PREDICCIÓN DEL ANÁLISIS SARA DE CRUDOS COLOMBIANOS APLICANDO ESPECTROSCOPIA FTIR-ATR Y MÉTODOS QUIMIOMETRICOS

LESLY VIVIANA MELÉNDEZ CORREA
ADRIANA LACHE GARCÍA

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE QUÍMICA
LABORATORIO DE ESPECTROSCOPIA ATÓMICA Y MOLECULAR
BUCARAMANGA
2010

PREDICCIÓN DEL ANÁLISIS SARA DE CRUDOS COLOMBIANOS APLICANDO ESPECTROSCOPIA FTIR-ATR Y MÉTODOS QUIMIOMETRICOS

LESLY VIVIANA MELÉNDEZ CORREA
ADRIANA LACHE GARCÍA

Trabajo de grado para optar al título de
QUIMICO

DIRECTORES:
ZARITH PACHÓN – ECOPETROL ICP
ENRIQUE MEJÍA OSPINO - UIS
Químico, Ph. D.

COODIRECTOR:
JORGE ARMANDO ORREGO
Qco, M.Sc.

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE QUÍMICA
LABORATORIO DE ESPECTROSCOPIA ATÓMICA Y MOLECULAR
BUCARAMANGA
2010

A mi Madre y hermanas Por su apoyo incondicional
A mi familia.... Andrés y demás por su compañía y
colaboración.....Gracias

Viviana

A Dios, mi amigo, mi soporte espiritual, por darme todo lo que tengo y no dejarme caer nunca. . .

A mi Hija, Daniela Alejandra, por haber cambiado mi vida, que con sus ojitos y amor me dan la fuerza necesaria para estar de pie y con la cabeza en alto para enfrentar cualquier situación por difícil que sea. . .

Mis padres por ser los mejores y estar conmigo incondicionalmente, gracias por su paciencia y porque sin sus enseñanzas no estaría aquí ni sería quien soy ahora. . .

A Viviana, por ser mi amiga fiel que ha sido el soporte diario para llenar estás páginas que marcan el fin de una etapa y el comienzo de otra. . .

A mis amigos que siempre me han acompañado para llevar a buen término mi carrera universitaria. . .

A todos ellos les dedico esta Tesis. . .

Adriana

AGRADECIMIENTOS

Queremos que las próximas líneas sirvan como reconocimiento a las personas e instituciones que hicieron posible la realización de esta tesis. A todos ellas, queremos agradecer de todo corazón por su colaboración.

La más sincera gratitud al doctor Enrique Mejía Ospino por la oportunidad brindada, por su guía, sus valiosos aportes y la confianza depositada en nosotras para llevar a término este trabajo.

También queremos agradecer a la directora del Laboratorio de Química de Producción del ICP Zarith Pachón, al doctor Daniel Ricardo Molina y al Químico Jorge Armando Orrego (Co-director) quienes con sus aportes, conocimientos y valiosa colaboración contribuyeron en este proyecto.

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

A todos nuestros compañeros del Laboratorio de Espectroscopia Atómica y Molecular (LEAM) que de una u otra forma estuvieron involucrados en el desarrollo de este trabajo.
A nuestras familias y amigos, por su compañía y apoyado incondicional.

CONTENIDO

Pág

INTRODUCCIÓN 1

1.CONSIDERACIONES TEÓRICAS 23
1.1.PETRÓLEO 23
1.1.1.Generalidades 23
1.1.2.Clasificación del Petróleo 23
1.1.3.Composición Química del Petróleo 25
1.1.3.1.Saturados 25
1.1.3.2.Aromáticos 26
1.1.3.3.Resinas 27
1.1.3.4.Asfaltenos 27
1.1.4.Análisis SARA 27
1.1.4.1.Cromatografía de Exclusión (CE) 29

1.2.ESPECTROSCOPÍA DE INFRARROJO 29
1.2.1.Aspectos fundamentales 29
1.2.2.Regiones Espectrales 32
1.2.3.Tipos de medidas en el infrarrojo 32
1.2.3.1.Reflectancia Total Atenuada (ATR) 33
1.2.4.Interpretación de Espectros 36

1.3.QUIMIOMETRIA 37
1.3.1.Calibración multivariable 38
1.3.1.1.Clasificación de los métodos de calibración multivariable 38
1.3.1.2.Construcción de modelos de calibración multivariable 39
1.3.2.Técnicas de Pretratamientos de Datos 42
1.3.2.1.Suavizado espectral 42
1.3.2.2.Normalización 42
1.3.2.3.Corrección de línea base 43
1.3.2.4.Centralización 43
1.3.2.5.Derivación 44
1.3.3.Métodos basados en reducción de variables 44
1.3.3.1.Análisis Por Componentes Principales (PCA) 45
1.3.3.2.Regresión por Mínimos Cuadrados Parciales (PLS) 48

1.3.4.Factores a Incluir en el Modelo 51
1.3.4.1.Detección de outliers 51
1.3.4.2.Leverage 51
1.3.4.3.Estadístico T^2 de Hotelling 52
1.3.4.4.Errores de calibración y predicción 52
1.3.4.5.Residuales en la respuesta instrumental 53
1.3.4.6.Validación del modelo 55
1.3.4.7.Predicción de muestras desconocidas 55

1.4.APLICACIÓN DE LA ESPECTROSCOPIA FTIR-ATR EN EL PETRÓLEO. 56

2.PARTE EXPERIMENTAL 57

2.1.MUESTRAS 57
2.1.1.Tratamiento de muestras 59

2.2.ESPECTROSCOPIA MIR-ATR 61
2.2.1.Instrumentación 61
2.2.2.Verificación del desempeño del Espectrómetro 62
2.2.3.1.Análisis de los espectros obtenidos 63

2.3.PROCESAMIENTO DE DATOS 64

3.RESULTADOS Y ANÁLISIS 67

3.1.ANÁLISIS POR COMPONENTES PRINCIPALES (PCA) 67
3.1.1.Análisis por componentes principales para muestras livianas 72
3.1.2.Análisis por componentes principales para muestras pesados. 75

3.2.DESARROLLO DE MODELOS PLS 77
3.2.1.Calibración del modelo PLS para asfaltenos 78
3.2.2.Validación cruzada completa del modelo PLS para asfaltenos 81

4.RESULTADOS DE LOS MODELOS DE PREDICCIÓN 85

4.1.MODELOS DE PREDICCIÓN DE LA COMPOSICIÓN SARA PARA MUESTRAS LIVIANAS 85
4.1.1.Modelo PLS para la predicción de Saturados 85
4.1.2.Modelo PLS para la predicción de aromáticos 89
4.1.3.Modelo PLS para la predicción de resinas. 93
4.1.4.Modelo PLS para la predicción de asfaltenos 97

4.2.MODELOS DE PREDICCIÓN DE LA COMPOSICICÓN SARA PARA MUESTRAS PESADAS 101

4.2.1.Modelo PLS para la predicción de Saturados 101
4.2.2.Modelo PLS para la predicción de aromáticos 105
4.2.3.Modelo PLS para la predicción de resinas 109
4.2.4.Modelo PLS para la predicción de Asfaltenos 112

5.CONCLUSIONES 117

6.RECOMENDACIONES 119

7.REFERENCIAS BIBLIOGRAFICAS 120

LISTA DE FIGURAS

Pág

Figura1. Esquema simplificado de la separación del petróleo crudo en las fracciones SARA.....28

Figura 2. Perfil de energía potencial según el modelo del oscilador (a) armónico, (b) anharmónico.....30

Figura 3. Fenómenos de absorción, transmisión y reflexión de la radiación electromagnética al interaccionar con la materia.....33

Figura 4. Reflexión total interna y elemento de reflexión interna (IRE) utilizado en el sistema ATR.....34

Figura 5 a) Diagrama que ilustra dos componentes principales, PC1 Y PC2, para dos variables, X_1 y X_2. (b) Puntos referidos a los ejes de los componentes principales. Indica puntos de datos, su proyección sobre los ejes.....46

Figura 6. Notación matricial para la descomposición por componentes principales 47

Figura 7. Notación matricial extendida para la descomposición por componentes principales.....48

Figura 8. Descripción gráfica del método de regresión PLS.....49

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. 53

Figura 10. Gráfico del residual frente al leverage. (a) objetos con una varianza residual elevada se consideran outliers, (b) si además tienen un leverage alto son outliers peligrosos para el modelo, debido a que tienen mucha influencia sobre él. Las muestras con un alto leverage (c) son muestras influyentes y no necesariamente outliers.....54

Figura 11. Ejemplo de la comparación de espectros IR de muestras hidratadas (rosada) y no hidratadas (verde).....59

Figura 12. Comparación de los espectros de la muestra 28 antes (fucsia) y después de deshidratar (azul).....61

Figura 13. Sistema de caracterización MIR; a)Espectrómetro, b) Celda ATR ..62

Figura 14. Prueba de repetibilidad del Espectrómetro MIR 63

Figura 15. Espectros Originales de 50 muestras de Crudo. 64
Figura 16. Espectros Normalizados y derivados.....66

Figura 17. Gráfica de Scores de los primeros componentes principales de las 50 muestras 68

Figura 18. Estadístico T^2 aplicado a la gráfica de Scores de las 50 muestras..69

Figura 19. Grafica de Influencia para las 50 muestras.....70

Figura 20. Gráfico de X-Loadings de PC1 y PC2.....71

Figura 21. Estadístico T^2 aplicado a la gráfica de Scores de las muestras livianas 73

Figura 22. Grafica de Influencia para muestras livianas 74

Figura 23. Estadístico T^2 aplicado a la gráfica de Scores de muestras pesadas76

Figura 24. Grafica de Influencia para muestras pesadas.....77

Figura 25. Varianza explicada en el modelo PLS de asfaltenos 79

Figura 26. Gráfica de los errores RMSEC y RMSEP calculados en función de los PCs usados para el modelo PLS de asfaltenos.....80

Figura 27. Coeficientes de regresión para el primer componente principal del modelo PLS de asfaltenos 81

Figura 28. Curvas de calibración y validación del modelo de predicción de %W de asfaltenos 83

Figura 29. Descripción gráfica del modelo con tres componentes principales para la predicción del contenido %W de saturados en muestras livianas.....87

Figura 30. Descripción grafica del modelo con cuatro componentes principales para la predicción del contenido %W de aromáticos en muestras livianas.....91

Figura 31. Descripción grafica del modelo con cuatro componentes principales para la predicción del contenido %W de resinas en muestras livianas.....95

Figura 32. Descripción gráfica del modelo con cuatro componentes principales para la predicción del contenido %W de asfaltenos en muestras livianas.....98
Figura 33. Descripción gráfica del modelo con cinco componentes principales para la predicción del contenido %W de saturados en muestras pesadas.....103

Figura 34. Descripción grafica del modelo con 4 componentes principales para la predicción del contenido %W de aromáticos en muestras pesadas.....107

Figura 35. Descripción grafica del modelo con 5 componentes principales para la predicción del contenido %W de resinas en muestras pesadas.....110

Figura 36. Descripción gráfica del modelo con cinco componentes principales para la predicción del contenido %W de asfaltenos en muestras pesadas 114

LISTA DE TABLAS

Pág

Tabla1. Especificaciones Generales de clasificación de Crudos 24

Tabla 2. Regiones de absorción en el infrarrojo.....32

Tabla 3. Frecuencias de absorción de algunos grupos funcionales en la región MIR. 37

Tabla 4. Análisis SARA de los crudos analizados.....58

Tabla 5. Comparación de los %BSW de las muestras antes y después de deshidratar 60

Tabla 6. Cálculo de la RSD para tres números de onda característica aplicando diferentes procesamientos 65

Tabla 7. Varianza explicada por los componentes principales de las 50 muestras de crudo 67

Tabla 8. Varianza explicada por los componentes principales de las muestras livianas 72

Tabla 9. Varianza explicada por los componentes principales de las 25 muestras pesadas 75

Tabla 10. Parámetros estadísticos del modelo seleccionado para la predicción de %W de Asfaltenos.....78

Tabla 11. Validación del modelo de predicción de %W de asfaltenos de las 50 muestras de calibración 82

Tabla 12. Prueba de repetibilidad del modelo de predicción de %W de asfaltenos 84

Tabla 13. Parámetros estadísticos del modelo desarrollado para la predicción de %W de saturados.....86

Tabla 14. Validación cruzada del modelo de predicción de %W de saturados de las muestras livianas 88

Tabla 15. Prueba de repetibilidad del modelo de predicción de %W de saturados 89

Tabla 16. Parámetros estadísticos del modelo generado para la predicción de %W de aromáticos en muestras livianas.....89

Tabla 17. Validación cruzada del modelo de predicción del %W de aromáticos de las muestras livianas.....92

Tabla 18. Prueba de repetibilidad del modelo de predicción de %W de aromáticos 93

Tabla 19. Parámetros estadísticos del modelo generado para la predicción de %W de resinas en muestras livianas.....93

Tabla 20. Validación cruzada del modelo de predicción del %W de resinas de las muestras livianas. 96

Tabla 21. Prueba de repetibilidad del modelo de predicción de %W de resinas en muestras livianas 96

Tabla 22. Parámetros estadísticos del modelo generado para la predicción de %W de asfaltenos en crudos livianos.....97

Tabla 23. Validación cruzada del modelo de predicción de %W de asfaltenos de las muestras livianas.....99

Tabla 24. Prueba de repetibilidad del modelo de predicción de %W de asfaltenos 100

Tabla 25. Prueba de Validación externa. Predicción de la composición SARA en crudos livianos 101

Tabla 26. Parámetros estadísticos del modelo generado para la predicción de %W de saturados en crudos pesados 101

Tabla 27. Validación cruzada del modelo de predicción de %W de saturados de las muestras pesadas.....104

Tabla 28. Prueba de repetibilidad del modelo de predicción de %W de saturados 104

Tabla 29. Parámetros estadísticos del modelo generado para la predicción de %W de aromáticos en muestras pesadas 105

Tabla 30. Validación cruzada del modelo de predicción del %W de aromáticos de las muestras pesadas. 108

Tabla 31. Prueba de repetibilidad del modelo de predicción de %W de aromáticos 108

Tabla 32. Parámetros estadísticos del modelo generado para la predicción de %W de resinas en muestras pesadas 109

Tabla 33. Validación cruzada del modelo de predicción del %W de resinas de las muestras pesadas.....111

Tabla 34. Prueba de repetibilidad del modelo de predicción de %W de resinas112

Tabla 35. Parámetros estadísticos del modelo generado para la predicción de %W de asfaltenos.....113

Tabla 36. Validación cruzada del modelo de predicción de %W de asfaltenos de las muestras pesadas 115

Tabla 37. Prueba de repetibilidad del modelo de predicción de %W de asfaltenos 116

Tabla 38. Prueba de Validación externa. Predicción de la composición SARA en crudos pesados.....116

RESUMEN

TÍTULO
PREDICCIÓN DEL ANÁLISIS SARA DE CRUDOS COLOMBIANOS APLICANDO ESPECTROSCOPIA FTIR-ATR Y MÉTODOS QUIMIOMÉTRICOS[*].

AUTORES
LESLY VIVIANA MELÉNDEZ CORREA.
ADRIANA LACHE GARCÍA[**].

PALABRAS CLAVES
ESPECTROSCOPIA MIR-ATR, CRUDO, PCA, PLS.

DESCRIPCIÓN

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

Los modelos de predicción fueron desarrollados empleando el rango espectral de 4000 a 690cm^{-1}. La validación mostró resultados satisfactorios para la predicción del análisis SARA de crudos. Para cada fracción, se obtuvieron errores estándar de predicción (SEP) para muestras livianas de 1.9, 1.7, 1.3 y 0.4 y errores estándar de predicción (SEP) para muestras pesadas de 2.5, 1.6, 3.6, y 1.4 respectivamente. En todos los casos el coeficiente de correlación (R^2) entre los valores de referencia y predichos por los modelos fue superior a 0.95.

Los modelos de mejor desempeño fueron los desarrollados para resinas y asfaltenos en muestras livianas con una varianza explicada del 99 y 97%, y aromáticos y asfaltenos para muestras pesadas con una varianza explicada de 98% para los dos.

La metodología de caracterización propuesta por espectroscopia 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 solventes y mayor tiempo de análisis.

[*] Trabajo de Grado
[**] Facultad de Ciencias. Escuela de Química. Director: Enrique Mejía Ospino. Co director: Jorge Armando Orrego.

SUMMARY

TITLE :

PREDICTION OF THE ANALYSIS SARA OF COLOMBIANS OIL CRUDE USING FTIR-ATR SPECTROSCOPY AND CHEMOMETRICS METHODS *.

AUTHORS:

LESLY VIVIANA MELÉNDEZ CORREA **.

ADRIANA LACHE GARCI'A **.

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 690cm^{-1}. The validation showed satisfactory results for the prediction of the SARA analysis of crude. For each fraction, were obtained standard errors of prediction (SEP) for light samples of 1.9, 1.7, 1.3 and 0.4 and heavy samples of 2.5, 1.6,3.6, and 1.4 and heavy samples respectively. In all cases the coefficient of correlation (R^2) between the values of reference and predicted by the models was superior to 0.95.

The models of better performance were developed for resins and the asphaltenes in light samples with an explained variance of 99 and 97%, and aromatic and asphaltenes for heavy samples with an explained variance of 98% for both.

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.

* Work Degree

** Sciences Faculty. Chemistry School. Directreess: Enrique Mejia Ospino. Codirectress Jorge Armando Orrego

INTRODUCCIÓN

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 Asfaltenos 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 asfaltenos por la acción de n-alcanos como pentano o heptano. Posteriormente, la fracción desasfaltada, se separa por cromatografía de exclusión, con la ayuda de diferentes fases estacionarias y solventes de polaridad variada. De cada fracción eluída se remueve el solvente por evaporación, y el análisis SARA se completa por la determinación de los pesos de las fracciones. Todo el proceso tarda en promedio 2 días por muestra y durante el análisis se 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 tarda 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 multivariantes llamadas quimiométricas aplicadas a datos espectroscópicos ha permitido avanzar en la determinación de las propiedades fisicoquímicas de las muestras. El análisis por componentes principales (PCA) y la regresión por mínimos cuadrados parciales (PLS) son técnicas quimiométricas que permiten obtener modelos para la predicción de diferentes propiedades a partir de datos espectroscópicos [8]. Motivados por esto antecedentes en los laboratorios de Espectroscopia Atómica y Molecular de la universidad industrial de Santander y de Química de Producción del ICP, se desarrolló una metodología para la obtención del análisis SARA de crudos colombianos basada en el uso combinado de reflectancia total atenuada en el infrarrojo medio (FTIR-ATR) y métodos de calibración multivariable que permite operar en condiciones menos rigurosas que los procedimientos estándar con un menor costo de análisis.

1. CONSIDERACIONES TEÓRICAS

1.1. PETRÓLEO

1.1.1. Generalidades

El petróleo crudo es uno de los combustibles de mayor aplicación del presente siglo, convirtiéndose en la base económica de diversos sectores como la industria y el transporte, los cuales dependen directamente de los productos que de él se derivan. Se puede considerar como una mezcla compleja de cientos de hidrocarburos, que pueden incluir desde 1 hasta 60 átomos de carbono, junto con compuestos derivados de éstos, presentes en cantidades relativamente bajas, que pueden contener en su estructura azufre, nitrógeno, oxigeno y algunos elementos metálicos como níquel, vanadio, hierro y cobre. La apariencia y composición de un crudo varía ampliamente, aunque se considera que un crudo promedio contiene aproximadamente de 84 a 87 % de carbono, 11 a 14 % de hidrógeno, entre 1 a 3 % de azufre y menos del 1 % de nitrógeno, oxigeno, metales y sales. Estas diferencias en composición influyen de manera apreciable en los aspectos de diseño y localización de las plantas de refinación, en la determinación de los procesos de conversión, tratamiento requeridos y en la producción de los derivados de mayor demanda e importancia económica [9, 10, 11].

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 ligeros. 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, ecuación 1) cuyo valor varía entre 70 y 5; esta variabilidad de la densidad es consecuencia de composiciones en familias químicas muy diferentes [12].

$$°API = \frac{141,5}{Densidad\ relativa\ estándar\ a\ \ 60°F\ /\ 60°} - 131,5 \qquad (1)$$

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 ligeros o de baja gravedad específica presentan alto dicho valor. Los crudos con bajo contenido de carbono, alto contenido de hidrogeno 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 hidrógeno 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.

Tabla1. Especificaciones Generales de clasificación de Crudos

Aceite crudo	Densidad (g/cm3)	Gravedad API
Extrapesado	> 1.0	1
Pesado	1.0 - 0.92	10.0
Mediano	0.92 - 0.87	22.3 - 31.1
Ligero	0.87 - 0.83	31.1 – 39
Superligero	< 0.83	> 39

Fuente: WAUQUIER, J. P. El refino 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 asfaltenos (fracción SARA). La fracción de saturación está compuesta por hidrocarburos saturados no polares lineales, ramificados y cíclicos. Los aromáticos, que contienen uno o más anillos aromáticos, son más polarizables. Las otras dos fracciones, resinas y asfaltenos, tienen sustituyentes polares. La distinción entre los dos, es que los asfaltenos son insolubles en un exceso de heptano (o pentano), mientras que las resinas son miscibles en estos solventes. Este sistema de clasificación es útil porque identifica las fracciones del crudo que se refieren a la estabilidad de asfaltenos y por lo tanto permite la identificación de los crudos con potencial para generar problemas por asfaltenos [13]. A continuación se da una pequeña descripción de la composición de estas fracciones [14]:

1.1.3.1. Saturados

Saturado 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 2.000.

- **Hidrocarburos alifáticos saturados o 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 otros cuatro átomos (carbono o hidrógeno); 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 las isoparafinas o isoalcanos.

- **Hidrocarburos cíclicos saturados, cicloalcanos o naftenos:** En estos hidrocarburos hay una ciclación total o parcial del esqueleto carbonado. El número de átomos de carbono del anillo formado puede ser variable. Tienen 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 cinco o seis átomos de carbono. En estos anillos, cada átomo de hidrógenos puede ser sustituido por una cadena parafínica recta o ramificada, llamada alquilo.

1.1.3.2. Aromáticos

Comprenden los compuestos nafteno-aromaticos 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. Consisten 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 hidrógeno. El aromático más simple es el benceno C_6H_6. Incluye mono-aromáticos y aromáticos policíclicos.

- **Hidrocarburos aromáticos** Son hidrocarburos cíclicos poliinsaturados que están presentes en una gran proporción en los crudos de petróleo. La presencia en su fórmula de uno o más ciclos con tres dobles enlaces conjugados les confiere unas notables propiedades. Así, los primeros compuestos (benceno, tolueno, xilenos) son materias primas fundamentales de la petroquímica (además contribuyen igualmente a aumentar el número de octano de las gasolinas) mientras que los homólogos superiores son en general nefastos (problemas de medio ambiente, de sanidad pública, deterioro de la actividad de los catalizadores por su capacidad.

1.1.3.3. Resinas

Son moléculas con un fuerte carácter aromático, al igual que los asfaltenos tienen una elevada proporción de hidrógeno y carbono, contienen pequeñas cantidades de oxigeno, azufre y nitrógeno. Son sólidos negros, brillantes, quebradizos y su naturaleza es muy polar. Las resinas constituyen el componente polar no volátil del petróleo, que es soluble en n-alcanos e insoluble en propano líquido.

1.1.3.4. Asfaltenos

Son sólidos amorfos de color marrón oscuro o negro, solubles en n-heptano. Están constituidos, además de carbono, hidrógeno, nitrógeno, azufre y oxígeno, que dan lugar a ciclos tiofénicos y piridínicos que contienen elevada polaridad. Los asfaltenos son considerados generalmente como hidrocarburos aromáticos altamente polares de elevado peso molecular. Una representación de su estructura consiste en láminas aromáticas apiladas, enlazadas entre sí por los electrones π de los dobles enlaces del anillo bencénico. El rendimiento en asfaltenos y su constitución varían con la naturaleza del disolvente utilizado. La gran complejidad de esta fracción no hace posible la formación de estructuras moleculares concretas.

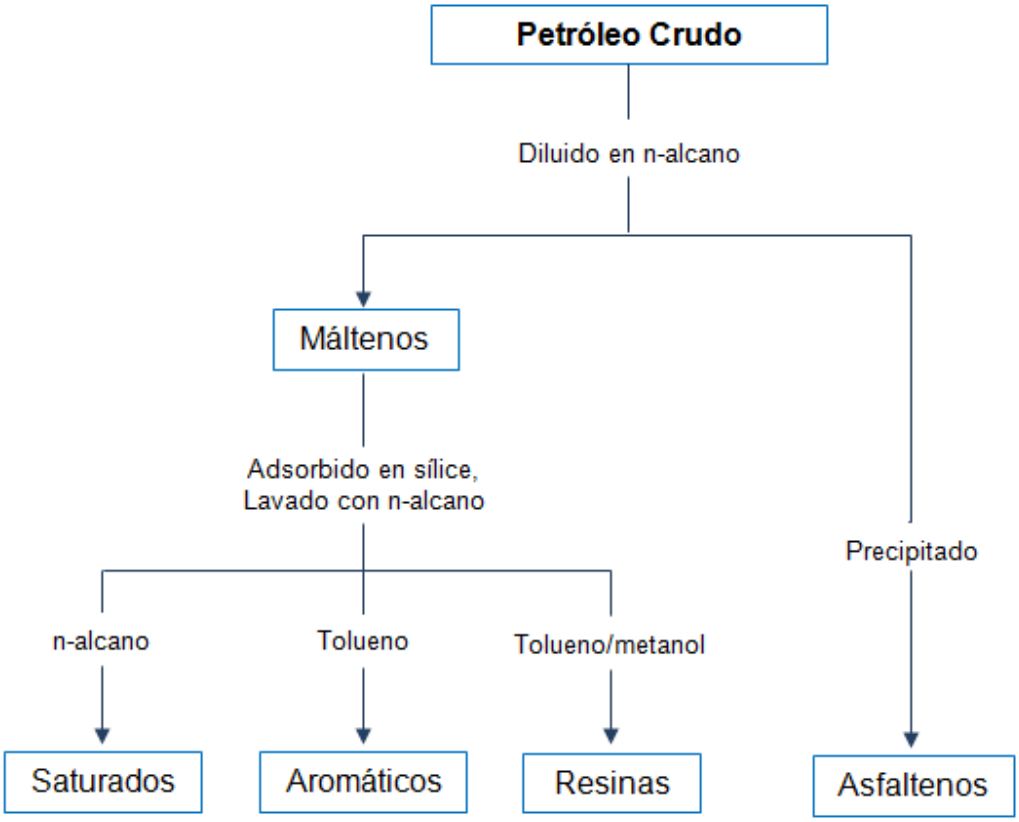
1.1.4. Análisis SARA

El gran número de compuestos que forman parte de los crudos hace necesaria la separación en grupos más homogéneos mediante técnicas de fraccionamiento para su identificación. La mayoría de los procedimientos existentes hacen una primera separación de los crudos mediante precipitación con hidrocarburos saturados de bajo peso molecular (n-heptano o n- pentano). A la fracción insoluble se le denomina *asfaltenos* y a la soluble *maltenos*. Estos últimos a la vez pueden dividirse en saturados, nafteno-aromáticos y polares o resinas (fracción SAR, figura 1).

Para determinar la composición porcentual del contenido de saturados, aromáticos, resinas y asfaltenos en un crudo se han estandarizado diferentes

métodos entre los cuales se encuentra la cromatografía de capa fina con detección de ionización de llama (TLC-FID), cromatografía liquida de alta resolución (HPLC), cromatografía de exclusión (CE), entre otras. Estas técnicas requieren del uso de diferentes métodos para el pretratamiento de la muestra, así como de solventes para hacer posible su análisis; esto hace que el estudio de una muestra sea tedioso y requiera de un tiempo prolongado (2 a 3 días) [15].

Figura1. Esquema simplificado de la separación del petróleo crudo en las fracciones SARA



Fuente. KAMRAN, Kbarzadeh. Los asfaltenos: problemáticos pero ricos en potencial.

La técnica con la que se determino la composición SARA de los crudos usados para realizar los modelos predictivos (técnica de referencia), fue la cromatografía de exclusión, está se realiza siguiendo la norma ASTM D2007 [16]. A continuación se encuentra una breve descripción.

1.1.4.1. Cromatografía de Exclusión (CE)

La cromatografía liquida está precedida por una precipitación de los asfaltenos por la acción de n-alcanos como pentano o heptano, por lo que el método cromatográfico realiza una separación SAR (maltenos) mediante una columna mixta de sílice seguida de alúmina. La elución de los hidrocarburos saturados se realiza con n-heptano y tolueno, la de los hidrocarburos aromáticos con una mezcla 2:1 en volumen de n-heptano y tolueno, y las resinas con una mezcla 1:1:1 de diclorometano, tolueno y metanol. El rendimiento de cada una de las fracciones depende de su respectivo volumen de retención, que a su vez depende del adsorbente elegido y del poder de elución de los disolventes [17].

1.2. ESPECTROSCOPÍA DE INFRARROJO

1.2.1. Aspectos fundamentales

La espectroscopia molecular se basa en la interacción entre la radiación electromagnética y las moléculas. Dependiendo 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), esta interacción será de diferente naturaleza: excitación de electrones, vibraciones moleculares y rotaciones moleculares [18]. 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 solo es posible observarlas específicamente en el caso de muestras gaseosas. En el caso del estudio del espectro infrarrojo (IR) de muestras sólidas y líquidas sólo se tienen en cuenta los cambios entre estados de energía vibracional [19, 20].

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

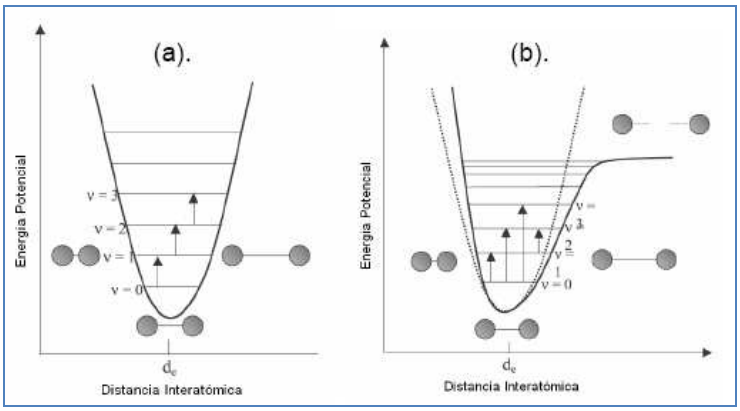
$$E_{vib} = \left[\upsilon + \frac{1}{2} \right] \frac{h}{2\pi} \sqrt{\frac{\kappa}{\mu}} \qquad (2)$$

Donde υ es el número cuántico vibracional, h es la constante de Planck, κ es la constante de fuerza y μ es la masa reducida del sistema. Este modelo permite únicamente las transiciones entre niveles energéticos adyacentes ($\Delta\upsilon$ = ± 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).

$$\Delta\upsilon = E_{\upsilon^2} - E_{\upsilon^1} = \Delta\upsilon h\nu \qquad (3)$$

Donde v 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 del oscilador (a) armónico, (b) anharmónico.



Fuente: PASKINI, C. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications.

Aunque el modelo del oscilador armónico es una buena aproximación, no explica muchos aspectos sobre el comportamiento molecular. El modelo del oscilador anharmónico 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 υ, y se puede calcular por medio de la ecuación (4).

$$\Delta E_{vib} = h\nu\left[1-(2\upsilon+\Delta\upsilon+1)y\right] \qquad (4)$$

Donde "y" es el factor de anharmonicidad. Esta anharmonicidad permite las transiciones entre estados de energía vibracional no consecutivos, ($\Delta\upsilon$ = ±2, ±3,…), generando las bandas de absorción conocidas como sobretonos 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 sobretono 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 sólo se observan las bandas de absorción correspondientes a las frecuencias de vibración fundamental en el MIR (υ) y a los dos primeros sobretonos en el NIR ($2\upsilon, 3\upsilon$) [21].

Una molécula poliatómica (n átomos) tiene 3n-6 modos 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 sobretonos en el infrarrojo. Los modos de vibración que se pueden producir incluyen: cambios en la distancia de enlace (elongaciones o stretching, que pueden ser simétricas o asimétricas) y cambios en el ángulo de enlace, o bending (simétricos en el plano, asimétricos en el plano, simétricos fuera del plano y asimétricos fuera del plano) [22].

1.2.2. Regiones Espectrales

Aunque el espectro de infrarrojo se extiende desde 10 a 14300cm^{-1} desde un punto de vista funcional se divide en tres zonas: IR lejano, donde se producen las absorciones debido a cambios rotacionales, el IR medio (MIR o simplemente IR), donde tiene lugar las vibraciones fundamentales y el IR cercano (NIR), donde se producen absorciones debidas a sobretonos y combinaciones de las bandas fundamentales. La tabla 2 muestra el rango en el espectro electromagnético al que corresponde cada región IR.

Tabla 2. Regiones de absorción en el infrarrojo

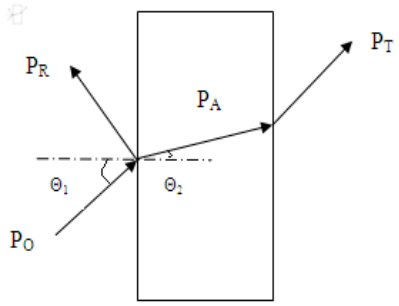
Región Infrarroja	Longitud de Onda	Número de Onda, (cm-1)	Transición característica
Cercano (NIR)	780 a 2500 nm	12800 a 4000	Sobretonos y Combinaciones
Medio (MIR)	2.5 a 50 µm	4000 a 200	Vibraciones fundamentales
Lejano (FIR)	50 a 1000 µm	200 a 10	Rotaciones

Fuente: SKOOG, A. Douglas. Principios de Análisis Instrumental.

1.2.3. Tipos de medidas en el infrarrojo

Cuando la radiación incide en la muestra (figura 3), ésta puede sufrir diferentes fenómenos: absorción, transmisión y reflexión. La intensidad de la luz trasmitida 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. Fenómenos de absorción, transmisión y reflexión de la radiación electromagnética al interaccionar con la materia.



$$P_O = P_A + P_T P_R$$

Fuente: MACHO. A, Santiago. Metodologías analíticas basadas en espectroscopia 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 [23]. A continuación se describe la técnica que se ha utilizado en la tesis.

1.2.3.1. Reflectancia Total Atenuada (ATR)

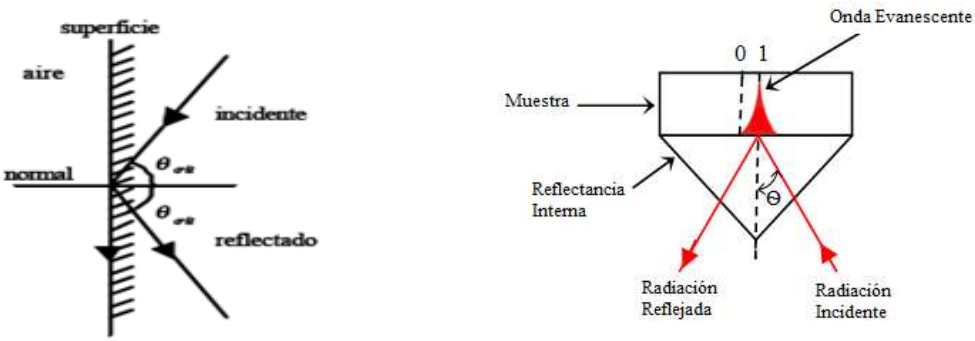
El principio de esta medida se basa en el fenómeno de la reflexión total interna y la transmisión de la luz a través de un cristal con un elevado índice de refracción [24]. La radiación penetra (unos μm) más allá de la superficie del cristal donde se produce la reflexión total, en forma de onda evanescente [25]. Si en el lado exterior del cristal se coloca un material absorbente (muestra), la luz que viaja a través del cristal se verá atenuada (por ello el nombre de la técnica) y se puede registrar el espectro de la muestra.

El ángulo de la luz incidente y la geometría del cristal facilitan que se produzcan sucesivas reflexiones en sus caras internas (figura 4). El espectro medido tiene

una apariencia similar al espectro de trasmisión excepto por ciertas variaciones en la intensidad en función de la longitud de onda que se producen.

La intensidad de la onda evanescente es sustraída de la intensidad del haz que continúa reflejándose hasta que sale del cristal al detector del equipo IR, y de esta manera generar un espectro IR [26, 27]. Para un uso adecuado de la celda ATR (figura 4), es necesario tener en cuenta los siguientes factores experimentales y cómo ellos afectan el espectro final:

Figura 4. Reflexión total interna y elemento de reflexión interna (IRE) utilizado 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-pyrrolidone-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).

$$\theta_c = \sin^{-1}\left(\frac{n_2}{n_1}\right) \qquad (5)$$

Donde n_2 es el índice de refracción de la muestra, n_1 el índice de refracción del cristal y θ_c 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 esto 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)

$$d_p = \frac{\lambda}{2\pi\left(n_1^2 \sin^2\theta - n_2^2\right)^{\frac{1}{2}}} \qquad (6)$$

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 reflexiones dentro del cristal ATR*: En celdas de reflexión simple, el haz se hace incidir una vez con la muestra, mientras en celdas de múltiple reflexión la profundidad efectiva de penetración aumenta y la señal ATR es de mayor intensidad.

- *Calidad del contacto entre la muestra y el cristal*: se refiere 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 sólidos 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 [28].

1.2.4. Interpretación de Espectros

Asignación de bandas: En el espectro infrarrojo medio, entre 4000 y 200cm^{-1} (región de frecuencias de grupo) se observan una serie de bandas asignadas a vibraciones de sólo dos átomos de la molécula. En este caso la banda de absorción se asocia únicamente a un grupo funcional y a la estructura molecular completa, aunque hay influencias estructurales que provocan desplazamientos significativos en la frecuencia de la vibración. Estas vibraciones derivan de grupos que contienen hidrógeno (C-H, O-H y N-H) o grupos con dobles y triples enlaces aislados. Entre 1300 y 400cm^{-1} (fingerprint región) la asignación a grupos funcionales determinados es más difícil debido a la multiplicidad de bandas, pero es una zona de espectro muy útil para la identificación de compuestos específicos [25]. La tabla 3 muestra un cuadro resumen de las frecuencias de absorción de los grupos funcionales más comunes en el IR medio.

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

Enlace	Tipo de compuesto	Intervalo de frecuencias, cm^{-1}	Intensidad
C-H	Alcanos	2.850 – 2.970	Fuerte
		1.340 – 1.470	Fuerte
C-H	Alquenos	3.010 – 3.095	Media
		675 – 995	Fuerte
C-H	Alquinos	330	Fuerte
C-H	Anillos aromáticos	3.010 – 3.100	Media
		690 – 900	Fuerte
O-H	Alcoholes y fenoles (monómeros)	3.590 – 3.650	Variable
	Alcoholes y fenoles (unidos por puentes de H)	3.200 – 3.600	Variable
	Ácidos Carboxílicos(monómero)	3.500 – 3.650	Media
	Ácidos carboxílicos (unidos por puente de H)	2.500 – 2.700	Ancha
N –H	Aminas, amidas	3.300 – 3.500	Media
C=C	Alquenos	1.610 – 1.680	Variable
C=C	Anillos aromáticos	1.500 – 1.600	Variable
C≡C	Alquinos	2.100 – 2.260	Variable
C-N	Aminas y amidas	1.180 – 1.360	Fuerte
C≡N	Nitrilos	2.210 – 2.280	Fuerte
C-O	Alcoholes, éteres, A. carboxílicos, esteres	1.050 – 1.300	Fuerte
C=O	Aldehídos, cetonas, a. carboxílicos, esteres	1.690 – 1.760	Fuerte
NO_2	Nitroderivados	1.500 – 1.570	Fuerte
		1.300 – 1.370	Fuerte

Fuente: SKOOG, A. Douglas. Principios de Análisis Instrumental.

1.3. QUIMIOMETRIA

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 grupos de investigación 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 especifico 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) [29].

1.3.1. Calibración multivariable

La calibración multivariable se puede definir como la actividad de encontrar relaciones entre una o más variables de respuesta “y” y una matriz de variables predictoras “x”, de manera que se cumpla la ecuación (7) [30].

$$y=g(x) \qquad (7)$$

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 lo cual puede existir más de una posibilidad de ajuste de los datos diferenciándose básicamente en la complejidad de la función y en sus parámetros estadísticos.

1.3.1.1. Clasificación de los métodos de calibración multivariable:

Aunque no existe un criterio unificado de los métodos de calibración multivariable, el propuesto por Martens & Naes, [31] ha sido uno de los más aceptados. Su clasificación se basa en los siguientes aspectos fundamentales:

- Según la relación entre las variables dependiente e independiente: ésta relación puede ser descrita mediante un modelo lineal o no lineal.

- Según la forma de encontrar la relación entre las variables: pueden 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 las mezclas de los componentes.

- Según la variable que se defina como dependiente o independiente: si la calibración sigue el criterio directamente 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, [32] sugiere las siguientes etapas para la construcción de modelos de calibración multivariable a partir de mediciones espectrales sobre el analito de interés:

Selección de muestras para la calibración: se debe contar con muestras altamente representativas que contenga la máxima variabilidad física y química esperada en las muestras para las cuales será aplicado el modelo.

Caracterización de muestras de calibración: la caracterización se debe realizar por un método de referencia previamente establecido, el cuál sea altamente confiable y haya sido evaluado estadísticamente.

Toma de espectros infrarrojo: incluye la selección de condiciones experimentales óptimas de adquisición espectral y tratamientos previos de acondicionamiento de la muestra.

Calculo del modelo matemático: en esta etapa se realizan pretratamiento de la señal y aplicación de técnicas de regresión sobre los datos espectrales. Los pretratamientos incluyen: el suavizado para reducir el ruido en 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 diferente orden para extraer información detallada que no puede ser observada en el espectro normal.

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

Validación del modelo 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 validación externa. Si se emplea 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 el desarrollo de métodos de calibración multivariable aplicado a técnicas espectroscópicas de análisis instrumental tiene su origen en el algebra matricial. Si “S” espectros de calibración son medidos a “W” discretas longitudes de onda, es posible construir una matriz de datos espectrales “X” de dimensiones “W x S” que contiene un espectro en columna. De la misma manera es posible construir un vector “Y” de dimensión “Sx1” que contiene los valores de referencia de las muestras de calibración. El objetivo de la calibración multivariable es calcular un vector “p” de dimensión “Wx1” que resuelva la ecuación (8).

$$Y = X^{t} p + e \qquad (8)$$

Donde "X^t" es la transpuesta de la matriz "X" y "e" es un vector de dimensión "S x 1", llamado vector de error. Este último vector se calcula como la diferencia entre los vectores de referencia y los valores estimados por el modelo. Generalmente el vector "p" se estima minimizando la suma de los cuadrados de los errores mediante la ecuación (9).

$$e^t e \;= /e^2/ = \left(y - X^t p\right)^t (y - X^t p) \qquad (9)$$

Ya que normalmente "X" no es una matriz cuadrada, la ecuación (9) no puede ser resuelta directamente. Una alternativa para solucionar esto es determinar la matriz pseudoinversa de "X", "X^+", y calcular el vector de predicción "p" mediante ecuación (10).

$$X^+ y = \left(X \; X^t\right)^{-1} X \; y = p \qquad (10)$$

Habitualmente los espectros de calibración son medidos sobre un amplio rango de frecuencias o longitudes de onda. Esto ocasiona que el número de valores de absorbancia por espectro, W, exceda el número de espectros de calibración, S, haciendo laborioso, y en algunos casos imposible, la estimación del vector p. En este caso es necesario reducir la dimensionalidad de la matriz "X", generalmente mediante un análisis por componentes principales (PCA), y realizar posteriormente una regresión multivariable [33].

1.3.2. Técnicas de Pretratamientos de Datos

1.3.2.1. Suavizado espectral

El 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 a la misma escala, puede ser:

- Normalización por rangos

En esta transformación se normaliza un espectro Xi calculando el área bajo la curva del espectro. Se intenta corregir el espectro de longitud de la trayectoria indeterminada cuando no hay forma de medirla, o aislar a un grupo de un componente constante.

- 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, neutralizando así la influencia de los factores ocultos.

Es equivalente a la sustitución de las variables originales por un perfil centrado alrededor de 1: sólo los valores relativos de las variables que se utilizan para describir la muestra, y la información correspondiente a su nivel absoluto se ha abandonado. Esto se indica en el caso concreto cuando todas las variables se miden en la misma unidad, y sus valores se supone que es proporcional a un factor que no puede ser directamente tomado en cuenta en el análisis.

- Máxima normalización

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

- Propiedad de un máximo de muestras normalizadas:

Si todos los valores son positivos: el valor máximo se convierte en 1.

Si todos los valores son negativos: el valor mínimo se convierte en -1.

Si el signo de los cambios de valores en la curva: o bien el valor máximo se convierte en 1 o el mínimo valor se convierte en -1.

1.3.2.3. Corrección de línea base

La corrección de la línea base es un tipo de pretratamiento que intenta corregir determinadas tendencias en la línea base que aporta el ruido a la señal. Existen varios tipos de corrección de línea base según el efecto que se desea corregir. Un tipo de corrección 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 (Xm) del conjunto de calibración (de cada columna de la matriz), y restar este valor a cada punto (Xi,m).

$$x'i\ ,m = xi\ ,m – xm \qquad (11)$$

siendo x'i,m el dato centrado, xi,m el dato de la fila i y la columna m antes del centrado, xm media de la columna m ($xm = \Sigma xi\ ,m / I$) . La propiedad fundamental de los datos centrados es que el valor medio de cada una de las variables es igual a cero.

1.3.2.5. Derivación

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

- Derivadas (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 suavizado que determina el número de variables adyacentes que se usarán en la estimación de la aproximación polinómica utilizada en la derivación. El algoritmo de Norris, a diferencia del anterior, solo permite el cálculo de derivadas de primer orden [34].

Una desventaja del uso de las derivadas es que disminuyen el valor de la relación señal-ruido, por esta razón, se recomienda realizar un suavizado de la señal antes de la diferenciación de los datos. Otra desventaja es que en ocasiones los modelos de calibración obtenidos mediante datos espectrales tratados con primera o segunda derivada, son menos robustos frente a cambios instrumentales, como por ejemplo derivas de la longitud de onda, que ocurren a lo largo del tiempo, por lo que habría que revisar las calibraciones [35, 36].

1.3.3. Métodos basados en reducción de variables

Estos métodos se basan en que la información contenida en las variables de la señal puede estar contenida 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 espectro 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 autoescalan previamente, uno de estos métodos es PLS (regresión parcial por mínimos cuadrados).

1.3.3.1. Análisis Por Componentes Principales (PCA)

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….Z_n que sean combinaciones lineales de las variables originales X_1, X_2…..X_n, que describen cada muestra, es decir,

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

$$Z_2 = a_{21}X_1 + a_{22}X_2 + a_{23}X_3 + a_{2n}X_n \quad \text{etc.}$$

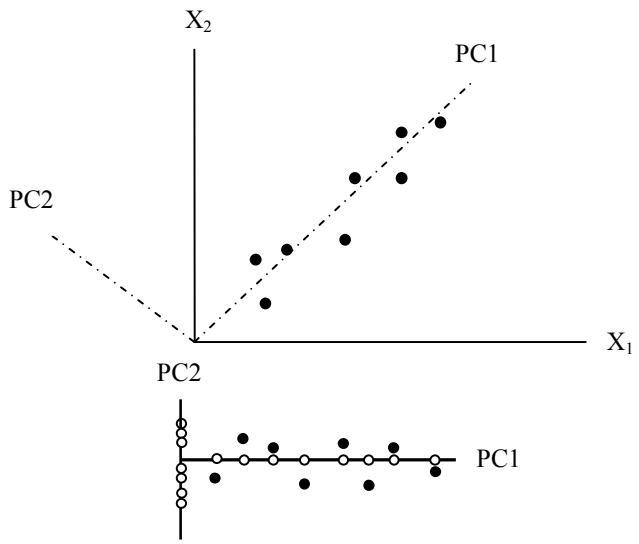
Los coeficientes a_{11}, a_{12}, etc., se eligen de manera que las nuevas variables, a diferencia de las variables originales, no se encuentren correlacionadas unas con otras. De esta forma se obtienen n nuevas variables en lugar de las n originales, y en consecuencia hay una reducción en el conjunto de datos. Sin embargo las componentes principales se eligen de manera que la primera componente principal (PC1), Z_1, recoge la mayor parte de la variación 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 haya correlación significativa el número de PCs útiles será mucho menor que el número de variables originales.

La figura 5 aclara el método cuando sólo hay dos variables y, por tanto, sólo dos componentes principales. En la figura 5a las componentes principales se muestran mediante líneas de trazos suspensivos. Las componentes principales forman ángulos rectos unas con otras, propiedad conocida como **ortogonalidad**. La figura 5b muestra los puntos referidos a estos dos nuevos ejes y también la proyección de los puntos sobre PC1 y PC2. 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.

Figura 5 a) Diagrama que ilustra dos componentes principales, PC1 Y PC2, para dos variables, X_1 y X_2. (b) Puntos referidos a los ejes de los componentes principales. ● Indica ○ puntos de datos, su proyección sobre los ejes.

a)

b)



Fuente: MILLER N. James. Estadística y Quimiometría para Química Analítica.

La figura 5 muestra que el PCA es equivalente a una rotación de los ejes originales, de tal manera que PC1 se encuentra en la dirección de la máxima variación, pero manteniendo el ángulo entre los ejes. Con más de dos variables no resulta posible ilustrar el método gráficamente pero de nuevo se puede pensar en

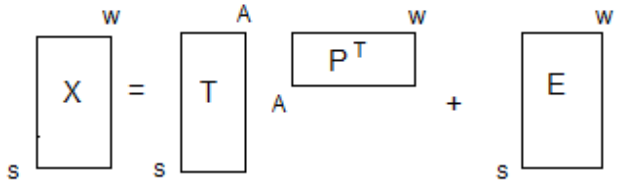
el PCA como una rotación de los ejes de tal manera que PC1 se encuentre en la dirección de máxima variación, PC2 se encuentre en la dirección de la siguiente mayor variación y así sucesivamente [35].

Cuando se emplean 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 para S muestras medidas a W longitudes de onda, se realiza una descomposición por componentes principales que proporciona una aproximación a la matriz X como un producto de dos matrices (ecuación 13): la matriz de puntuaciones (scores) T y la matriz de cargas (loadings) P.

$$X = TPT + E \qquad (13)$$

Donde E es la matriz de residuos de dimensiones S x W.
La matriz T contiene información pertinente a las relaciones entre muestras y está constituida por S filas, que corresponden al número de muestras u objetos, y A columnas, 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

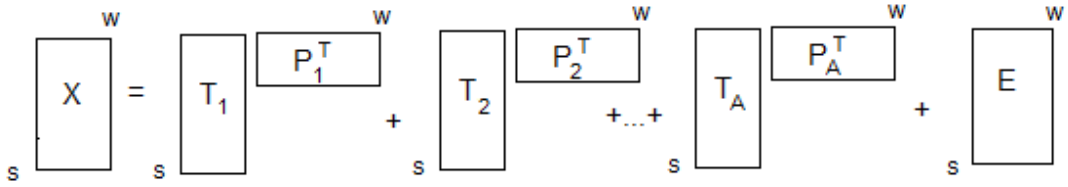


Fuente: Adaptación. Grupo de Quimiometría y Cualimetría de Tarragona, España. Quimiometría: Una disciplina útil para el análisis químico.

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

$$X = t_1 p_1^T + t_2 p_2^T + \ldots + t_a p_a^T + E \qquad (14)$$

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



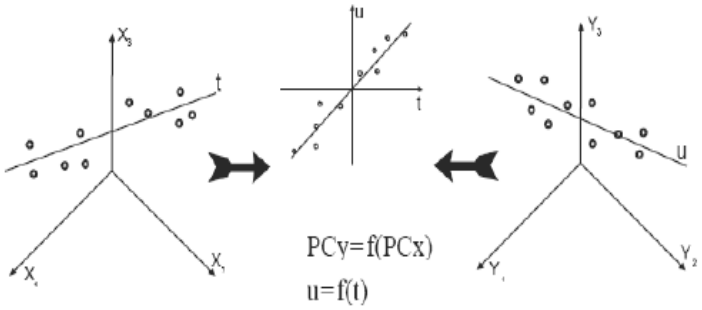
Fuente: Adaptación. Grupo de Quimiometría y Cualimetría de Tarragona, España. Quimiometría: Una disciplina útil para el análisis químico.

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 [37].

1.3.3.2. Regresión por Mínimos Cuadrados Parciales (Regresión PLS)

La regresión por mínimos cuadrados parciales de estructuras latentes (PLS) es un método matemático que modela simultáneamente las matrices X y Y para encontrar un conjunto de variables latentes (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 componentes principales, estas nuevas variables en X y Y se pueden representar como un producto de matrices según se muestra en las ecuaciones (15) y (16).

$$X = TPT + E = \sum t_a p_a^T + E \qquad (15)$$

$$Y = UQT + F = \sum u_a q_a^T + F \qquad (16)$$

Donde:

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

P y Q son las matrices de carga (loadings) 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).

$$u_a = b_a t_a \qquad (17)$$

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

$$b_a = \frac{u_a^T t_a}{t_a^T t_a} \qquad (18)$$

Los coeficientes *ba* 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).

$$Y = TBQ^T + F \qquad (19)$$

En resumen, PLS es utilizado 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 en 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 scores (puntuaciones) para cada uno de los espacios X y Y.

En 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 maximizando el cuadrado de la covarianza entre Y y las posibles combinaciones lineales de X. De este modo se obtienen las variables latentes que contienen la información de la correlación entre las matrices X y Y, presentando una relación más directa con la respuesta, de manera que los primeros componentes principales concentran mayor información predictiva [37, 38].

1.3.4. Factores a Incluir en el Modelo

1.3.4.1. Detección de outliers

Una de las ventajas de los métodos multivariables sobre los tradicionales univariantes, es la capacidad que tienen de detectar la observación u observaciones inconsistentes 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 de 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 [39]. 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:

$$h_i = \frac{1}{I} + t_i^T (T^T T)^{-1} t_i \qquad (20)$$

Donde t_i representa el vector de scores de la muestra i, T, la matriz de scores 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 [40].

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 grande que la que le correspondería si solo influyeran variaciones aleatorias. La interpretación de este grafico es la misma que cualquier gráfico univariante; las muestras fuera de control poseen un valor de T^2 superior al límite, y aparecen más allá de la línea de control [41].

1.3.4.4. Errores de calibración y predicción

Se debe determinar el tamaño óptimo del 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 muestras, 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 I_p muestras que no han participado en la calibración:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{I} (\overline{C}_i - C_i)^2}{I_p}} \qquad (21)$$

Donde C_i y $\hat{C}_i$ representan 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, RMSECV) [42, 43].

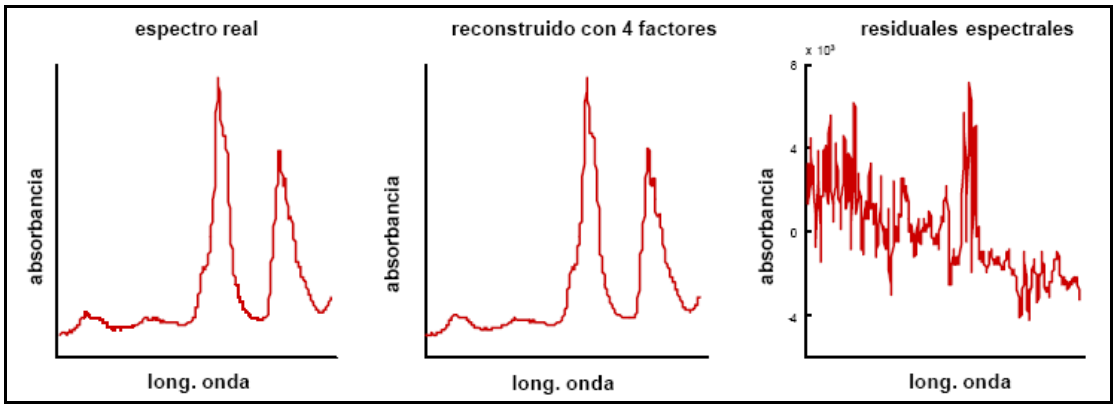
$$RMSECV = \sqrt{\frac{\sum_{i=1}^{I} (\overline{C}_i - C_i)^2}{I}} \qquad (22)$$

1.3.4.5. Residuales en la respuesta instrumental

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

$$E = R - TP^T \qquad (23)$$

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 espectroscopia de infrarrojo y calibración multivariante. Aplicación a la industria petroquímica

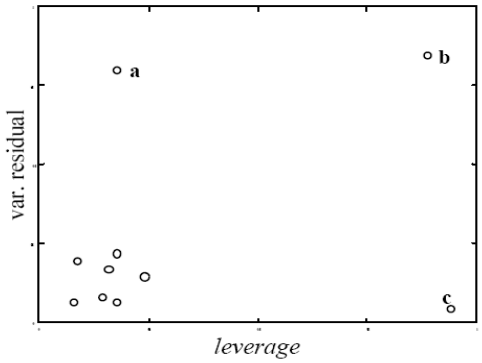
Los residuales en la respuesta se pueden utilizar de varias formas. La más habitual es, para el error en la respuesta de la muestra i, $\mathbf{e}_i$, realizar una suma de cuadrados extendida a las J longitudes de onda y dividir por los grados de libertad (df) adecuados, para obtener una desviación estándar de la muestra i, $s(\mathbf{e}_i)^2$.

Residuales en la propiedad. En la etapa de establecimiento del modelo se dispone del valor de la propiedad de interés determinado por el método de referencia. Los residuales en la propiedad compara el valor predicho por el modelo multivariable $\hat{C}_i$ con el valor considerado verdadero, C, que proporciona el método de referencia.

$$F = C - \hat{C}_i \qquad (24)$$

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 unas muestras hipotéticas (figura 10).

Figura 10. Gráfico del residual frente al *leverage*. (a) objetos con una varianza residual elevada se consideran *outliers,* (b) si además tienen un *leverage* alto son *outliers* peligrosos para el modelo, debido a que tienen mucha 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 la propiedad de interés, por lo que deben ser cuantitativa o cualitativamente validados.

La validación consiste en el análisis de un grupo de muestras independiente al utilizado en la calibración y comprueba que no existe un error sistemático (bias) entre las predicciones que realiza el modelo y los valores proporcionados por el método de referencia. También se mide el grado de concordancia entre las predicciones del modelo y los valores del método de referencia.

1.3.4.7. Predicción de muestras 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 utilizando 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 pueden utilizar medidas del *leverage* de las muestras, y del residual espectral. Herramientas para el control estadístico multivariante, como el estadístico T^2 de Hotelling y el estadístico *Q*, , que se pueden 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 residuales 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 predicción de estas muestras puede diferir significativamente del valor verdadero.

1.4 APLICACIÓN DE LA ESPECTROSCOPIA FTIR-ATR EN EL PETRÓLEO

La técnica de espectroscopia 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 [44].

Los trabajos realizados usando espectroscopia infrarroja media en muestras de petróleo, han demostrado que es la región del IR que presenta menor solapamiento de bandas lo que genera mejores resultados. Entre las aplicaciones más sobresalientes de esta técnica se encuentra 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, presentan bandas características de aromáticos, cadenas parafínicas y estructuras nafténicas 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, 45, 46].

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 [16, 17, 45] resinas [5,17], gasolinas [46], queroseno [28], diesel [47] y petróleo crudo [1, 3, 4, 40] muestran el potencial que ha desarrollando la espectroscopia de FTIR para esta clase de analitos.

2. PARTE EXPERIMENTAL

2.1. MUESTRAS

Se estudió cincuenta crudos procedentes de diferentes zonas de explotación petrolífera de Colombia que fueron provistos por el Instituto Colombiano de Petróleo (ICP), junto con el análisis SARA (tabla 4), el cuál fue determinado por, precipitación de los asfaltenos y posterior cromatografía de exclusión; este proceso fue desarrollado en el laboratorio de Geoquímica del ICP aplicando la norma ASTM D- 4124.

Tabla 4. Análisis SARA de los crudos analizados

Muestra Nº	S (%w)	A (%w)	R (%w)	A (%w)
1	30,71	31,94	33,78	3,57
2	35,77	33,73	27,09	3,41
3	36,81	35,18	24,90	3,11
4	40,02	30,26	26,85	2,87
5	38,12	34,83	24,95	2,10
6	33,57	30,56	33,57	2,30
7	36,73	34,79	26,52	1,96
8	39,18	30,53	24,42	5,87
9	37,15	30,96	30,40	1,49
10	42,55	28,21	25,44	3,80
11	42,50	30,00	26,88	0,62
12	46,82	28,09	24,75	0,34
13	45,22	25,03	29,07	0,68
14	38,40	32,49	28,80	0,32
15	41,63	24,64	33,13	0,60
16	43,89	31,46	24,14	0,51
17	38,50	32,53	28,55	0,42
18	46,66	24,52	28,47	0,36
19	46,61	26,99	26,17	0,23
20	39,41	29,56	30,32	0,72
21	49,17	21,67	20,00	9,15
22	34,14	29,87	25,61	10,38
23	62,28	24,91	11,80	1,01
24	66,21	21,19	11,48	1,12
25	49,86	20,43	23,70	6,01
26	34,41	25,03	27,53	13,04
27	42,52	25,05	25,48	6,95
28	27,23	23,34	35,79	13,63
29	35,20	24,43	28,02	12,35
30	34,00	27,82	25,97	12,21
31	28,59	22,32	36,26	12,83
32	59,73	31,51	8,60	0,17
33	42,73	30,51	25,79	0,97
34	53,99	26,11	18,89	1,01
35	20,85	37,31	26,89	14,95
36	17,83	36,71	28,84	16,62
37	15,01	32,69	34,39	17,91
38	20,34	41,23	25,29	13,15
39	28,49	40,62	16,34	14,55
40	31,30	35,53	21,99	11,18
41	30,54	34,36	24,81	10,29
42	28,55	34,17	26,88	10,39
43	20,46	36,13	37,87	5,55
44	21,27	40,35	32,17	6,21
45	32,09	35,63	29,04	3,24
46	26,57	38,18	28,87	6,37
47	36,91	26,45	30,85	5,79
48	27,36	28,01	39,08	5,55
49	31,03	26,50	34,91	7,56
50	33,23	25,92	34,56	6,30
Promedio	*36,99*	*30,10*	*26,98*	*5,93*
Max	*66,21*	*41,23*	*39,08*	*17,91*
Min	*15,01*	*20,43*	*8,60*	*0,17*

2.1.1. Tratamiento de muestras

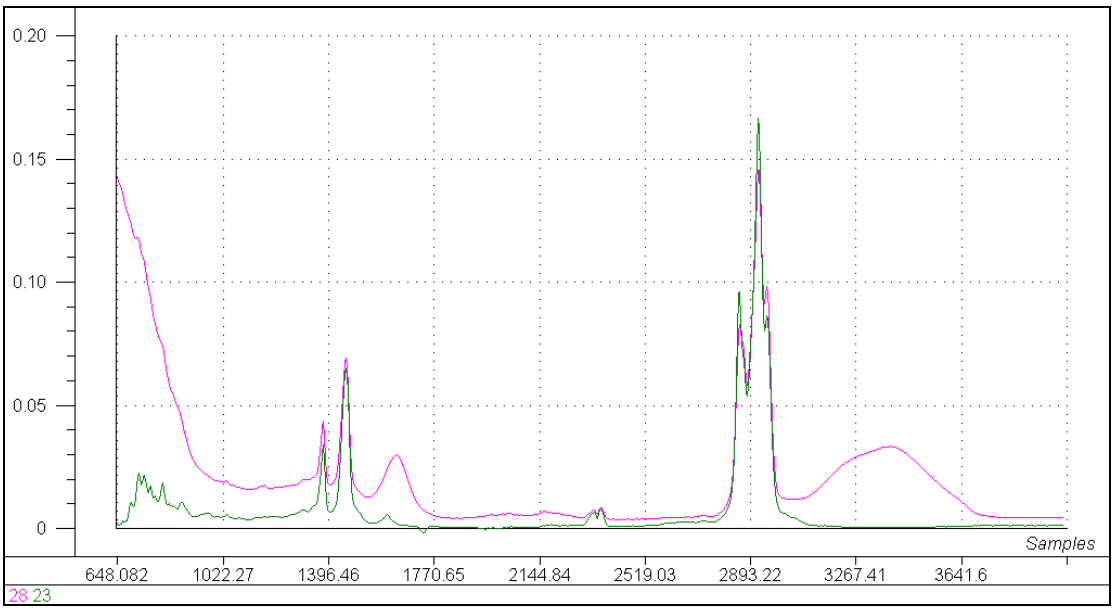
Debido a que los crudos contienen agua en diferentes proporciones tanto de emulsiones regulares (emulsiones tipo agua en aceite) como agua libre (agua no emulsionada, que se separa en menos de 5 minutos del crudo); se realizó una primera toma de espectros para determinar que muestras presentaban mayor hidratación, ya que el contenido de agua se puede observar en un espectro IR, porque ésta presenta momento dipolar no nulo y todos sus modos vibracionales son activos. En el MIR presenta:

Tensión simétrica = 3651.5cm^{-1}

Flexión simétrica = 1595 cm^{-1}

Como se puede observar en la figura 11. El efecto del agua en el espectro IR de las muestras es grande, por lo que fue necesario realizar una deshidratación del agua libre por centrifugación en aquellas muestras que presentaron mayor contenido de ésta, para minimizar los errores en el tratamiento de los datos.

Figura 11. Ejemplo de la comparación de espectros IR de muestras hidratadas (rosada) y no hidratadas (verde).



Se determinó por inspección de los espectros que las muestras 28, 29 y 30 presentaban mayor proporción de hidratación que las demás, por lo que se les realizó una deshidratación por centrifugación.

a) Determinación del porcentaje de hidratación (%BSW) inicial: Esta prueba se aplicó para comparar el % de hidratación antes y después de deshidratar. Proceso:

En un tubo de centrífuga milimetrado se agregó 15ml de varsol, 2 gotas de rompedor (mezcla de solventes) que permite separar las fases y finalmente se agregó crudo hasta llegar a un total de 35ml de volumen. Se centrifugó durante 10 minutos a 1800rpm. Se observó el volumen de agua presente en el fondo del tubo y se determinó el *%BSW* por la ecuación (25).

$$\%BSW = ml\, H_2O * 50 \qquad (25)$$

b). Deshidratación: En tres tubos de centrífuga de 100ml se agregaron los crudos 28, 29 y 30 y se centrifugaron por 15 minutos a 1500 rpm; a continuación se extrajo de cada tubo aproximadamente 70ml y se eliminó el agua depositada en el fondo, una vez extraída el agua se centrifugó nuevamente. Este procedimiento se repitió 3 veces para cada crudo.

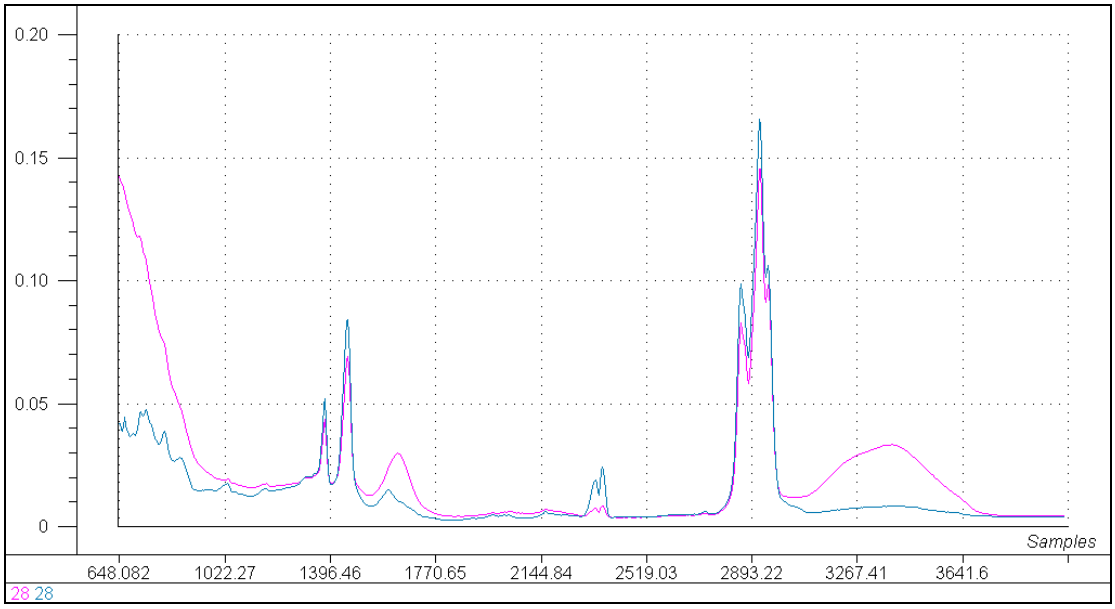
Al crudo deshidratado obtenido se determinó el %BSW. Tabla 5.

Tabla 5. Comparación de los %BSW de las muestras antes y después de deshidratar

Muestra	% BSW inicial	%BSW final	% Deshidratación
28	30	10	66.6
29	50	25	50
30	45	25	44.4

Después de deshidratar las muestras se repitió la toma de espectros, observando el efecto de la deshidratación. (Ver Figura 12).

Figura 12. Comparación de los espectros de la muestra 28 antes (fucsia) y después de deshidratar (azul).



2.2. ESPECTROSCOPIA MIR-ATR

2.2.1. Instrumentación

Se empleó el espectrómetro IR Prestige-21 de la marca Shimadzu, (figura 13a) equipado con una celda ATR PIKE miracle (figura 13b) con cristal de diamante de reflexión simple de índice de refracción 2,4 y un detector con control de temperatura (DLATGS) que permite cubrir las regiones del infrarrojo medio en el rango de 4000 a 650 cm^{-1}, del Laboratorio de Química de Producción del ICP. El equipo cuenta, adicionalmente, con un divisor de haz (Beam splitter) de germanio cubierto con KBr, este último es usado porque no absorbe en la región IR y un sistema de recirculación de nitrógeno para control de la atmósfera; además están

disponibles una serie de accesorios para la celda, que permiten el análisis de muestras liquidas y sólidas. El equipo fue operado mediante la aplicación IR Solution.

Figura 13. Sistemas de caracterización MIR

a. Espectrómetro

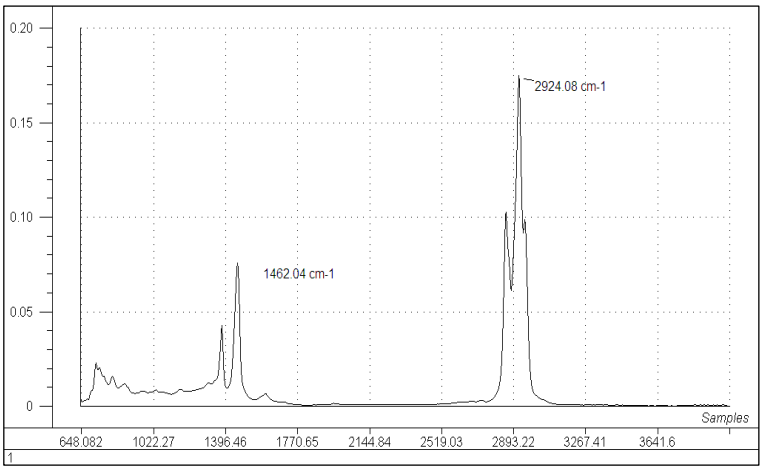
b. Celda ATR

2.2.2. Verificación del desempeño del Espectrómetro

El desempeño del espectrómetro MIR fue verificado realizando una prueba de repetibilidad en las medidas de absorbancia; establecida dentro de los protocolos de funcionamiento del equipo.

La repetibilidad del espectrómetro se determinó a partir del cálculo de las desviaciones estándar para las absorbancias medidas a 2924.08cm^{-1} y 1462.04cm^{-1} en 9 espectros adquiridos sobre la muestra de crudo 1 perteneciente a las muestras de calibración. Las desviaciones halladas para cada pico de absorción fueron inferiores a 0.001 (figura 14), asegurando, de esta manera la repetibilidad de medición del espectrómetro.

Figura 14. Prueba de repetibilidad del Espectrómetro MIR



Espectro	A 1462.04	A 2924.08
1	0.0726	0.1719
2	0.0727	0.172
3	0.0727	0.1718
4	0.0727	0.172
5	0.0726	0.172
6	0.0727	0.172
7	0.0727	0.172
8	0.0726	0.1719
9	0.0727	0.172
Medida	0.07	0.17196
Desviación	$4.83E^{-5}$	$6.99E^{-5}$

2.2.3. Adquisición de espectros MIR

Los espectros MIR fueron tomados en el Laboratorio de Química de Producción del Instituto Colombiano del Petróleo. Cada espectro fue el resultado de las siguientes condiciones: 32 barridos realizados en el rango de 4000 – 650 cm^{-1}, resolución de 8 cm^{-1}, ángulo de incidencia del haz IR 45°, velocidad de espejo del interferómetro de 2.8 mm/s a una temperatura de 25°C; estas condiciones se utilizaron debido a que fueron las que mostraron la mejor relación señal/ruido en los espectros.

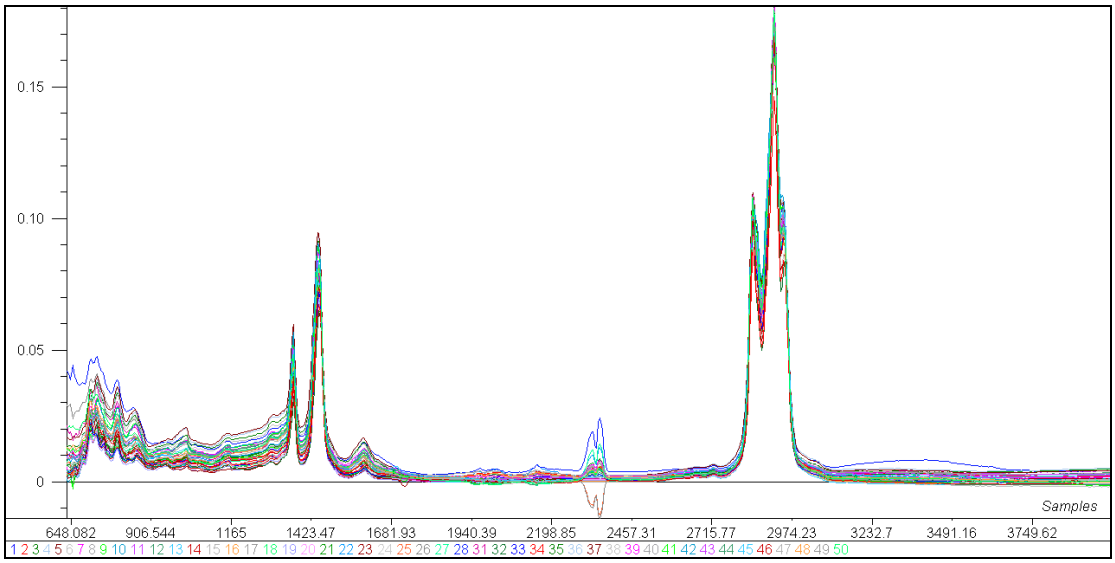
2.2.3.1. Análisis de los espectros obtenidos

En la figura 15 se muestran los 50 espectros IR medio de las muestras de crudo. Las bandas más intensas se encuentran cerca de 3000 cm^{-1}: la tensión de C-H en CH_3 a 2956 cm^{-1} y 2872 cm^{-1}, la tensión de C-H en CH_2 a 2926 cm^{-1} y 2853 cm^{-1}. También tienen una intensidad importante las bandas a 1495 cm^{-1} y 1385 cm^{-1}, que corresponden a torsiones de tijera (simétrica y 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 tomará en cuenta en los procesos multivariables.

Figura 15. Espectros Originales de 50 muestras de Crudo.



2.3. PROCESAMIENTO DE DATOS

Los datos espectroscópicos obtenidos fueron analizados en el software Unscrambler 9.7 licenciado por Ecopetrol S.A. 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 entre otros.

Debido a que el software Unscrambler presenta diferentes opciones para el pretratamiento de los datos como tres tipos de normalización, segunda y primera derivada entre otras [48]. Como primera medida se realizó un análisis exploratorio

de los datos espectrales determinando el valor de la desviación estándar media (RSD) a tres número de onda sobresalientes de los espectros después de aplicar los procesos opcionales, como se muestra en la tabla 6.

Tabla 6. Cálculo de la RSD para tres números de onda característica aplicando diferentes procesamientos

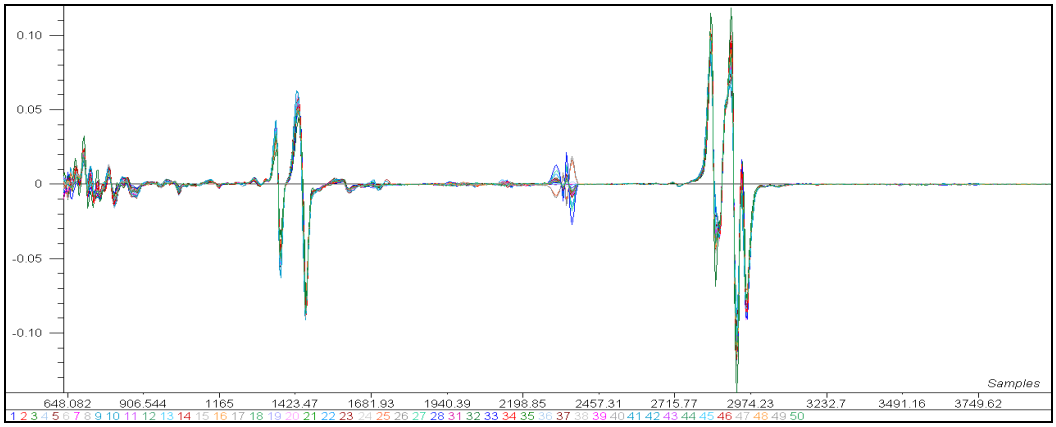
Número de Onda (cm-1)	721.377	1465.9	2927.94
Datos Originales	26.93	7.74	3.73
Normalización Máxima	26.62	6.12	1.36
Normalización Rango	26.50	6.38	2.03
Normalización Media	45.98	12.84	16.96
Primera Derivada	16.44	9.57	10.32
Segunda Derivada	24.86	12.01	11.58
primera Derivada + N. rangos	15.69	5.06	3.16

Debido a que el ruido está asociado con la dispersión de los valores de RSD [49], se puede deducir que los pretratamientos que hacen la mejor corrección de los espectros es el uso de la normalización por rangos seguida de una derivación de primer orden (derivación Norris, Figura 16), ya que esta combinación generó los menores valores de RSD para las bandas evaluadas (Tabla 6).

La normalización por rangos 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 del valor máximo menos el mínimo.

La derivación Norris 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 16. Espectros Normalizados y derivados



La corrección de la línea base y el suavizado de los espectros fueron pretratamientos omitidos debido a que el software IR solution los hace automáticamente como se puede observar en los espectros originales (figura 15).

3. RESULTADOS Y ANÁLISIS

A continuación se describe el desarrollo y los principales resultados obtenidos en las etapas de calibración y validación de los modelos de predicción.

3.1. ANÁLISIS POR COMPONENTES PRINCIPALES (PCA)

En primer lugar se realizó una descomposición en componentes principales del conjunto total de espectros corregidos por el pretratamiento realizado anteriormente. Este conjunto de 50 muestras está comprendido por crudos de diferentes campos de explotación petrolífera de Colombia.

Con el PCA se definieron los datos en los que los modelos de predicción se fundamentaron (número de componentes), la complejidad de este y la forma de validarlo. (Tabla 7)

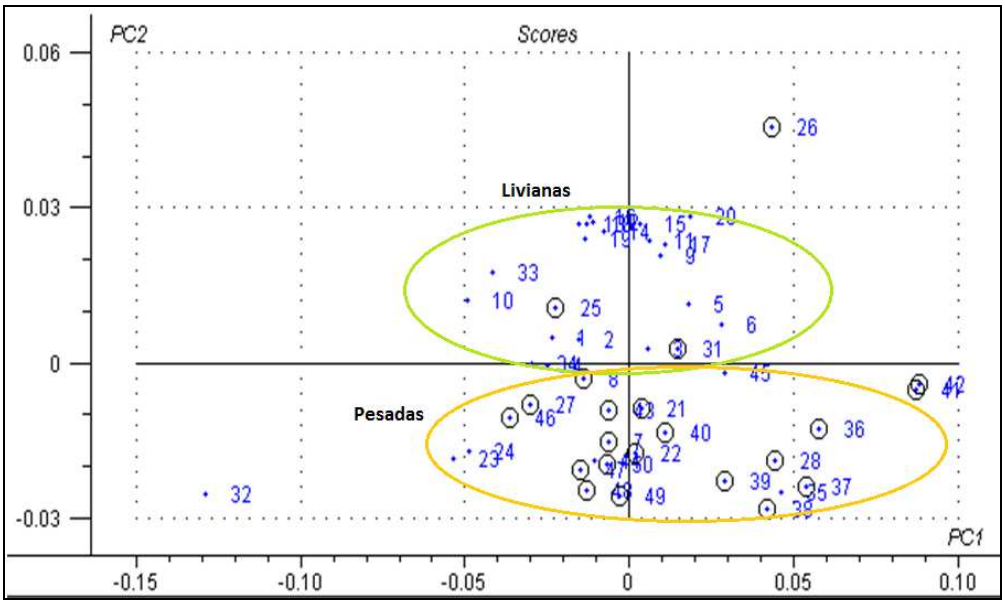
Tabla 7. Varianza explicada por los componentes principales de las 50 muestras de crudo

Componente principal	Varianza Explicada (%)	Varianza Explicada Acumulada (%)
PC 1	58.46	58.46
PC 2	16.02	74.48
PC 3	12.09	86.57
PC 4	4.016	90.58
PC 5	1.861	92.45
PC 6	1.586	94.03
PC7	0.984	95.04
PC8	0.95	95.99

Observando el porcentaje de varianza de los datos espectrales que explica los primeros 8 componentes principales se determinó que 7 es el número máximo de componentes a utilizar en los modelos de predicción ya que a partir del PC8 el aumento de la varianza explicada acumulada es menor a 0.95%.

La representación de los dos primeros componentes principales (PC1 y PC2) se muestran en la figura 17.

Figura 17. Gráfica de Scores de los primeros componentes principales de las 50 muestras

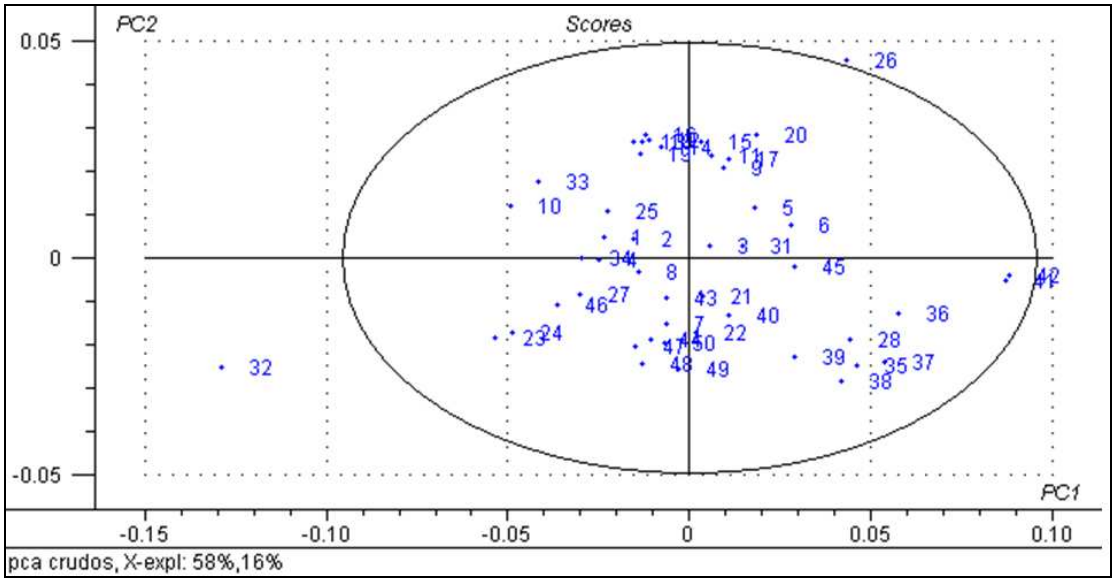


La figura de scores, es un gráfico de dispersión en dos dimensiones de los componentes principales PC1 y PC2 en los cuales se recoge el 58% y el 16% de varianza respectivamente, lo que indica que estos dos componentes contienen el 74% de la varianza total de las intensidades espectrales. Además éste permite establecer las relaciones entre las muestras según su cercanía en el plano PC1-PC2, ya que cuanto más cerca están las muestras al eje de coordenadas del grafico, más 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 (muestras marcadas) se agruparon en el tercer y cuarto cuadrante de la gráfica de scores. De acuerdo con lo anterior se pudo establecer que las muestras se pueden dividir en dos grandes grupos; pesadas y livianas según su composición de asfaltenos.

Mediante la aplicación de Hotelling T^2 elipse, se determinó si las muestras 26, 32, 41 y 42 eran potenciales outliers (puntos fuera de la elipse o muestras atípicas), debido a que estas se ubicaron fuera de los dos grupos establecidos en el gráfico de scores (Figura 18).

Figura 18. Estadístico T^2 aplicado a la gráfica de Scores de las 50 muestras

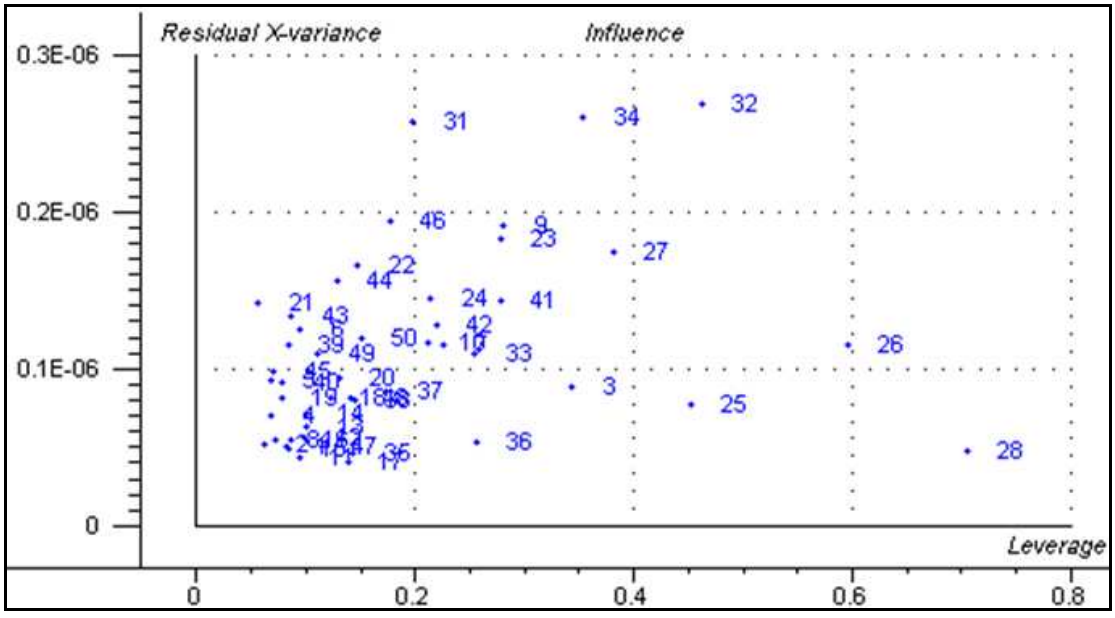


Con el estadístico T^2 Hotelling se comprobó que las muestras 26 y 32 son outliers, la primera es outliers posiblemente por efecto de propiedades fisicoquímicas diferentes a la composición SARA, ya que ésta cae dentro del promedio; y la

segunda posiblemente porque es la muestra que presenta el mínimo valor de composición de resinas y asfaltenos 8.6 y 0.17 respectivamente.
Las muestras 41 y 42 quedaron dentro de la elipse, demostrando, que a pesar de estar alejadas de las demás, éstas presentan similitud con algunas muestras en la composición de asfaltenos (pesadas).

El gráfico de Influence (figura 19) permite clasificar de forma más detallada los valores atípicos (outliers), influyentes y peligrosos (dangerous outliers) de las muestras analizadas.

Figura 19. Grafica de Influencia para las 50 muestras

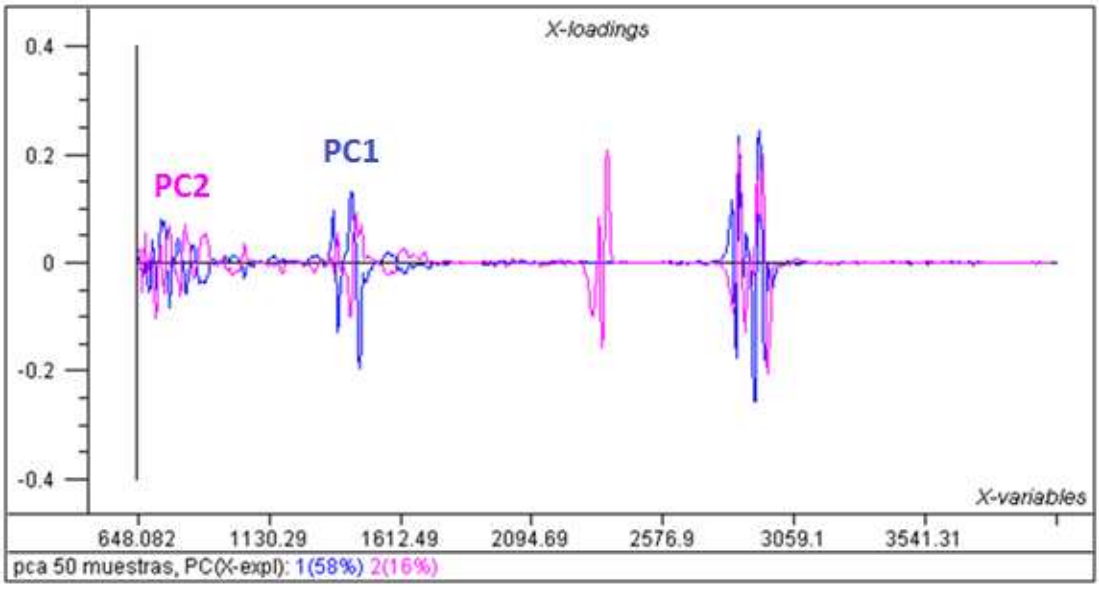


Como las muestras 31 y 34 presenta alta varianza residual (parte superior izquierda del gráfico) se definen como probables outliers. Las muestras que presentan alto leverage, es decir las que extienden a la derecha de la gráfica son influyentes, lo que significa que son las que mejor describen el modelo de predicción; para este caso se observó la muestra 28.

Las muestras que presentan alta varianza residual y alto leverage como la muestra 32 se definen como dangerous outliers, lo que significa que no serán descritas por el modelo que describe la mayoría de las muestras. Éstas distorsionan el modelo con el fin de estar mejor descritas, lo que significa que el modelo tenderá a centrarse en la diferencia entre estas muestras y las demás, en lugar de describir más características generales comunes a todas las muestras.

La gráfica de X- loadings (figura 20) se utilizó para determinar los números de onda de mayor importancia los cuales se encuentran aproximadamente en los rangos de 690-1770 cm^{-1} y de 2646 – 3320cm^{-1} del espectro IR. Las señales presentes en 648 – 690 cm^{-1} y en 2333cm^{-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 20. Gráfico de X-Loadings de PC1 y PC2



La gráfica de X- loading también permite visualizar que regiones espectrales explican los vectores PC1 y PC2 en mayor proporción. Como se puede observar el

PC1 es el que presenta mayor intensidad en el rango de 1600-1300cm^{-1} y de 3000-2800 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 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.

Debido a la diferente naturaleza de los crudos se estableció que lo más recomendable es dividir las 50 muestras en los dos grupos determinados en los resultados anteriores y generar el PCA para cada uno.

3.1.1. Análisis por componentes principales para el grupo de muestras livianas

Se determinó según el contenido de asfaltenos que el grupo de las muestras livianas está comprendido por 25 muestras que son: 1-7, 9-20, 23, 24, 32-34 y 45.

El porcentaje de varianza explicado por cada componente principal se muestra en la tabla 8.

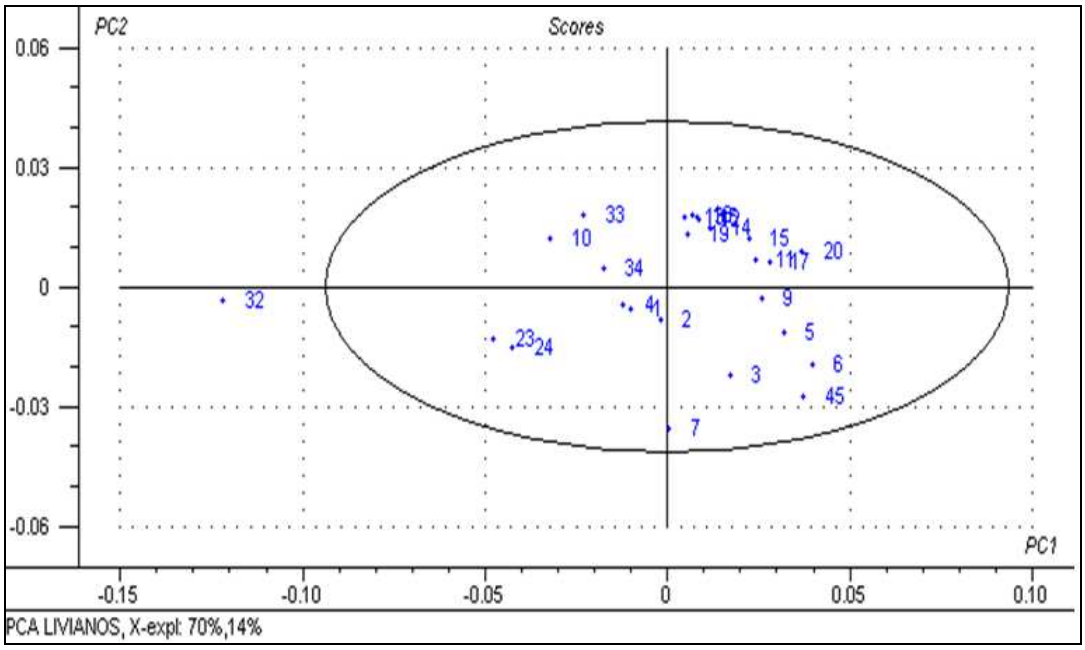
Tabla 8. Varianza explicada por los componentes principales de las muestras livianas

Componente principal	Varianza Explicada (%)	Varianza Explicada Acumulada (%)
PC 1	70.34	70.34
PC 2	14.18	84.52
PC 3	2.85	87.37
PC 4	2.2	89.57
PC 5	1.17	90.78
PC 6	1.17	91.95

Observando el porcentaje de varianza de los datos espectrales que explica los primeros 6 componentes principales se determinó que 5 es el número máximo de componentes a utilizar en los modelos de predicción ya que a partir del PC6 el aumento de la varianza explicada acumulada es bajo.

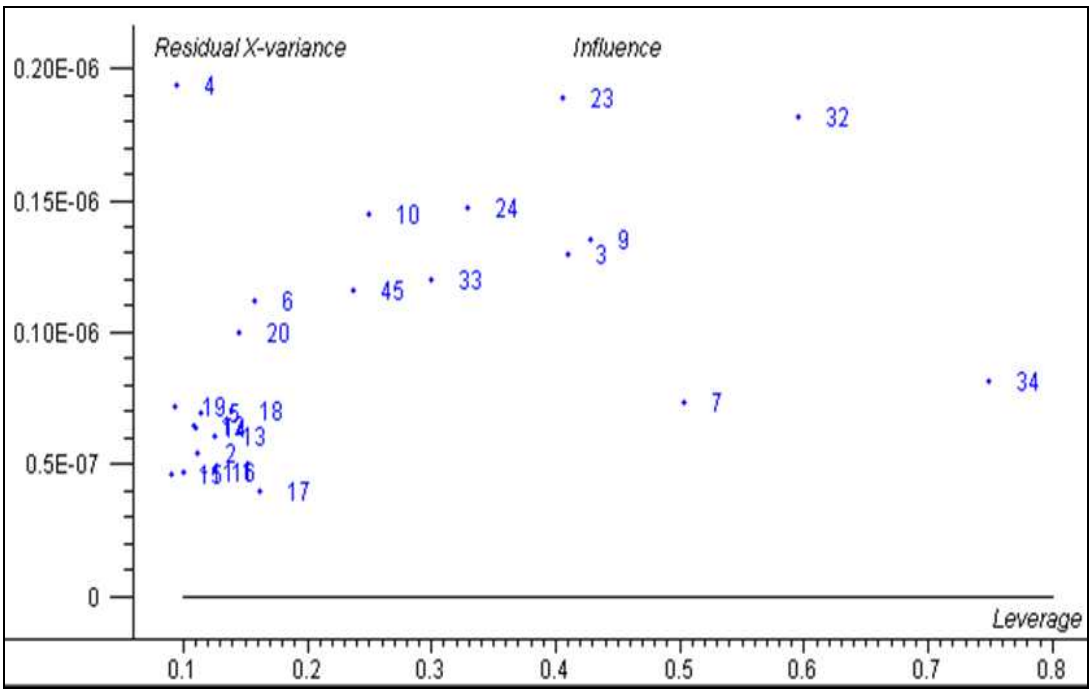
El gráfico de scores con la aplicación del estadístico de Hotelling T^2 (Figura 21) permite visualizar que la muestra 32 se aleja considerablemente de las demás; analizando las propiedades disponibles (composición SARA) se estableció que esto se debe a que dicha muestra presenta la menor composición porcentual en resinas y asfaltenos (límites inferiores del rango de calibración), por lo que se define como un potencial outliers, y es posible que deba sacarse de la calibración de alguno de los modelos de predicción.

Figura 21. Estadístico T^2 aplicado a la gráfica de Scores de las muestras livianas



El gráfico de Influence (figura 22) permite clasificar de forma más detallada los valores atípicos (outliers), influyentes y peligrosos (dangerous outliers) de las muestras analizadas.

Figura 22. Grafica de Influencia para muestras livianas



Como la muestra 4 presenta alta varianza residual (parte superior izquierda del gráfico) se define como un probable outlier, posiblemente por efecto de propiedades fisicoquímicas diferentes a la composición SARA , ya que esta se encuentra dentro del promedio para todas las fracciones. 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 describirá el modelo de predicción; para este caso se observó las muestras 7 y 34.

Las muestras que presentan alta varianza residual y alto leverage como las muestras 23 y 32 se definen como dangerous outliers, lo que significa que no

serán descritas por el modelo que describe la mayoría de las muestras, éstas distorsionan el modelo con el fin de estar mejor descritas.

3.1.2. Análisis por componentes principales para el grupo de muestras pesados.

Se determinó según el contenido de asfaltenos, que el grupo de las muestras pesadas está comprendido por 25 muestras que son: 8, 21,22, 25-28, 29-31, 35-44, 46-50.

Observando el porcentaje de varianza de los datos espectrales que explica los primeros 8 componentes principales se determinó que 6 es el número máximo de componentes a utilizar en los modelos de predicción ya que a partir del PC6 el aumento de la varianza explicada acumulada es bajo.(Tabla 9)

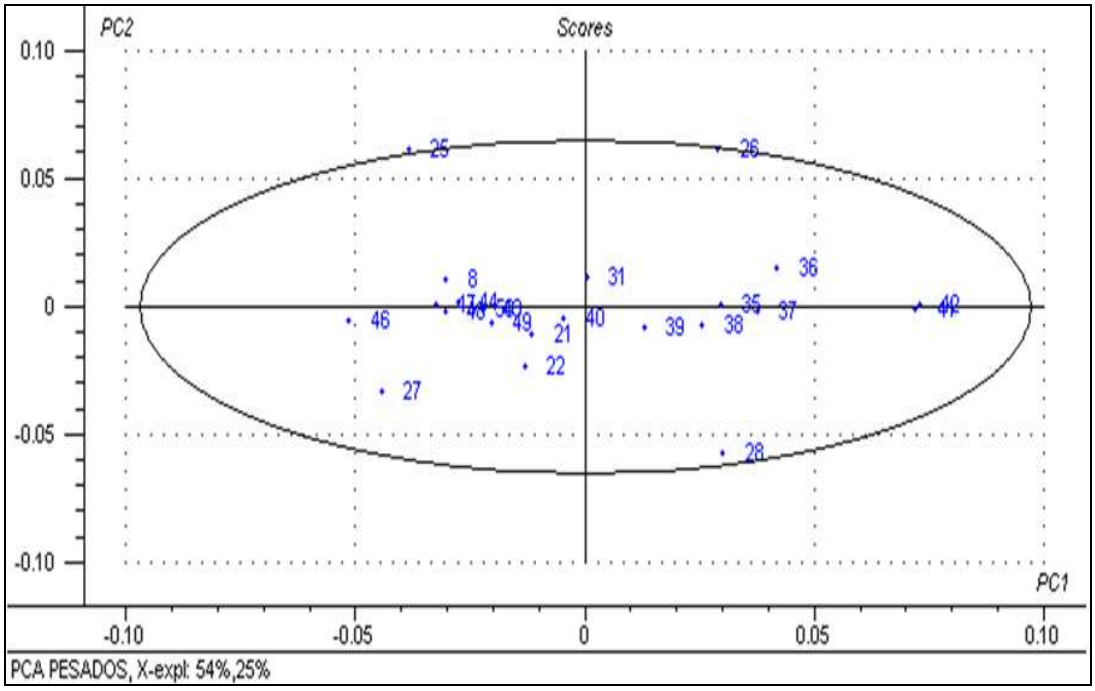
Tabla 9. Varianza explicada por los componentes principales de las 25 muestras pesadas

Componente principal	Varianza Explicada (%)	Varianza Explicada Acumulada (%)
PC 1	54,14	54,14
PC 2	25.03	79,17
PC 3	10.01	89.18
PC 4	3.4	92.66
PC 5	1.91	94,57
PC 6	1.53	96.10
PC 7	1.04	97,14
PC8	0,66	97,8

Mediante el gráfico de scores con la aplicación del estadístico de Hotelling T^2 (Figura 23) se observó que las muestras 25 y 26 se encuentra sobre la elipse, y no fue posible con esto determinar si son potenciales outliers que deban sacarse de las muestras de calibración de alguno de modelos de predicción; por ende se

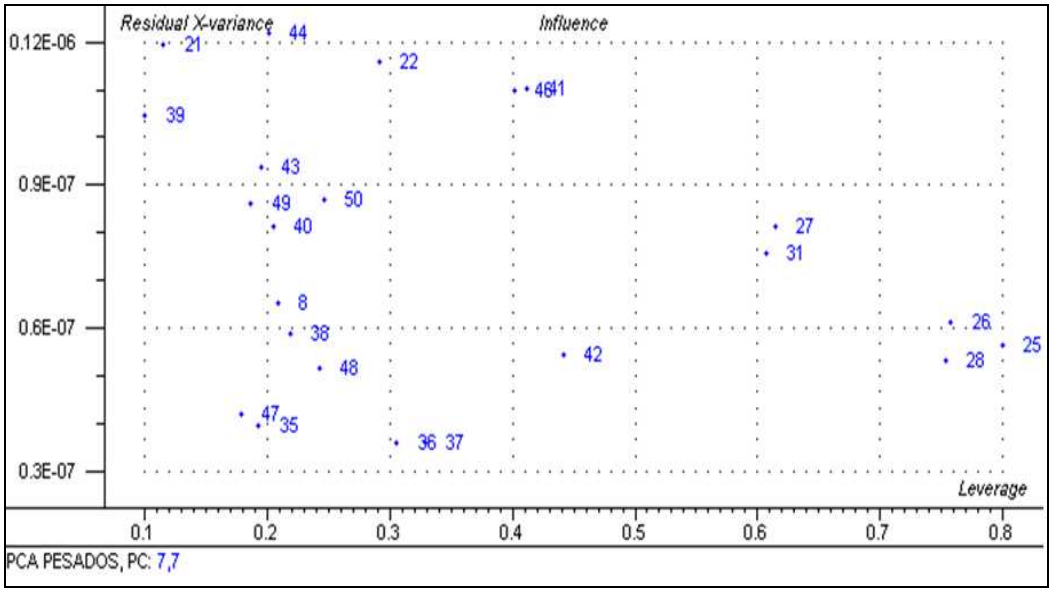
realizó el análisis en la grafica de influence (figura 24) para verificar dicho comportamiento de estas muestras.

Figura 23. Estadístico T^2 aplicado a la gráfica de Scores de muestras pesadas



En la gráfica de leverage se observa que las muestras 21, 39, 44 presentan alta varianza residual (parte superior izquierda del gráfico) se definen como probables outliers. Las muestras que se extienden a la derecha de la gráfica son influyentes, lo que significa que son las que mejor describirá el modelo de predicción para pesados; en este caso las muestras 25, 26 y 28; descartando de esta manera que la muestra 25 y 26 fueran outliers.

Figura 24. Grafica de Influencia para muestras pesadas



En la figura 24 no se observó muestras con alta varianza residual y alto leverage, es decir, no hay posibles dangerous outliers.

3.2. DESARROLLO DE MODELOS PLS

Mediante el algoritmo de regresión PLS, y empleando el programa de análisis multivariado THE UNSCRAMBLER versión 9.7 (CAMO), se evaluó el potencial de la espectroscopia MIR-ATR para predecir la composición en %W de la fracción SARA presente en crudos colombianos.

El desarrollo de los modelos de predicción abarcó dos grandes etapas: calibración y validación. A continuación se describe en detalle el procedimiento empleado, tomando como ejemplo el desarrollo del modelo para determinar el % W de asfaltenos, usando las 50 muestras disponibles en la calibración, las cuales presentan un rango de 0.17 a 17.41 en %W de asfaltenos. Para las demás fracciones se siguió la misma metodología, por tanto serán mostrados los resultados más relevantes en la sección 4.

3.2.1. Calibración del modelo PLS para asfaltenos

Teniendo en cuenta el número óptimo de componentes principales y las regiones espectrales más influyentes, fueron desarrollados modelos de predicción variando el número de muestras de calibración en función del análisis PCA descrito en la sección 3.1. La selección del modelo de predicción se realizó evaluando parámetros estadísticos como el RMSEP, RMSEC y la varianza explicada. (Tabla 10).

Tabla 10. Parámetros estadísticos del modelo seleccionado para la predicción de %W de Asfaltenos

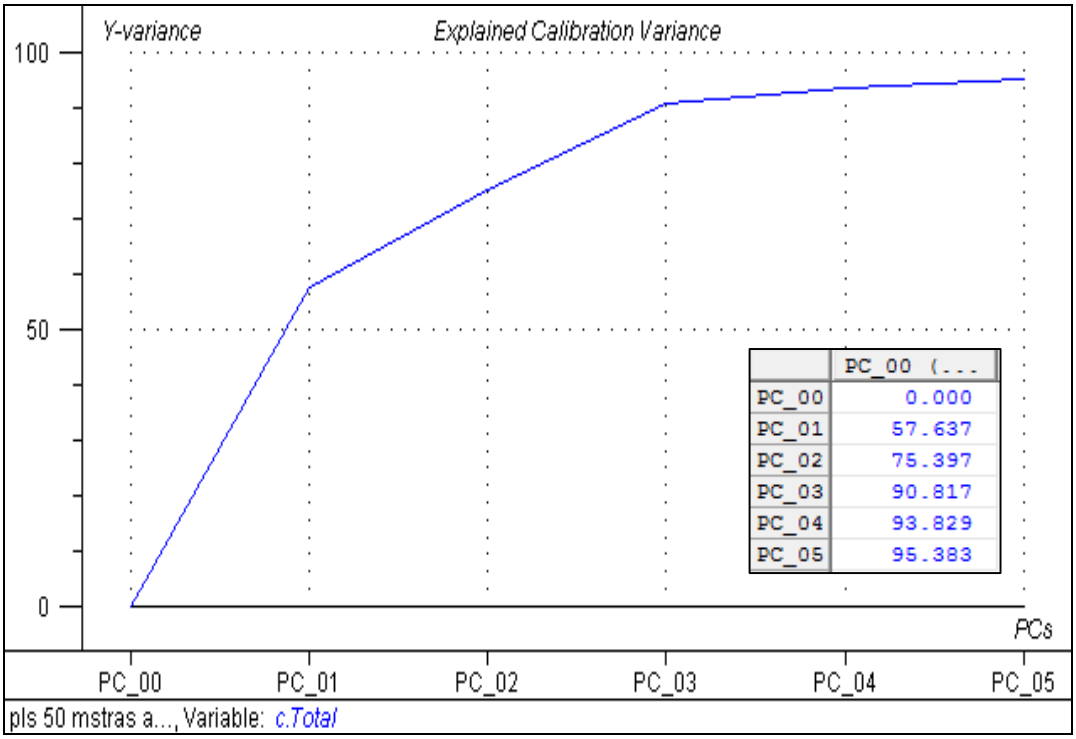
Rangos usados (cm^{-1})	Muestras excluidas	Componentes	Varianza Explicada	RMSEC (%)	RMSEP (%)
686.801-1770.65 2646.54-3321	32, 26	5	98	1.08	1.5

Éste modelo (tabla 10) fue seleccionado por presentar el menor % de error tanto de predicción (RMSEP) como de calibración (RMSEC) y la mayor varianza explicada por los componentes (98%).

Las muestras 26 y 32 fueron excluidas del modelo de predicción, por diferentes razones como ser potenciales outliers, entre otras descritas anteriormente en el análisis por componentes principales (PCA).

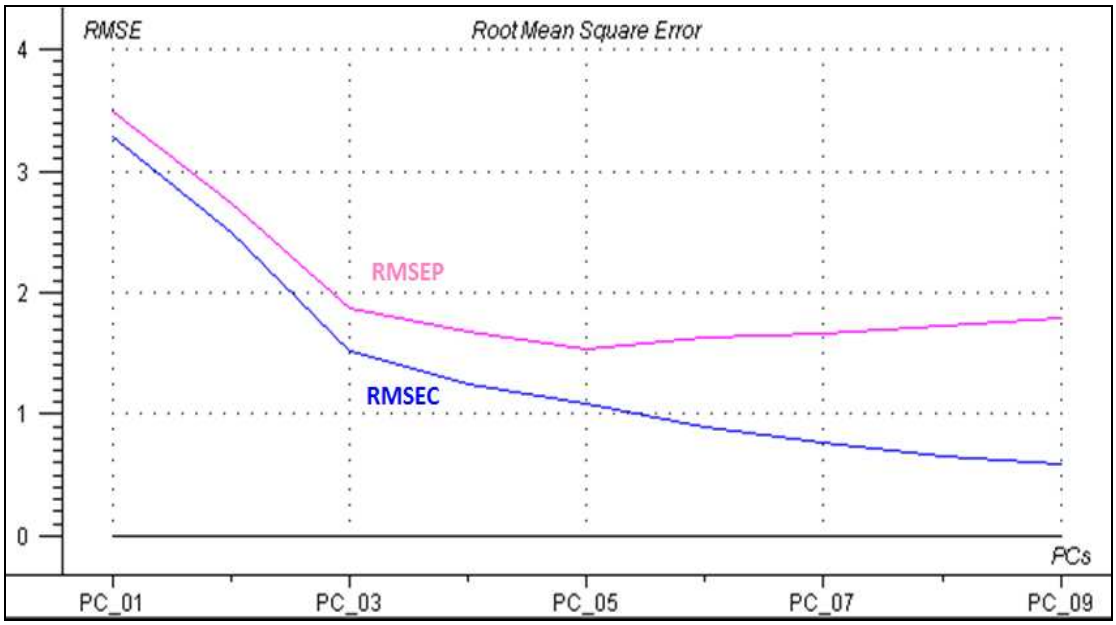
El modelo de predicción explica en las regiones espectrales comprendidas entre 690-1770cm^{-1} y 2646-3321cm^{-1}, más del 98% en la variabilidad de los datos (variable Y) a partir del espectro MIR, con un mínimo de cinco componentes principales. (Figura 25)

Figura 25. Varianza explicada en el modelo PLS de asfaltenos



Otra gráfica que permite visualizar el software, con la que se comprobó el número óptimo de componentes principales (PCs) para generar el modelo de predicción, es la de la raíz cuadra del error de predicción (RMSEP) y calibración (RMSEC). Figura 26; 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 muestras dentro del grupo de calibración para desarrollar el modelo de predicción y posteriormente evaluarlo sobre la muestra restantes.

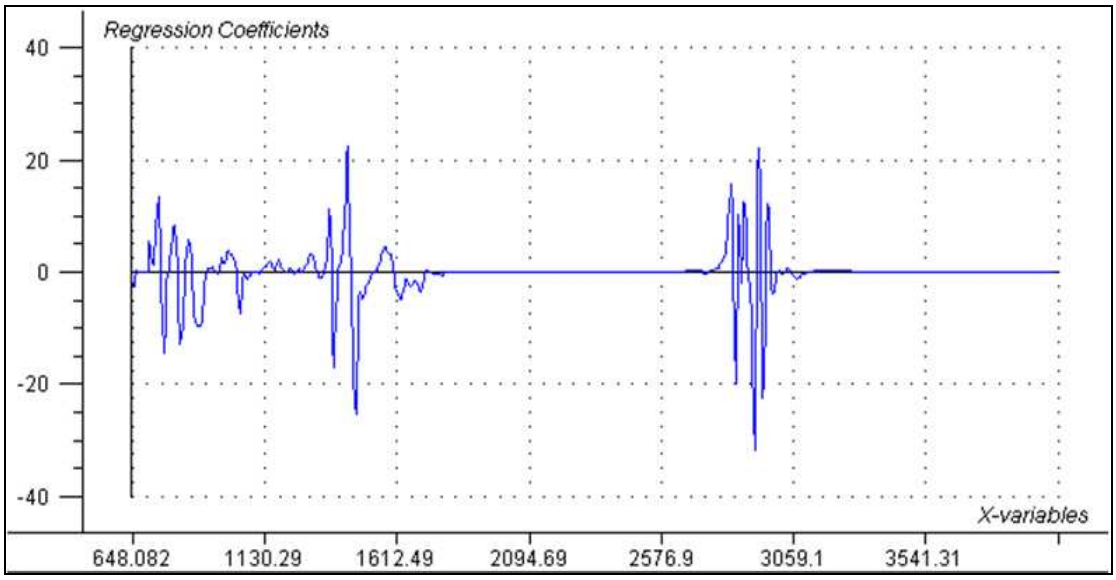
Figura 26. Gráfica de los errores RMSEC y RMSEP calculados en función de los PCs usados para el modelo PLS de asfaltenos



Como se observa en la figura 26, se logró comprobar que 5 es el número óptimo de PCs usados en el modelo de predicción de la composición porcentual en peso de asfaltenos, ya que para este caso, a pesar que el RMSEC disminuye a medida que aumenta el número de variables latentes, el RMSEP aumenta a partir del PC5, teniendo este último como valor numérico de 1.54.

Coeficientes 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 27 muestra los coeficientes de regresión hallados para la primera componente principal ya que este explica la mayor variabilidad en los datos.

Figura 27. Coeficientes de regresión para el primer componente principal del modelo PLS de asfaltenos



En esta gráfica se identificaron claramente dos regiones espectrales como las de mayor influencia sobre los PCs y que concuerdan con las regiones de mayor absorbancia en los espectros MIR obtenidos para las muestras de crudo: la región de 690-1770cm^{-1} (región 1) y la región de 2646 – 3321cm^{-1} (región 2). Aunque la figura 27 muestra algunas frecuencias de absorción negativas, estas no se omitieron del desarrollo del modelo, debido a que los datos tratados están previamente derivados.

3.2.2. Validación cruzada completa del modelo PLS para asfaltenos

La validación del modelo PLS para asfaltenos 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.

La tabla 11 muestra que los valores predichos por el modelo propuesto y los valores de referencia determinados por cromatografía de exclusión están cercanos

entre sí, aunque en nueve (9, 14-18, 20, 29, 30 y 45) de las 50 muestras utilizadas para la etapa de validación, el erro relativo calculado entre los valores predichos y de referencia de la composición de asfaltenos, es superior al 50%.

Tabla 11. Validación del modelo de predicción de %W de asfaltenos de las 50 muestras de calibración

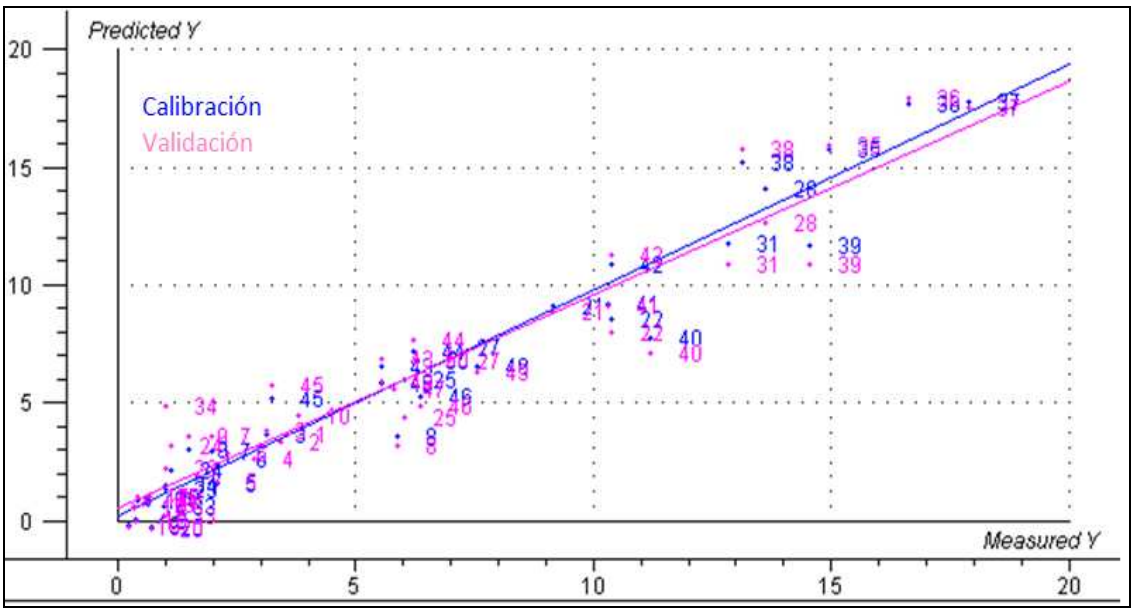
crudo	Referencia (%p/p)	Predicho (% p/p)	Residual	Error Relativo
1	3.57	3.67	-0.1	2.66
3	3.11	3.68	-0.57	18.2
4	2.87	2.64	0.23	7.91
5	2.1	1.62	0.48	23
6	2.3	2.6	-0.3	13.13
7	1.96	2.93	-0.97	49.54
9	1.49	3.02	-1.53	50.66
10	3.8	4.5	-0.7	18.47
11	0.62	0.74	-0.12	18.55
12	0.34	0.54	-0.2	60
13	0.68	1	-0.32	46.91
14	0.32	0.57	-0.25	78.44
15	0.6	0.97	-0.37	61.17
16	0.51	0.72	-0.21	41.96
17	0.42	0.85	-0.43	50.58
18	0.36	0.06	0.3	82.76
19	0.23	-0.17	0.4	26.96
20	0.72	-0.27	0.99	63.06
21	9.15	9.1	0.05	0.59
22	10.38	8.53	1.84	17.77
23	1.01	1.33	-0.32	31.73
24	1.12	2.12	-0.99	88.31

Crudo	Referencia (%p/p)	Predicho (% p/p)	Residual	Error Relativo
25	6.01	6.02	-0.01	0.15
26	13.04	13.48	-0.44	3.4
27	6.95	7.39	-0.44	6.36
28	13.63	14.06	-0.43	3.14
29	12.35	15.52	3.17	25.7
30	12.21	19.66	7.45	61.09
31	12.83	11.72	1.11	8.64
33	0.97	0.64	0.33	33.79
34	1.01	1.52	-0.51	50.74
35	14.95	15.76	-0.81	5.41
36	16.62	17.71	-1.09	6.54
37	17.91	17.75	0.16	0.89
38	13.15	15.2	-2.05	15.63
42	10.39	10.85	-0.46	4.42
43	5.55	6.52	-0.97	17.58
44	6.21	7.16	-0.95	15.29
45	3.24	5.21	-1.97	60.77
46	6.37	5.26	1.11	17.49
47	5.79	5.6	0.19	3.24
48	5.55	5.84	-0.3	5.33
49	7.56	6.59	0.97	12.88
50	6.3	6.82	-0.52	8.27

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 0.95 para la curva de validación. (Figura 28).

Los resultados hallados en esta etapa demostraron que el modelo para la predicción de %W de asfaltenos no es satisfactorio para todas las muestras.

Figura 28. Curvas de calibración y validación del modelo de predicción de %W de asfaltenos



La repetibilidad de la metodología propuesta en la predicción de asfaltenos se determinó seleccionando del grupo de muestras de validación tres muestras al azar. Cada uno de los espectros fue adquirido bajo las mismas condiciones instrumentales utilizadas en la calibración del modelo. Sobre las señales espectrales obtenidas en cada lectura se aplicó el modelo de PLS con cinco componentes principales para la predicción de %W de asfaltenos. La desviación estándar calculada en los valores predichos por el modelo fue inferior a 0.036 (Tabla 12), de esta manera se demostró que la repetibilidad en la predicción del contenido de %W de asfaltenos a partir del espectro de MIR, fue satisfactoria.

Tabla 12. Prueba de repetibilidad del modelo de predicción de %W de asfaltenos

Lectura	Muestra 1	Muestra 21	Muestra 25
1	3.665	9.100	6.017
2	3.607	9.120	6.014
3	3.598	9.098	6.020
Promedio	3.623	9.106	6.017
Desviación estándar	0.036	0.012	0.003

Los resultados del modelo de predicción descritos anteriormente para las 50 muestras iniciales confirmaron cuantitativamente los resultados cualitativos del PCA, en los cuales se determinó que existen dos agrupaciones de las muestras, en livianas y pesadas; ya que en la regresión PLS, a pesar de obtener un modelo de predicción con bajos errores de calibración y validación, no se obtuvo una predicción satisfactoria de la composición %W de asfaltenos de todas las muestras analizadas.

Por lo anterior se desarrollaron 8 modelos de predicción de la composición SARA; 4 para el grupo de 25 muestras clasificadas como pesadas por su alto contenido de asfaltenos y 4 para las restantes muestras de crudo clasificadas como livianas. Estos modelos se realizaron y analizaron siguiendo la misma metodología descrita en el desarrollo del modelo de %W de asfaltenos para el total de muestras disponibles.

4. RESULTADOS DE LOS MODELOS DE PREDICCIÓN

A continuación se presentan los principales resultados obtenidos en las etapas de calibración y validación de los modelos de predicción, para las diferentes propiedades de interés (composición SARA), siguiendo el procedimiento mostrado en la sección 3.2

4.1. MODELOS DE PREDICCIÓN DE LA COMPOSICIÓN SARA PARA MUESTRAS LIVIANAS

Para determinar los cuatro modelos de predicción de cada fracción del análisis SARA a partir de las 25 muestras de crudos livianos, se tomó en cuenta el análisis por componentes principales (PCA) desarrollado para dichas muestras en la sección 3.1.1.

4.1.1. Modelo PLS para la predicción de Saturados

El rango de calibración en el que se desarrollaron los análisis fue de 66.21 – 30.71% W de saturados.

La tabla 13 muestra el error estándar de calibración (RMSEC), de validación (RMSEP) y la varianza explicada para el modelo de predicción seleccionado con 3 PCs, por mostrar el mejor desempeño en la etapa de calibración de acuerdo a los tres parámetros evaluados.

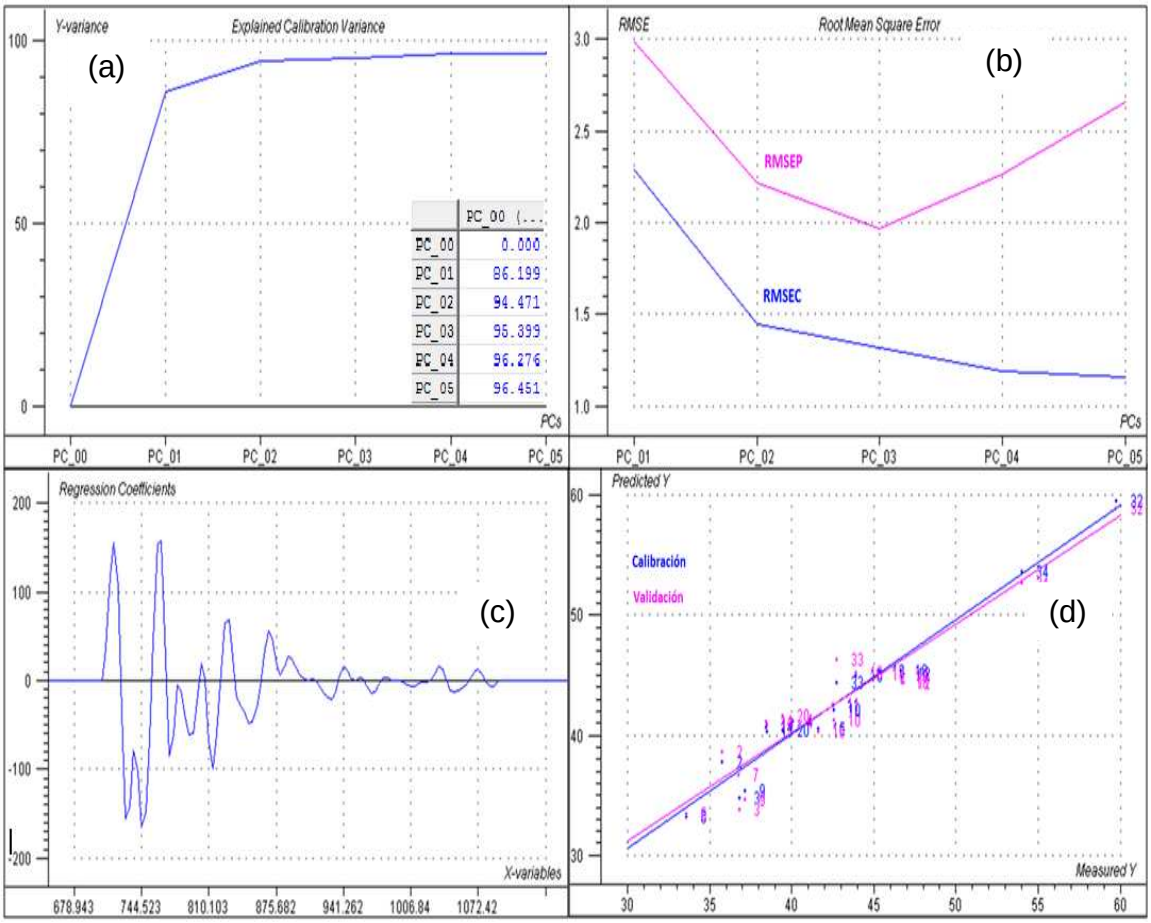
Tabla 13. Parámetros estadísticos del modelo desarrollado para la predicción de %W de saturados

Rango usado (cm^{-1})	Muestras excluidas	Componentes	Varianza Explicada	RMSEC (%)	RMSEP (%)
690-1075	1,5,23,24,45	3	96	1.3	1.9

Las muestras 23 y 24 fueron excluidas del modelo de predicción por presentar los mayores porcentajes en la composición de saturados, 62.28 y 66.21 respectivamente, lo que indica que son muestras influenciables (sección 3.1.1). La muestra 45 se excluyó del modelo por presentar composición SARA alejada de los promedios calculados (tabla 4); finalmente las muestras 1 y 5 fueron excluidas, porque al hacerlo por prueba y error mejoró el modelo, posiblemente por efecto de propiedades fisicoquímicas diferentes de la composición SARA.

El modelo seleccionado con 3 PCs explica, a partir de la señal MIR en la región de 690-1072cm^{-1}, más del 96% en la variabilidad de los datos de contenido de saturados (figura 29a). El error estándar en la etapa de calibración y validación cruzada disminuye apreciablemente hasta el tercer componente donde alcanza valores de 1.3% y 1.9% respectivamente (figura 29b). De los coeficientes de regresión calculados para el PC1, que explica la mayor variabilidad en los datos (86%), se determinó que la región de 690-1075cm^{-1} presenta un efecto positivo para la predicción de saturados (figura 29 c). La validación cruzada del modelo mostró un desempeño favorable para la predicción de %W de saturados, 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 0.95 para la curva de validación (figura 29d).

Figura 29. Descripción gráfica del modelo con tres componentes principales para la predicción del contenido %W de saturados en muestras livianas



Como se observa en la tabla 14 en la etapa de validación, para 2 de las 25 muestras evaluadas se obtuvieron porcentajes de error relativo superior al 5%, siendo 5.76% el mayor porcentaje de error, que lo presenta la muestra 14.

Tabla 14. Validación cruzada del modelo de predicción de %W de saturados de las muestras livianas

Muestras	Referencia (%W)	Predicción (W%)	Residual	Error relativo (%)
2	35.77	37.74	1.96	-5.49
3	36.81	34.73	2.08	5.65
4	40.02	41.09	1.07	-2.68
6	33.57	33.21	0.36	1.08
7	36.73	36.72	0.01	0.03
9	37.15	35.41	1.75	4.70
10	42.55	42.13	0.42	0.99
11	42.50	42.50	0.00	0.00
12	46.82	45.15	1.67	3.57
13	45.22	45.32	0.10	-0.22
14	38.40	40.61	2.21	-5.76
15	41.63	40.57	1.06	2.54
16	43.89	45.02	1.13	-2.58
17	38.50	40.32	1.82	-4.73
18	46.66	44.85	1.81	3.88
19	46.61	45.32	1.29	2.76
20	39.41	40.46	1.05	-2.66
32	59.73	59.56	0.17	0.28
33	42.73	44.36	1.63	-3.80
34	53.99	53.63	0.36	0.66

La repetibilidad de los resultados obtenidos (tabla 15), en el modelo para la predicción de la composición %W de saturados, mostró una desviación estándar inferior a 0.02; de esta manera se determinó que la repetibilidad en la predicción del contenido de saturados a partir del espectro MIR, asegura una buena predicción en todas las muestras.

Tabla 15. Prueba de repetibilidad del modelo de predicción de %W de saturados

Lectura	Muestra 7	Muestra 13	Muestra 45
1	36.72	45.32	31.359
2	36.73	45.30	31.35
3	36.70	45.29	31.37
Promedio	36.72	45.30	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, calculado con las muestras no incluidas en el modelo. Ambos valores, 1.35% y 1.9% respectivamente, se encuentran por debajo del 2% de error, demostrando junto con los resultados de repetibilidad obtenidos (tabla 15), que el modelo desarrollado presenta un desempeño satisfactorio en la predicción de la composición %W de saturados en muestras de crudos livianos.

4.1.2. Modelo PLS para la predicción de aromáticos

El rango de calibración en el que se desarrollaron los análisis fue de 35.63 – 21.19% W de aromáticos.

La tabla 16 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.

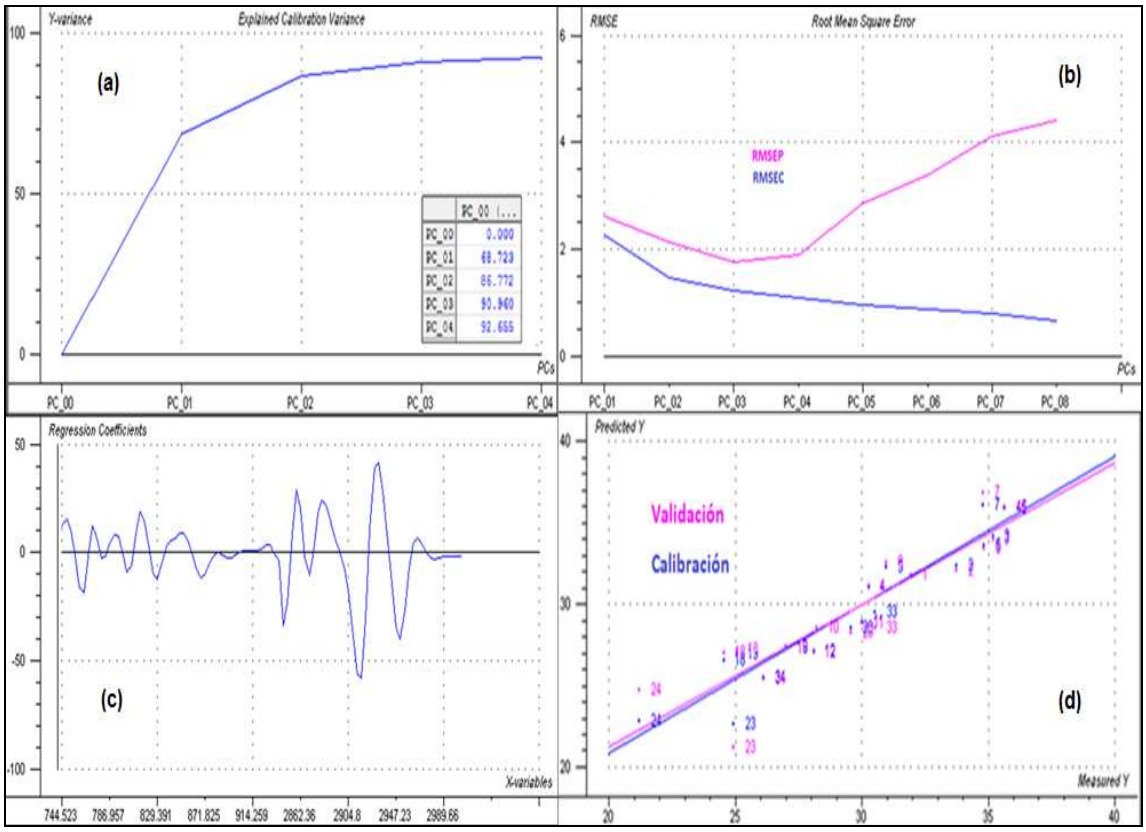
Tabla 16. Parámetros estadísticos del modelo generado para la predicción de %W de aromáticos en muestras livianas

Rangos usados (cm-1)	Muestras excluidas	Componentes	Varianza Explicada	RMSEC (%)	RMSEP (%)
744-937 2846-3005	6,12,17, 32	4	93	1.21	1.77

Las muestras 6, 12 y 17 fueron excluidas del modelo de predicción seleccionado, por presentar baja variabilidad en su composición de aromáticos con otras muestras (valores muy cercanos). (Ver tabla 4).
Además por prueba y error se determinó que para el modelo de predicción de %W de aromáticos, la muestra 32 se comporta como outliers (aumenta el porcentaje de error de predicción del modelo).

El modelo seleccionado con 4 PCs explica, a partir de la señal MIR en la región 744-937 cm^{-1} y de 2846-3005 cm^{-1} más del 93 % en la variabilidad de los datos de contenido de aromáticos (figura 30a). El error estándar en la etapa de calibración y validación cruzada disminuye apreciablemente hasta el cuarto componente, donde alcanza 1.21% y 1.77% respectivamente (figura 30b). De los coeficientes de regresión calculados para el PC1, que explica la mayor variabilidad en los datos (55%), se determinó que la región 744-937 cm^{-1}, 2846-3005 cm^{-1} presenta un efecto positivo para la predicción de aromáticos (figura 30c). La validación cruzada del modelo mostró un desempeño favorable para la predicción de %W aromáticos, ya que el coeficiente de correlación entre los valores de referencia y de predicción fue hallado superior al 0.95 para la curva de calibración y superior al 0.90 para la curva de validación (figura 30d).

Figura 30. Descripción grafica del modelo con cuatro componentes principales para la predicción del contenido %W de aromáticos en muestras livianas



Como se puede observar en la tabla 17 en la etapa de validación para 4 de las 25 muestras evaluadas se obtuvieron porcentajes de error relativo superior al 10%. La muestra 16 con 15,91% presenta el mayor porcentaje de error con un valor residual de 5.00 posiblemente esto se deba a la influencia de propiedades fisicoquímicas diferentes a la composición SARA, ya que se observó en la tabla 4 de los análisis SARA de referencia que la muestra 16 cae dentro del promedio de dichas propiedades.

Tabla 17. Validación cruzada del modelo de predicción del %W de aromáticos de las muestras livianas

Muestras	Referencia %W	Predicción %W	Residual	Error relativo (%)
1	31.94	31.82	0.12	0.37
2	33.73	32.36	1.37	4.07
3	35.18	34.32	0.86	2.44
4	30.26	31.02	0.76	2.52
5	34.83	33.96	0.87	2.51
7	34.79	36.10	1.31	3.77
9	30.96	32.40	1.44	4.65
10	28.21	28.23	0.02	0.07
11	30.00	28.99	1.01	3.37
13	25.03	26.64	1.61	6.44
14	32.49	27.67	4.82	14.84
15	24.64	28.35	3.71	15.05
16	31.46	26.46	5.00	15.91
18	24.52	26.26	1.74	7.10
19	26.99	27.06	0.07	0.27
20	29.56	28.63	0.93	3.14
23	24.91	21.36	3.55	14.24
24	21.19	21.44	0.25	1.19
33	30.51	29.81	0.70	2.28
34	26.11	25.32	0.79	3.02
45	35.63	36.16	0.53	1.49

La repetibilidad de los resultados obtenidos (tabla 18), en el modelo, mostró 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 18. Prueba de repetibilidad del modelo de predicción de %W de aromáticos

Lectura	Muestra 10	Muestra 2	Muestra 19
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 midió 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 1,21% y 1,77% respectivamente, se encuentran por debajo del 2% de error, demostrando junto con los resultados de repetibilidad obtenidos (tabla 18), que el modelo desarrollado presenta un desempeño satisfactorio en la predicción de la composición %W de aromáticos en muestras de crudos livianos.

4.1.3. Modelo PLS para la predicción de resinas.

El rango de calibración en el que se desarrollaron los análisis fue de 33.78 – 8.60 % W de resinas.

La tabla 19 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.

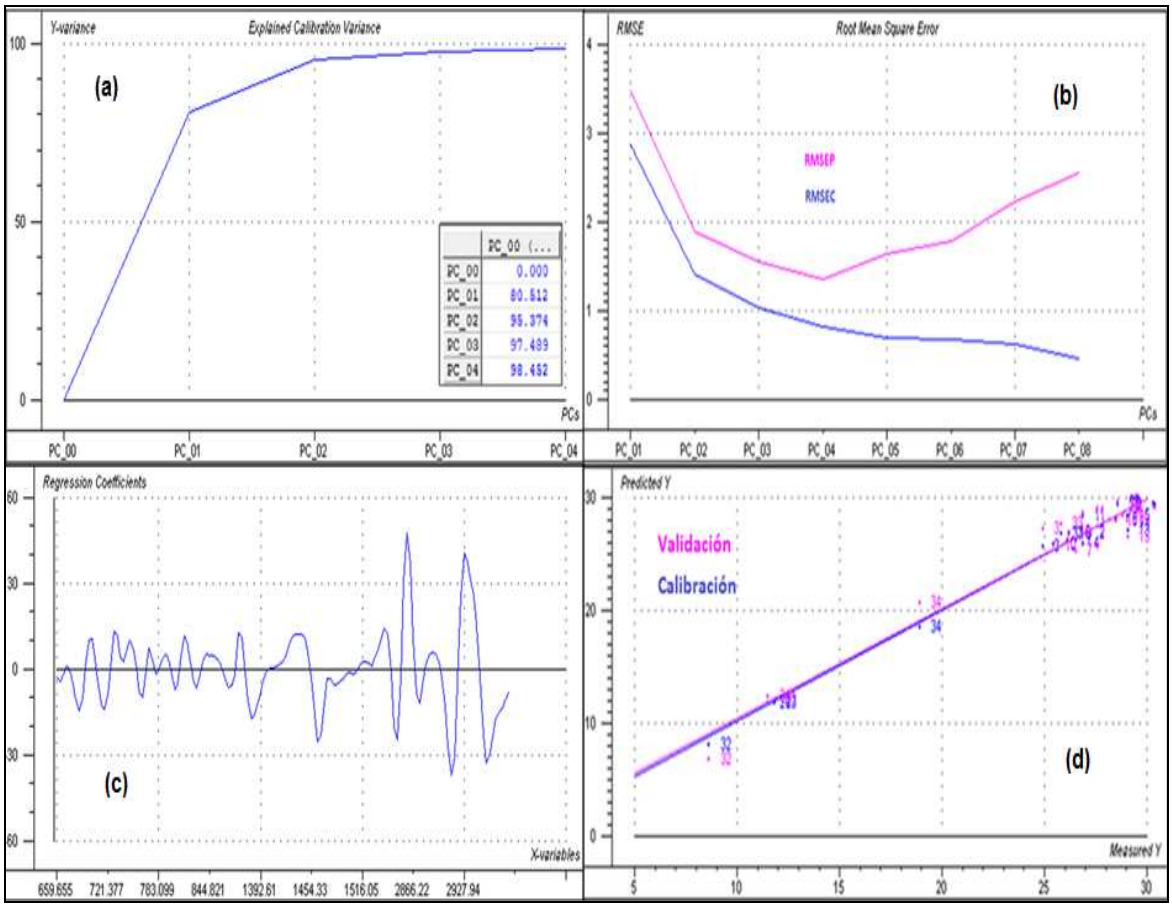
Tabla 19. Parámetros estadísticos del modelo generado para la predicción de %W de resinas en muestras livianas

Rangos usados (cm^{-1})	Muestras excluidas	PC	Varianza Explicada	RMSEC (%)	RMSEP (%)
690-875 1365-1527 2819-2981	1, 5, 6, 12, 15, 16	4	99	0.81	1.35

Las muestras 1, 6 y 15 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 33.78, 33.57 y 33.13 respectivamente; las muestras 5, 12 y 16 fueron excluidas por la misma razón, ya que presentan una composición porcentual en resinas de 24.95, 24.75 y 24.14 respectivamente. (Ver tabla 4)

El modelo explica, a partir de la señal MIR en la región 659-875 cm^{-1}, 1365-1527 cm^{-1} y de 2819-2981 cm^{-1} más del 99 % en la variabilidad de los datos de contenido de resinas (figura 31a). El error estándar en la etapa de calibración y validación cruzada disminuye apreciablemente hasta el cuarto componente, donde alcanza 0.81% y 1.35% respectivamente (figura 31b). De los coeficientes de regresión calculados para el PC1, que explica la mayor variabilidad en los datos (78%), se determinó que la región 659-875 cm^{-1}, 1365-1527 cm^{-1} y de 2819-2981 cm^{-1} presenta un efecto positivo para la predicción de resinas (figura 31c). La validación cruzada del modelo mostró un desempeño favorable para la predicción de %W resinas, 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 31d).

Figura 31. Descripción grafica del modelo con cuatro componentes principales para la predicción del contenido %W de resinas en muestras livianas



Como se puede observar en la tabla 20 en la etapa de validación para una de las 25 muestras evaluadas se obtuvieron porcentajes de error relativo superior al 6%. La muestra 13 con 6.63% presenta el mayor porcentaje de error con un valor residual de 1.93 posiblemente esto se deba a la influencia de propiedades fisicoquímicas diferentes a la composición SARA. En la tabla 4 de los análisis SARA de referencia se puede observar que la muestra 13 cae dentro del promedio de dichas propiedades.

Tabla 20. Validación cruzada del modelo de predicción del %W de resinas de las muestras livianas.

Muestras	Referencia %W	Predicción %W	Residual	Error relativo (%)
2	27.09	27.00	0.10	0.35
3	24.90	25.77	0.87	3.50
4	26.85	26.10	0.75	2.80
7	26.52	26.31	0.21	0.79
9	30.40	29.54	0.86	2.84
10	25.44	25.97	0.53	2.07
11	26.88	28.42	1.54	5.74
13	29.07	27.14	1.93	6.63
14	28.80	28.53	0.27	0.93
17	28.55	29.54	0.99	3.45
18	28.47	28.21	0.26	0.90
19	26.17	26.87	0.70	2.66
20	30.32	29.62	0.70	2.30
23	11.80	11.86	0.05	0.47
24	11.48	11.90	0.42	3.61
32	8.60	8.20	0.40	4.62
33	25.79	26.89	1.10	4.27
34	18.89	18.68	0.21	1.12
45	29.04	28.52	0.52	1.78

La repetibilidad de los resultados obtenidos (tabla 21), en el modelo seleccionado (con 4 componentes principales) para la predicción de la composición %W de resinas, mostró 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 MIR, asegura una buena repetición en todas las muestras.

Tabla 21. Prueba de repetibilidad del modelo de predicción de %W de resinas en muestras livianas

Lectura	Muestra 2	Muestra 7	Muestra 23
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 midió 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 0.81%y 1,35% respectivamente, se encuentran por debajo del 2% de error, demostrando junto con los resultados de repetibilidad obtenidos (tabla 21), que el modelo desarrollado presenta un desempeño satisfactorio en la predicción de la composición %W de resinas en muestras de crudos livianos.

4.1.4. Modelo PLS para la predicción de asfaltenos

El rango de calibración en el que se desarrollaron los análisis fue de 3.8 – 0.17 % W de asfaltenos.

La tabla 22 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 22. Parámetros estadísticos del modelo generado para la predicción de %W de asfaltenos en crudos livianos

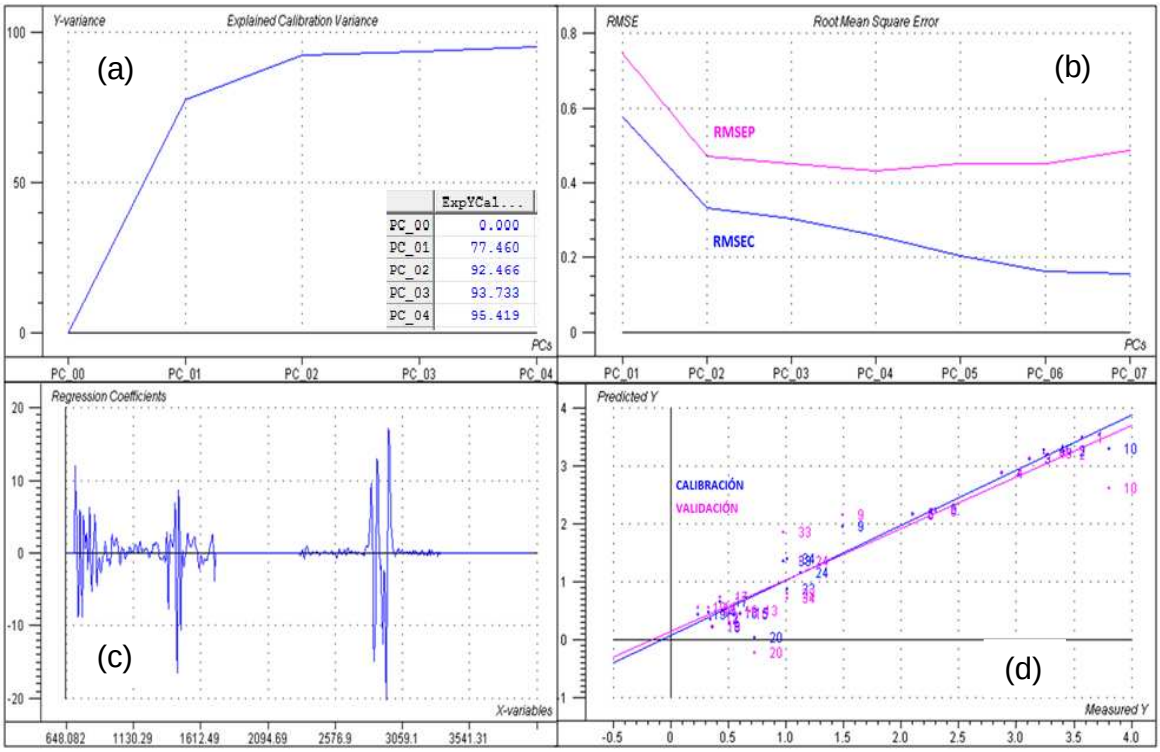
Rangos usados (cm^{-1})	Muestras excluidas	Componentes	Varianza Explicada	RMSEC (%)	RMSEP (%)
690-1072 2303-3321	7,11,32	4	97	0.25	0.43

Las muestras 7 y 11 fueron excluidas del modelo de predicción seleccionado, por presentar baja variabilidad en su composición de asfaltenos con las muestras 5 y 15 respectivamente. (Ver tabla 4).

La muestra 32 fue excluida del modelo por presentar un valor de composición en asfaltenos muy alejado de los demás, 0.17% siendo este el valor límite del rango de calibración; con lo cual se comprobó que la muestra 32 se comporta como outliers.

El modelo seleccionado con 4 PCs explica, a partir de la señal MIR en las regiones de 690-1072cm^{-1} y 2303-3321cm^{-1}, más del 96% en la variabilidad de los datos de contenido de asfaltenos (figura 32a). El error estándar en la etapa de calibración y validación cruzada disminuye apreciablemente hasta el cuarto componente donde alcanza los valores mínimos de 0.25% y 0.43% respectivamente (figura 32b). De los coeficientes de regresión calculados para el PC1, que explica la mayor variabilidad en los datos (77%), se determinó que las dos regiones del espectro usadas presentan un efecto positivo para la predicción de saturados (figura 32c). La validación cruzada del modelo mostró un desempeño favorable para la predicción de %W de asfaltenos, 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 0.94 para la curva de validación (figura 32d).

Figura 32. Descripción gráfica del modelo con cuatro componentes principales para la predicción del contenido %W de asfaltenos en muestras livianas



Como se observar en la tabla 23 en la etapa de validación, para 4 de las 25 muestras evaluadas se obtuvieron porcentajes de error relativo superior al 10%, siendo %17.3 el mayor porcentaje de error, que lo presenta la muestra 19, esto se debió muy posiblemente porque su contenido de asfaltenos se aleja del promedio del contenido de las demás muestras (5.57).

Tabla 23. Validación cruzada del modelo de predicción de %W de asfaltenos de las muestras livianas

Muestras	Referencia (%W)	Predicción (W%)	Residual	Error relativo (%)
1	3.57	3.62	0.05	1.82
2	3.41	3.42	0.01	4.16
3	3.11	3.127	0.02	-0.61
4	2.87	2.856	0.01	-0.70
5	2.10	2.18	0.08	-3.62
6	2.30	2.27	0.03	1.96
9	1.49	1.58	0.09	-5.91
10	3.80	3.24	0.56	13.11
12	0.34	0.35	0.01	-3.53
13	0.68	0.766	0.09	-12.65
14	0.32	0.29	0.03	9.37
15	0.60	0.586	0.01	2.33
16	0.51	0.55	0.04	-8.63
17	0.42	0.44	0.02	8.70
18	0.36	0.34	0.02	5.56
19	0.23	0.27	0.04	17.3
20	0.72	0.65	0.07	9.72
23	1.01	1.032	0.02	-2.18
24	1.12	1.15	0.03	-2.68
33	0.97	0.85	0.12	12.37
34	1.01	1.05	0.04	-3.96
45	3.24	3.28	0.03	-1.08

La repetibilidad de los resultados obtenidos (tabla 24), en el modelo seleccionado (con cuatro 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 24. Prueba de repetibilidad del modelo de predicción de %W de asfaltenos

Lectura	Muestra 4	Muestra 15	Muestra 45
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	0.01	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, calculado con las muestras no incluidas en el modelo (validación externa). Ambos valores, 0.23% y 0.45% respectivamente, se encuentran por debajo del 2% de error, demostrando junto con los resultados de repetibilidad obtenidos (tabla 24), que el modelo desarrollado presenta un desempeño satisfactorio en la predicción de la composición %W de asfaltenos en muestras de crudos livianos.

Prueba de validación externa: Se realizó la predicción de la composición SARA de algunas muestras excluidas de los modelos de calibración desarrollados para crudos livianos, comprobando que éstos predicen satisfactoriamente cada fracción, ya que el mayor porcentaje de error fue de 3.2 para resinas y asfaltenos. (Tabla 25).

Tabla 25. Prueba de Validación externa. Predicción de la composición SARA en crudos livianos

Muestra	Fracción	Referencia (%W)	Predicción (%W)	Error Relativo (%)
5	Saturados	38.12	37.97	0.39
6	Aromáticos	2.3	2.26	1.74
12	Resinas	24.75	23.95	3.2
11	Asfaltenos	0.62	0.60	3.2

4.2. MODELOS DE PREDICCIÓN DE LA COMPOSICICÓN SARA PARA MUESTRAS PESADAS

Para determinar los cuatro modelos de predicción de cada fracción del análisis SARA a partir de las 25 muestras de crudos pesados, se tomó en cuenta el análisis por componentes principales (PCA) desarrollado para dichas muestras en la sección 3.1.2.

4.2.1. Modelo PLS para la predicción de Saturados

El rango de calibración en el que se desarrollaron los análisis fue de 49.86 – 15.01 % W de saturados.

La tabla 26 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.

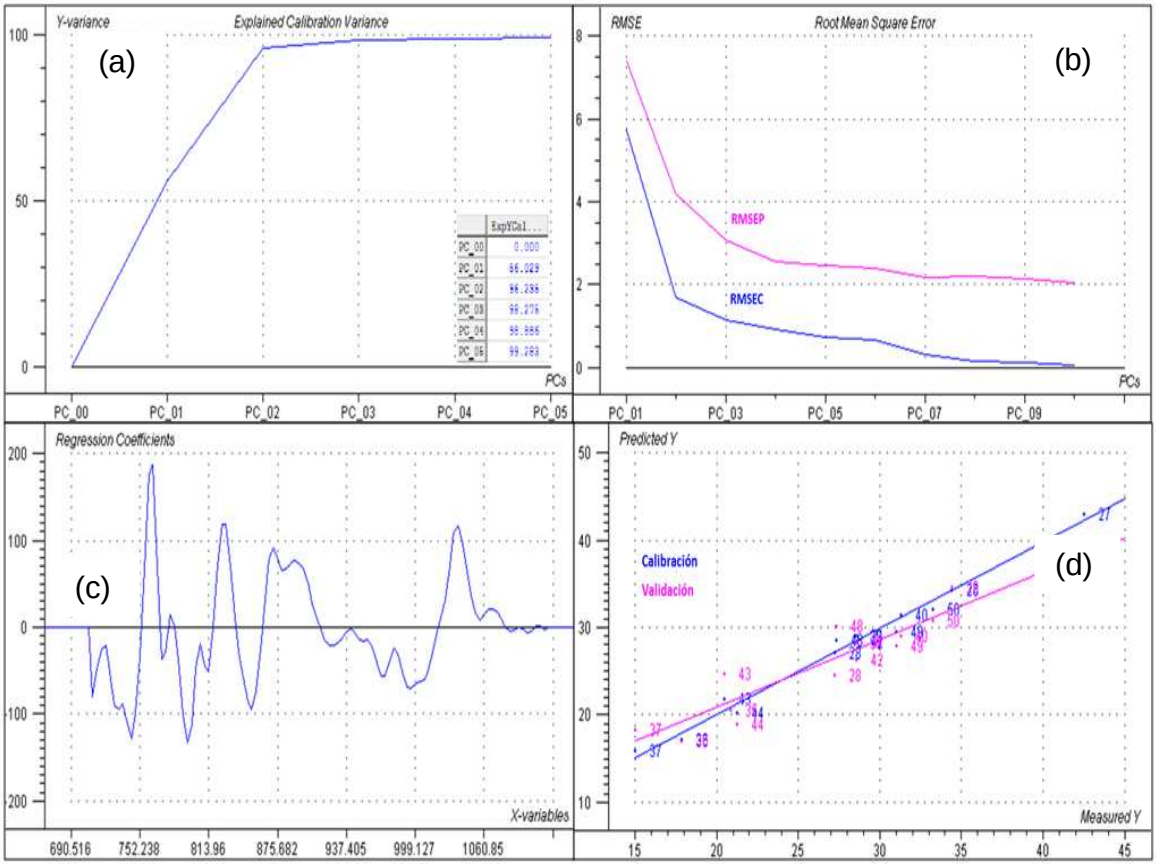
Tabla 26. Parámetros estadísticos del modelo generado para la predicción de %W de saturados en crudos pesados

Rango usado (cm^{-1})	Muestras excluidas	Componentes	Varianza Explicada	RMSEC (%)	RMSEP (%)
690- 1099	8, 21, 22, 41, 46 47	5	99	1.1	2.5

Las muestras excluidas por prueba y error del modelo de predicción seleccionado, fueron elegidas por presentar alta varianza residual (ser posibles outliers, Ver sección 3.1.2). Excluir la muestra 8 mejoró el modelo posiblemente porque presenta propiedades fisicoquímicas extremas diferentes de la composición SARA.

El modelo seleccionado con 5 PCs explica, a partir de la señal MIR en la región de 690-1099cm^{-1}, más del 99% en la variabilidad de los datos de contenido de saturados (figura 33a). El error estándar en la etapa de calibración y validación cruzada disminuye apreciablemente hasta el quinto componente donde alcanza valores de 1.1% y 2.5 % respectivamente (figura 33b). De los coeficientes de regresión calculados para el PC1, que explica la mayor variabilidad en los datos (50%), se determinó que la región de 690-1099cm^{-1} presenta un efecto positivo para la predicción de saturados (figura 33 c). La validación cruzada del modelo mostró un desempeño favorable para la predicción de %W de saturados, 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 a 0.95 para la curva de validación (figura 33d).

Figura 33. Descripción gráfica del modelo con cinco componentes principales para la predicción del contenido %W de saturados en muestras pesadas



Como se observar en la tabla 27 en la etapa de validación, para una de las 16 muestras evaluadas se obtuvo un porcentaje de error relativo superior al 10%, siendo éste %9.27, que lo presenta la muestra 37, esto se debió muy posiblemente por efecto de otras propiedades fisicoquímicas diferentes a la composición SARA, ya que la composición % de las cuatro fracciones para esta muestra se encuentran dentro del promedio (tabla 4).

Tabla 27. Validación cruzada del modelo de predicción de %W de saturados de las muestras pesadas

Muestra	Referencia (%W)	Predicción (%W)	Residual	Error relativo (%)
25	49.86	49.98	0.12	-0.24
26	34.41	35.73	1.32	-3.85
27	42.52	42.04	0.49	1.14
28	27.23	27.00	0.23	0.86
31	28.59	28.98	0.39	-1.37
35	20.85	20.43	0.42	2.01
36	17.83	17.04	0.79	4.45
37	15.01	16.40	1.39	-9.27
39	28.49	27.90	0.59	2.08
40	31.30	30.41	0.89	2.83
42	28.55	28.78	0.23	-0.82
43	20.46	21.88	1.42	-6.95
44	21.27	20.15	1.13	5.29
48	27.36	28.45	1.09	-3.99
49	31.03	29.39	1.64	5.28
50	33.23	33.43	0.20	-0.60

La repetibilidad de los resultados obtenidos (tabla 28), en el modelo seleccionado (con cinco componentes), mostró una desviación estándar inferior a 0.17; de esta manera se determinó que la repetibilidad en la predicción del contenido de saturados a partir del espectro MIR, asegura una buena predicción en todas las muestras.

Tabla 28. Prueba de repetibilidad del modelo de predicción de %W de saturados

Lectura	Muestra 28	Muestra 39	Muestra 50
1	27.00	27.90	33.43
2	27.22	28.24	33.35
3	27.15	28.12	33.40
Promedio	27.12	28.09	33.39
Desviación estándar	0.11	0.17	0.04

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, calculado con las muestras no incluidas en el modelo. Ambos valores, 1.1% y 2.5% respectivamente, se encuentran por debajo del 2.5% de error, demostrando junto con los resultados de repetibilidad obtenidos (tabla 28), que el modelo desarrollado presenta un desempeño satisfactorio en la predicción de la composición %W de saturados en muestras de crudos pesados.

4.2.2. Modelo PLS para la predicción de aromáticos

El rango de calibración en el que se desarrollaron los análisis fue de 41.23 – 20.43 % W de aromáticos.

La tabla 29 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 29. Parámetros estadísticos del modelo generado para la predicción de %W de aromáticos en muestras pesadas

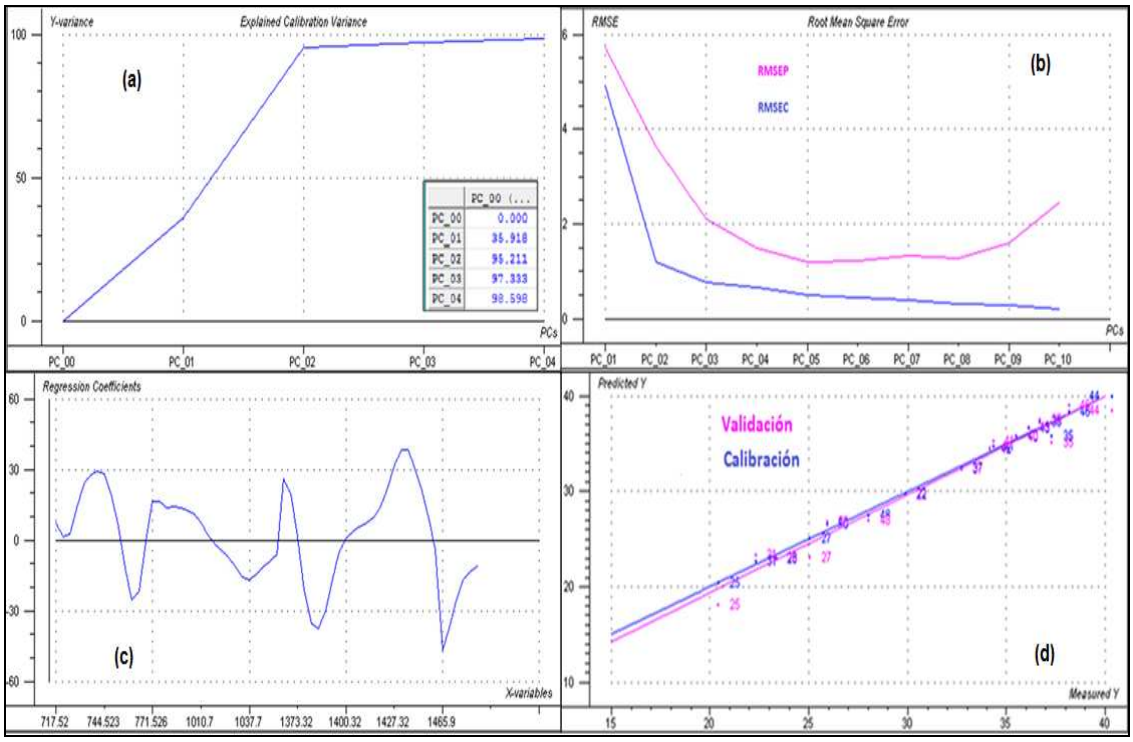
Rangos usados (cm^{-1})	Muestras excluidas	Componentes	Varianza Explicada	RMSEC (%)	RMSEP (%)
717-775 991-1053 1365-1450 1466-1485	8, 21, 26, 38, 39	4	98	0.72	1.68

La muestra 21 y 39 fueron excluidas del modelo por presentar alta varianza residual (posibles outliers), la muestra 26 fue excluida por presentar alta influencia en el modelo, así como la muestra 38 que presenta el mayor valor en la composición de aromáticos (41.23, límite superior del rango de calibración);

finalmente excluir la muestra 8 mejoró el modelo posiblemente porque presenta propiedades fisicoquímicas extremas diferentes de la composición SARA.

El modelo seleccionado con 4 PCs explica, a partir de la señal MIR en la región 717-775 cm^{-1}, 991-1053 cm^{-1}, 1365-1450 cm^{-1} y 1466-1485 cm^{-1} más del 98 % en la variabilidad de los datos de contenido de aromáticos (figura 34a). El error estándar en la etapa de calibración y validación cruzada disminuye apreciablemente hasta el cuarto componente, donde alcanza 0.72% y 1.68% respectivamente (figura 34b). De los coeficientes de regresión calculados para el PC1, que explica la mayor variabilidad en los datos (78%), se determinó que la región 659-875 cm^{-1}, 1365-1527 cm^{-1} y de 2819-2981 cm^{-1} presenta un efecto positivo para la predicción de resinas (figura 34c). La validación cruzada del modelo mostró un desempeño favorable para la predicción de %W aromáticos, 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 0.96 para la curva de validación (figura 34d).

Figura 34. Descripción grafica del modelo con 4 componentes principales para la predicción del contenido %W de aromáticos en muestras pesadas



Como se puede observar en la tabla 30 en la etapa de validación para una de las 18 muestras evaluadas se obtuvo un porcentaje de error relativo superior al 10%, siendo éste %11.75, que lo presenta la muestra 47, esto se debió muy posiblemente por efecto de otras propiedades fisicoquímicas diferentes a la composición SARA, ya que la composición % de las cuatro fracciones para esta muestra se encuentran dentro del promedio (tabla 4).

Tabla 30. Validación cruzada del modelo de predicción del %W de aromáticos de las muestras pesadas.

Muestras	Referencia %W	Predicción %W	Residual	Error relativo (%)
22	29.87	29.59	0.83	0.94
25	20.43	20.01	1.44	2.06
27	25.05	25.32	1.40	1.07
28	23.34	23.21	1.52	0.54
31	22.32	23.13	1.00	3.64
35	37.31	36.07	0.89	3.31
36	36.71	37.82	1.17	3.03
37	32.69	32.89	1.18	0.61
40	35.53	35.66	0.78	0.35
41	34.36	34.10	1.40	0.76
42	34.17	33.83	1.37	0.98
43	36.13	36.42	1.19	0.81
44	40.35	39.60	1.28	1.85
46	38.18	38.75	1.48	1.48
47	26.45	29.56	0.93	11.75
48	28.01	26.77	0.99	4.42
49	26.50	27.97	1.29	5.54
50	25.92	25.72	0.96	0.77

La repetibilidad de los resultados obtenidos (tabla 31), en el modelo seleccionado (con 4 componentes principales) para la predicción de la composición %W de aromáticos, mostró una desviación estándar inferior a 0.64; de esta manera se determinó que la repetibilidad en la predicción del contenido de aromáticos a partir del espectro MIR, asegura una buena predicción en todas las muestras.

Tabla 31. Prueba de repetibilidad del modelo de predicción de %W de aromáticos

Lectura	Muestra 27	Muestra 37	Muestra 43
1	24,92	33.81	33.77
2	25.02	34.32	33.87
3	24.90	33.05	32.98
Promedio	24.96	33.73	33.54
Desviación estándar	0.08	0.64	0.49

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, calculado con las muestras no incluidas en el modelo. Ambos valores 0.72% y 1,68% respectivamente, se encuentran por debajo del 2% de error, demostrando junto con los resultados de repetibilidad obtenidos (tabla 31), que el modelo desarrollado presenta un desempeño satisfactorio en la producción de la composición %W de aromáticos en muestras de crudos pesados.

4.2.3. Modelo PLS para la predicción de resinas

El rango de calibración en el que se desarrollaron los análisis fue de 39.08 –16.34 % W de resinas.

La tabla 32 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.

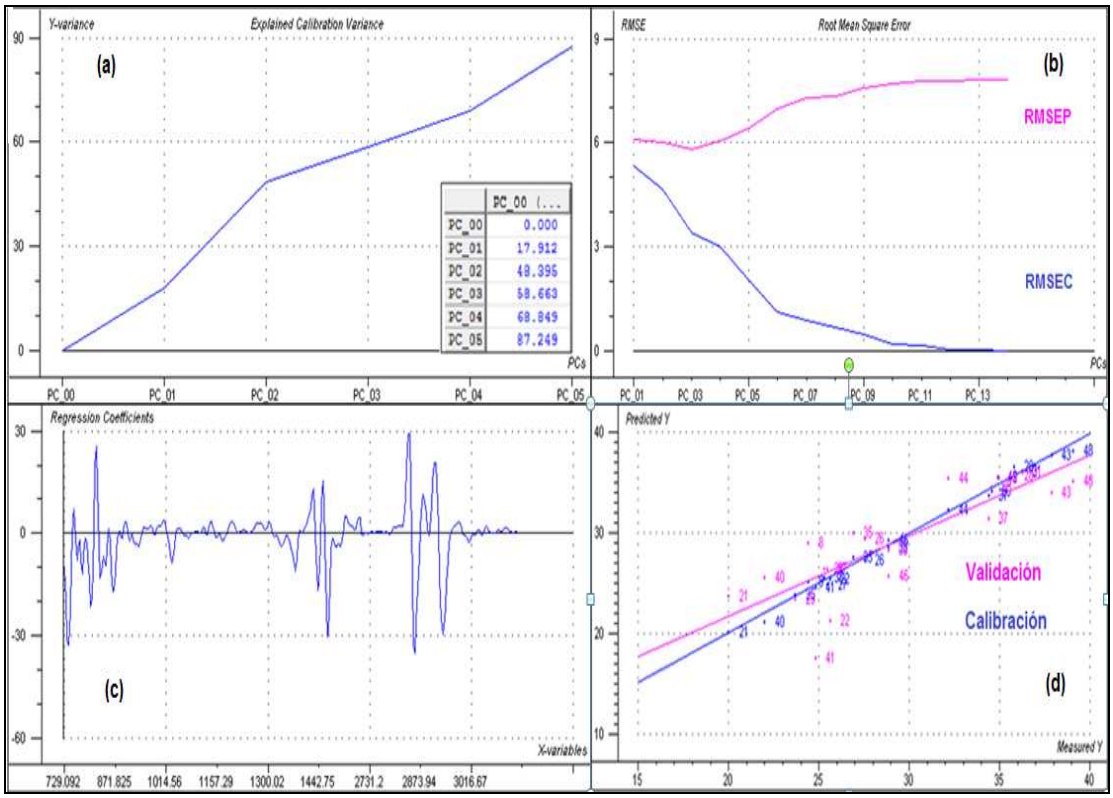
Tabla 32. Parámetros estadísticos del modelo generado para la predicción de %W de resinas en muestras pesadas

Rango usado (cm^{-1})	Muestras excluidas	Componentes	Varianza Explicada	RMSEC (%)	RMSEP (%)
732-1543	8, 21, 26, 38, 39	5	87	1.09	3.68

La muestra 21 y 39 fueron excluidas del modelo por presentar alta varianza residual (posibles outliers), las muestras 26 y 38 fueron excluidas por presentar alta influencia en el modelo, finalmente excluir la muestra 8 mejoró el modelo posiblemente porque presenta propiedades fisicoquímicas extremas diferentes de la composición SARA.

El modelo seleccionado con 5 PCs explica, a partir de la señal MIR en la región 732-1543 cm^{-1} más del 87 % en la variabilidad de los datos de contenido de resinas (figura 35a). El error estándar en la etapa de calibración y validación cruzada disminuye hasta el quinto componente, donde alcanza 1.091% y 3.68% respectivamente (figura 35b). De los coeficientes de regresión calculados para el PC1, se observó que no se explica la mayor variabilidad en los datos (29%), (figura 35c). La validación cruzada del modelo mostro un desempeño favorable para la predicción de %W resinas, 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 al 0.93 para la curva de validación (figura 35d).

Figura 35. Descripción grafica del modelo con 5 componentes principales para la predicción del contenido %W de resinas en muestras pesadas



Como se puede observar en la tabla 33 en la etapa de validación para 3 de las 15 muestras evaluadas se obtuvieron porcentajes de error relativo superior al 10%. La muestra 40 con 13.25% presenta el mayor porcentaje de error con un valor residual de 3.09 posiblemente esto se deba a la influencia de propiedades fisicoquímicas diferentes a la composición SARA. En la tabla 4 de los análisis SARA de referencia se observa que la muestra 40 cae dentro del promedio de dichas propiedades.

Tabla 33. Validación cruzada del modelo de predicción del %W de resinas de las muestras pesadas.

Muestras	Referencia %W	Predicción %W	Residual	Error relativo (%)
22	25.61	22.68	2.949	11.44
25	23.70	23.48	4.126	0.93
27	25.48	26.00	3.507	2.03
28	35.79	36.29	3.659	1.41
31	36.26	37.89	3.511	4.50
35	26.89	29.47	2.103	9.61
36	28.84	28.43	2.51	1.43
37	34.39	32.34	2.569	5.96
40	21.99	24.90	3.091	13.25
43	37.87	36.30	2.806	4.16
46	28.87	28.44	3.155	1.50
47	30.85	34.34	2.065	11.33
48	39.08	36.12	2.401	7.58
49	34.91	32.64	2.29	6.50
50	34.56	35.35	2.684	2.27

La repetibilidad de los resultados obtenidos (tabla 34), en el modelo seleccionado (con 5 componentes principales), mostró una desviación estándar inferior a 0.16; de esta manera se determinó que la repetibilidad en la predicción del contenido de

resinas a partir del espectro MIR, asegura una buena predicción en todas las muestras.

Tabla 34. Prueba de repetibilidad del modelo de predicción de %W de resinas

Lectura	Muestra 25	Muestra 38	Muestra 46
1	23.48	25.22	28.44
2	23.55	24.23	28.67
3	23.69	25.30	28.37
Promedio	23.57	24.92	28.49
Desviación estándar	0.11	0.60	0.16

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, calculado con las muestras no incluidas en el modelo. El valor del RMSEC (1.09%) se encuentra por debajo del 2 % de error; a diferencia del RMSEP que tomó valores por encima (3.68%), esto puede atribuirse a la incertidumbre de los datos aromáticos de partida, ya que la elusión de la fracción AR (aromáticos y resinas) en muchos casos no puede resolverse con facilidad; a pesar de este último, el modelo desarrollado para la composición %W de resinas en muestras de crudo pesado presenta un desempeño satisfactorio.

4.2.4. Modelo PLS para la predicción de Asfaltenos

El rango de calibración en el que se desarrollaron los análisis fue de 17.91 –5.55 % W de asfaltenos.

La tabla 35 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.

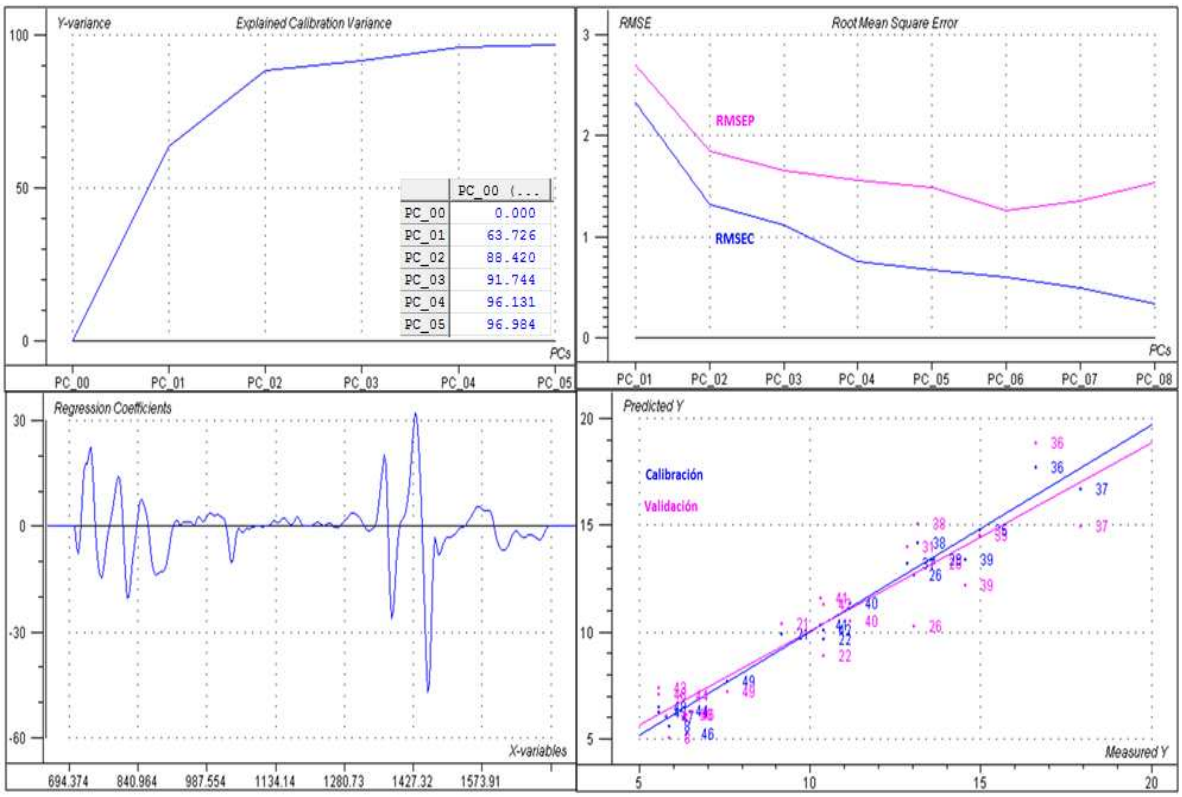
Tabla 35. Parámetros estadísticos del modelo generado para la predicción de %W de asfaltenos

Rango usado (cm^{-1})	Muestras excluidas	Componentes	Varianza Explicada	RMSEC (%)	RMSEP (%)
690-1072	25,27	5	98	0.6	1.4

Las muestras 25 y 27 fueron excluidas del modelo por prueba y error, debido a que presenta alta influencia. (Ver sección 3.1.2)

El modelo seleccionado con 5 PCs explica, a partir de la señal MIR en la región de 690-1072cm^{-1}, más del 98% en la variabilidad de los datos de contenido de asfaltenos (figura 36a). El error estándar en la etapa de calibración y validación cruzada disminuye apreciablemente hasta el quinto componente donde alcanza valores de 0.6% y 1.4 % respectivamente (figura 36b). De los coeficientes de regresión calculados para el PC1, que explica la mayor variabilidad en los datos (78%), se determinó que la región de 690-1072cm^{-1} presenta un efecto positivo para la predicción de asfaltenos (figura 36 c). La validación cruzada del modelo mostró un desempeño favorable para la predicción de %W de asfaltenos, 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 0.94 para la curva de validación (figura 36d).

Figura 36. Descripción gráfica del modelo con cinco componentes principales para la predicción del contenido %W de asfaltenos en muestras pesadas



Como se observa en la tabla 36 en la etapa de validación, para tres de las 21 muestras evaluadas se obtuvo un porcentaje de error relativo superior al 10%, siendo este 17.7% el mayor porcentaje de error, que lo presenta la muestra 46, esto se debió muy posiblemente por efecto de otras propiedades como él %W de saturados y resinas, ya que la composición porcentual de estos se alejan del promedio. (Ver tabla 4).

Tabla 36. Validación cruzada del modelo de predicción de %W de asfaltenos de las muestras pesadas

Muestra	Referencia (%W)	Predicción (%W)	Residual	Error relativo (%)
8	5.87	5.58	0.29	5.01
21	9.15	9.92	0.77	-8.46
22	10.38	9.70	0.68	6.59
26	13.04	12.71	0.33	2.57
28	13.63	13.41	0.22	1.59
31	12.83	13.25	0.42	-3.24
35	14.95	14.76	0.19	1.29
36	16.62	17.75	1.13	-6.79
37	17.91	16.70	1.21	6.73
38	13.15	14.17	1.02	-7.76
39	14.55	13.42	1.13	7.78
40	11.18	11.37	0.19	-1.71
41	10.29	10.32	0.03	-0.33
42	10.39	10.08	0.32	3.03
43	5.55	6.26	0.71	-12.74
44	6.21	6.33	0.12	-1.90
46	6.37	5.24	1.13	17.71
47	5.79	5.99	0.20	-3.49
48	5.55	6.49	0.94	-16.94
49	7.56	7.68	0.12	-1.64
50	6.30	6.15	0.15	2.40

La repetibilidad de los resultados obtenidos (tabla 37), en el modelo seleccionado (con cinco componentes), mostró una desviación estándar inferior a 0.025; 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 37. Prueba de repetibilidad del modelo de predicción de %W de asfaltenos

Lectura	Muestra 28	Muestra 41	Muestra 50
1	13.41	10.32	6.15
2	13.39	10.35	6.18
3	13.43	10.30	6.17
Promedio	13.41	10.32	6.167
Desviación estándar	0.02	0.025	0.015

La habilidad de predicción global del modelo de predicción se midió 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, 0.6% y 1.4% respectivamente, se encuentran por debajo del 2% de error, demostrando junto con los resultados de repetibilidad obtenidos (tabla 37), que el modelo desarrollado presenta un desempeño satisfactorio en la predicción de la composición %W de asfaltenos en muestras de crudos pesados.

Prueba de Validación externa: Se realizó la predicción de la composición SARA de algunas muestras excluidas de los modelos de calibración desarrollados para crudos pesados, comprobando que éstos predicen satisfactoriamente cada fracción, ya que el mayor porcentaje de error fue de 5.17 para asfaltenos. (Tabla 38).

Tabla 38. Prueba de Validación externa. Predicción de la composición SARA en crudos pesados

Muestra	Fracción	Referencia (%W)	Predicción (%W)	Error Relativo (%)
8	Saturados	39.18	37.99	3.03
21	Aromáticos	21.67	20.56	5.12
26	Resinas	27.53	26.69	3.05
27	Asfaltenos	6.95	6.59	5.17

5. CONCLUSIONES

La espectroscopia MIR-ATR en combinación con técnicas de calibración multivariable como el PLS se ha mostrado como una técnica alternativa, a la cromatografía de exclusión molecular para la determinación del contenido en saturados, aromáticos, resinas y asfaltenos en muestras de crudo; con la ventaja de evitar el pretratamiento de la muestra e incrementar la reproducibilidad en los resultados.

Luego de revisar la repetibilidad de la toma de espectros de las muestras de crudo, se pudo establecer que se presenta una disminución considerable de la incertidumbre (medida como la RSD) cuando a los espectros se les somete a una normalización por rangos seguida de una derivación de primer orden (derivación Norris).

El análisis por componentes principales permitió identificar las agrupaciones naturales de los espectros de las muestras y el número de componentes principales optimo a incluir en el modelo de predicción por contener la mayor información de las variables originales; además se determinó que las regiones del espectro de mayor importancia se encuentran en los rangos de 650 – 1600cm^{-1} y de 2800-3000cm^{-1}.

La técnica de regresión parcial por mínimos cuadrados (PLS) aplicada sobre el espectro de MIR-ATR mostró un desempeño satisfactorio para predecir la composición porcentual de saturados, aromáticos, resinas y asfaltenos en los crudos estudiados, demostrando que los datos espectroscópicos pueden

proporcionan información detallada de estas y otras propiedades de las muestras analizadas si se aplica un adecuado tratamiento quimiométrico.

Los modelos tuvieron que ser ajustados por eliminación de un número alto de outliers debido a que los datos de referencia de la composición porcentual en peso del análisis SARA tienen implícito el porcentaje de error sistemático del método usado para su determinación.

Los modelos desarrollados así como su aplicación, deben hacerse en muestras de la misma naturaleza (pesada o liviana). Adicionalmente, las señales espectroscópicas deben ser adquiridas en el espectrómetro empleado bajo las mismas condiciones y parámetros instrumentales. La aplicación de los modelos predictivos a muestras de naturaleza diferente puede llevar a resultados erróneos, como se mostró en el modelo desarrollado para el grupo de las 50 muestras iniciales.

La viabilidad de la aplicación de la calibración multivariante a partir de medidas espectroscópicas en el ámbito industrial depende en gran medida de la estabilidad de los modelos desarrollados, así como en su adaptación a nuevas situaciones experimentales.

6. RECOMENDACIONES

Los resultados del estudio anteriormente presentado muestran el gran potencial de la espectroscopia MIR y técnicas multivariables para la determinación de la composición SARA de crudos colombianos. Aunque los resultados de validación cruzada muestran un buen desempeño de los modelos desarrollados, se recomienda que éstos sean evaluados y alimentados con nuevas muestras para hacerlos más robustos y asegurar que siempre estarán actualizados.

Realizar validación cruzada con otras técnicas espectroscópicas como espectroscopia fotoacústica (PAS) y fluorescencia que permitan confirmar la precisión del método FTIR-ATR.

La implementación de la espectroscopia MIR acoplada con celda ATR y el uso de técnicas multivariables en crudos y fracciones de este, para desarrollar modelos de predicción de nuevas propiedades fisicoquímicas como densidad, gravedad API, punto de fluidez, entre otras, las cuales son de gran interés para su caracterización.

7. REFERENCIAS BIBLIOGRAFICAS

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 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.

4. KHARRAT, A.M; ZACHARIA, J; CHERIAN, V.J and ANYATONWU, A. Issues with comparing SARA methodologies. Energy and Fuels, 2007, 21 (6), 3618-3621.

5. HONGFU, Y; XIAOLI, C; HAORAN, L and YUPENG, X. Determination of multi-properties 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 crudo extrapesado Carabobo. Revista Latinoamericana de Metalurgia y materiales. 2007. Vol 2. P83-94.

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 304-11.

8. KALLEVIK, H. Characterization of crude oil and model oil emulsions by means of near infrared spectroscopy and multivariate analysis.

9. SPEIGHT, J. G. The chemistry and technology of petroleum. Cuarta edición: Laramie: 2006. Chemical Industries Series. P 916.

10. ALBOUDWARJ, H; FELIX, J; TAYLOR, S. La importancia del petróleo pesado. Oilfield Review, 2006. P 38-58.

11. ALTGELT, K.H. AND BODUSZYNSKI, M. M. Composition and analysis of heavy petroleum fractions. Editorial Marcel Dekker. Segunda Edición. New York. 1994. P 495.

12. CONAWAY, C. The petroleum industry: A Nontechnical al guide. Tulsa: Pennwell Publishing Co. 1999.

13. WAUQUIER, J.P. El refino del petróleo. Petróleo Crudo, Productos petrolíferos, Esquemas de fabricación. En: Composición de los petróleos crudos y de los productos petrolíferos. España.: 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. ISLAS, F. E; GONZALEZ, E and LIRA-GALEANA C. Comparisons between open column chromatography and HPLC SARA fractions in petroleum. Energy Fuels, 2005. Vol 19, Nº 5.P 2080-2088.

16. AMERICAN SOCIETY FOR TESTING AND MATERIALS. Standard Test Method for Separation of Asphalt into four fractions. ASTM D4124-09

17. ABDEL, M.K; ZACHARIA, J; CHERIAN, J. Issues with Comparing SARA Methodologies. Energy & Fuels, 2007. Vol 21. P 3618-3621.

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

19. OSBOME, B. G; FEAM, T; HINDLE, P. H. Practical NIR spectroscopy with applications in food and beverage analysis. 2nd ed. England: Longman Scientific & Technical. 1993.

20. HOLLAS, J. M. Modern Spectroscopy. 2nd ed. Chichester, England: John Wiley & Sons. 1992.

21. PASQUINI, C. Near Infrared spectroscopy fundamentals, practical aspects and analytical applications. En: J. Braz. Chem Soc. 2003. Vol 14, N°2. p.198-219

22. SILVERSTEIN, R. M; and WEBSTER, F. X. Spectrometric Identification of Organic Compounds, 6a ed. New York.: John Wiley & Sons, 1998.

23. R UBINSON, K. A; RUBINSON, J. F. Análisis instrumental, Prentice Hall Hispanoamericana S.A. Madrid.: Editorial Pearson Educación, 2001.

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

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

26. 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

27. ZHANG, Z; EWING, G.E. Attenuated partial internal reflection infrared spectroscopy. En: Analytical Chemistry. 2002. Vol 74 N°11. P. 2578-2583.

28. 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 orgánica. 2002. 20-29p.

29. MARTENS, H; NAES, T. Multivariate calibration. New York: Jhon Wiley & Sons, 1989. P 438.

30. GELADI, P. Some recent trends in the calibration literatura. En: Chemometrics and intelligent Laboratory Systems. Vol. 60 (2002). P. 211 – 2

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

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

33. GELADI, P. Notes on the History and Nature of Partial Least Squares (PLS) modeling. Journal of Chemometrics, 1988. Vol 2. P 231-245.

34. STEINER, J; TERMONIA, Y; DELTOUR, J; Anal. Chemometrics. Vol 44. P. 1906, 1972.

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

36. GEMPERLINE, P. Practical guide to chemometrics. Segunda edición. Editorial Taylor y Francis Grupo. 2006. P. 541

37. FERNÁNDEZ, C.M. Quimiometría. Editorial Publicaciones Universidad de Valencia, 2005. P.423.

38. VILCA, J.C. Generalizaciones de mínimos cuadrados parciales con aplicación en clasificación supervisada. Universidad de puerto rico. Mayagüaez. 2004.

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

40. 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

41. EISENHART, M; HASTAY, W; WALLIS, W.A.. Techniques of statistical analysis. En H.Hotelling. Multivariate Quality Control. New York. Mc Graw-Hill, 1947.

42. WOLD, S. Technometrics. 1978, Vol 20 P 397.

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

44. FREDERICKS Peter, y, RINTOUL Llewellyn. Vibrational Spectroscopy: Instrumentation for Infrared and Raman Spectroscopy. Queensland University of Technology, Brisbane, Queensland, Australia

45. 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.

46. PEREIRA C. Rita, SKROBOT L. Vinicius, CASTRO R. Eusta´quio, FORTES C. P Isabel, and, PASA M. D. Vaˆnya. Determination of Gasoline Adulteration by Principal Components Analysis-Linear Discriminant Analysis Applied to FTIR Spectra. Energy & Fuels 2006, 20, 1097-1102.

47. SOARES. P, REZENDE. F. Multivariate calibration by variable selection for blends of raw soybean oil/ biodiesel from different sources using Fourier Transform Infrared Spectroscopy (FTIR) spectra data. Energy & Fuels 2008.

48. THE UNSCRAMBLER 9.7. METHODS. Software para diseño de experimentos y análisis multivariado. (En línea). http://www.camo.com [citado el 13 de abril de 2010].

49. ORREGO, J.A. Estudio de la estructura de cinco carbones colombianos por Espectroscopia Fotoacústica en el infrarrojo medio. *Tesis de maestría*, maestro en Química. Bucaramanga. Universidad Industrial de Santander. Facultad de ciencias. Departamento de Química. 2008, 38p.