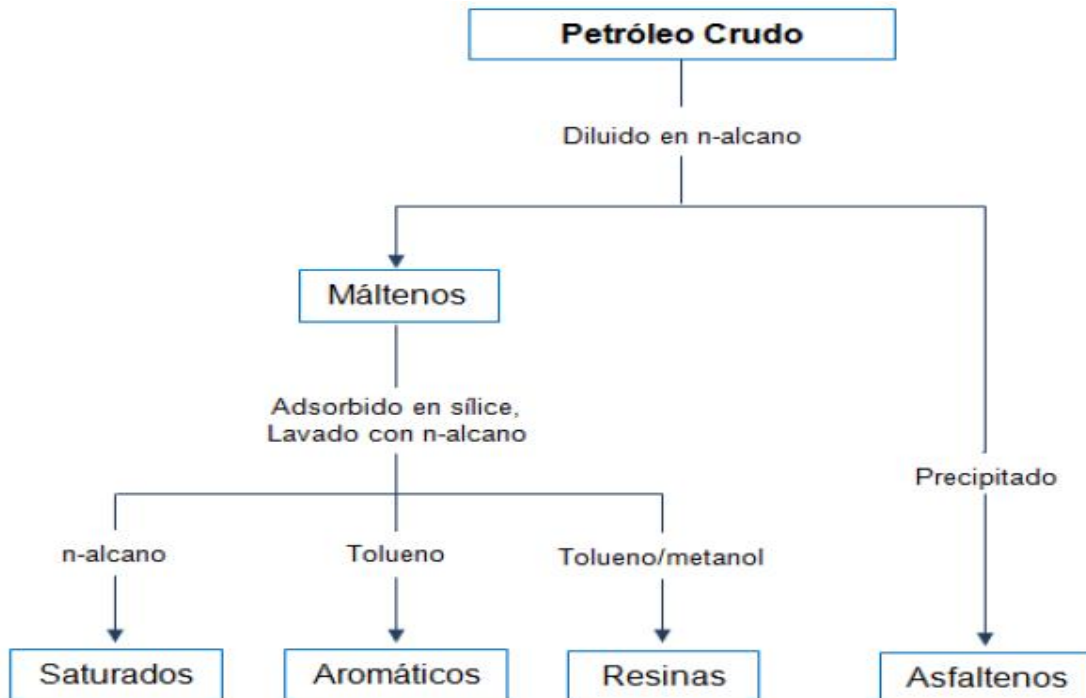
Si los pesos moleculares de los asfáltenos son grandes, entonces cada molécula debe estar constituida por número considerable de anillos fusionados separados. Si los pesos moleculares son pequeños, entonces cada molécula tiene solo 1 o quizás 2 sistemas de anillos^[25].

1.1.4 Fraccionamiento S.A.R.A.

El análisis de las fracciones de saturados, aromáticos, resinas y asfáltenos del petróleo se denomina (análisis SARA), expresados como porcentajes en peso. La fortaleza del método radica en que analiza la muestra de forma completa, desde los compuestos livianos hasta los pesados, y de este modo permite comparar todo los petróleos con base en un estándar consistente. Estas técnicas requieren del uso de diferentes métodos para el pretratamiento de la muestra, así como de solventes para hacer posible su análisis.

El fraccionamiento SARA hace en primer lugar una separación mediante precipitación con hidrocarburos saturados de bajo pesos molecular (n-heptano, o n-pentano) obteniéndose dos fracciones, a la fracción insoluble se le denomina asfáltenos y a la soluble máltenos; estos últimos a la vez contienen varios tipos de hidrocarburos que pueden dividirse en saturados, aromáticos y resinas (figura 1).

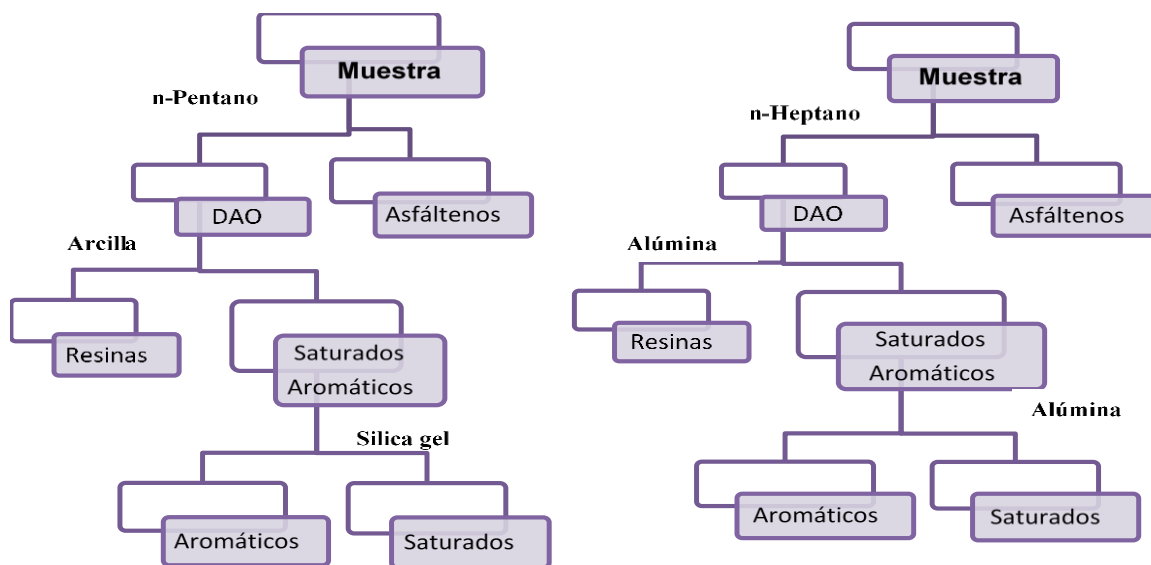
Figura 1. Esquema simplificado para la determinación de las fracciones SARA.



La ASTM ofrece varios procedimientos para el análisis SARA: la norma ASTM D2007 ^[26] y la ASTM D4124 ^[29], mostrados en la figura 1.a, junto con la norma

ASTM D2549^[27] que usa la cromatografía líquida para la separación de las fracciones de saturados y aromáticos^[28].

Figura 1.a procedimiento estándar para el análisis SARA según las normas ASTM.



La porción de crudo que es soluble en pentano y heptano se conoce como máltenos este a su vez puede fraccionarse por percolación a través de cualquier material superficialmente activo como alúmina o arcilla, la fracción más fuertemente absorbida es la resina, la fracción que eluye con el tolueno o el benceno se define como aromáticos, mientras que la que lo hace por acción del heptano se conoce como saturados.

1.2 ESPECTROSCOPIA DE INFRARROJO

1.2.1 Aspecto fundamentales.

La espectroscopia molecular se basa en la interacción entre la radiación electromagnética y las moléculas. Dependen de la región del espectro en la que se trabaje y por tanto la energía de la radiación utilizada (caracterizada por su longitud o número de onda), ésta interacción será de diferente naturaleza: excitación de electrones, vibraciones moleculares y rotación moleculares^[30]. La molécula, al absorber la radiación infrarroja, cambia su estado de energía vibracional y rotacional. Las transiciones entre dos estados rotacionales requieren muy poca energía, por lo que es posible observarla específicamente en el caso de muestras gaseosas. En el caso del estudio del espectro infrarrojo (IR) de muestras sólidas y líquidas solo se tienen en cuenta los cambios entre estados de energía vibracional^[31,32].

Matemáticamente la energía de los estados vibracionales en una molécula diatómica se puede describir por el modelo del oscilador armónico (figura 2) mediante la (ecuación 2).

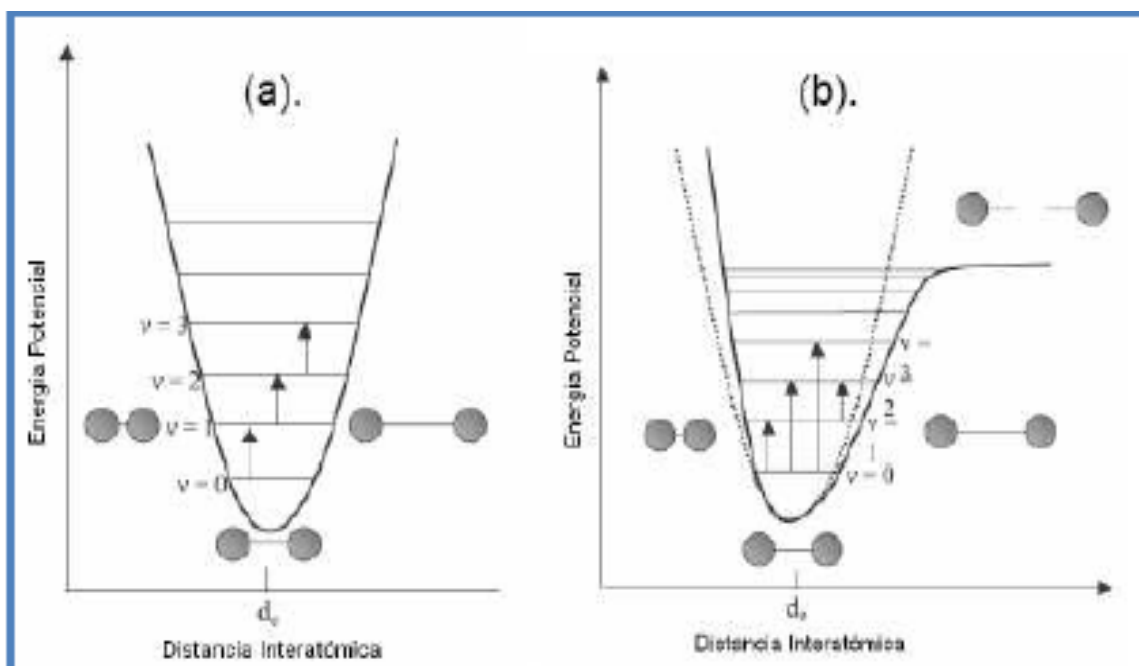
$$E_{vib} = \left[v + \frac{1}{2} \right] \frac{h}{2\pi} \sqrt{\frac{K}{\mu}} \quad (\text{ecuación 2})$$

Donde v es el número cuántico vibracional, h es la constante de Planck, K es la constante de fuerza y μ es la masa reducida del sistema. Este modelo permite únicamente las transiciones entre niveles energéticos adyacentes ($\Delta v = \pm 1$), y asume que los niveles energéticos se encuentran igualmente espaciados siendo su diferencia de energía siempre la misma. Esta diferencia energética puede ser calculada por la (ecuación 3).

$$\Delta v = E_{v^2} - E_{v^1} = \Delta v h \nu \quad (\text{ecuación 3})$$

Donde ν es la frecuencia de vibración fundamental del enlace que produce una banda de absorción en la región del infrarrojo medio.

Figura 2. Perfil de energía potencial según el modelo de oscilador (a) armónico (b) anarmónico.



Aunque el modelo del oscilador armónico es una buena aproximación, no explica muchos aspectos sobre el comportamiento molecular. El modelo del oscilador anarmónico representa de manera más aproximada estos aspectos y asume que los niveles energéticos no se encuentran igualmente espaciados (figura 7b) de esta manera la diferencia energética disminuye a medida que aumenta v y se puede calcular por medio de la (ecuación 4).

$$\Delta E_{vib} = hv [1 - (2v + \Delta v + 1)\gamma] \quad (\text{ecuación 4})$$

Donde γ es el factor de anharmonicidad. Esta anharmonicidad permite las transiciones entre estados de energía vibracionales no consecutivos, ($\Delta v = \pm 2, \pm 3 \dots$) generando las bandas de absorción conocidas como sobre tonos las cuales son, aproximadamente, múltiplos de las frecuencias fundamentales de vibración aunque su intensidad es mucho menor la intensidad de las bandas para el primer sobre tono puede ser, dependiendo del enlace, diez a cien veces menor que la frecuencia fundamental. Aunque teóricamente son posibles transiciones entre cualquier par de niveles energéticos, experimentalmente solo se observan las bandas de absorción correspondientes a las frecuencias de vibración fundamental en el MIR (v) y a los dos primeros sobre tonos en el NIR ($2v, 3v$)^[33].

Una molécula poliatómica (n átomos) tiene $3n-6$ modo de vibración diferentes ($3n-5$ si la molécula es lineal). Cada uno de estos modos de vibración viene representado por una curva de energía potencial diferente y da lugar a una banda fundamental y sus correspondientes sobre tonos en el infrarrojo. Los modos de vibración que se pueden producir incluyen cambios en la distancia de enlaces (elongaciones que pueden ser simétricas o asimétricas) y cambios en el ángulo de enlace (simétricos en el plano; asimétricos en el plano, simétricos fuera del plano y asimétricos fuera del plano)^[34].

1.2.2 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, donde se produce las absorciones debido a cambios rotacionales, el IR medio (MIR o simplemente IR), donde tiene lugar las variaciones fundamentales y el IR cercano (NIR), donde se producen absorciones debidas a sobre tonos y combinaciones de las bandas fundamentales. Las tablas 2 muestra el rango en el espectro electromagnético al que corresponde cada región IR.

Tabla 2 regiones de absorción en el infrarrojo.

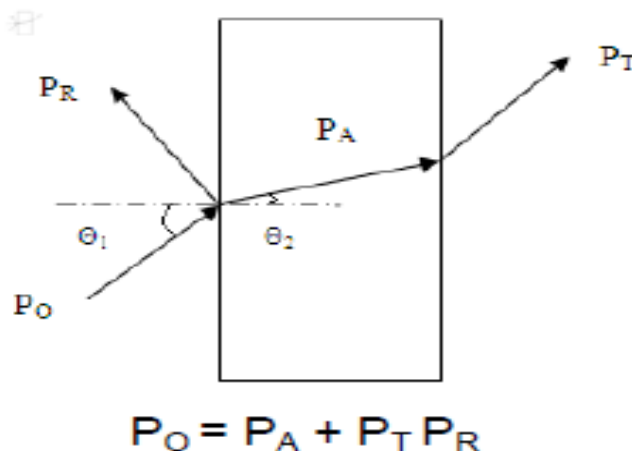
| Región infrarrojo | Longitud de onda | Numero de onda cm-1 | Transición característica |
|-------------------|------------------|---------------------|-----------------------------|
| Cercano (NIR) | 780 a 2500nm | 12800 a 4000 | Sobre tonos y combinaciones |
| Medio (MIR) | 2,5 s 50 um | 4000 a 200 | vibraciones fundamentales |
| Lejano (FIR) | 50a 1000 um | 200 a 10 | rotaciones |

Fuente SKOOG.A. Douglas principios de análisis instrumental.

1.2.3 Tipos de medidas en el infrarrojo.

Cuando la radiación incidente incide en la muestra (figura 3), esta puede sufrir diferentes fenómenos: absorción, transmisión y reflexión. La intensidad de la luz transmitida a través de la muestra (P_T) es menor que la intensidad incidente (P_0). Una parte de esta intensidad incidente se ha reflejado (P_R), mientras que otra parte ha sido absorbida por la sustancia (P_A).

Figura 3. Fenómeno de absorción, transmisión y reflexión de la radiación electromagnética al interactuar con la materia.



Fuente MACHO, A Santiago: metodologías analíticas basadas en espectros de infrarrojo y calibración multivariante. Aplicación a la industria petroquímica.

La medida más común en el infrarrojo es la que se basa en la absorción (o la intensidad transmitida), aunque también se han desarrollado espectroscopias basadas en el fenómeno de la reflexión como son la reflectancia total atenuada y

la reflectancia difusa ^[35]. A continuación se describen las técnicas que se utilizan en la tesis.

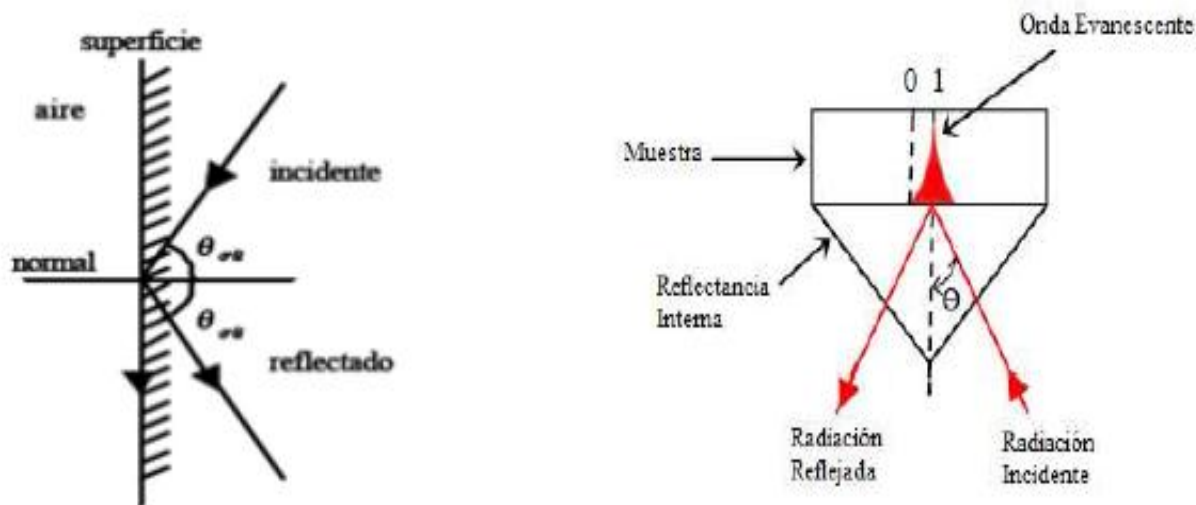
1.2.3.1 Reflectancia total atenuada.

El principio de esta medida se basa en el fenómeno de la reflexión total interna y la transmisión de la luz a través de un cristal con un elevado índice de refracción ^[35]. La radiación penetra más allá de la superficie del cristal donde se produce la reflexión total, en forma de onda evanescente, la luz que viaja a través del cristal se verá atenuada y se puede registrar el espectro de la muestra.

El ángulo de la luz incidente y la geometría del cristal facilitan que se produzca sucesivas reflexiones en sus caras internas (figura 4). El espectro medio tiene una apariencia similar al espectro de trasmisión excepto por cierta variación en la intensidad en función de la longitud de onda que se producen.

La intensidad de las ondas evanescentes es sustraída de la intensidad del haz que continua reflejándose hasta que sale del cristal al detector del equipo IR, y de esta manera genera un espectro IR ^[36,37]. Para un uso adecuado de la celda ATR (figura 4), es necesario tener en cuenta los siguientes factores experimentales y como ellos afectan el espectro final.

Figura 4. Reflexión total interna y elemental de reflexión interna (IRE) utilizando en el sistema ATR.



Fuente adaptado de CAI, M, F and SMART, R, B. comparison of seven West Virginia coals with their N-metyl-2-pyrroidone-soluble extracts and residues.

Índice de refracción del cristal ATR y la muestra: los índices de refracción de la muestra y el cristal gobiernan el fenómeno ATR en virtud de la (ecuación 5).

$$\theta = \sin^{-1} \left[\frac{n_2}{n_1} \right] \quad (\text{ecuacion 5})$$

Donde n_2 es el índice de refracción de la muestra, n_1 el índice de refracción del cristal y θ es el ángulo crítico. Cuando excedemos el ángulo crítico, se puede observar un espectro ATR puro. Si el ángulo crítico no es conocido, observaremos un resultado combinado entre ATR y reflectancia externa. Esto ocurre si el ángulo de incidencia del rayo es muy bajo, si el índice de refracción del cristal es muy bajo, si el índice de refracción de la muestra es muy alto o una combinación estos tres factores. En la mayoría de los casos este problema no es observado. Una manera de corregir estos es aumentando el ángulo de incidencia a un valor cercano al ángulo crítico.

- **Profundidad de penetración:** es la distancia requerida para que la amplitud del campo eléctrico caiga e^{-1} de su valor en la superficie y está dada por la (ecuación 6).

$$d_p = \frac{\lambda}{2\pi(n_1^2 \sin^2 \theta - n_2^2)^{1/2}} \quad (\text{ecuacion 6})$$

Donde λ es la longitud de onda y θ el ángulo de incidencia del rayo IR relativo a una perpendicular desde la superficie del cristal. Una profundidad de penetración típica oscila entre 0.5 y 5 μm . Como se muestra en la figura 4 de representación del fenómeno ATR, la intensidad de la onda evanescente decae rápidamente desde la superficie del cristal ATR.

- **Número de reflexión dentro del cristal ATR.** En celdas de reflexión simple el haz se hace incidir una vez con la muestra mientras en la celda de múltiples reflexiones la profundidad efectiva de penetración aumenta efectivamente de penetración aumentan y la señal ATR es de mayor intensidad.
- **Calidad del contacto entre la muestra y el cristal.** Se refieren al contacto íntimo de la muestra y el cristal ATR para evitar problemas de ruido por la detección de gases atmosféricos y de dispersión de la radiación. En este punto juega un papel importante el estado físico de las muestra. En líquidos es posible garantizar un contacto adecuado con el cristal. Mientras que en el sólido o muestras pulverizadas, se requiere de una prensa que acerque lo suficiente la muestra al cristal.

La información presente en un espectro MIR puede ser usada para estimar la concentración de un componente o para estimar una propiedad física cuando estas reflejan cambios significativos en las características espectrales generadas por la muestra. Para lograr esto, es necesario recurrir a diversos métodos multivariados de análisis que se encuentran agrupados en una rama de la química analítica la quimiometría, encargada de usar técnicas matemáticas y estadísticas

para extraer información relevante de datos analíticos, en este caso de la información espectral obtenida en la región del infrarrojo medio ^[38].

1.2.4 Interpretación del espectro.

Asignación de bandas: en los espectros de infrarrojo medio, entre 4000 y 200 cm^{-1} (región de frecuencia de grupo) se observan una serie de bandas asignadas a vibración de solo dos átomos de la molécula. En este caso la banda de absorción se asocia únicamente a un grupo funcional y a la estructura molecular compleja, aunque hay influencia estructural que provocan desplazamientos significativos en la frecuencia de la vibración. Estas vibraciones derivan de grupos que contiene hidrógeno (C-H- O-H y N-H) o grupos con doble o triple enlace aislados. Entre 1300 y 400 cm^{-1} la asignación a grupos funcionales determinados es más difícil debido a la multiplicidad de bandas, pero es una zona de espectros muy útil para la identificación de compuestos específicos ^[39]. La tabla 3 muestra un cuadro resumen de la frecuencia de absorción de los grupos funcionales más comunes en el IR medio.

Tabla 3. Frecuencia de absorción de algunos grupos funcionales en la región MIR.

| enlaces | tipos de compuestos | intervalo de frecuencia cm^{-1} | intensidad |
|-----------------|---------------------|-----------------------------------|------------|
| C-H | alcanos | 2850-2970 | fuerte |
| C-H | alquenos | 1340-1470 | media |
| C-H | Alquinos | 3010-3095 | fuerte |
| C-H | anillos aromáticos | 675-995 | media |
| O-H | alcoholes y fenoles | 330 | variable |
| N-H | aminas, amidas | 3300-3500 | media |
| C=C | alquenos | 1610-1680 | variable |
| C=C | anillo aromáticos | 1500-1600 | variable |
| C≡C | alquinos | 2100-2260 | variable |
| C-N | aminas, amidas | 1180-1360 | fuerte |
| C≡N | nitrilos | 2210-2280 | fuerte |
| C-O | éteres | 1050-1300 | fuerte |
| C=O | aldehídos, cetonas | 1690-1760 | fuerte |
| NO ₂ | nitro derivados | 1300-1370 | fuerte |

Fuente: SKOOG, A, Douglas principios de análisis instrumental.

1.2 QUIMIOMETRÍA

El desarrollo de la quimiometría, como herramienta de análisis, clasificación y calibración multivariable, se da por la imposibilidad de describir y modelar sistemas químicos mediante la estadística univariada tradicional. Sus primeras aplicaciones se dieron por un grupo de investigadores en el área de química analítica a finales de la década de 1960 con el fin de analizar datos dependientes de más de una variable simultáneamente. Para el caso específico de la espectroscopia MIR, raramente se puede emplear una única longitud de onda (análisis univariado) para fines cuantitativos. En la mayoría de los casos se requiere emplear varias o todas las variables espectrales con el fin de obtener información suficiente para el desarrollo de un procedimiento analítico dado (análisis multivariado).

1.3.1 Calibración multivariable.

La calibración multivariable se puede definir como la actividad de encontrar relación entre uno más variables de respuesta “Y” y una matriz de variables predictoras “X” de manera que se cumpla la (ecuación 7)

$$y = g(x) \quad (\text{ecuación 7})$$

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 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 la 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.

1.3.1.1 Clasificación de los métodos de calibración multivariable.

Aunque no exista un criterio unificado de los métodos de calibración multivariable, el propuesto por Martens & Naes ^[40], ha sido uno de los más aceptados. Su clasificación se basa en los siguientes aspectos fundamentales:

- **Según la relación entre las variables dependientes e independientes:** esta relación puede ser descrita mediante un modelo lineal o no lineal.
- **Según la forma de encontrar la relación entre la variable:** puede ser métodos directos, donde los parámetros de calibración se calculan directamente a partir de la señal de cada uno de los analitos en forma individual, o métodos indirectos, donde tales parámetros se calculan a partir de las señales analíticas de la mezcla de los componentes.

- **Según la variable que se defina como dependiente o independiente:** si la calibración sigue el criterio directo relacionado con la ley de Beer donde la señal analítica actúa como variable dependiente y la concentración como variable independiente se tiene un método de calibración directa. En caso contrario se tiene un método de calibración inversa.

1.3.1.2 Construcción de modelos de calibración multivariable.

La norma ASTM E-1655 ^[41], sugiere la siguiente etapa para la construcción de modelos de calibración multivariables a partir de mediciones espectrales sobre el analito de interés:

- **Selección de muestras para la calibración:** la caracterización se debe realizar pretratamiento de la señal y aplicación de técnicas de regresión sobre los datos espectrales. Los pre-tratamientos incluyen: el suavizado para reducir el ruido de los datos, la normalización para lograr que los datos estén aproximadamente a la misma escala, la centralización para evitar que ciertos puntos tengan más peso que otros en el modelo y la derivación de diferentes órdenes para extraer información detallada que no pueda ser observada en el espectro normal.

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

Validación del método de calibración: 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 variación externa. Si se emplean muestras incluidas en la calibración del modelo se tiene el procedimiento de validación cruzada.

Implementación del modelo al análisis de muestras desconocidas: en esta etapa final el modelo se instala como herramienta de análisis de rutina y se realizan chequeos periódicos para evaluar su desempeño.

El fundamento matemático para desarrollo del método de calibración multivariable aplicado a técnicas espectroscópicas de análisis instrumental tiene su origen en el álgebra matricial. Si "S" espectros de calibración son medidos "W" discretas longitudes de onda, es posible construir una matriz de datos espectrales "X" de dimensiones "W*X" que contiene un espectro en columna. De la misma manera es posible construir un vector "Y" de dimensiones "S*1" que contiene los valores de referencia de las muestras de calibración. El objetivo de la calibración multivariable es calcular un vector "P" de dimensiones "W*1" que resuelva la (ecuación 8)

$$Y = X^t p + e \quad (\text{ecuacion 8})$$

Donde “X” es la transpuesta de la matriz “X” y “e” es un vector de dimensiones “S*1” llamado vector de error. Este vector se calcula como la diferencia entre los vectores de referencia y los valores estimados por modelo. Generalmente el vector “P” se estima minimizando la suma de los cuadrados de los errores mediante la (ecuación 9).

$$e^t e = [e^2] = (y - X^t P)^t (y - X^t P) \quad (\text{ecuacion 9})$$

Ya que normalmente “X” no es una matriz cuadrada, la (ecuación 9) no puede ser resultado directamente. Una alternativa para solucionar esto es determinar la matriz pseudoinversa de “X”, “X” y calcular el vector de predicción “P” mediante (ecuación 10).

$$X^+ y = (X X^t)^{-1} X y = P \quad (\text{ecuacion 10})$$

Habitualmente los espectros de calibración son medidos sobre un “X”, generalmente mediante un análisis por componentes principales (PCA), y realizar posteriormente una regresión multivariable.

1.3.2 Técnicas de pretratamiento de datos.

1.3.2.1 Suavizado espectral.

Le suavizado espectral se aplica en aquellos casos que la relación señal/ruido es pequeña y por medio de algoritmos matemáticos aplicados al espectro se reduce el ruido suavizando la señal. Los métodos de suavizado más habituales son los basados en filtros de savitzky golay y transformadas de Fourier.

1.3.2.2 Normalización.

Esta se usa para lograr que los datos estén aproximadamente en la misma escala puede ser:

Normalizada por rangos:

En esta transformada se normaliza un espectro Xi calculando el área bajo la curva del espectro.

La media de normalización:

Este es el caso más clásico de la normalización, consiste en dividir cada fila de una matriz de datos por su media.

Máxima normalización:

Esta es una alternativa a la normalización clásica que divide cada fila por su valor máximo absoluto en lugar de la media.

Propiedades 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

1.3.2.4 Correlación de la línea base.

La correlación de la línea base es un tipo de pretratamiento que intenta corregir determinadas tendencias en la línea base que aportan el ruido a la señal. Existen varios tipos de corrección de línea base según el efecto que se desea corregir. Un tipo de correlación es el que modela la línea base como una función simple de longitud de onda y sustrae esta función a todos los datos espectrales.

1.3.2.4 Centralización.

La centralización evita que ciertos puntos tengan más peso que otros en el modelo, consiste en calcular el valor medio de cada variable (X_m) del conjunto de calibración y reste este valor cada punto (X_{im}).

1.3.2.5 Derivación.

La derivada tiene como función extraer la información detallada que no puede ser observada en el espectro normal y esta puede ser de diferente orden:

Derivada (primera y segunda):

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. El primero, permite calcular derivadas de primer orden o mayor incluyendo un factor de suavización que determina el número de variaciones 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^[42].

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 menor robusto 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 la calibración^[43,44].

1.3.3 Método basado en reducción de variables.

Estos métodos se basan en que la información contenida en la variables de la señal pueden estar contenidas en un número menor de variables sin que haya

pérdida de información relevante. El proceso de calibración se realiza, no sobre los datos originales. Sino sobre estas nuevas variables, simplificando el modelo y la interpretación de los resultados.

Este tipo de métodos de calibración son de espectros completo no presentan problemas de colinealidad ni las consecuencias derivadas de ella, por estas razones, la tendencia actual es la utilización de métodos de calibración basados en una reducción de variables previamente al cálculo del modelo, generalmente al igual que en el análisis de componentes principales (PCA), los procedimientos de reducción de variables no son realizados sobre los datos originales sino que se centran o auto escalan previamente, uno de estos métodos es (PLS).

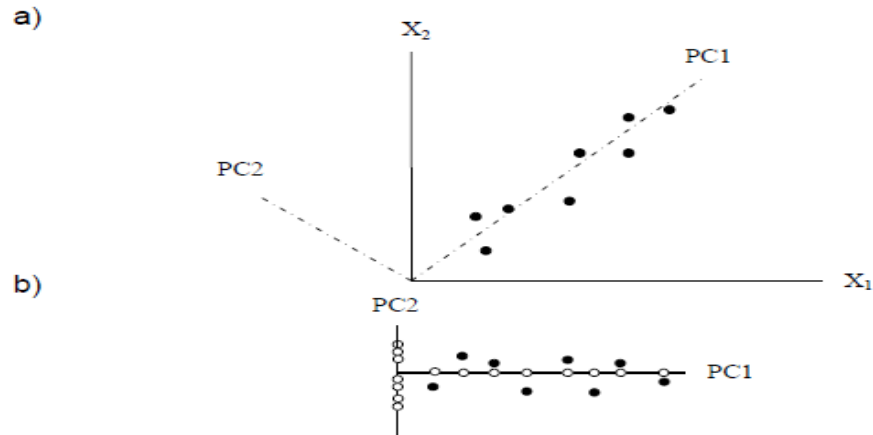
1.3.3.1 Análisis por componentes principales.

El PCA es una técnica para reducir la cantidad de datos cuando está presente la correlación. No es una técnica útil cuando las variables no están correlacionadas la idea del PCA es encontrar componentes principales Z_1, Z_2, \dots, Z_n que sean combinaciones lineales de la variable original X_1, X_2, \dots, X_n que describen cada muestra, es decir,

$$Z_1 = a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + a_{14}X_4$$

Los coeficientes $a_{11} \dots a_{12}$ se eligen de manera que las nuevas variables a diferencia de las variables originales, no se encuentren correlacionadas unas con otras. De esta forma se obtiene n nuevas variables en lugar de la n originales, en consecuencia hay una reducción en el conjunto de datos. Sin embargo las componentes principales se eligen de manera que la primera componente principal (PC1) Z_1 recoger parte de la variaciones que hay en el conjunto de datos la segunda (PC2) Z_2 , recoge la siguiente mayor parte de la variación y así sucesivamente. Por consiguiente cuando hay correlación significativa el número de PC útiles será mucho menor que el número de variables originales.

Figura 5 aclaración del método cuando solo hay dos variables y, por tanto, solo dos componentes principales. En la figura 5a los componentes principales forman ángulos rectos unas con otras, propiedad conocida como ortogonalidad. Figura 5b muestra los puntos referidos a estos dos nueve ejes y también la proyección de los puntos sobre PC-1 y PC-2 se puede ver que en este caso Z_1 recoge la mayor parte de la variación y así sería posible reducir la cantidad de datos a manejar trabajando con Z_1 en una dimensión en lugar de trabajar en dos dimensiones con X_1 y X_2 .



Fuente: MILLER N James Estadística y Quimiometria para química Analítica.

La figura 5 muestra que el PCA es equivalente a una rotación de los originales, de tal manera que PC-1 se encuentre en la dirección de la máxima variación, pero manteniendo el ángulo entre los ejes con más de dos variables no resulta posible el método gráficamente pero de nuevo se puede pensar en el PCA como una rotación de los ejes de tal manera que PC-1 se encuentre en la dirección de máxima variación, PC-2 se encuentra en la dirección de la siguiente mayor variación y así sucesivamente ^[43].

Cuando se emplea métodos espectroscópicos, como es en este caso, cada muestra genera respuestas en cientos o miles de longitudes de onda. A partir de la matriz "X", construida de la información espectral obtenida de S muestras medidas W longitudes de onda, se realiza una descomposición por componentes principales que proporcionan una aproximación a la matriz X como un producto de dos matrices (ecuación 13): la matriz de puntuación (scores) T y la matriz de cargas (loadings) P.

$$X = TPT + E \quad (\text{ecuación 13})$$

Donde E es la matriz de residuos de dimensiones S x W

La matriz contiene información pertinente a las relaciones entre muestras y está constituida por S filas, que corresponden al número de componentes principales. La matriz P explica la relación existente entre variables originales y está constituida por A filas y W columnas (figura 6)

Figura 6. Notación matricial para la descomposición por componentes principales Fuente adaptación. Grupo de quimiometría de Tarragona, España, quimiometría, una disciplina útil para el análisis químico.

$$\begin{array}{c} \text{w} \\ \boxed{X} \\ \text{s} \end{array} = \begin{array}{c} \text{A} \\ \boxed{T} \\ \text{s} \end{array} \begin{array}{c} \text{w} \\ \boxed{P^T} \\ \text{A} \end{array} + \begin{array}{c} \text{w} \\ \boxed{E} \\ \text{s} \end{array}$$

El producto TPT se puede representar como la suma de A términos de la forma tapa T (ecuación 14), que corresponde a cada una de las columnas y filas de las matrices T y P prospectivamente (figura 7) cada uno de dichos términos se denomina factor o componente principal.

$$X = t_1 p_1 T + t_2 p_2 T + \dots + t_0 p_0 T + E \quad (\text{ecuación 14})$$

Figura 7. Notación matricial extendida para la descomposición por componentes principales.

$$\begin{array}{c} \text{w} \\ \boxed{X} \\ \text{s} \end{array} = \begin{array}{c} \text{w} \\ \boxed{T_1} \\ \text{s} \end{array} \begin{array}{c} \text{w} \\ \boxed{P_1^T} \\ \text{A} \end{array} + \begin{array}{c} \text{w} \\ \boxed{T_2} \\ \text{s} \end{array} \begin{array}{c} \text{w} \\ \boxed{P_2^T} \\ \text{A} \end{array} + \dots + \begin{array}{c} \text{w} \\ \boxed{T_A} \\ \text{s} \end{array} \begin{array}{c} \text{w} \\ \boxed{P_A^T} \\ \text{A} \end{array} + \begin{array}{c} \text{w} \\ \boxed{E} \\ \text{s} \end{array}$$

Fuente: Adaptación. Grupo de quimiometría y cualimetría de Tarragona, España. Quimiometría: una disciplina útil para el análisis químico.

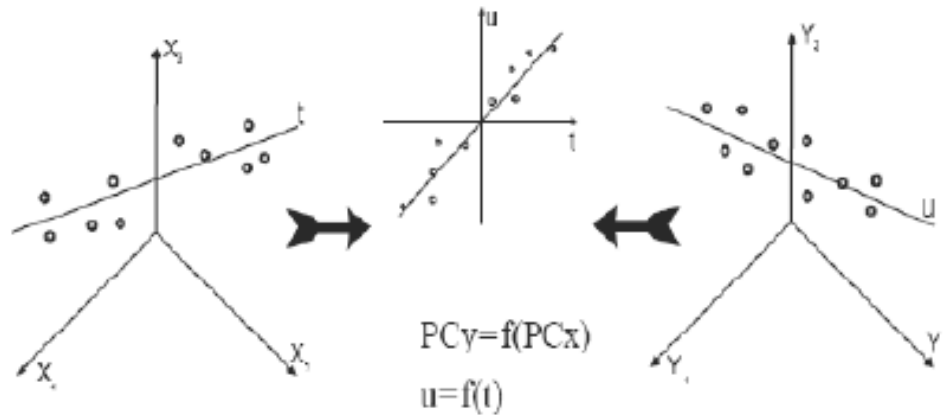
Los componentes principales se determinan con base en el criterio de varianza máxima. El primer componente es aquel que explica una mayor cantidad de la información contenida en la matriz X. Los componentes sucesivos explican cada vez menos información de los datos originales. En resumen, el análisis por componentes principales es un método que tiene como objetivo reducir la complejidad de una matriz de datos a partir de combinaciones lineales de las variables originales ^[45].

1.3.3.2 Regresión por mínimos cuadrados parciales (regresión PLS).

La regresión por mínimos cuadrados parciales de estructuras parciales de estructuras laterales (PLS) es un método matemático que modelo

simultáneamente las matrices X y para encontrar un conjunto de variable latente (VL) en X que mejor predicen las variables latentes en Y (figura 8).

Figura 8. Descripción gráfica del método de regresión PLS.



Como en el caso del análisis de componente principal, estas nuevas variables en X y se pueden representar como un producto de matrices según se muestra en las ecuaciones (15) y (16).

$$X = TPT + E = \sum t_0 p_0^r + E \quad (\text{ecuación 15})$$

$$Y = UQT + F = \sum u_0 q_0^r + E \quad (\text{ecuación 16})$$

Dónde:

T y U son las matrices de puntuación (scores) de X y respectivamente;

P y Q son las matrices de carga de X y Y respectivamente;

E y F son los residuos. Las variables originales X y Y se pueden relacionar mediante los scores de cada una de las nuevas variables latentes como lo muestra la (ecuación 17).

$$U_0 = b_0 t_0 \quad (\text{ecuación 17})$$

Donde b_0 es el coeficiente de regresión para cada variable latente obtenido a través de la (ecuación 18).

$$b_0 = \frac{U_0^r t_0}{t_0^r t_0} \quad (\text{ecuación 18})$$

Los coeficientes b_0 hallados para cada componente son agrupados 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 calculada por medio de la (ecuación 19).

$$Y = TBQ^r + F \quad (\text{ecuación 19})$$

En resumen, PLS se utilizó para estudiar la estructura de covarianza entre los espacios correspondientes a dos matrices X y Y, para predecir un conjunto de variables dependientes a partir de un conjunto grande de variables independientes. Cada dirección de estos espacios es representada por una componente principal o variable latente, ya que la suposición básica de todos los modelos PLS, es que el sistema o proceso estudiado depende de un número pequeño de variables latentes. Así al finalizar los cálculos de PLS, se obtienen loadings y score (puntuaciones) para cada uno de los espacios X y Y.

El PLS, a diferencia de la regresión por componentes principales (PCR), las variables latentes son determinadas considerando conjuntamente X y Y. Para la regresión PLS, cada componente se obtiene minimizando el cuadro de la covarianza entre Y y las posibles combinaciones lineales de X. De este modo se obtiene la variable latente que contiene la información de la correlación entre las matrices X y Y, presentando una relación más directa que los primeros componentes principales concentran mayor información predictiva^[45, 46].

1.3.4 Factores a incluir en el modelo.

1.3.4.1 Detección de outliers.

Una de las ventajas de los métodos multivariantes sobre los tradicionales univariantes, es la capacidad que tiene de detectar la observación u observación inconsistente con el resto de los datos. En la etapa de establecimiento del modelo se puede utilizar información de la influencia de los objetos en el conjunto de calibración (leverage) y los residuales, tanto en la propiedad de interés como en la respuesta instrumental. La detección de los outliers en esta etapa es importante porque la inclusión de estas muestras discrepantes en el modelo degrada su capacidad predictiva.

1.3.4.2 Leverage.

Es una medida de la posición (o influencia) de una muestra en relación al modelo^[47]. Las muestras con un elevado valor de leverage están muy alejadas del centro del modelo, por lo que tendrán una influencia muy alta sobre el mismo. Este valor se calcula como:

$$h = \frac{1}{I} + t_i^r (T^r T)^{-1} t_i \quad (\text{ecuación 20})$$

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. Se propone diferentes niveles umbral, los más aceptados son dos o tres veces el leverage medio de calibración,

que es igual a $1 + A/I$, siendo A el número de componentes principales o factores utilizados en el modelo ^[48].

1.3.4.3 Estadístico T^2 de hotelling.

Fue propuesto originalmente por Hotelling y mide la variación de cada muestra dentro del modelo PCA. Se calcula como la suma de los cuadrados de los scores. El gráfico T^2 monitoriza la distancia de una nueva medida al valor de referencia en el espacio reducido de los factores PCA. Permite detectar si la variación incluida en los componentes principales considerados es más grandes que la que le correspondería si solo incluyeran variaciones aleatorias. La interpretación de este gráfico es la misma que cualquier gráfico univariante; las muestras fueron de control poseen un valor de T^2 superior al límite, y aparecen más allá de la línea de control ^[49].

1.3.4.4 Errores de calibración y predicción.

Se debe determinar el tamaño óptimo de modelo. Esta elección se basa en el cálculo de un error de predicción medio para modelos que incluyen cada vez más factores (1.2...A) y en el estudio de la evolución de este error de predicción medio. Si se dispone de un conjunto independiente de muestra, no utilizado en la calibración, se puede calcular la raíz cuadrada del error medio de predicción (root-mean-square error of prediction, RMSEP), para el conjunto de muestras que no han participado en la calibración:

$$RMSEP = \sqrt{\frac{\sum_{i=0}^i (C_i - c_i)^2}{I_p}} \quad (\text{ecuación 21})$$

Donde C_i y c_i representa la concentración determinada de forma independiente y la concentración predicha por el modelo respectivamente. Si no se dispone de un conjunto independiente se puede utilizar el método de validación cruzada o cross-validation, en la que sucesivamente se va dejando una parte de la muestra fuera del conjunto de calibración, se realiza el modelo con las muestras restantes y se predicen las muestras descartadas.

Este proceso permite obtener predicciones independientes sin renunciar al uso de toda la información (muestras) disponibles en el conjunto de calibración. En este caso se obtiene un error de predicción similar al RMSEP, que se denomina raíz cuadrada del error medio de validación cruzada (o root-mean-square error of cross-validation, RMSEV) ^[50,51].

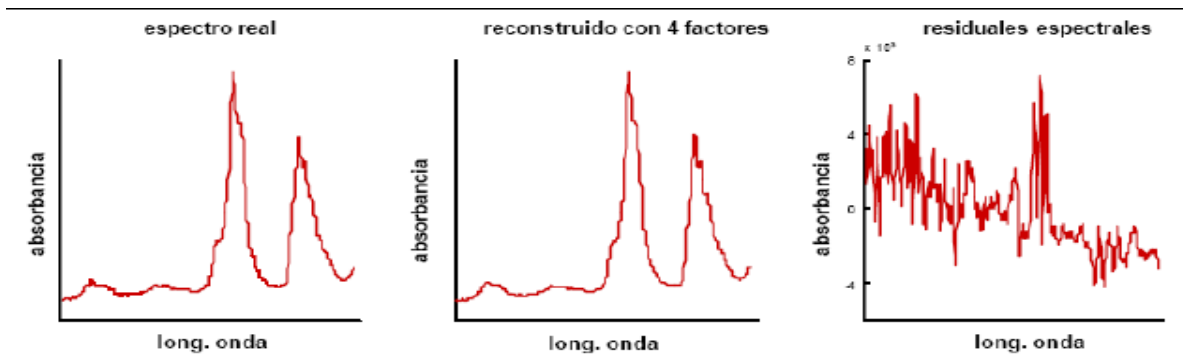
$$RMSEV = \sqrt{\frac{\sum (C_i - c_i)^2}{I}} \quad (\text{ecuación 22})$$

1.3.4.5 Residuales en la respuesta instrumental.

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

$$E = R - TP^T \quad (\text{ecuación 23})$$

Figura 9. Ejemplo del cálculo del residual de un espectro MIR. Al espectro original se le resta el espectro reconstruido con 4 factores para obtener el residual espectral.



Fuente: MACHO, A, Santiago. Metodologías analíticas basadas en espectroscopía de infrarrojo y calibración multivariante. Aplicación a la industria petroquímica.

Los residuales en la respuesta se pueden utilizar de varias formas. La más habitual es, para el error en la respuesta de la muestra i , e_i realizar una suma de cuadrados extendida a las j longitudes de onda y dividir por los grados de libertad (df) adecuados, para obtener una desviación estándar de la muestra i , $s(e_i)$.

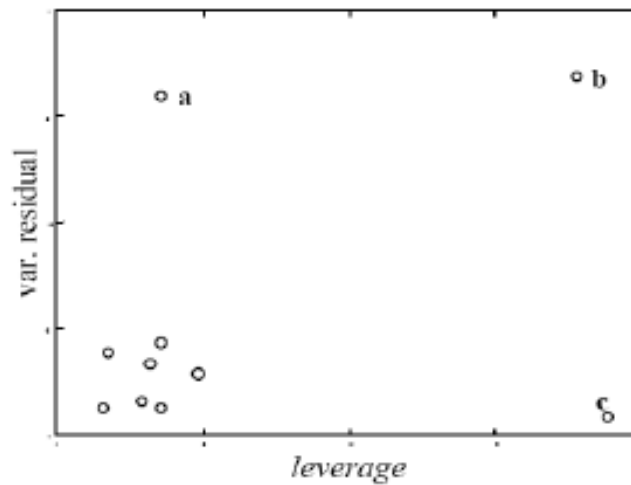
Residuales en la propiedad. En la etapa de establecimiento del modelo se dispone del valor de la propiedad de interés determinado por el método se dispone del valor de la propiedad de interés determinado por el método de referencia. Los residuales en la propiedad comparan el valor predicho por el modelo multivariable C_i con el valor considerado verdadero, C , que proporciona el método de referencia.

$$F = C - C_i \quad (\text{ecuación 24})$$

Muchas veces la detección de outliers se realiza combinando estas herramientas como en el gráfico que se representa el residual (espectral o de propiedad) frente al leverage de una muestra hipotética (figura 10).

Figura 10. Grafico del residual frente al leverage. (a) Objetos con varianza residual elevada se considera outliers, (b) si además tiene un leverage alto son

outliers peligrosos para el modelo, debido a que tiene muchas influencia sobre él. Las muestras con un alto leverage (c) son muestras influyentes y no necesariamente outliers.



Fuente: KALIVAS H. John practical guide to chemometrics, second edition

1.3.4.6 Validación del modelo.

Los métodos de calibración sesgados, como PCR o PLS, no se apoyan directamente en un modelo teórico y puede incorporar variabilidad de los datos no necesariamente relacionada con las propiedades de interés, por lo que deben ser cuantitativa o cualitativamente validados.

Las validaciones consisten en el análisis de un grupo de muestras independiente al utilizar en la calibración y comprueba que no existe un error sistemático (bias) entre las predicciones que realizan el modelo y los valores proporcionados por el método de referencia. También se mide el grado de concordancia entre las predicciones del método y los valores del método de referencia.

1.3.4.7 Predicción de muestra desconocidas.

Después de que el modelo ha sido aceptado, este se puede usar para el análisis de nuevas muestras. En esta etapa se deben seguir usando los test para detectar muestras discrepantes, outliers, con el fin de detectar la presencia de extrapolaciones al modelo, presencia de nuevas interferencias, fallos instrumentales etc. En este caso se puede utilizar medidas de leverage de las muestras, y del residual espectral herramienta para el control estadístico multivariante, como el estadístico T^2 de hotelling y el estadístico Q, que se puede utilizar también para la detección de outliers, ya que proporcionan una información similar al leverage (el T^2) y al residual espectral (el estadístico Q) los residuos de la propiedad de interés no están disponibles ya que estas muestras no han sido analizadas por el método de referencia. La detección de los outliers en esta etapa es muy importante porque la pérdida de esta muestra puede diferir significativamente del valor verdadero.

1.4 APLICACIÓN DE LA ESPECTROSCOPIA FTIR-ATR EN EL PETRÓLEO

La técnica de espectroscopía infrarroja con reflectancia total atenuada (FTIR-ATR) acoplada y no acoplada con modelos quimiométricos es una técnica que se ha implementado en gran parte del mundo para analizar muestras complejas debido a la efectividad que ha demostrado ^[52].

Los trabajos realizados usando espectroscopía infrarroja media en la muestra de petróleo, han demostrado que es la región del IR que presenta menor solapamiento de banda lo que genera mejor resultado. Entre las aplicaciones más sobresalientes de esta técnica se encuentran la determinación de la composición química de las fracciones analizadas, ya que esta varía dependiendo del método de obtención, por ejemplo; los espectros MIR de la fracción obtenida después de aplicar refinación por solventes, presenten bandas características de aromáticos, cadenas parafinicas y estructuras naftenicas e isoparafinas a diferencia de los espectros obtenidos después de aplicar craqueo, que generalmente presentan bandas características de asfáltenos y ausencia de bandas propias de aromáticos ^[7,53,54].

La aplicación de técnicas multivariadas, a la espectroscopia FTIR permitió la interpretación cuantitativa de las señales espectrales, prediciendo y determinando parámetros y propiedades fisicoquímicas de gran interés en las fracciones del petróleo. Análisis realizados en asfáltenos, resinas, gasolinas, queroseno, diésel, y petróleo crudo muestra el potencial que ha desarrollado la espectroscopia de FTIR para esta clase de analitos.

2. PARTE EXPERIMENTAL

2.1 MUESTRAS

Se analizaron 20 muestras de crudos con diferentes características que fueron provistos por el instituto colombiano de petróleo (ICP), se proporcionó el análisis SARA (tabla 4) el cual fue determinada por, precipitación de los asfaltenos este procedimiento se realizó según las normas ASTM D-4124.

Datos estándar de las 20 muestras de crudos utilizados como parámetros para la determinación de los modelos de predicción del análisis S.A.R.A. la cual fue determinada por, precipitación de los asfaltenos este procedimiento se realizó según las normas ASTM D-4124.

Tabla 4. Análisis S.A.R.A de los crudos.

| S.A.R.A | (%) peso saturados | (%) peso aromáticos | (%) peso resinas | (%) peso asfaltenos |
|----------|--------------------|---------------------|------------------|---------------------|
| crudo 1 | 17,61 | 25,16 | 15,69 | 2,69 |
| crudo 2 | 30,96 | 29,72 | 13,19 | 3,09 |
| crudo 3 | 31,09 | 29,18 | 12,33 | 3,55 |
| crudo 4 | 23,77 | 8,31 | 1,96 | 0,7 |
| crudo 5 | 22,48 | 8,13 | 2,24 | 0,38 |
| crudo 6 | 30,65 | 25,74 | 11,57 | 4,46 |
| crudo 7 | 30,98 | 30,41 | 16,34 | 2,13 |
| crudo 8 | 31,41 | 29,01 | 11,54 | 1,7 |
| crudo 9 | 35,37 | 33,96 | 7,9 | 10,9 |
| crudo 10 | 32,32 | 29,78 | 13,22 | 3,29 |
| crudo 11 | 32,04 | 28,1 | 13,08 | 3,75 |
| crudo 12 | 32,32 | 29,78 | 13,22 | 3,29 |
| crudo 13 | 28,49 | 9,38 | 1,97 | 0,29 |
| crudo 14 | 12,92 | 40,87 | 19,35 | 17,33 |
| crudo 15 | 31,09 | 29,18 | 12,33 | 3,55 |
| crudo 16 | 32,87 | 28,09 | 6,4 | 0,61 |
| crudo 17 | 23,77 | 8,31 | 1,96 | 0,7 |
| crudo 18 | 28,49 | 9,38 | 1,97 | 0,29 |
| crudo 19 | 31,41 | 29,01 | 11,54 | 1,7 |
| crudo 20 | 30,98 | 30,41 | 16,34 | 2,13 |
| promedio | 28.55 | 24.65 | 10.28 | 3.32 |
| máximo | 35.37 | 30.41 | 16.34 | 17.33 |
| mínimo | 12.92 | 8.13 | 1.96 | 0.29 |

2.2 Espectroscopia MIR-ATR.

2.2.1 Instrumentación.

Se empleó el espectrómetro IR Thermo Scientific Nicolet IS50 espectrómetro (figura 12a) 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}). El detector en el ATR es DLATGS.

ATR PIKE miracle (figura 11b) 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} .

Figura 11. Sistemas de caracterización MIR.



a) Espectrómetro



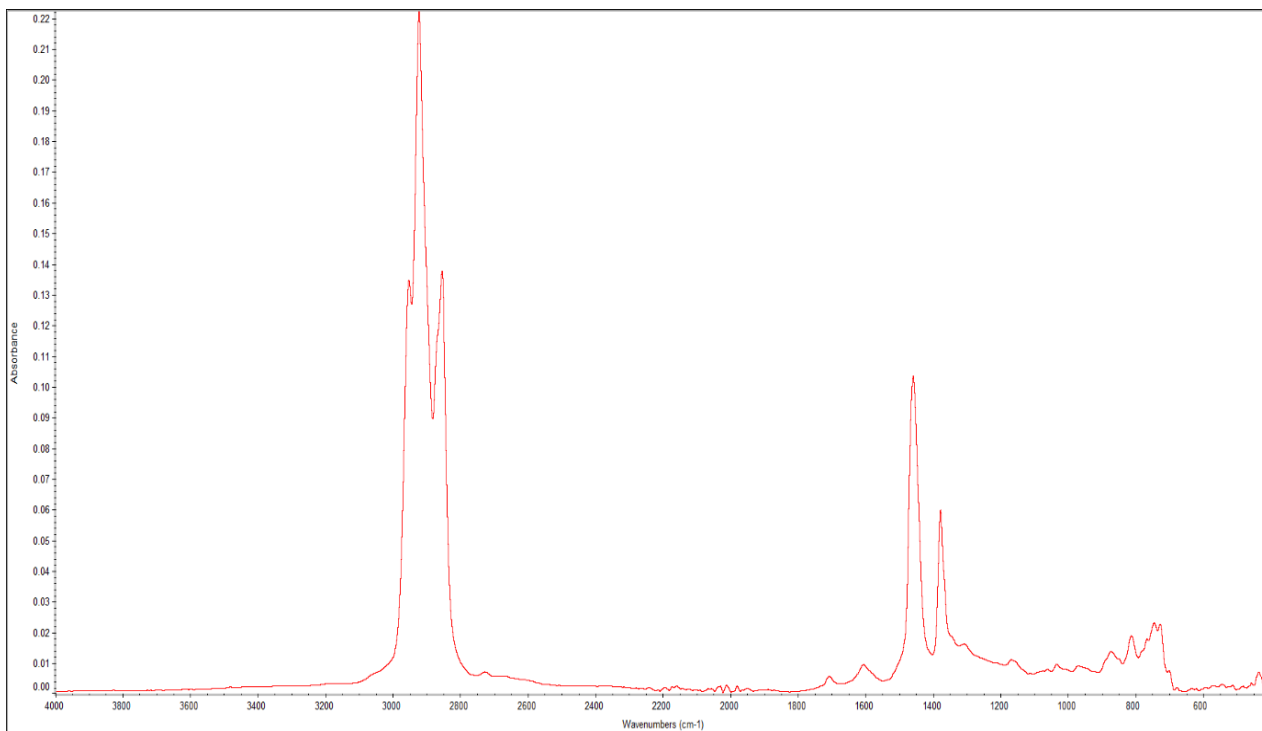
b) celda ATR

2.2.2. Verificación del desempeño del espectrómetro MIR.

Para medir el desempeño del equipo IR se realizó una prueba establecida dentro de los protocolos de funcionamiento del equipo. Para este caso se determinó la repetibilidad en las medidas de absorbancia.

La repetibilidad del espectrómetro se determinó a partir del cálculo de las desviaciones estándar para las absorbancias medidas a 2924.08 cm^{-1} y 1462.04 cm^{-1} en 9 espectros adquiridos sobre la muestra de crudo 5 perteneciente a la muestra de calibración. Las desviaciones halladas para cada pico de absorción fueron inferiores a 0.002 (figura 12), asegurando, de esta manera la repetibilidad de medición del espectrómetro.

Figura 12. Prueba de repetibilidad del espectrómetro MIR.



| espectros | 2924,56 cm^{-1} | 1462,76 cm^{-1} |
|---------------------|-------------------|-------------------|
| 5.1 | 0,2193737 | 0,09638 |
| 5.2 | 0,2254996 | 0,0925 |
| 5.3 | 0,2193064 | 0,09802 |
| 5.4 | 0,2218017 | 0,09854 |
| 5.5 | 0,202747 | 0,0939 |
| 5.6 | 0,2194744 | 0,09662 |
| 5.7 | 0,220055 | 0,09837 |
| 5.8 | 0,21905 | 0,0971819 |
| 5.9 | 0,2165558 | 0,0973661 |
| desviación estándar | 6,29E-03 | 0,00206 |

2.2.3 Adquisición del espectro MIR.

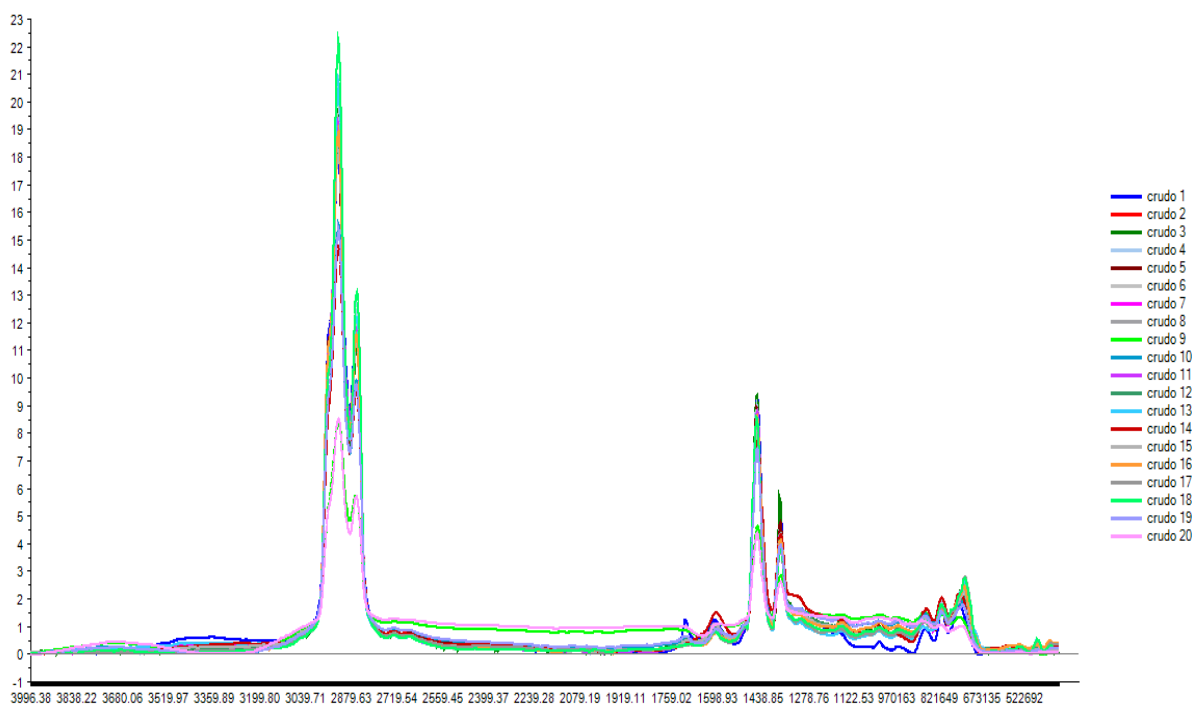
Los espectros MIR fueron tomados en el laboratorio de espectroscopia atómica molecular (LEAM) cada espectro fue el resultado de las siguientes condiciones: 32 barridos realizados en el rango de 4000 – 400 cm^{-1} , resolución de 4 cm^{-1} , ángulo de incidencia del haz IR 45°, velocidad del espejo interferómetro de 2.8mm/s a una temperatura de 25 C, estas condiciones mejoran la relación señal / ruido de los espectros.

2.2.3.1 Características de los espectros obtenidos

En la figura 13 se muestran los 20 espectros MIR de las muestras de crudos, obtenidos utilizando el espectrómetro IR Thermo Scientific Nicolet IS50 y analizadas en el software UNSCRAMBLER versión 10.3. Las bandas más intensas se encontraron cerca de 3000 cm^{-1} , la tensión de C-H en CH_3 a 2956 cm^{-1} , la tensión de C-H en CH_2 a 2926 cm^{-1} y 2853 cm^{-1} . También tiene una intensidad importante las bandas a 1495 cm^{-1} y 1385 cm^{-1} , que corresponden a torsiones de tijeras (simétricas asimétrica) del grupo CH_2 . A bajas frecuencias se encuentran otras bandas aunque de asignación más compleja como las bandas de deformación de C-H aromático en el plano y fuera de él.

La señal que se observa a 2349 cm^{-1} corresponde a una tensión asimétrica del CO_2 que proviene del medio por lo que esta señal no se toma en cuenta en los procesos multivariables.

Figura 13. Espectro original de 20 muestras de crudo.



2.3 PROCESAMIENTO DE DATOS

Los datos espectroscópicos obtenidos se analizaron en el software UNSCRAMBLER versión 10.3 donde inicialmente se les realizó un procesamiento

para disminuir la fuente de variabilidad de los espectros por efectos del ruido, movimiento de la línea base, diferencias de escala y otros.

Debido a que el software UNSCRAMBLER presenta diferentes opciones de normalización para el tratamiento de los datos como 6 tipos de normalización, segunda y primera derivada entre otras ^[55], como primera medida se realiza un análisis exploratorio de los datos espectrales determinando el valor de la desviación estándar media (RSD) a tres números de onda sobresalientes de los espectros después de aplicar los procesos opcionales como se muestra en la tabla 5.

Tabla 5. Cálculo de los RSD para tres números de onda característicos aplicando diferentes procedimientos.

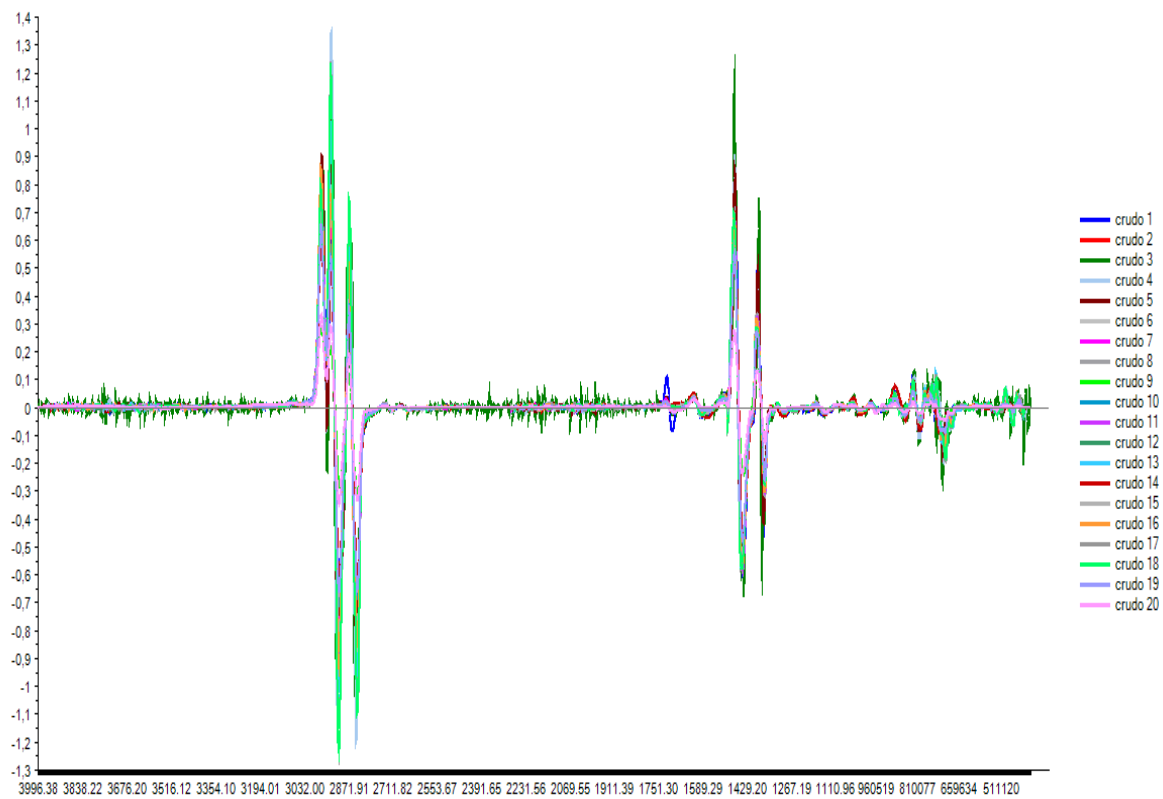
| Numero de onda (cm^{-1}) | 782.36 | 1462.76 | 2924.56 |
|---------------------------------|--------|---------|---------|
| Datos originales | 35.93 | 12.30 | 5.23 |
| Normalización por área | 35.72 | 8.23 | 2.35 |
| Normalización máxima | 35.49 | 9.57 | 12.30 |
| Normalización por rango | 40.18 | 10.36 | 23.10 |
| Normalización media | 37.25 | 7.74 | 9.23 |

Debido a que el ruido está asociado con la dispersión de los valores de RSD ^[56], se puede deducir que el tratamiento que hacen la mejor corrección de los espectros es el uso de la normalización por área seguida del suavizado Savitzky Golay y una derivación de primer orden (derivación Savitzky Golay figura 14), ya que esta combinación generó los menores valores de RSD para las bandas evaluadas (tabla 5).

La normalización por rango tiene como función hacer que los datos espectroscópicos de todas las muestras estén aproximadamente a la misma escala y consiste en dividir cada punto del espectro por su rango es decir en la resta de valores máximos menos el mínimo.

La derivación Savitzky Golay tiene como función diferenciar mejor picos solapados eliminar desplazamientos lineales y cuadráticos de la línea base, además de reducir el efecto de dispersión debido al tamaño de las partículas: consiste en derivar el promedio de cada tres puntos (para este caso) del espectro.

Figura 14. Espectro normalizado y derivado.



3. ANALISIS DE RESULTADOS

3.1 ANALISIS POR COMPONENTES (PCA)

3.1.1 Análisis por componentes principales para crudos livianos.

Es una técnica utilizada para reducir la dimensionalidad de un conjunto de datos. Intuitivamente la técnica sirve para hallar las causas de la variabilidad de un conjunto de datos y ordenarlas por importancia. Técnicamente, el PCA busca la proyección según la cual los datos queden mejor representados en términos de mínimos cuadrados.

El PCA construye una transformación lineal que escoge un nuevo sistema de coordenadas para el conjunto original de datos denominados componentes principales (PC) el cual explica la máxima variabilidad de las muestras; el segundo PC se escoge de forma que sea ortogonal al primero mostrando la máxima variabilidad de las muestras una vez restada la explicada por el primer PC.

Primero se realizó una descomposición en componentes principales del conjunto total de los 20 espectros de los crudos. Con el PCA se definieron los datos en los que los modelos de predicción se fundamentaron (números de componentes), la complejidad de estos y la forma de variarlos. (Tabla 6).

Tabla 6. Varianza explicada por los componentes principales de los 20 muestras de crudos.

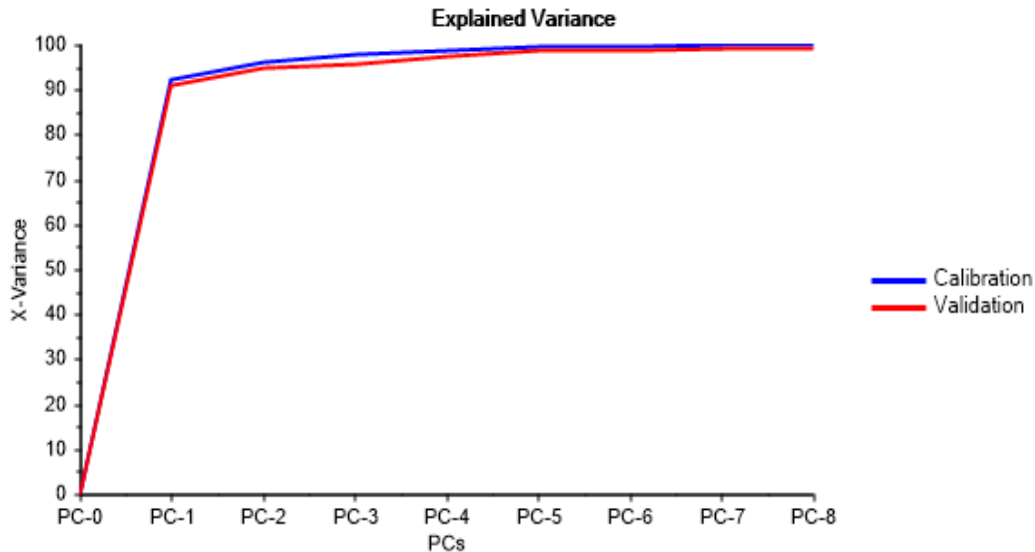
| (PCA) | Varianza explicada (%) | Varianza explicada acumulada (%) |
|-------|------------------------|----------------------------------|
| PC- 1 | 90,70 | 90,70 |
| PC- 2 | 4,56 | 95,27 |
| PC- 3 | 1,81 | 97,07 |
| PC- 4 | 1,34 | 98,42 |
| PC- 5 | 0,77 | 99,19 |
| PC- 6 | 0,31 | 99,50 |
| PC- 7 | 0,13 | 99,64 |
| PC-8 | 0,092 | 99,73 |

Observaciones el porcentaje de varianza de los datos espectrales que explica los primeros 8 componentes se determinó que 7 es un número máximo de componentes a utilizar en los modelos de predicción ya que a partir de PC-8 el número de la varianza explicada acumulada es menor a 0.092%.

El modelo de predicción explica en las regiones espectrales comprendidas entre $690-1075\text{cm}^{-1}$ y $2646-3321\text{cm}^{-1}$, más del 98% en la variabilidad de los datos

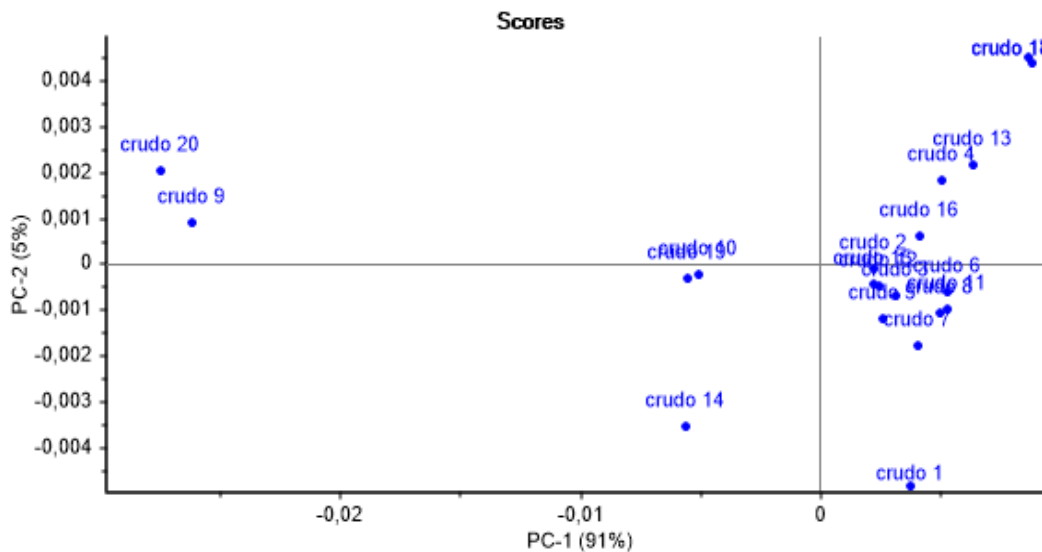
(variable Y) a partir del espectro MIR con cinco componentes principales. (Figura 15).

Figura 15. Varianza explicada en el modelo PLS de saturados.



La representación en dos dimensiones de los dos primeros componentes principales (PC1 y PC2) se muestra en la figura 16.

Figura 16. Gráfica de Scores en dos dimensiones para las 20 muestras.



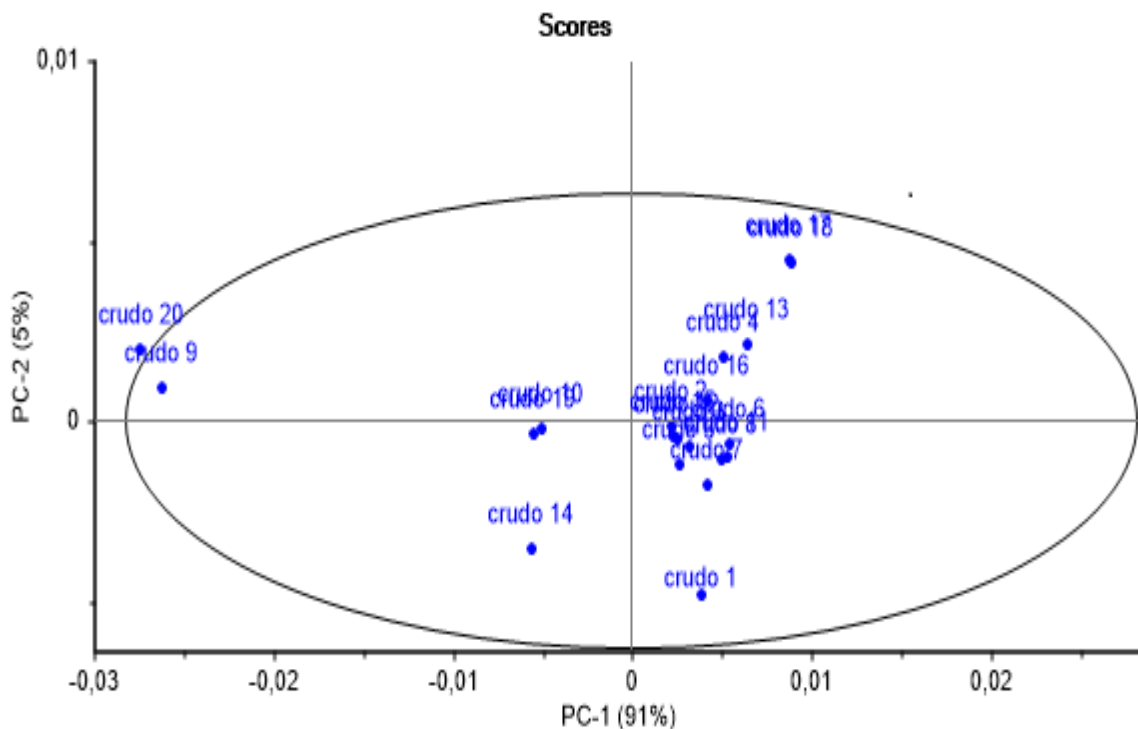
La figura de score. es un gráfico de dispersión en dos dimensiones de los componentes principales PC-1 y PC-2 en los cuales se recoge el 90.70% y el 4.56% de varianza respectivamente, lo que indica que esto dos componentes contienen el 96% de la varianza total de las intensidades espectrales además este permite establecer las relaciones entre las muestras según su cercanía en el plano

PC-1 y PC-2 ya que cuanto más cerca están las muestras al eje de coordenadas del gráfico, mas similares son con respecto a los dos componentes en cuestión.

Comparando las muestras cercanas con las propiedades disponibles (composición SARA), se determinó que la mayoría de las muestras con alta composición porcentual en peso de asfaltenos se agrupan en el tercer y cuarto cuadrante de la gráfica de scores. De acuerdo con lo anterior se pudo establecer que la muestra se puede dividir en dos grandes grupos; pesados y livianos según la composición de asfaltenos.

Mediante la aplicación de hotelling T^2 elipse, se determinó si las muestra 20 era un potencial outliers (puntos fuera de la elipse o muestras atípicas), debido a que éstas se ubicaron fuera de los dos grupos establecidos en el gráfico de scores (figura 17).

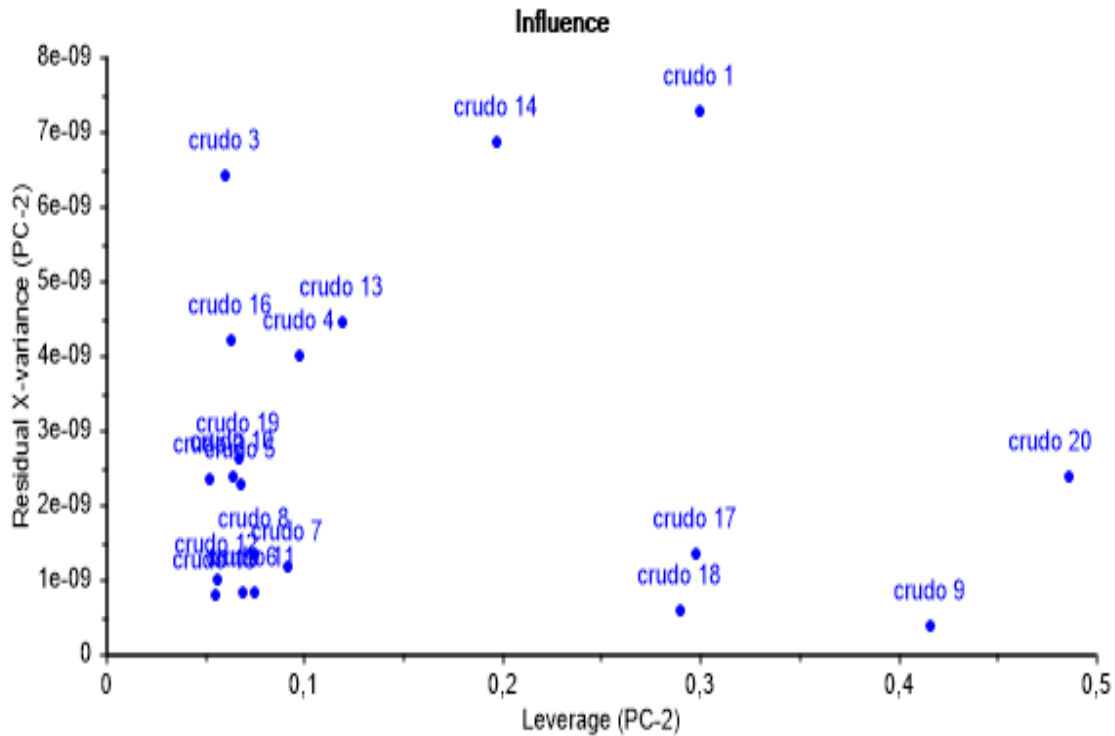
Figura 17. Estadístico T^2 aplicado ala grafica de score de las 20 muestras.



Las muestras 20 y 9 quedaron dentro del promedio la elipse, demostrando, que a pesar de estar alejadas de las demás, estas presentan similitud con algunas muestras en la composición de asfaltenos.

El gráfico de influence (figura 18) permite clasificar la forma más detallada los valores atípicos (outliers), influyentes y peligrosa (dangerous outliers) de las muestras analizadas.

Figura 18. Grafica de influencia para las 20 muestras.

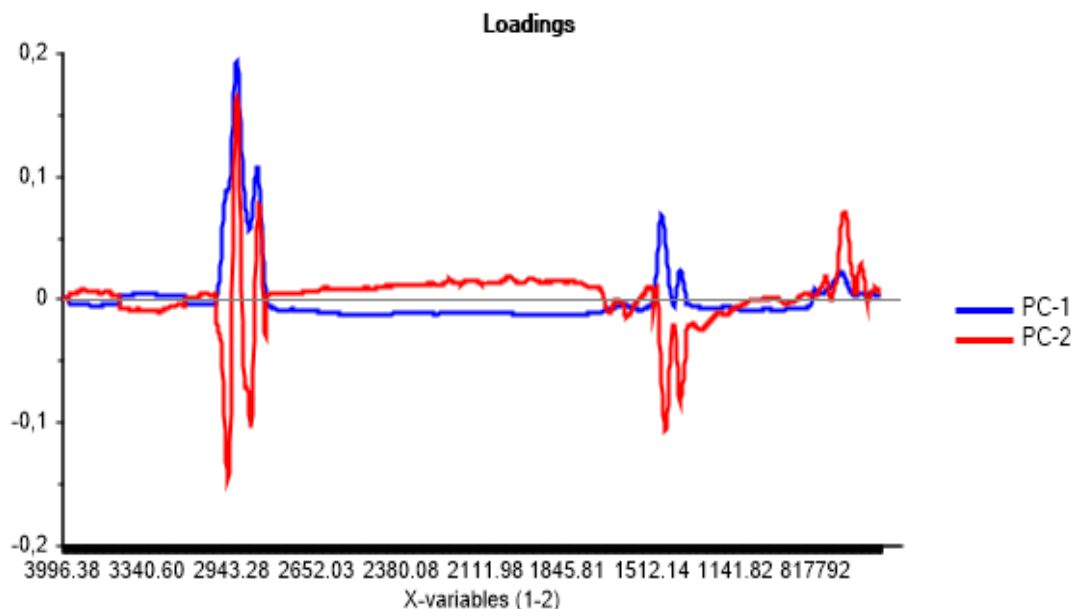


Como las muestras 3 presenta alta varianza residual (parte superior izquierda del gráfico) se definen como probables outliers. Las muestras que presentan alto leverage, es decir las que extienden a la derecha de la gráfica son influyentes, lo que significa que son las que mejor describen el modelo de predicción; para este caso se observó la muestra 20 y 9.

La muestra que representa alta varianza residual y alto leverage como la muestra 1 y 14 se definen como dangerous outliers, lo que significa que no serán descritas por el modelo que describe la mayoría de las muestras. Estas distorsionan el modelo con el fin de estar mejor descrita, entre estas muestras y las demás, en lugares de describir más características generales comunes a todas las muestras.

La grafica de X-loadings (figura 19) se utilizó para determinar los números de onda de mayor importancia los cuales se encuentran aproximadamente en los rangos de $690-1770\text{ cm}^{-1}$ y de $2646-3320\text{ cm}^{-1}$ del espectro IR las señales presentes en $648-690\text{ cm}^{-1}$ y en 2333 cm^{-1} no se tomaron en cuenta debido a que estas pertenecen a vibraciones del agua emulsionada presente en los crudos y de CO_2 del medio respectivamente.

Figura 19. Gráfico de X-loadings de PC1 y PC2.



La grafica X-loadign también permite visualizar que regiones espectrales aplican los vectores PC1 y PC2 en mayor proporción. Como se puede observar el PC1 es el presenta mayor intensidad en el rango de $1600-1300\text{ cm}^{-1}$ y de $3000-2800\text{ cm}^{-1}$ que corresponde a la región de vibraciones de saturados aromáticos y saturados respectivamente. El PC2 muestra mayor intensidad en el rango de $920-650\text{ cm}^{-1}$ que corresponde a la región de aromáticos. Esta tendencia se mantiene si se realiza el PCA sobre todo el grupo de muestras o sobre una parte de ellas, variando solamente las intensidades.

Conclusión del (PCA) El análisis de componentes principales permite identificar las agrupaciones naturales de las intensidades espectros de las muestras y el número de componentes principales óptimos a incluir en el modelo de predicción por contener la mayor información de las variables originales.

3.2 DESARROLLO DE MODELOS PLS

Mediante el algoritmo de regresión PLS, y empleando el programa de análisis multivariado THE UNSCRAMBLER versión 10.3, se evaluó el porcentaje de la espectroscopia MIR-ATR para predecir la composición de %w de la fracción SARA presente en crudos colombianos.

El desarrollo de los modelos de predicción abarcó dos grandes etapa: calibración y validación (interna y externa).

- Calibración y validación interna
- Validación externa

4. RESULTADO DE LOS MODELOS DE PREDICCIÓN

4.1. Modelo PLS para la predicción de saturados.

El rango en el que se desarrollaron los análisis fue de 35.55 – 12.92 %W de saturados.

La tabla 7 muestra el error estándar de calibración (RMSEC), de validación (RMSEP) y la varianza explicada para el modelo de predicción seleccionado con 5 variables latentes (VL) por mostrar el mejor desempeño en la calibración de acuerdo a los parámetros evaluados.

Tabla 7. Parámetros estadísticos del modelo desarrollado para la predicción %W de saturados.

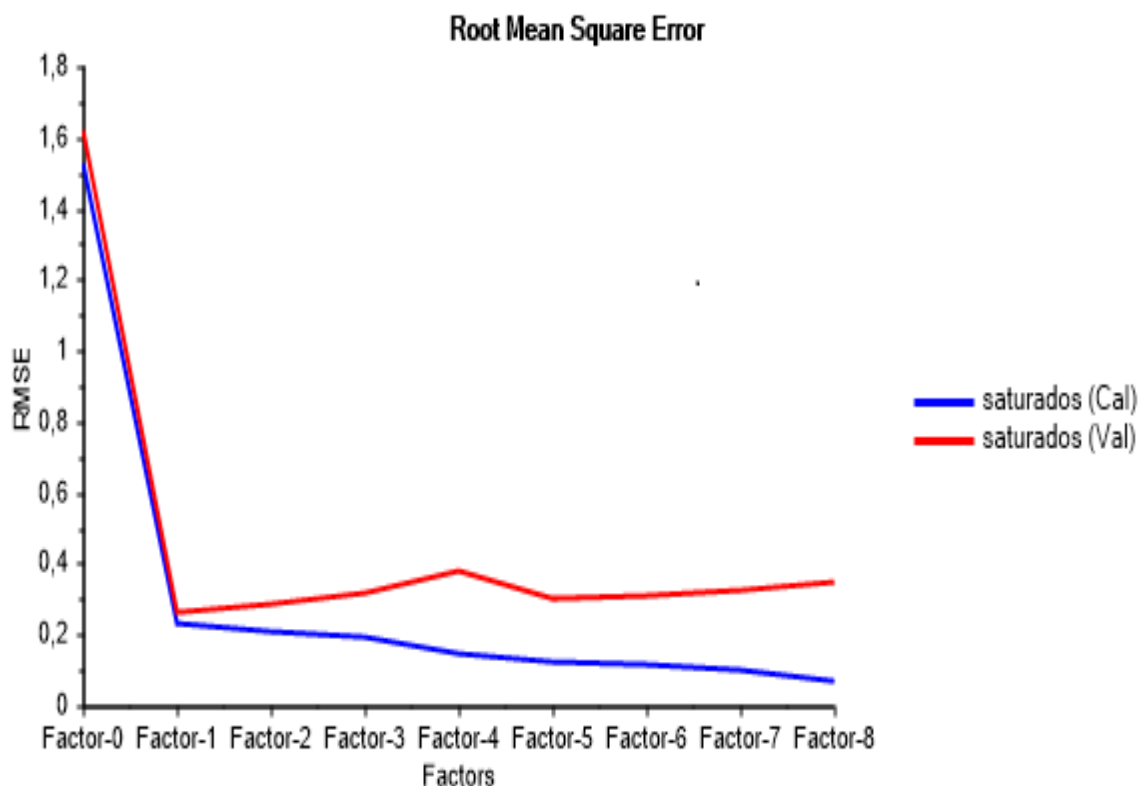
| Rango (cm-1) | muestras excluidas | Variables latentes (VL) | varianza explicada | RMSEC (%) | RMSEP (%) | R^2 |
|--------------|--------------------|-------------------------|--------------------|-----------|-----------|-------|
| 690 - 1075 | 1, 3, 14 | 5 | 98% | 0.12 | 0.30 | 0.99 |

Las muestras 1 y 14 fueron excluidas del modelo de predicción por presentar los menores porcentajes en la composición de saturados, 17.61 y 12.92 respectivamente, lo que indica que son muestras influenciadas (sección 3.1.1). La muestra 1 tiende a un comportamiento dangerous outliers (alto varianza residual y alto leverage) y las muestra 3 y 14 son potencial outliers ver figura 18 (gráfica de influencia).

La muestra 20 y 9 a pesar de presentar comportamiento outliers, ver la figura 17 (gráfica Estadístico T^2 aplicado a la gráfica de score) presentan alto leverage, es decir son influyentes, lo que significa que son las que mejor describen el modelo de predicción.

Otra gráfica que permite visualizar el software, con la que se comprobó el número óptimo de variables latentes (VL) para generar el modelo de predicción, es la de la raíz cuadrada del error de validación (RMSEP) y calibración (RMSEC), figura 20; ya que, el RMSEC es una medida de la desviación estándar de los residuales obtenidos por la diferencia entre los valores de referencia y predichos por el modelo para las muestras de calibración; y el RMSEP se basa en un algoritmo iterativo que selecciona muestra dentro del grupo de calibración para desarrollar el modelo de predicción y posteriormente evaluarlo sobre la muestra restantes.

Figura 20. Grafica de los errores RMSEC y RMSEP para el modelo PLS de saturados.

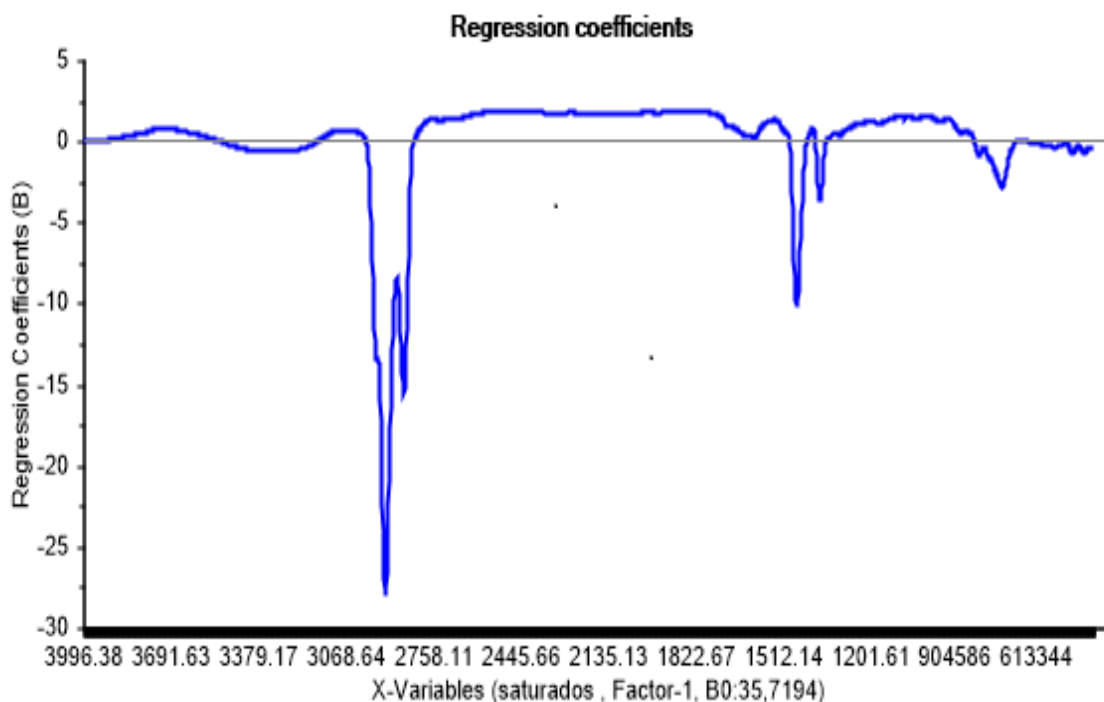


Como se observó en la figura 20, se logra comprobar que 1 es el número óptimo de PC usados en el modelo de predicción de la composición porcentual en peso de saturados, ya que para este caso a pesar que el RMSEC disminuye a medida que aumenta el número de variables latente, el RMSEP aumenta a partir del PC5, teniendo este último como valor numérico de 0.30.

Coefficiente de regresión

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 permitió visualizar los coeficientes hallados para cada componente utilizado. La figura 21, muestra los coeficientes que multiplican cada longitud de onda hallados para el primer componente ya que este explica la mayor variabilidad en los datos.

Figura 21. Coeficiente de regresión para el primer componente principal del modelo PLS de saturados.

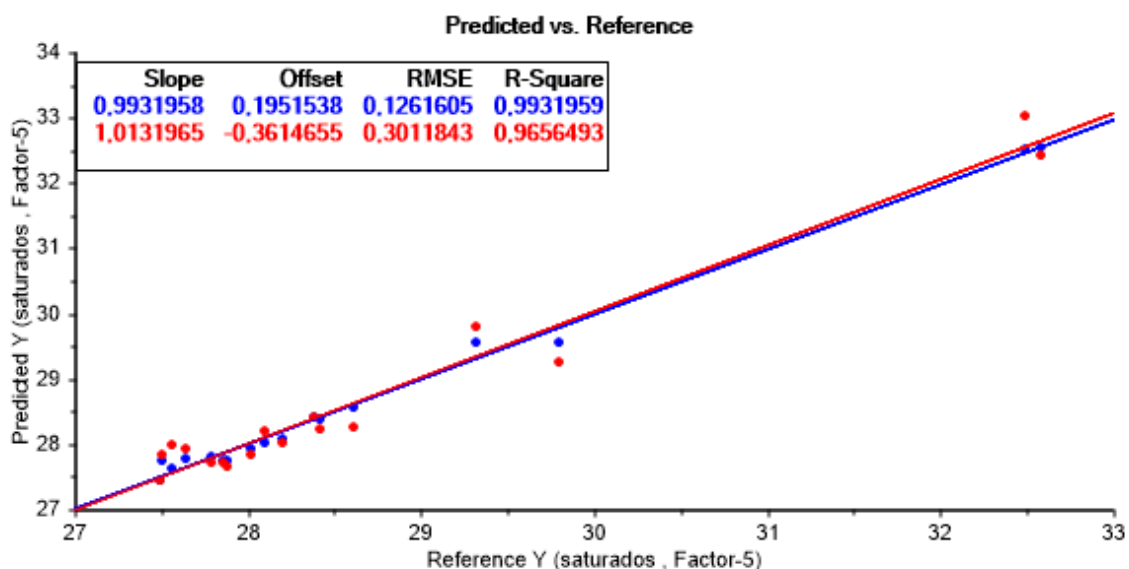


En esta gráfica se identificaron claramente dos regiones espectrales como las de mayor influencia y que concuerdan con las regiones de mayor absorbancia en los espectros MIR obtenidos para las muestras de crudo: la región de $690-1770\text{ cm}^{-1}$ (región 1) y la región de $2646-3321\text{ cm}^{-1}$ (región 2).

La validación cruzada del modelo PLS para saturados se realizó empleando el grupo de muestras de crudo utilizadas para generar el modelo de calibración sobre los datos completos de los espectros MIR pre- tratados se aplicó el modelo.

El coeficiente de correlación entre los valores de referencia y de predicción fue hallado superior al 0.99 para la curva de calibración y superior a 0.96 para la curva de validación (figura 21).

Figura 22. Curva de calibración y validación del modelo de predicción de %W de saturados.



La repetibilidad de la metodología propuesta en la predicción de saturados se determinó seleccionando del grupo de muestras de validación tres muestras al azar. Cada uno de los espectros fue adquirido bajo las mismas condiciones espectrales obtenidas en cada lectura se aplicó el modelo de PLS con cinco componentes principales para la predicción de %W de asfáltenos. La desviación estándar calculada en los valores predichos por el modelo fue inferior a 0.02 (Tabla 8), de esta manera se demostró que la repetibilidad en la predicción del contenido de %W de saturados a partir del espectro de MIR, fue satisfactoria.

Tabla 8. Prueba de repetibilidad del modelo de predicción de %W de saturados.

| toma | crudo 5 | Crudo 15 | Crudo 20 |
|---------------------|---------|----------|----------|
| 1 | 36,72 | 45,32 | 31,35 |
| 2 | 36,73 | 45,3 | 31,35 |
| 3 | 36,7 | 45,3 | 31,37 |
| promedio | 36,72 | 45,2 | 31,36 |
| desviación estándar | 0,02 | 0,02 | 0,01 |

La habilidad de predicción global del modelo se midió a través de los valores RMSEC, calculado para las muestras de calibración en la validación cruzada y el

RMSEP, calculados con las muestras no incluidas en el modelo. Ambos valores, 0.12% y 0.30% respectivamente, se encuentran por debajo del 2% de error. Demostró junto con los resultados de repetitividad obtenidos (tabla 8), que el modelo desarrollado presentará un desempeño satisfactorio en la predicción de la composición %W de saturados en muestra de crudo livianos.

4.2. Modelos PLS para predicción de aromáticos

El rango de calibración en el que desarrollaron los análisis fue de 30.41- 8.13 %W de aromáticos.

La (tabla 9) muestra el error estándar de calibración (RMSEC) de validación (RMSEP) y la varianza explicada para el modelo de predicción seleccionado de una serie de pruebas por mostrar el mejor desempeño en la etapa de calibración de acuerdo a los tres parámetros evaluados.

Tabla 9. Parámetros estadísticos del modelo generado para la predicción de %W de aromáticos en muestras livianas.

| rango (cm-1) | muestras excluidas | Variables latentes (VL) | varianza explicada | RMSEC (%) | RMSEP (%) | R^2 |
|--------------|--------------------|-------------------------|--------------------|-----------|-----------|-------|
| 690 – 1075 | | | | | | 0.99 |
| 2846-3005 | 1, 4 , 13 , 14 | 5 | 98% | 0.61 | 1.70 | |

Las muestras 4 y 13 fueron excluidas del modelo de predicción seleccionado, por presentar baja variabilidad en su composición de aromáticos entre ellas y con otras muestras, siendo estos valores de 8.31 y 9.38 respectivamente (Ver tabla 4).

Además se determinó que para el modelo de predicción de %w de aromáticos, la muestra 1 se comporta con dangerous outliers y la muestra 14 se comporta como potencial outliers aumenta el porcentaje de error de predicción del modelo ver figura 18 (grafica de influencia).

El coeficiente de correlación entre los valores de referencia y de predicción fue hallado superior al 0.99 para la curva de calibración (figura 23) y superior a 0.96 para la curva de validación (figura 24) para el modelo de aromáticos. El error estándar en la etapa de calibración y validación disminuye apreciablemente hasta el quinto componente, donde alcanza 0.61% y 1.70% respectivamente (figura 25). de los coeficientes de regresión calculados para el PC1, que explica la mayor variabilidad en los datos (91%), se determinó que la región $744-875\text{ cm}^{-1}$, $1365-1527\text{ cm}^{-1}$ y $2846-3005\text{ cm}^{-1}$ representa un efecto positivo para la predicción de aromáticos (figura 26) la variación del modelo mostro un desempeño favorable para la predicción de %W aromáticos, ya que coeficiente de corrección entre los valores de referencia y de predicción fue hallada superior al 0.99 para la curva de calibración y superior al 0.96 para la curva de validación (figura 27).

Figura 23. Curva de calibración del modelo de predicción de %W de aromáticos.

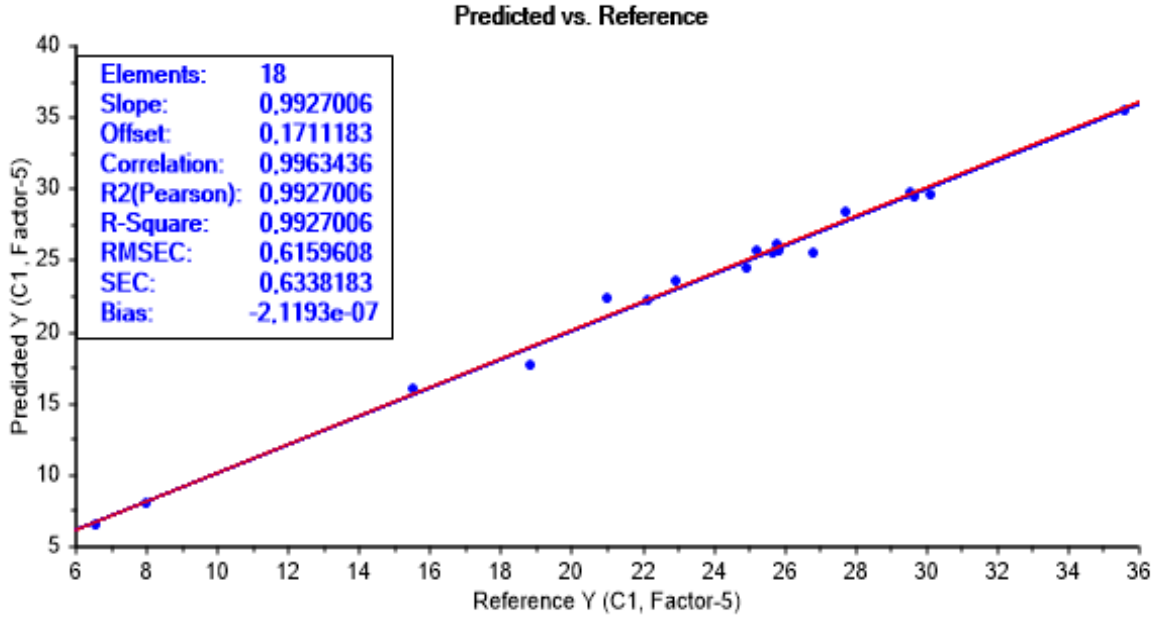


Figura 24. Curva de validación del modelo de predicción de %W de aromáticos.

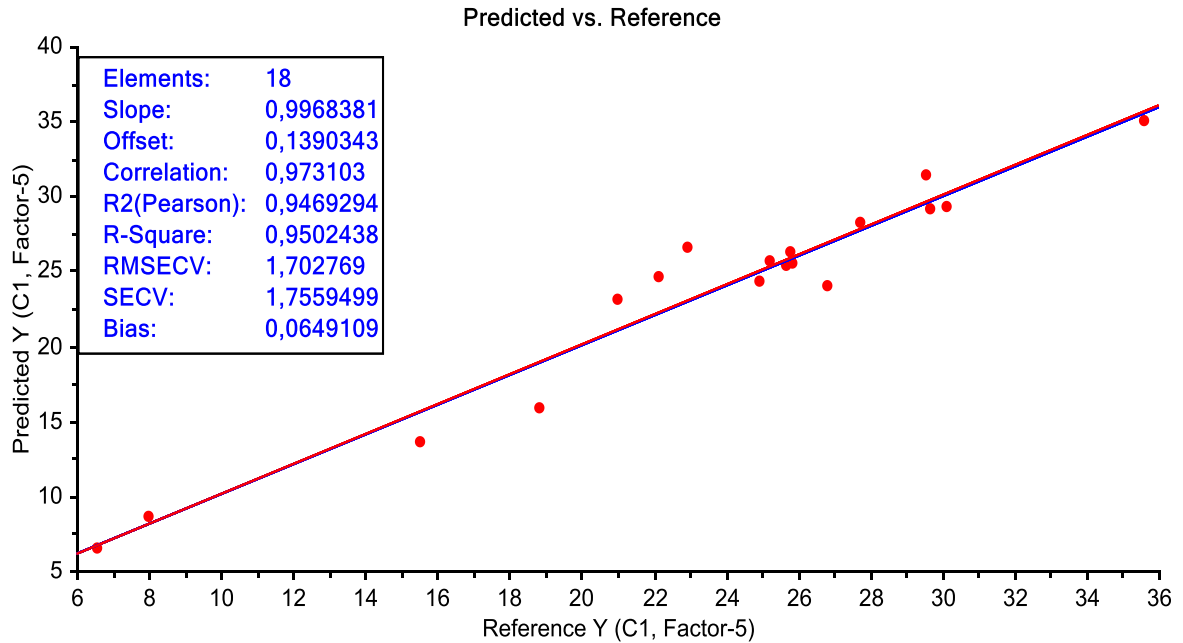


Figura 25. Gráfica de los errores RMSEC y RMSEP para el modelo PLS de aromáticos.

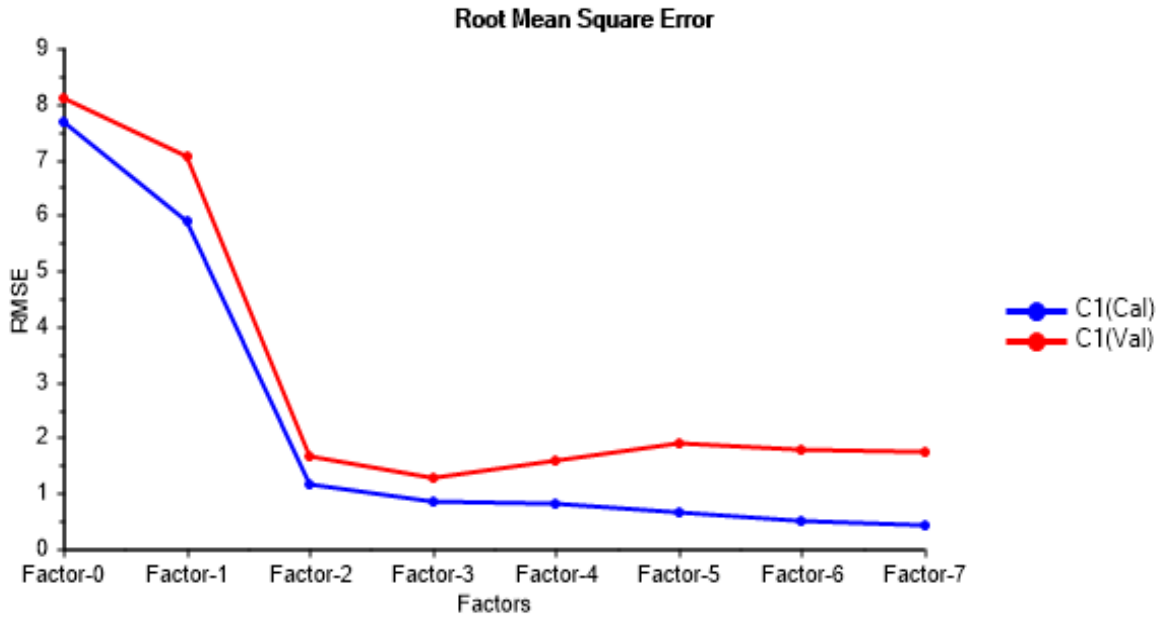


Figura 26. Coeficiente de regresión para el primer componente para el modelo PLS de aromáticos.

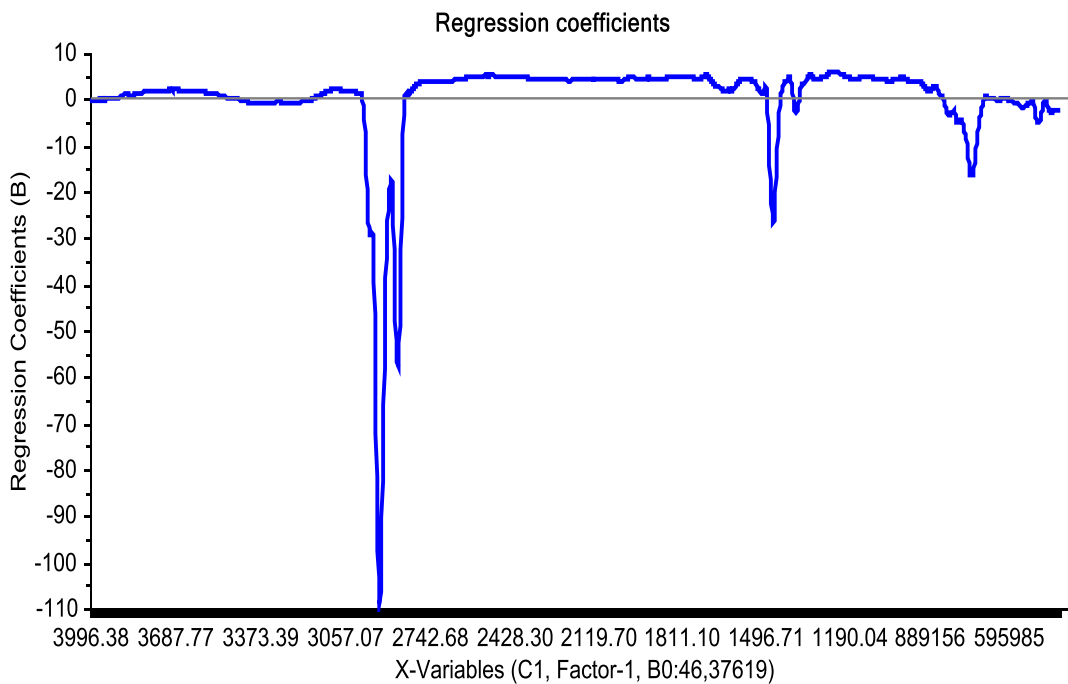
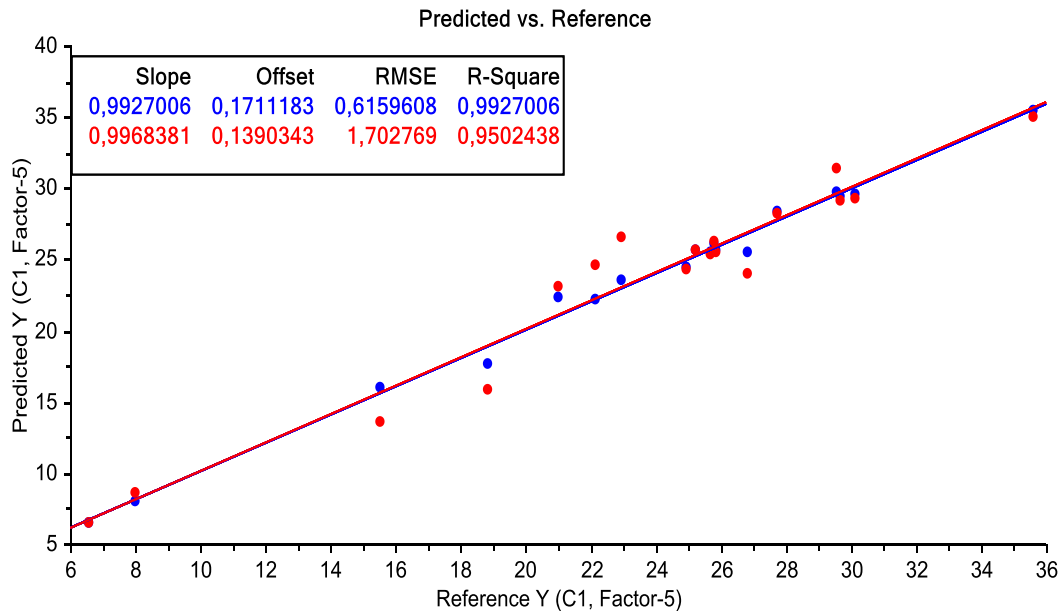


Figura 27. Curva de calibración y validación del modelo de predicción de %W de aromáticos.



La repetitividad de los resultados observados obtenidos (tabla 10), en el modelo, mostro una desviación estándar inferior a 0.03 de esta manera se determinó que la repetibilidad en la predicción del contenido de aromáticos a partir del espectro MIR, asegura una buena repetición en todas las muestras.

Tabla 10. Prueba de repetibilidad del modelo de predicción de %W de aromáticos.

| toma | crudo 2 | Crudo 8 | Crudo 17 |
|---------------------|---------|---------|----------|
| 1 | 28.23 | 31.82 | 27.06 |
| 2 | 28.24 | 31.85 | 27.00 |
| 3 | 28.20 | 31.87 | 27.02 |
| promedio | 28.22 | 31.85 | 27.03 |
| desviación estándar | 0,02 | 0,03 | 0,03 |

La habilidad de predicción global del modelo se mide a través de los valores RMSEC, calculado para las muestras de calibración en la validación cruzada y el RMSEP, calculado con las muestras no incluidas en el modelo. Ambos valores 0.61% y 1.70% respectivamente, se encuentran por debajo del 2% de error, demostrando junto con los resultados de repetibilidad obtenidos (tabla 10), que el

modelo desarrollado presenta un desempeño satisfactorio en la predicción de la composición %W de aromáticos en muestras de crudos livianos.

4.3. Modelo PLS para la predicción de resinas.

El rango de calibración en el que se desarrollaron los análisis fue de 16.34 – 1.96 %w de resinas.

La tabla 11 muestra el error estándar de calibración (RMSEC), de validación (RMSEP) y la varianza explicada para el modelo de predicción seleccionado de una serie de pruebas, por mostrar el mejor desempeño en la etapa de calibración de acuerdo a los tres parámetros evaluados.

Tabla 11. Parámetros estadísticos del modelo generado para la predicción de %W de resinas en muestras livianas

| rango (cm-1) | muestras excluidas | Variables latentes (VL) | varianza explicada | RMSEC (%) | RMSEP (%) | R^2 |
|-------------------------------------|--------------------|-------------------------|--------------------|-----------|-----------|-------|
| 690 – 875 1365-1527 2819-2981 | 1, 5, 4,14, | 3 | 97% | 0.23 | 0.68 | 0.99 |

La muestra 1 y 14 se excluidas del modelo de predicción seleccionado, por presentar comportamiento dangerous outliers y potencial outliers respectivamente, La muestra 4 y 5 se excluidas por presentar baja variabilidad en su composición de resinas, entre ellas y con otras muestras, siendo estos valores 1.96 y 2.26 respectivamente (ver tabla 4).

El coeficiente de correlación entre los valores de referencia y de predicción fue hallado superior al 0.99 para la curva de calibración (figura 28) y superior a 0.99 (figura 29) para la curva de validación para el modelo de resinas. El error estándar en la etapa de calibración alcanzo 0.23% y 0.68% respectivamente (figura 30). De los coeficientes de regresión calculados para el PC-1, que explican mayor variabilidad en los datos (91%) se denominó que la región $659-875\text{ cm}^{-1}$, $1365-1527\text{ cm}^{-1}$ y de $2819-2981\text{ cm}^{-1}$ presentan un efecto positivo para la predicción de las resinas (figura 31) ya que el coeficiente de correlación entre los valores de referencia y de predicción fue hallado superior al 0.99 para la curva de calibración y superior al 0.97 para la curva de validación (figura 32).

Figura 28. Curva de calibración del modelo de predicción de %W de resinas.

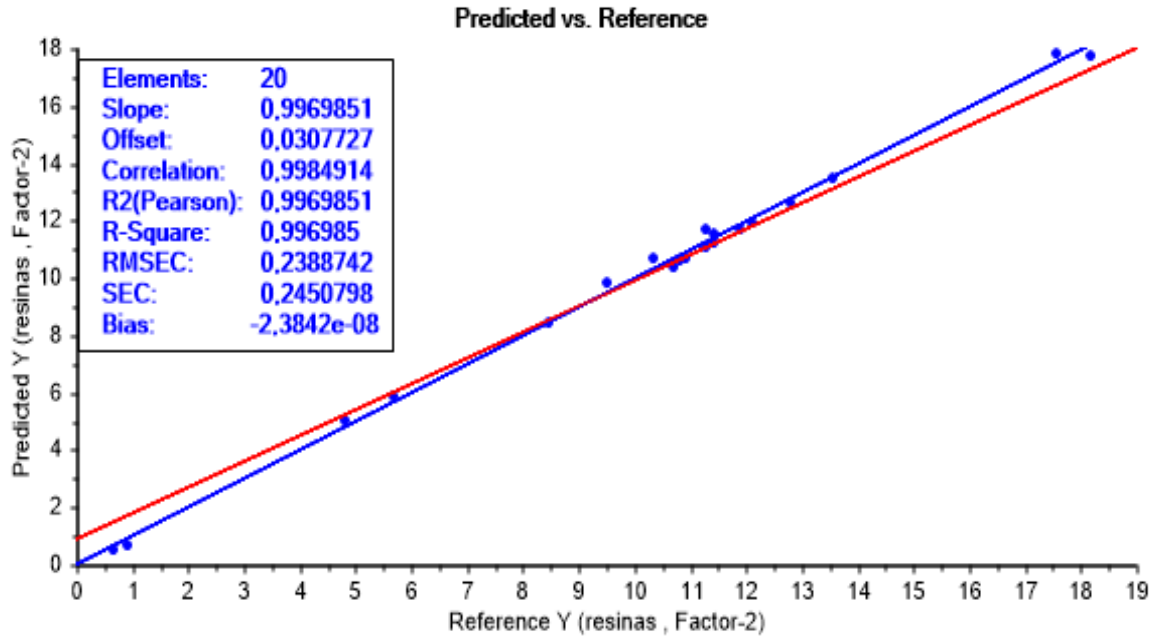


Figura 29. Curva de validación del modelo de predicción de %W de resinas.

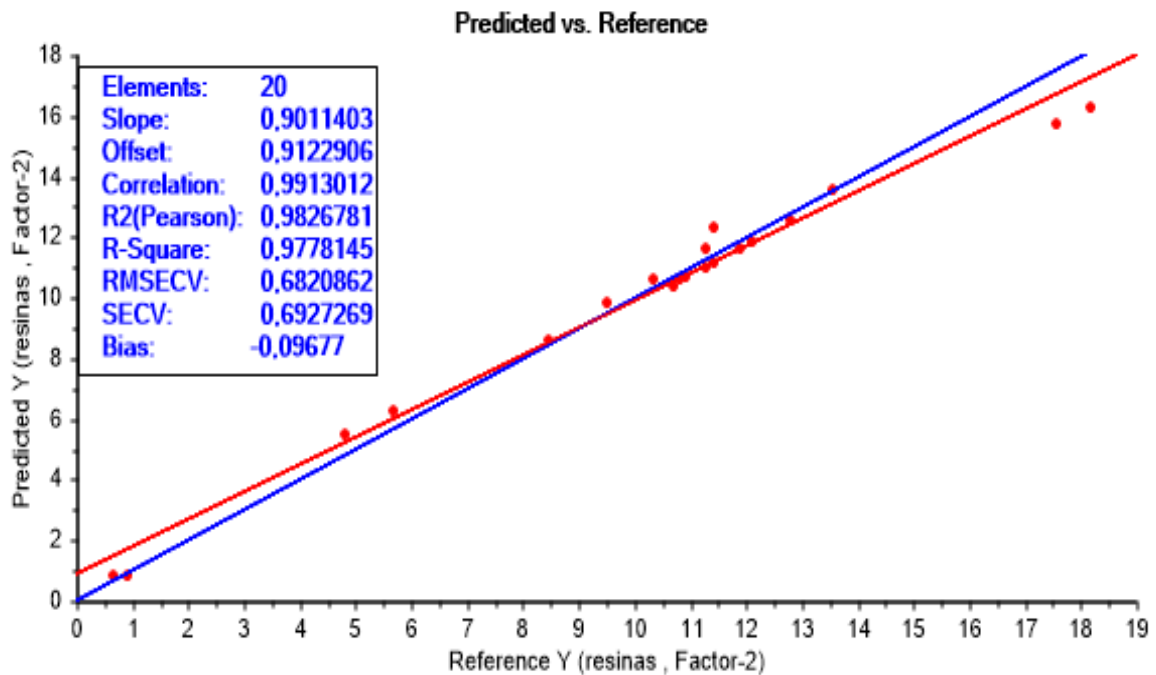


Figura 30. Gráfica de los errores RMSEC y RMSEP para el modelo PLS de resinas.

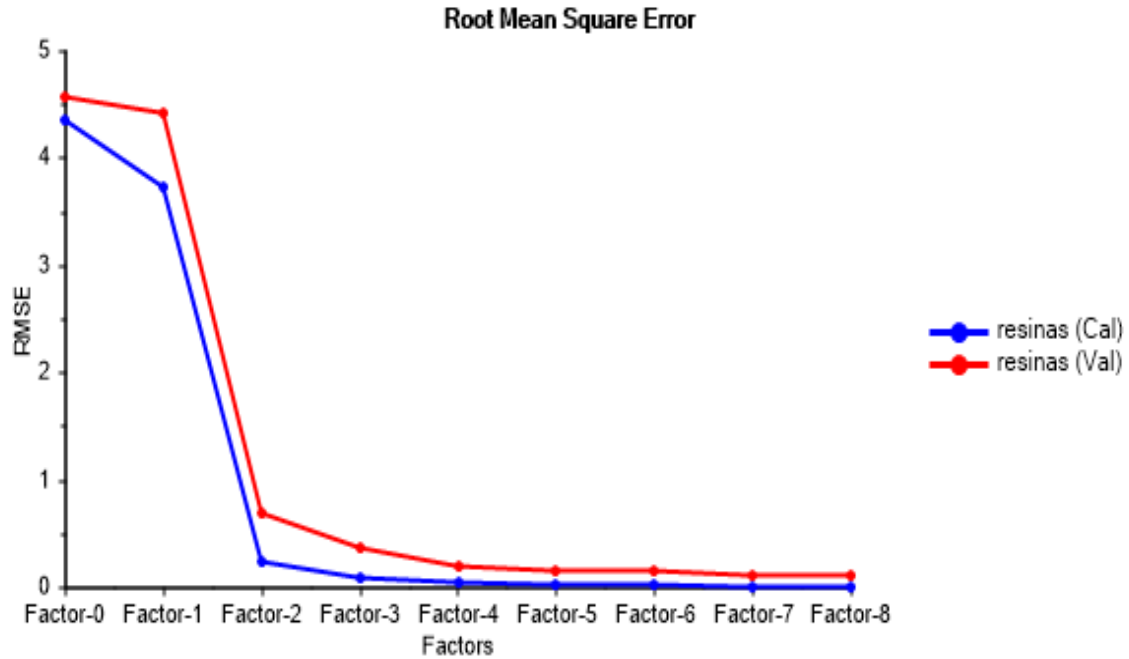


Figura 31. Coeficiente de regresión para el primer componente del modelo PLS de resinas.

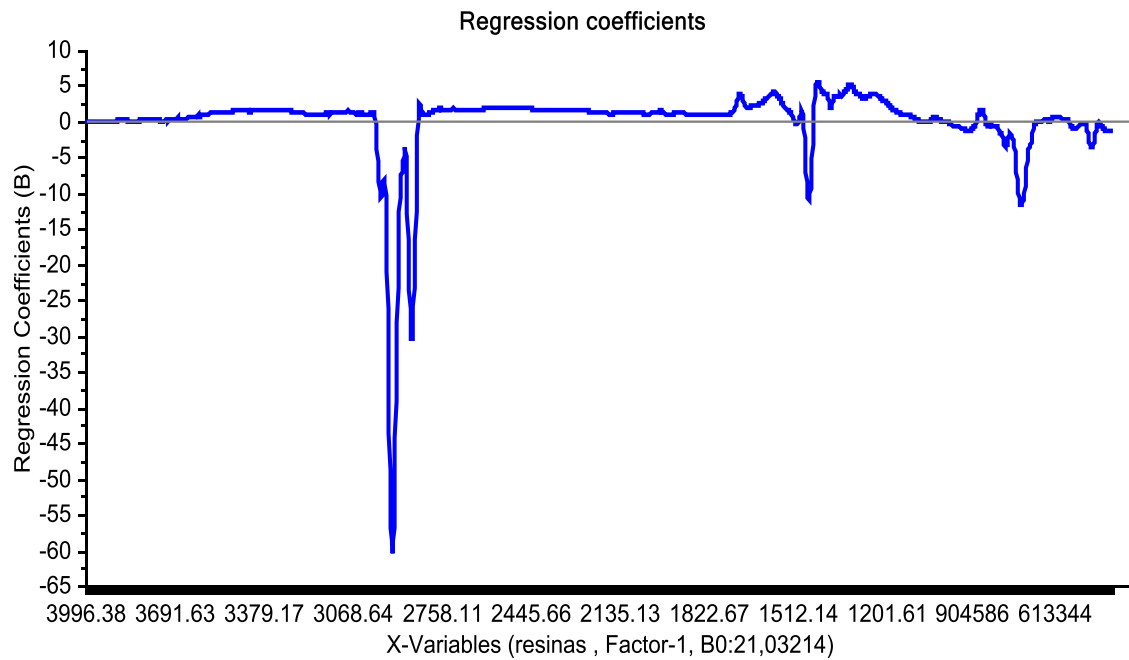
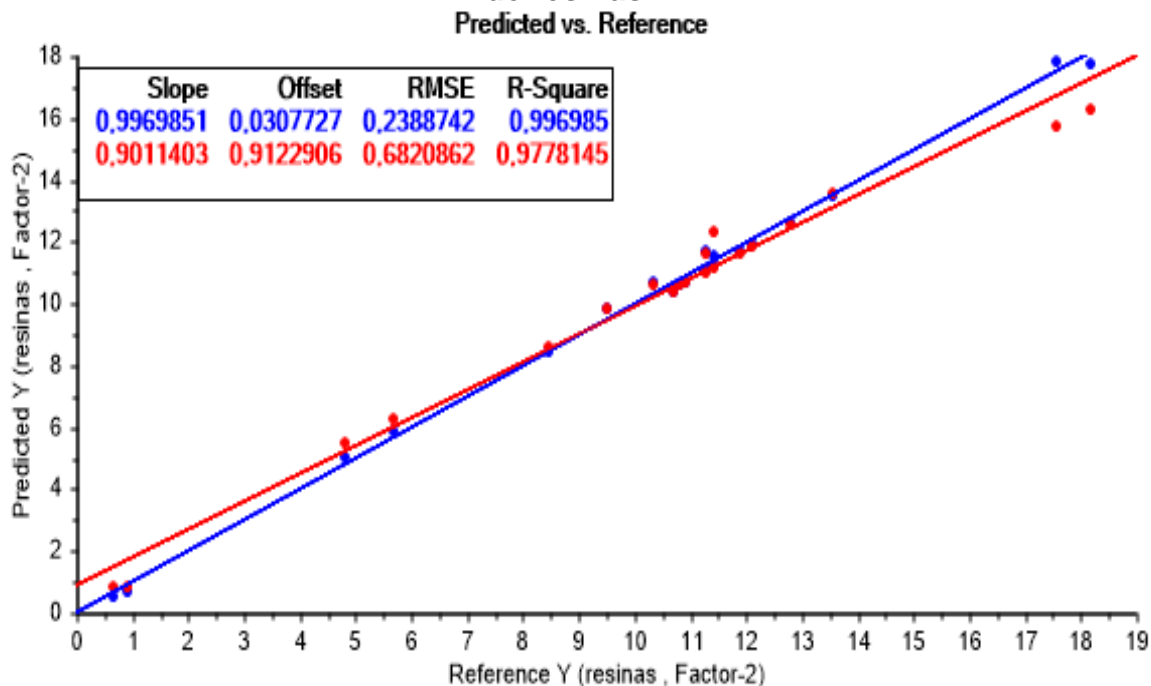


Figura 32. Curva de calibración y validación del modelo de predicción de %W de resinas.



La repetibilidad de los resultados obtenidos (tabla 12), en el modelo seleccionado con 2 componentes principales) para la predicción de la composición %W de resinas, mostro una desviación estándar inferior a 0.06 de esta manera se determinó que la repetibilidad en la predicción del contenido de resinas a partir del espectro IR asegura una buena repetición en todas las muestras.

Tabla 12. Prueba de repetibilidad del modelo de predicción de %W de resinas en muestras livianas.

| toma | crudo 2 | Crudo 8 | Crudo 17 |
|---------------------|---------|---------|----------|
| 1 | 27.00 | 26.31 | 11.86 |
| 2 | 27.05 | 26.40 | 11.83 |
| 3 | 27.03 | 26.29 | 11.78 |
| promedio | 27.03 | 26.33 | 11.82 |
| desviación estándar | 0,03 | 0,06 | 0,04 |

La habilidad de predicción global del modelo se mido a través de los valores RMSEC, calculado para las muestras de calibración en la validación cruzada y el RMSEP, calculado con las muestras no incluidas en el modelo. Ambos valores 0.23% y 0.68% respectivamente, se encuentran por debajo del 2% de error, demostrando junto con los resultados de repetibilidad obtenidos (tabla 12), que el modelo desarrollado presenta un desempeño satisfactorio en la predicción de la composición %W de resinas en muestras de crudos livianos.

4.4. Modelo PLS para predicción de asfáltenos

El rango de calibración en el que se desarrollaron los análisis fue de 17.36- 0.29% W de asfáltenos.

La tabla 13 muestra el error estándar de calibración (RMSEC), de validación (RMSEP) y la varianza explicada para el modelo desempeño en la etapa de calibración de acuerdo a los tres parámetros evaluados.

Tabla 13. Parámetros estadísticos del modelo generado para la predicción de %W de asfáltenos en muestras livianas.

| rango (cm-1) | muestras excluidas | Variabes latentes (VL) | varianza explicada | RMSEC (%) | RMSEV (%) | R ² |
|--------------|--------------------|------------------------|--------------------|-----------|-----------|----------------|
| 690 – 1072 | | | | | | 0.98 |
| 2303-3321 | 1,4, 5,13,14 | 6 | 98% | 0.48 | 1.53 | |

La muestra 4 y 5 fueron excluidas del modelo de predicción seleccionado, por presentar baja variabilidad en su composición de asfáltenos La muestra 13 fue excluida del modelo por presentar un valor de composición en asfáltenos muy alejado de los demás 0.38% (ver tabla 4), lo que indica que son muestras influenciada siendo este el valor límite del rango de calibración. La muestras 1 fue excluida del modelo de predicción por tener comportamiento dangerous outliers y 14 fueron excluida del modelo de predicción por ser potenciales outliers.

El coeficiente de correlación entre los valores de referencia y de predicción fue hallado superior al 0.99 para la curva de calibración (figura 33) y superior a 0.95 para la curva de validación (figura 34) para el modelo de asfaltenos. El error estándar en la etapa de calibración y validación cruzada disminuye apreciablemente hasta el cuarto componente donde alcanza los valores mínimos de 0.48 % y 153% respectivamente (figura 35). De los coeficientes de regresión calculados para el PC1, que explica la mayor variabilidad en los datos (91%), se determinó que las dos regiones del espectro usadas presentan un efecto positivo para la predicción de asfáltenos (figura 36). la validación cruzada del modelo mostro un desempeño favorable para la predicción de %W de asfáltenos, ya que el coeficiente de correlación entre los valores de referencia y de predicción fue hallado superior al 0.98 para la curva de calibración y superior a 0.90 para la curva de validación (figura 37).

Figura 33. Curva de calibración del modelo de predicción de %W de asfaltenos.

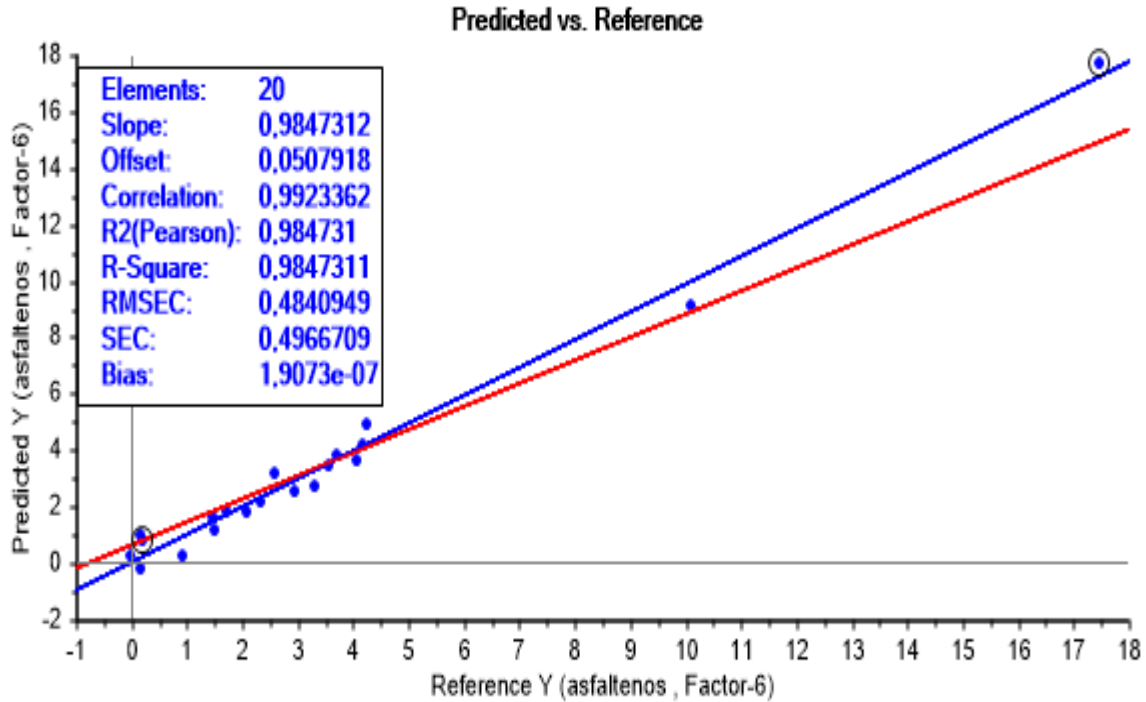


Figura 34. Curva de validación del modelo de predicción de %W de asfaltenos.

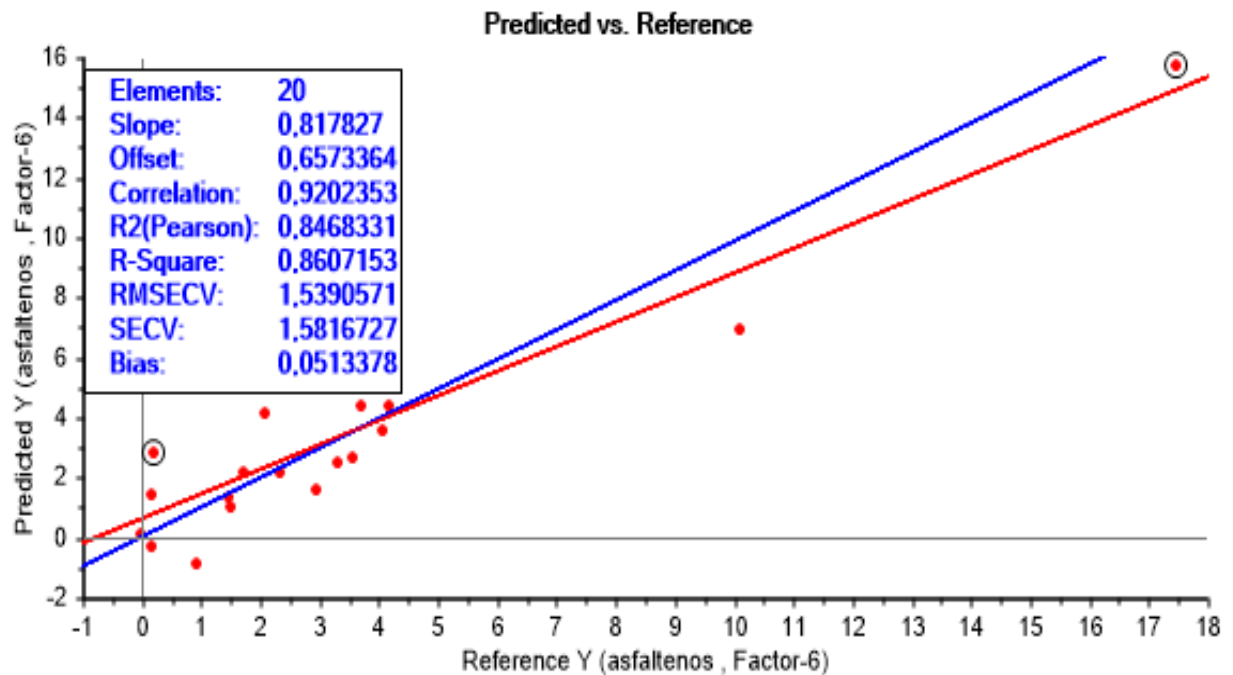


Figura 35. Gráfica de los errores RMSEC y RMSEP para el modelo PLS de asfaltenos.

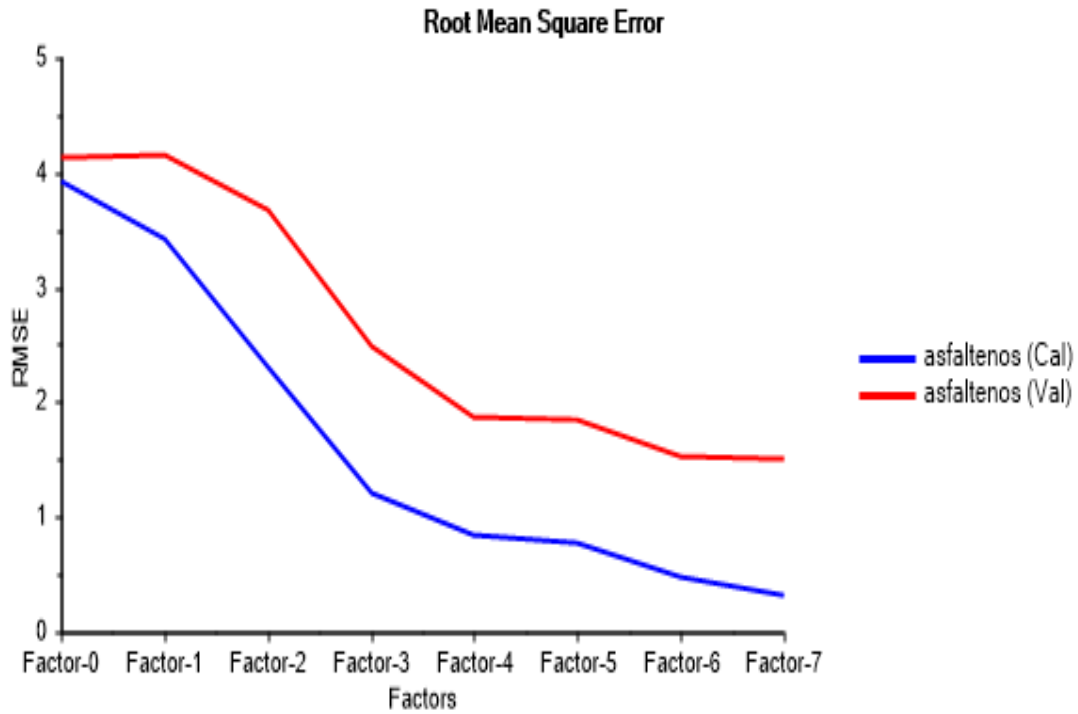


Figura 36. Coeficiente de regresión para el primer componente del modelo PLS asfaltenos.

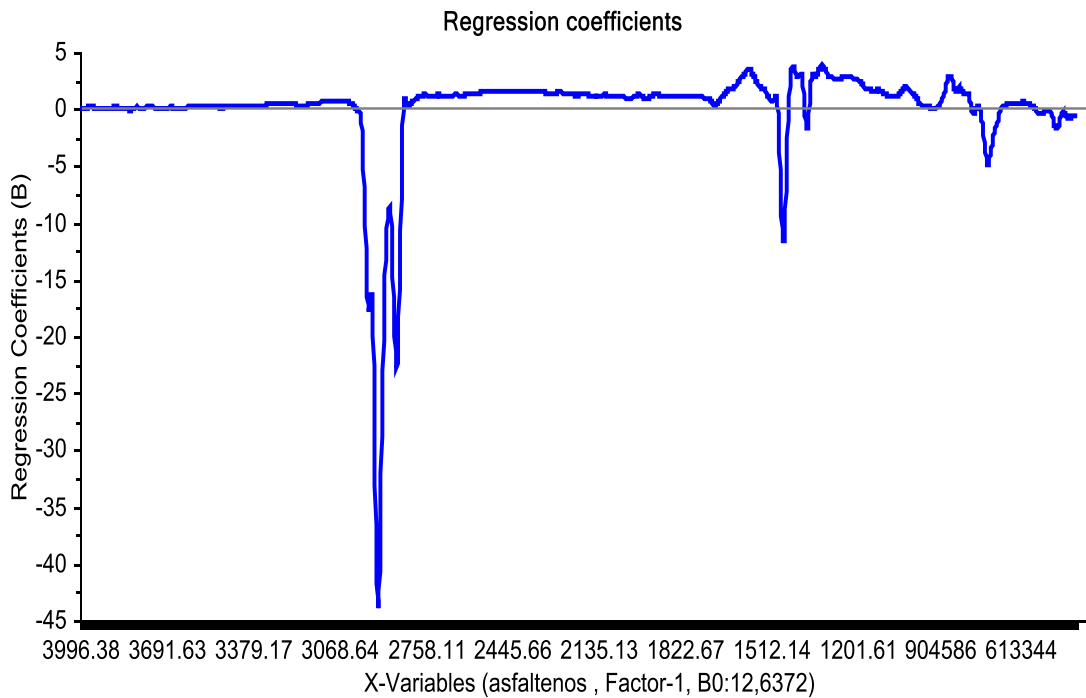
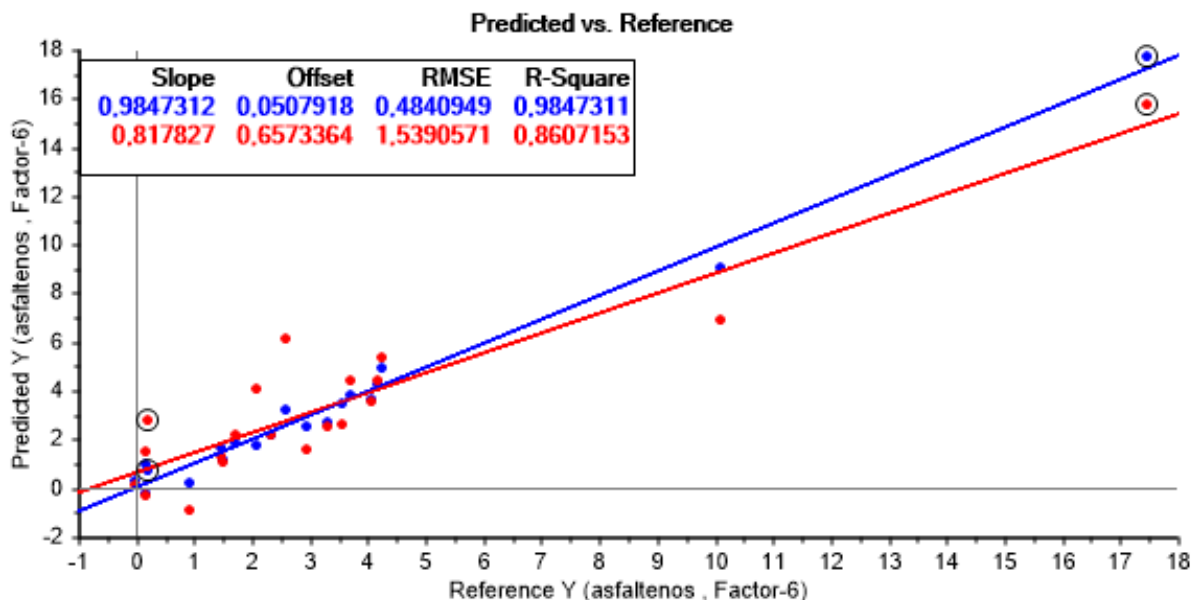


Figura 37. Curva de calibración y validación del modelo de predicción de %W de asfaltenos.



La respetibilidad de los resultados obtenidos (tabla 14), en el modelo seleccionado (con 6 componentes) para la predicción de la composición %W de asfaltenos.

Mostró una desviación estándar inferior a 0.05; de esta manera se determinó que la repetibilidad en la predicción del contenido de asfaltenos a partir del espectro MIR, asegura una buena predicción en todas las muestras.

Tabla 14. Prueba de repetibilidad del modelo de predicción de %W de asfaltenos en muestras livianas.

| toma | crudo 6 | Crudo 11 | Crudo 18 |
|---------------------|---------|----------|----------|
| 1 | 2.85 | 0.586 | 3.28 |
| 2 | 2.87 | 0.59 | 3.20 |
| 3 | 2.90 | 0.6 | 3.29 |
| promedio | 2.87 | 0.59 | 3.26 |
| desviación estándar | 0,03 | 0,8 | 0,05 |

La habilidad de predicción global del modelo se midió a través de los valores RMSEC, calculado para las muestras de calibración en la validación cruzada y el RMSEP, calculados con las muestras no incluidas en el modelo (validación extrema). Ambos valores, 0.48% y 1.53% respectivamente, se encuentran por debajo del 2% de error. Demostró junto con los resultados de repetibilidad

obtenidos (tabla 14), que el modelo desarrollado presentara un desempeño satisfactorio en la predicción de la composición %W de asfáltenos en muestra de crudo livianos.

5. CONCLUSIONES

La utilización de la espectroscopia MIR-ATR en combinación con técnicas de calibración multivariable es una técnica alternativa, a la cromatografía de exclusión molecular para la determinación del contenido de saturados, aromáticos y asfáltenos en muestras de crudos, con la ventaja de evitar el pretratamiento de la muestra, no utilizar solventes tóxicos para el medio ambiente y aumentar la reproducibilidad en los resultados.

El pretratamiento realizado a los datos con el software UNSCRAMBLER versión 10.3 presenta disminuciones considerables a la incertidumbre al someter a los espectros a normalización, suavizado y derivación.

Los análisis de componentes principales (PCA) permitieron identificar de manera cuantitativa el parentesco de los 20 espectros correspondientes a las 20 muestras de crudos y definir el número de componentes principales que mejora la incertidumbre en los datos.

El análisis por regresión parcial por mínimos cuadrados (PLS) permitieron identificar de manera cualitativa el parentesco de los 20 espectros correspondientes a las 20 muestras de los crudos para poder determinar la composición de saturados aromáticos asfáltenos y resinas, el modelo mostro un desempeño satisfactorio para la predicción composicional de las ya mencionadas propiedades.

En el análisis de los PCA y del PLS se realizó exclusión de muestras por mostrar comportamiento que afectaba la predicción correcta del modelo, algunas muestras fueron excluidas por presentar alta varianza residual, potenciales outliers otras por presentar alta varianza residual y al mismo tiempo alto leverage denominadas dangers outliers y otras por ensayo y error para mejorar la predicción del modelo.

Los análisis se realizaron en muestras con propiedades similares o de la misma naturaleza y manteniendo las mismas condiciones y parámetros.

CITAS BIBLIOGRÁFICAS

- [1] ZHAN, G; SHAOHUI, G; SUOQI, Z; GUANGXU, Y; LANQI, S and Chen, L Alkyl Side Chains Connected to Aromatic Units in Dagang Vacuum Residue and Its Supercritical Fluid Extraction and Fractions (SFEFs). *Energy and Fuels*. 2009, 23, 374-385.
- [2] ACEVEDO, S; CASTRO, A; NEGRIN, J. G; FERNANDEZ, A; ESCOBAR, G and PISCITELLI, V. Relations between Asphaltene Structures and Their Physical and Chemical Properties: The Rosary-Type Structure. *Energy and Fuels*. 2007, 21 2165-2175.
- [3] ASKE, N; KALLEVIK, H and SJOBLUM, J. Determination of saturate, aromatic, resin and asphaltenic (SARA) components in crude oils by means of infrared and Near-infrared spectroscopy. *Energy and Fuels*. 2001, 15, 1304-1312.
- [4] KHARRAT, A.M; ZACHARIA, J; CHERIAN, V.J and ANYATONWU, A. Issues with comparing SARA methodologies. *Energy and Fuels*, 2007, 21 (6), 3818-3621.
- [5] HONGFU, Y; XIAOLI, C; HAORAN, L and YUPENG, X. determination of multiproperties of residual oils using mid-infrared attenuated total reflection spectroscopy. *Fuel*, 2006 vol. 85. P 1720-1728.
- [6] BORGES, B; ACEVEDO, S. Caracterización estructural de distintas fracciones aisladas de crudos extrapesado Carabobo. *Revista latinoamericana de metalurgia y materiales*. 2007. Vol 2 P 83-94.
- [7] SASTRY, S; Chopra, A; Sarpal, S; Jain, K; Srivastava, P and bhatnagar K. determination of physicochemical properties and carbon-type analysis of base oils using Mid-IR spectroscopy and partial least-squares regression analysis *Energy & Fuels* 1998 Vol 12 p 304-311.
- [8] KALLEVIK, H. characterization of crude oil and model oil emulsions by means of near infrared spectroscopy and multivariable analysis.
- [9] Spiecker, P. M., Gawrys, K. L., Trail, C. B. & Kilpatrick, P. K. Effects of petroleum resins on asphaltene aggregation and water-in-oil emulsion formation. *Colloids Surfaces A Physicochem. Eng. Asp.* 220, 9–27 (2003).
- [10] Trejo, F.; Rana, M. S.; Ancheyta J. Thermogravimetric determination of coke from asphaltenes, resins and sediments and coking kinetics of heavy crude asphaltenes. *Catalysis Today* 150 (2010) 272-278.
- [11] Kharrat, A. M, Zacharia, J, Cherian, V. J. & Anyatonwu, A Issues with comparing SARA methodologies. *Energy and Fuels* 21, 3618-2185 (2008).

- [12] CONAWAY, C. The petroleum industry: A Nontechnical al guide. Tulsa; pennwell publishing Co. 1999.
- [13] WAUQUIER, J.P. El refino del petrolio. Petr leo crudo, producto petrol fero Esquema de fabricaci n. En: composici n de los petrolios crudos y de los productos petrol feros. Espa a: Editorial Diaz de santos 2004 p 1-15.
- [14] YEN. T.F. Petroleum Analysis. En: use of the data – the Structure of petroleum asphaltene and its significance. Energy sources. 2006 Vol 1 p 447- 463.
- [15] Hamadou, R.; Khodja, M.; Kartout, M.; Jada, A. Permeability reduction by asphaltenes and resins deposition in porous media. Fuel (2008) 87 2178-2185.
- [16] Gonzalez, E. B.; Groenzin, H.; Lira-Galeana, C.; Mullins, O. C. The Overriding Chemical Principles that Define Asphaltenes. Energy & Fuels 2001, 15, 972-978.
- [17] Schabron, J. F.; Rovani, J. F. On-column precipitation and re-dissolution of asphaltenes in petroleum residua. Fuel 87 (2008) 165-176.
- [18] Schneider, M.; Andrews, A. B. Mitra-Kirtley, S.; Mullins, O. C. Asphaltene Molecular Size by Fluorescence Correlation Spectroscopy. Energy & Fuels 2007, 21, 2875-2882.
- [19] Sheu, E. Y.; Mullins, O. Frequency-Dependent Conductivity of Utah Crude Oil Asphaltene and Deposit. Energy & Fuels 2004, 18, 1531-1534.
- [20] Andrews, A. B.; Guerra, R. E.; Mullins, O. C.; Sen, P. N. Diffusivity of Asphaltene Molecules by Fluorescence Correlation Spectroscopy. J. Phys. Chem. A 2006, 110, 8093-8097.
- [21] GONZALEZ, S.; Castillo, J.; Fern ndez, A.; Hung, J.; Absorbance and fluorescence spectroscopy on the aggregation behavior of asphaltene–toluene solutions. Fuel 83 (2004) 1823-1828.
- [22] Strausz, O. P.; Lown, S. E. M.; Izquierdo, M.; A Critique of Asphaltene Fluorescence Decay and Depolarization-Based Claims about Molecular Weight and Molecular Architecture. Energy & Fuels 2008, 22, 1156-1166.
- [23] Morales, Y. R.; Wu. X.; Mullins, O. C. Electronic Absorption Edge of Crude Oils and Asphaltenes Analyzed by Molecular Orbital Calculations with Optical Spectroscopy. Energy & Fuels 2007, 21, 944-952.
- [24] Mullins, O. C. Triplet Electronic Spin States of Crude Oils and Asphaltenes. Energy & Fuels 2011, 25, 2065-2075.

[25] Groenzin, H.; Mullins, O. C. Molecular Size and Structure of Asphaltenes from Various Sources. *Energy & Fuels* 2000, 14, 677-684.

[26] AMERICAN SOCIETY FOR TESTING AND MATERIALS. Standard Test for characteristic Groups in Rubber Extender and Processing Oils and Other Petroleum- derived Oils by the Clay-gel Absorption Chromatographic Method. ASTM D 2007-03 (2008).

[27] AMERICAN SOCIETY FOR TESTING AND MATERIALS. Standard Test Method for separation of Representative Aromatics and Nonaromatics Fractions of High-boiling Oils by Elution Chromatography. ASTM D2549-02 (2012)

[28] Melendez, L. V; Lache, a Orrego, J. A; pachon, Z; Mejia, E. prediction of the SARA aanalysis of colombian crude oils using ATR-FTIR spectroscopy and chemometric methods *Journal of Petroleum Science and Engineering* 90-91 (2012) 132-140.

[29] AMERICAN SOCIETY FOR TESTING AND MATERIALS. Standard Test Method for Separation of Asphalt into Four Fractions. ASTM D 4124-09

[30] SCHWEDT, G JOHN. *The essential guide to analytical chemistry*, Chichester: Wiley & Sons. 1997.

[31] OSBOME, B. G; FEAM, T; HINDLE; P. H. *Practical NIR spectroscopy with applications in food and beverage analysis*. And ed. England: longman Scientific & Technical 1993.

[32] HOLLAS, J. M. *Modern Spectroscopy*. And ed. Chichester, England: John wiley & Sons. 1992.

[33] PASQUINI, C. *Near infrared spectroscopy fundamentals, practical aspects and analytical applications*, En; J, *Chenm Soc.* 2003. Vol 14, N2. P. 198-219.

[34] SILVERSTEIN, R. M; and WEBSTER, F. X. *Spectrometric identification of Organic compounds*, 6^a Ed. New York: John Wiley & sons 1998.

[35] R UBINSON, K. A; RUBINSON, J. F. *Analisis intrumental*, Prentice Hall Hispanoamericana S.A. Madrid: Editotial Pearson Educacion, 2001.

[35] NORRIS, K.H. *Multivariate analysis of raw maerials, in chemistry and world food supplies*. Oxford: Shemilt ed. Pergamon press 1983.

[36] RODRIGUEZ-SAONA, L.E; KOCA, N; HARPER, W. J. *Rapid Determination of swiss Cheese Composition by Fourier Transform infrared/Attenuated Total reflectance Spectroscopy*. En; department of Food science and Technology 2005. Vol 89 N 5 p 1407-1412.

[37] ZHANG, Z; Ewing G.E. Attenuated partial internal reflection infrared spectroscopy. En: Analytical Chemistry. 2002 vol 74 N11 P 2578-2583.

[38] APARICIO M S. Metodologías analíticas basadas en espectroscopia de infrarrojo y calibración multivariante aplicación a la industria petroquímica Tarragona. Tesis doctoral, Universidad de Rovira y Virgili. Departamento de química analítica y química organica 2002, p 22-29.

[39] KELLER, R; MERMET; J; OTTO; M; Analytical Chemistry. New York John Wiley & Sons. 1998.

[40] MARTENS, H; MARTENS, M. Multivariate analysis of quality an introduction. Segunda edición; Editorial John Wiley & Sons 2001 P 445.

[41] AMERICAN SOCIETY FOR TESTING AND MATERIALS. Standard practices for infrared multivariate quantitative analysis. Philadelphia: ASTM, 2005. 29h (ASTM E 1655).

[42] STERINER, J; TERMONIA, Y; DELTOUR, J; Anal. Chemometrics. Vol 44. P. 1906, 1972.

[43] MILLER, J.N; MILLER, J.C. Estadística y Quimiometría para química analítica. Cuarta ed. España; Prentice Hall, 2002 P 595.

[44] VILCA, J.C. Generalizaciones de mínimos cuadrados parciales con aplicación en clasificación supervisada. Universidad de Puerto Rico. Mayagüez. 2004.

[45] FERNANDEZ, C.M. Quimiométrica. Editorial Publicaciones Universidad de Valencia, 2005 P 423.

[46] GERPERLINE, P. Practical guide to chemometrics. Segunda edición. Editorial Taylor y Francis Grupo. 2006. P. 541.

[47] EGAN, W. J; MORGAN, L.S. Anal. Chem. 1998. Vol 70. P 2372.

[48] ABBAS, O; REBUFA, C; DUPUY, N; PERMANYER, A. KISTER Assessing petroleum oils biodegradation by chemometric analysis of spectroscopic data 2008. Vol 12. P 1-15.

[49] EISENHART, M; HASTAY, W; WALLS, WA. Techniques of statistical analysis. En H Hotelling Multivariate Quality Control. New York Mc Graw- Hill, 1947.

[50] WOLD; S Technometrics. 1978. Vol 20 P 397.

[51] STONE, M.J Staist Soc. 1973. Vol 36. P 111.

[52] FREDERICKS Peter; Y. RINTOUL Llewellyn Vibrational Spectroscopy Instrumentation for infrared and Raman Spectroscopy Queensland University of Technology; Brisbane, Australia.

[53] WILT K Brian and WELCH T. William. Determination of Asphaltenes in Petroleum Crude oils by Fourier Transform infrared Spectroscopy. Energy & Fuels 1998, 12, 1008-1012.

[54] PEREIRA C: Rita SKROBOT L. Vinicius. CASTRO R Eustaquio. FORTES C P Isabel; and PASA M D Vanya. Determination of Gasoline Adulteration by Principal Components Analysis-Linear Discriminal Analysis Applied to FTIR Spectral Energy & Fuels 2006; 201097-1102.

[55] THE UNSCRAMBLER 10.3. METHODS. Software para diseño de experimentos y análisis multivariado.

[56] ORREGO; J.A. Estudio de la estructura de cinco carbones colombianos por Espectroscopia Fotoacustica en el infrarrojo medio. Tesis de maestria, maestro en Química, Bucaramanga Universidad Industrial de Santander facultad de ciencias Departamento de química. 2008, 38p.

BIBLIOGRAFÍA

ABBAS, O; REBUFA, C; DUPUY, N; PERMANYER, A. KISTER Assessing petroleum oils biodegradation by chemometric analysis of spectroscopic data 2008. Vol 12. P 1-15.

ACEVEDO, S; CASTRO, A; NEGRIN, J. G; FERNANDEZ, A; ESCOBAR, G and PISCITELLI, V. Relations between Asphaltene Structures and Their Physical and Chemical Properties: The Rosary-Type Structure. Energy and Fuels. 2007, 21 2165-2175.

AMERICAN SOCIETY FOR TESTING AND MATERIALS. Standard practices for infrared multivariate quantitative analysis. Philadelphia: ASTM, 2005. 29h (ASTM E 1655).

AMERICAN SOCIETY FOR TESTING AND MATERIALS. Standard Test for characteristic Groups in Rubber Extender and Processing Oils and Other Petroleum- derived Oils by the Clay-gel Absorption Chromatographic Method. ASTM D 2007-03 (2008).

AMERICAN SOCIETY FOR TESTING AND MATERIALS. Standard Test Method for separation of Representative Aromatics and Nonaromatics Fractions of High-boiling Oils by Elution Chromatography. ASTM D2549-02 (2012).

ANDREWS, A. B.; Guerra, R. E.; Mullins, O. C.; Sen, P. N. Diffusivity of Asphaltene Molecules by Fluorescence Correlation Spectroscopy. J. Phys. Chem. A 2006, 110, 8093-8097.

APARICIO M S. Metodologías analíticas basadas en espectroscopia de infrarrojo y calibración multivariante aplicación a la industria petroquímica Tarragona. Tesis doctoral, Universidad de Rovira y Virgili. Departamento de química analítica y química organica 2002, p 22-29.

ASKE, N; KALLEVIK, H and SJOBLOM, J. Determination of saturate, aromatic, resin and asphaltenic (SARA) components in crude oils by means of infrared and Near-infrared spectroscopy. Energy and Fuels. 2001, 15, 1304-1312.

BORGES, B; ACEVEDO, S. Caracterización estructural de distintas fracciones aisladas de crudos extrapesado Carabobo. Revista latinoamericana de metalurgia y materiales. 2007. Vol 2 P 83-94.

CONAWAY, C. The petroleum industry: A Nontechnical al guide. Tulsa; pennwell publishing Co. 1999.

- EGAN, W. J; MORGAN, L.S. Anal. Chem. 1998. Vol 70. P 2372.
- EISENHART, M; HASTAY, W; WALLS, WA. Techniques of statistical analysis. En H Hotelling Multivariate Quality Control. New York Mc Graw- Hill, 1947.
- FERNANDEZ, C.M. Quimiometrica. Editorial Publicaciones Universidad de Valencia, 2005 P 423.
- FREDERICKS Peter; Y. RINTOUL Llewellyn Vibrational Spectroscopy Instrumentation for infrared and Raman Spectroscopy Queensland University of Technology; Brisbane; Australia.
- GERPERLINE, P. Practical guide to chemometrics. Segunda edición. Editorial Taylor y Francis Grupo. 2006. P. 541.
- GONZALEZ, E. B.; Groenzin, H.; Lira-Galeana, C.; Mullins, O. C. The Overriding Chemical Principles that Define Asphaltenes. Energy & Fuels 2001, 15, 972-978.
- GONZALES, S.; Castillo, J.; Fernández, A.; Hung, J.; Absorbance and fluorescence spectroscopy on the aggregation behavior of asphaltene-toluene solutions. Fuel 83 (2004) 1823-1828.
- GROENZIN, H.; Mullins, O. C. Molecular Size and Structure of Asphaltenes from Various Sources. Energy & Fuels 2000, 14, 677-684.
- HAMADOU, R.; Khodja, M.; Kartout, M.; Jada, A. Permeability reduction by asphaltenes and resins deposition in porous media. Fuel (2008) 87 2178-2185.
- HOLLAS, J. M. Modern Spectroscopy. And ed. Chichester, England: John Wiley & Sons. 1992.
- HONGFU, Y; XIAOLI, C; HAORAN, L and YUPENG, X. determination of multiproperties of residual oils using mid-infrared attenuated total reflection spectroscopy. Fuel, 2006 vol. 85. P 1720-1728.
- KALLEVIK, H. characterization of crude oil and model oil emulsions by means of near infrared spectroscopy and multivariable analysis.
- KHARRAT, A.M; ZACHARIA, J; CHERIAN, V.J and ANYATONWU, A. Issues with comparing SARA methodologies. Energy and Fuels, 2007, 21 (6), 3818-3621.
- KHARRAT, A.M; ZACHARIA, J; CHERIAN, V.J and ANYATONWU, A Issues with comparing SARA methodologies. Energy and Fuels 21, 3618-2185 (2008).
- KELLER, R; MERMET; J; OTTO; M; Analytical Chemistry. New York John Wiley & Sons. 1998.

MARTENS, H; MARTENS, M. Multivariate analysis of quality an introduction. Segunda edición; Editorial John Wiley & Sons 2001 P 445.

MELLENDEZ, L. V; Lache, a Orrego, J. A; pachon, Z; Mejia, E. prediction of the SARA aanalysis of colombian crude oils using ATR-FTIR spectroscopy and chemometric methods Journal of Petroleum Science and Engineering 90-91 (2012) 132-140.

MILLER, J.N; MILLER, J.C. Estadística y Quimiometria para química analítica. Cuarta ed. España; Prentice Hall, 2002 P 595.

MORALES, Y. R.; Wu. X.; Mullins, O. C. Electronic Absorption Edge of Crude Oils and Asphaltenes Analyzed by Molecular Orbital Calculations with Optical Spectroscopy. Energy & Fuels 2007, 21, 944-952.

MULLINS, O. C. Triplet Electronic Spin States of Crude Oils and Asphaltenes. Energy & Fuels 2011, 25, 2065-2075.

NORRIS, K.H. Multivariate analysis of raw maerials, in chemistry and world food supplies. Oxford: Shemilt ed. Pergamon press 1983.

ORREGO; J.A. Estudio de la estructura de cinco carbonos colombianos por Espectroscopia Fotoacustica en el infrarrojo medio. Tesis de maestria, maestro en Quimica, Bucaramnga Universidad Industrial de Santander facultad de ciencias Departamento de química. 2008, 38p.

OSBOME, B. G; FEAM, T; HINDLE; P. H. Practical NIR spectroscopy with applications in food and beverage analysis. And ed. England: longman Scientific & Technical 1993.

PASQUINI, C. Near infrared spectroscopy fundamentals, practical aspects and analytical applications, En; J, Chemm Soc. 2003. Vol 14, N2. P. 198-219.

PEREIRA C: Rita SKROBOT L. Vinicius. CASTRO R Eusta quio. FORTES C P Isabel; and PASA M D Vanya. Determination of Gasoline Adulteration by Principal Components Analysis-Linear Discriminal Analysis Applied to FTIR Spectral Energy & Fuels 2006; 201097-1102.

R UBINSON, K. A; RUBINSON, J. F. Analisis intrumental, Prentice Hall Hispanoamericana S.A. Madrid: Editotial Pearson Educacion, 2001.

RODRIGUEZ-SAONA, L.E; KOCA, N; HARPER, W. J. Rapid Determination of swiss Cheese Composition by Fourier Transform infrared/Attenuated Total reflectance Spectroscopy. En; department of Food science and Technology 2005. Vol 89 N 5 p 1407-1412.

SASTRY, S; Chopra, A; Sarpal, S; Jain, K; Srivastava, P and bhatnagar K. determination of physicochemical properties and carbon-type analysis of base oils using Mid-IR spectroscopy and partial least-squares regression analysis *Energy & Fuels* 1998 Vol 12 p 304-311.

SCHABRON, J. F.; Rovani, J. F. On-column precipitation and re-dissolution of asphaltenes in petroleum residua. *Fuel* 87 (2008) 165-176.

SCHNEIDER, M.; Andrews, A. B. Mitra-Kirtley, S.; Mullins, O. C. Asphaltene Molecular Size by Fluorescence Correlation Spectroscopy. *Energy & Fuels* 2007, 21, 2875-2882.

SCHWEDT, G JOHN. The essential guide to analytical chemistry, Chichester: Wiley & Sons. 1997

SHEU, E. Y.; Mullins, O. Frequency-Dependent Conductivity of Utah Crude Oil Asphaltene and Deposit. *Energy & Fuels* 2004, 18, 1531-1534.

SILVERSTEIN, R. M; and WEBSTER, F. X. Spectrometric identification of Organic compounds, 6^a Ed. New York: John Wiley & sons 1998.

SPIECKER, P. M., Gawrys, K. L., Trail, C. B. & Kilpatrick, P. K. Effects of petroleum resins on asphaltene aggregation and water-in-oil emulsion formation. *Colloids Surfaces A Physicochem. Eng. Asp.* 220, 9–27 (2003).

STERINER, J; TERMONIA, Y; DELTOUR, J; *Anal. Chemometrics*. Vol 44. P. 1906, 1972.

STRAUSZ, O. P.; Lown, S. E. M.; Izquierdo, M.; A Critique of Asphaltene Fluorescence Decay and Depolarization-Based Claims about Molecular Weight and Molecular Architecture. *Energy & Fuels* 2008, 22, 1156-1166.

STONE, M.J *Staist Soc.* 1973. Vol 36. P 111.

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

TREJO, F.; Rana, M. S.; Ancheyta J. Thermogravimetric determination of coke from asphaltenes, resins and sediments and coking kinetics of heavy crude asphaltenes. *Catalysis Today* 150 (2010) 272-278.

VILCA, J.C. Generalizaciones de mínimos cuadrados parciales con aplicación en clasificación supervisada. Universidad de Puerto Rico. Mayagüez. 2004.

WAUQUIER, J.P. El refino del petrolio. Petróleo crudo, producto petrolífero Esquema de fabricación. En: composición de los petrolios crudos y de los productos petrolíferos. España: Editorial Diaz de santos 2004 p 1-15.

WILT K Brian and WELCH T. William. Determination of Asphaltenes in Petroleum Crude oils by Fourier Transform intrared Spectrscopy. Energy & Fuels 1998, 12, 1008-1012.

WOLD; S Technometrics. 1978. Vol 20 P 397.

YEN. T.F. Petroleum Analysis. En: use of the data – the Structure of petroleum asphaltene and its significance. Energy sources. 2006 Vol 1 p 447- 463.

ZHAN, G; SHAOHUI, G; SUOQI, Z; GUANGXU, Y; LANQI, S and Chen, L Alkyl Side Chains Connected to Aromatic Units in Dagang Vacuum Residue and Its Supercritical Fluid Extraction and Fractions (SFEFs). Energy and Fuels. 2009, 23, 374-385.

ZHANG, Z; Ewing G.E. Attenuated partial internal reflection infrared spectroscopy. En: Analytical Chemistry. 2002 vol 74 N11 P 2578-2583.