

DESARROLLO DE MODELOS PARA PREDECIR EL ANÁLISIS SARA

**Desarrollo de Modelos para Predecir el Análisis SARA y Contenido de
carbón Conradson de Crudos Colombianos a partir de Espectroscopia
Raman, Espectrometría de Masas y Quimiometría**



Mónica Alexandra Flórez González

Trabajo de Investigación de Maestría en Física para Optar el Título de Magister en Física

Director

**Enrique Mejía Ospino
Ph.D Ciencias Químicas**

**Universidad Industrial de Santander
Facultad de Ciencias
Escuela de Física
Maestría en Física
Bucaramanga
2018**

A Mauricio, Andrés y María Alejandra

AGRADECIMIENTOS

Un especial agradecimiento a mi director Dr. Enrique Mejía Ospino, por sus valiosos aportes y la confianza que depositó en mí al encomendarme la elaboración de este trabajo.

Mi más sincera gratitud al Dr. Jáder Enrique Guerrero Bermúdez, por su generosa colaboración y valiosa orientación, fundamentales para la obtención de los resultados en este trabajo.

A los integrantes del Laboratorio de Espectroscopía atómico molecular LEAM y muy especialmente al profesor Rafael Cabanzo, Ximena Calderón Cascavita y Mónica Pinto por su colaboración en la etapa inicial de este trabajo.

Este trabajo hizo parte del proyecto No. FP44842-039-2015 de Colciencias y del convenio ICP Ecopetrol – UIS No. 5222395.

CONTENIDO

INTRODUCCIÓN	13
1. ASPECTOS TEÓRICOS	19
1.1 ASPECTOS GENERALES DEL PETROLEO CRUDO	19
1.1.2 Destilación del crudo	20
1.1.3 Procesos de conversión (craqueo).....	21
1.1.4 Procesos de mejoramiento	22
1.2 ANÁLISIS SARA	22
1.3 ESPECTROMETRIA DE MASAS.....	25
1.4 ESPECTROSCOPIA RAMAN.....	32
1.5 QUIMIOMETRIA.....	36
1.5.1 Calibración multivariable.....	38
1.6 ANTECEDENTES DE LA ESPECTROMETRÍA DE MASAS, ESPECTROSCOPIA RAMAN Y QUIMIOMETRIA EN EL ANÁLISIS DE CRUDOS	41
2. RESULTADOS EXPERIMENTALES.....	45
2.1 SELECCIÓN DE MUESTRAS	45
2.1.1 Caracterización de las muestras	45
2.1.2 Preparación de las muestras	46
2.2 INSTRUMENTACIÓN	47
2.2.1 Espectroscopía Raman	48
2.2.2 Espectrometría de masas.....	51
2.3 PARÁMETROS DE ADQUISICIÓN DE LOS ESPECTROS	51
2.3.1 Espectroscopía Raman.....	51

2.3.2 Espectrometría de masas.....	53
2.4 ADQUISICIÓN DE LOS ESPECTROS.....	54
2.4.1 Espectroscopía Raman.....	54
2.4.2 Espectrometría de masas.....	55
2.5 PRE-PROCESAMIENTO DE DATOS.....	56
2.5.1 Espectroscopía Raman.....	57
2.5.2 Espectrometría de masas.....	58
2.6 ANÁLISIS QUIMIOMÉTRICO DE DATOS.....	59
2.6.1 Análisis de residuales.....	59
2.6.2 Desarrollo del modelo PLS.....	62
3. CONCLUSIONES.....	89
4. RECOMENDACIÓN.....	90
REFERENCIAS BIBLIOGRAFICAS.....	91
ANEXOS.....	93

LISTA DE FIGURAS

Figura 1. Esquema análisis SARA.....	24
Figura 2. Esquema de un espectrómetro de masas	26
Figura 3. MALDI, ionización/desorción láser asistida con matriz	28
Figura 4. Esquema analizador TOF	30
Figura 5. Analizador en modo reflectron.....	32
Figura 6. Efecto Raman	33
Figura 7. Notación matricial para la descomposición de la matriz de regresores	39
Figura 8. Notación matricial para la descomposición de la matriz de predictores	40
Figura 9. Muestras diluidas en diclorometano.....	47
Figura 10. Microscopio Raman LabRam HR Evolution. HORIBA.....	48
Figura 11. Espectros Raman crudo 45	49
Figura 12. Espectrómetro de masas MALDI.TOF	51
Figura 13. Imagen de la muestra posterior a la adquisición del espectro preliminar.	52
Figura 14. Espectros obtenidos para las 67 muestras de crudo por método indirecto	55
Figura 15. Espectros Raman obtenidos para las 62 muestras de crudo	56
Figura 16. Espectros Raman para las muestras de crudo obtenidos por método indirecto después del preprocesamiento	58
Figura 17. Espectros de masas pre-procesados (corrección del corrimiento).....	59

Figura 18. Gráfica de influencia muestras en el modelo. Espectroscopía Raman.....	61
Figura 19. Gráfica de influencia de muestras en el modelo. Espectrometría de masas.....	61
Figura 20. Raíz del error cuadrático medio de validación cruzada (RMSECV) en función del número de componentes (variables latentes) para el modelo de predicción de la fracción de saturados.	65
Figura 21. Varianza explicada acumulativa de contenido de saturados	66
Figura 22. Coeficientes de regresión para el modelo de predicción PLSR de contenido de saturados	67
Figura 23. Curva de valores predichos contra valores medidos del contenido de saturados.....	68
Figura 24. Gráfica de residuales para el porcentaje del contenido de Saturados.....	70
Figura 25. Varianza explicada acumulativa de contenido de aromáticos.....	71
Figura 26. RMSECV, gráfica de coeficientes, modelo de calibración y gráfica de residuos para contenido de aromáticos	72
Figura 27. Varianza explicada acumulativa de contenido de resinas	74
Figura 28. RMSECV, gráfica de coeficientes, modelo de calibración y gráfica de residuales para contenido de resinas.....	75
Figura 29. Varianza explicada acumulativa de contenido de asfaltenos	76
Figura 30. RMSECV, gráfica de coeficientes, modelo de calibración y gráfica de residuales para contenido de asfaltenos.....	77

Figura 31. Varianza explicada acumulativa de contenido de carbón Conradson	78
Figura 32. RMSECV, gráfica de coeficientes, modelo de calibración y gráfica de residuales para contenido de carbón Conradson.....	79
Figura 33. Raíz del error cuadrático medio de validación cruzada (RMSECV) en función del número de componentes (variables latentes) para el modelo de predicción de la fracción de saturados.	81
Figura 34. Varianza explicada acumulativa de predictores y de contenido de saturados.....	82
Figura 35. Coeficientes de regresión para el modelo de predicción PLSR de contenido de saturados	83
Figura 36. Curva de valores predichos contra valores medidos del contenido de saturados.....	84
Figura 37. Gráfica de residuales para el porcentaje del contenido de Saturados.....	85
Figura 38. Curva de valores predichos contra valores medidos del contenido de aromáticos (B), resinas(C), asfaltenos (D) y carbón Conradson (E)	87
Figura 39. Grafica de residuales del contenido de aromáticos (B), resinas(C), asfaltenos (D) y carbón Conradson (E)	88

LISTA DE TABLAS

Tabla 1. Análisis SARA y contenido de carbón Conradson de las muestras de crudos	45
Tabla 2. Intensidad de espectros Raman crudo 45 a 1884.2 cm ⁻¹	50
Tabla 3. Máxima intensidad de los espectros Raman del crudo Santos 65 deshidratado a su respectivo número de onda	50
Tabla 4. Parámetros de adquisición de los espectros Raman.....	53
Tabla 5. Parámetros estadísticos del modelo de calibración para el contenido de saturados	68
Tabla 6. Parámetros estadísticos para el modelo de calibración del contenido de aromáticos....	72
Tabla 7. Parámetros estadísticos del modelo de calibración para el contenido de resinas.....	74
Tabla 8. Parámetros estadísticos del modelo de calibración para el contenido de asfaltenos	77
Tabla 9. Parámetros estadísticos del modelo de calibración para el contenido de carbón Conradson	79
Tabla 10. Parámetros estadísticos del modelo de calibración para el contenido de saturados	82
Tabla 11. Parámetros estadísticos del modelo de calibración para el contenido de saturados, aromáticos, resinas, asfaltenos y carbón Conradson.....	86

LISTA DE ANEXOS

ANEXO A: Análisis SARA y contenido de carbón conradson de referencia de las muestras de crudo empleadas en este trabajo.....93

ANEXO B: RMSECV y coeficientes de regresión de los modelos de predicción del contenido de aromáticos (B), resinas(C), asfaltenos (D) y carbón Conradson (E).....97

RESUMEN

TÍTULO: DESARROLLO DE MODELOS PARA PREDECIR EL ANÁLISIS SARA Y CONTENIDO DE CARBÓN CONRADSON DE CRUDOS COLOMBIANOS A PARTIR DE ESPECTROSCOPIA RAMAN, ESPECTROMETRÍA DE MASAS Y QUIMIOMETRÍA*

AUTOR: MÓNICA ALEXANDRA FLÓREZ GONZÁLEZ**

PALABRAS CLAVES: CRUDOS COLOMBIANOS, ESPECTROSCOPIA RAMAN, CALIBRACIÓN MULTIVARIABLE, ANALISIS SARA, CARBÓN CONRADSON, ESPECTROMETRIA DE MASAS.

DESCRIPCIÓN:

La determinación de las propiedades fisicoquímicas de los crudos es esencial para la optimización de los procesos de producción, con el objetivo de generar productos de alto valor de forma eficiente y mínimo impacto medioambiental. El conocimiento a nivel molecular y su relación con las características fisicoquímicas igualmente son fundamentales ya que permiten generar propuestas innovadoras en los procesos de refinación y transporte desde el pozo hacia las refinerías y puertos. La espectroscopía Raman y la espectrometría de masas generan un gran volumen de información a nivel molecular de crudos. El procesamiento de estas señales espectrales y la incorporación de técnicas quimiométricas, permite la construcción de modelos para predecir propiedades fisicoquímicas de los crudos. La determinación de estas propiedades facilita la toma de decisiones en procesos de refinación y la obtención de productos de alto valor. En este trabajo se utilizaron espectros Raman y espectros de masas de crudos de diferentes regiones de Colombia; los espectros fueron utilizados para predecir las fracciones de saturados, aromáticos, resinas, asfaltenos (análisis SARA) y residuos de carbón Conradson (CCR), mediante regresión por mínimos cuadrados parciales (*PLSR, partial least square regression*). Por otro lado, las propiedades fueron medidas cumpliendo con las normas establecidas por la *ASTM- American Society for Testing and Materials*. Los criterios de desempeño de los modelos atienden al valor del error cuadrático medio de la validación cruzada (*RMSECV*) y el coeficiente de determinación (R^2).

* Trabajo de Grado

**Facultad de Ciencias. Escuela de Física. Director: Dr. Enrique Mejía-Ospino.

ABSTRACT

TITLE: PREDICTIVE MODELS FOR SARA ANALYSIS AND CONRADSON CARBON CONTENT OF COLOMBIAN CRUDE OILS USING RAMAN SPECTROSCOPY, MASS SPECTROMETRY AND CHEMOMETRIC[†]

AUTHOR: MÓNICA ALEXANDRA FLÓREZ GONZÁLEZ[‡]

KEYWORDS: COLOMBIAN CRUDE OILS, RAMAN SPECTROSCOPY, MULTIVARIATE CALIBRATION, SARA ANALYSIS, CONRADSON CARBON RESIDUE, MASS SPECTROMETRY

DESCRIPTION:

Determination of physicochemical properties of crude oils is essential for the optimization of production processes, with the aim of generating high value products in an efficient manner and with minimal environmental impact. Knowledge at the molecular level and its relationship with physicochemical characteristics are also fundamental as they allow generating innovative proposals in the refining and transport processes from the oil well to the refineries and ports. Raman spectroscopy and mass spectrometry generates a large volume of information at the molecular level of crude oils, which processed using chemometric methods allows building models to predict their physicochemical properties, facilitating to make decisions in refining processes for obtain high-value products efficiently and minimal environmental impact. In this work, the Raman and mass spectra of crude oils from different regions of Colombia and Partial Least Square Regression (PLSR) were used to determine the fractions of saturates, aromatics, resins, asphaltenes (SARA analysis) and Conradson carbon residue (CCR), which were determined complying with the standards established by the American Society for Testing and Materials (ASTM). The values of the root mean square error of cross validation (RMSECV) and the coefficient of determination (R^2) were used to select the model.

[†] Work Degree

[‡] Faculty of Sciences. School of Physics. Director: Dr. Enrique Mejía-Ospino.

INTRODUCCIÓN

La producción petrolera en Colombia ha experimentado transformaciones que le han permitido ser exportador de crudo. Hacia finales del siglo pasado Colombia exportaba crudos con una calidad excelsa provenientes de los campos de Cusiana, Cupiagua y Caño Limón, los cuales llegaron a su máxima producción al finalizar el siglo. A medida que los suministros de los crudos más livianos y de fácil extracción continúan disminuyendo, como los campos de Caño Limón, Cusiana y Cupiagua, la demanda de petróleo sigue aumentando, el crudo más pesado que antes era poco atractivo para las compañías petroleras debido a los costos elevados asociados a sus procesos de extracción y refinación y a su menor valor económico comparado con el petróleo liviano, en la actualidad, se ha convertido en una alternativa que merece toda la atención de la industria petrolera.

Colombia tiene un gran potencial en reservas de crudo pesado; sus reservas están alrededor de 3.170 millones de barriles, razón por la cual el estudio de estos crudos es de supremo interés ya que su producción debe ser rentable. Se estima que de las reservas el 45% corresponde a crudos pesados y todo apunta a que en el 2018 representarán el 60% del total. Este potencial se encuentra principalmente en la cuenca de los Llanos, el valle del Magdalena medio, región del Putumayo y en la región de Yari-Caguán (Ministerio de minas y energía, 2013).

Cada tipo de crudo es único y es una mezcla compleja de miles de componentes. La mayoría de los componentes presentes en el petróleo crudo son hidrocarburos (componentes orgánicos compuestos por átomos de hidrógeno y carbono). Asimismo, se pueden encontrar no sólo carbono e hidrógeno, sino también pequeñas (pero importantes) cantidades de otros “hetero”

elementos, en particular azufre, nitrógeno y ciertos metales (por ejemplo, níquel, vanadio, etc.) (Speight, 1999)

Por otra parte, debido a que los crudos pesados son muy viscosos, ya que contienen moléculas grandes como los asfáltenos, su condición de transporte por las tuberías representa grandes dificultades logísticas y financieras. La solución más sencilla es adicionar nafta (fracción ligera) para disminuir la viscosidad del crudo y facilitar su transporte, pero ésta representa un costo adicional para las compañías.

La determinación de las propiedades fisicoquímicas de los crudos es esencial para la optimización de los procesos de producción, con el objetivo de generar productos de alto valor de forma eficiente y mínimo impacto medioambiental (Speight, 1999). El conocimiento a nivel molecular y su relación con las características fisicoquímicas igualmente son fundamentales ya que permiten generar propuestas innovadoras en los procesos de refinación y transporte desde el pozo hacia las refinerías y puertos.

En la actualidad el proceso típico de caracterización de los crudos y determinación de sus propiedades, incluye la prueba SARA (saturados, aromáticos, resinas y asfáltenos). Existen diversas metodologías diseñadas para conocer y cuantificar las fracciones de saturados, aromáticos, resinas y asfáltenos presentes en el crudo regidas por la *American Society Test and Material* (ASTM) entre los cuales encontramos, la cromatografía de adsorción basada en metodologías estándar (ASTM-D4124, D2007) (American Society Test and Materials (ASTM), 2009) (ASTM D2007-93 Annual Book of ASTM Standards, 1998). En la primera etapa de este proceso la muestra del crudo es diluida en n-alcano y después por un proceso de filtrado se retiran los asfáltenos y posteriormente a través de un proceso de cromatografía se separan las

fracciones de saturados, aromáticos y resinas (maltenos). Esta metodología genera muy altos costos debido a que requiere una gran cantidad de muestra y disolventes, los cuales no son amigables con el medio ambiente y finalmente un tiempo de aproximadamente de 48 horas para la obtención de los resultados (Melendez, Lache, Orrego, Pachon, & Mejia Ospino, 2012).

Debido a que los crudos pesados y sus fracciones poseen una alta complejidad composicional su caracterización es también compleja y obtener información rigurosa de estos crudos es muy difícil. En la actualidad varias técnicas espectroscópicas (NMR, NIR, FTIR, Raman y espectrometría de masas MS) se están utilizando para la caracterización y clasificación del petróleo, análisis de sus fracciones y productos del proceso de refinación (Melendez, Lache, Orrego, Pachon, & Mejia Ospino, 2012) (Filgueiras P. R., 2014) (Filgueiras P. R., 2016) (Ingvar Eide, 2005). Estas metodologías tienen enormes ventajas comparadas con las técnicas tradicionalmente usadas ya que generan una respuesta rápida, tienen un requerimiento mínimo de preparación de la muestra y no requiere grandes cantidades de solventes.

Técnicas de ionización como la desorción/ionización láser asistida por matriz (MALDI), ha hecho posible la ionización de macromoléculas como las que componen el crudo. MALDI acoplada a un analizador de masa tiempo de vuelo (TOF), constituyen una herramienta de Espectrometría de masas robusta para análisis de compuestos de alto peso molecular. En general MALDI-TOF/MS ha sido exitosa en el análisis de biopolímeros para lo cual ésta fue desarrollada inicialmente y posteriormente aplicada al análisis del crudo.

El uso de técnicas analíticas que permitan obtener información a nivel molecular del crudo de manera confiable, rápida y económica, se hace cada vez más necesario. La espectroscopia vibracional y la espectrometría de masas MALDI-TOF ofrecen gran volumen de

información que puede ser procesada por medio de métodos quimiométricos como PCA y PLS, con el objetivo de desarrollar modelos que permitan predecir con relativa rapidez las propiedades fisicoquímicas del crudo (Cooper, 1999) (Pudenzi, 2012).

En el presente trabajo se plantea una metodología para el desarrollo de modelos basados en técnicas quimiométricas como la regresión de mínimos cuadrados parciales (PLS) para determinar las fracciones de saturados, aromáticos, resinas, asfáltenos (análisis SARA) y contenido de carbón Conradson a partir del procesamiento de espectros de muestras de crudos colombianos adquiridos por espectroscopia Raman y espectrometría de masas MALDI-TOF/TOF.

1. Aspectos teóricos

1.1 Aspectos Generales del petróleo crudo

El petróleo es quizás la sustancia más importante consumida en la sociedad moderna, fundamental para su desarrollo. Proporciona no sólo materias primas para los plásticos y otros productos, sino también combustible para la energía, la industria, y el transporte. Desde una mirada química, su composición puede variar no sólo por la ubicación y la edad del yacimiento de petróleo, sino también con su profundidad. De hecho, dos pozos adyacentes pueden producir petróleo con muy diferentes características. Cada tipo de crudo es único y es una mezcla compleja de miles de componentes. La mayoría de los componentes presentes en el petróleo crudo son hidrocarburos (componentes orgánicos compuestos por átomos de hidrógeno y carbono). Asimismo, se pueden encontrar no sólo carbono e hidrógeno, sino también pequeñas (pero importantes) cantidades de otros “hetero” elementos, en particular azufre, nitrógeno y ciertos metales (por ejemplo, níquel, vanadio, etc.). (Speight, 1999)

1.1.1 Refinación. La obtención de productos valiosos (en especial GLP, gasolina, combustible pesado, combustible diésel, materias primas de petroquímicos, gasóleo de calefacción, aceite combustible y asfalto) a partir de diversos procesos y subprocesos de transformación del petróleo crudo con valor agregado, se denomina refinación. Este proceso se lleva a cabo en etapas: primero mediante un proceso de destilación se separa el crudo en diferentes fracciones (cada una con un punto de ebullición y distribución de carbono diferentes), posteriormente las fracciones obtenidas se someten a otros

procesamientos con el fin de obtener productos terminados, mediante una secuencia de transformaciones físicas y químicas (council, 2011) (Speight, 1999).

1.1.2 Destilación del crudo. La refinación comienza con la separación del petróleo por medio de un proceso de destilación. Las fracciones de crudo obtenidas están directamente relacionadas con las características del crudo procesado y se caracterizan por sus puntos de ebullición (Ministerio de minas y energía, 2013). Cada fracción obtenida de la unidad de destilación del crudo se define mediante un punto de ebullición único y a su vez está compuesta por un enorme número de distintos compuestos. Estas fracciones incluyen gases livianos, naftas, destilados, gasóleos y aceites residuales. Cada una de estas fracciones pasa por un proceso de refinación diferente para continuar su procesamiento. Los derivados de mayor importancia hacen referencia a los combustibles líquidos, cuyo uso principal es el de proveer la energía para el sector transporte (council, 2011).

Las naftas son materiales combustibles con punto de ebullición entre 28°C-177°C. Por lo general se envían a las unidades de mejoramiento (para mejorar el octano, controlar el azufre, etc.) y luego a la mezcla de combustibles. Los destilados, incluido el queroseno, por lo general atraviesan más tratamientos y luego se incorporan al combustible pesado, diésel y gasóleo de calefacción. El gasóleo pasa por las unidades de conversión, donde se fraccionan en flujos más livianos (gasolina, destilado). Finalmente, el aceite residual (o residuos) es dirigido a otras unidades de conversión, o se mezcla con el combustible industrial pesado y/o asfalto. Los residuos tienen relativamente escaso valor económico, en realidad un valor más bajo que el petróleo crudo del cual se extraen.

1.1.3 Procesos de conversión (craqueo). Los procesos de conversión provocan reacciones químicas que rompen moléculas de hidrocarburo de gran tamaño (bajo valor económico) dando origen a moléculas más pequeñas y livianas y reduce la proporción C/H. Las unidades de conversión forman el núcleo central de las operaciones de refinación modernas debido a que brindan flexibilidad operativa para mantener la elaboración de productos livianos y permiten el uso económico de los crudos sulfurados pesados. Los procesos de conversión de mayor interés son el craqueo catalítico fluidizado (FCC), el hidrocrqueo y la coquización (Speight, 1999).

El FCC opera a altas temperaturas y baja presión y emplea un catalizador para convertir el gasóleo pesado a partir de la destilación del crudo en gases livianos, materias primas de petroquímicos, mezcla de componentes de gasolina (nafta de FCC), y mezcla de componentes de combustible diésel. El hidrocrqueo ofrece altos rendimientos de productos livianos y una extensa flexibilidad. El hidrocrqueo presenta una notable ventaja respecto del FCC. La entrada de hidrógeno al hidrocrqueador no sólo causa reacciones de craqueo, sino también otras reacciones que extraen los heteroátomos, en especial el azufre, de los flujos hidrocrqueados (council, 2011).

La coquización es proceso de conversión térmica, no catalítico que descompone el aceite residual, el residuo más pesado que resulta de la destilación del crudo, en un rango de intermedios más livianos para continuar su procesamiento.

1.1.4 Procesos de mejoramiento. Los procesos de mejoramiento provocan reacciones químicas que combinan o reestructuran las moléculas en los flujos de bajo valor para producir otros de mayor valor, principalmente una mezcla de componentes de gasolina con alto octanaje y bajo contenido de azufre. Todos los procesos de mejoramiento de principal interés emplean catalizadores, incluyen moléculas de hidrocarburo pequeñas y se aplican a la producción de gasolina (council, 2011)

1.2 Análisis SARA

Es común emplear análisis por tipo de grupo de Hidrocarburos, ya que son necesarios y de mayor importancia para la industria. Con ellos se pueden hacer estudios del yacimiento, migración y maduración del crudo, procesos de degradación, procesamiento del crudo y efectos medioambientales. Un ejemplo de tal análisis por tipo de grupo de Hidrocarburos es el Análisis SARA, el cual separa los aceites crudos en cuatro clases químicas principales basándose en diferencias de solubilidad y polaridad: Saturados (S), Aromáticos (A), Resinas (R), y Asfáltenos(A). Estas fracciones son de gran importancia debido a que, mediante diversos estudios y diferentes análisis, se pueden predecir posibles acumulaciones de asfáltenos, que resulta en un problema que genera un gran incremento de costos en la industria, dado que afecta la producción y los procesos de refinamiento.

Los saturados son la fracción menos polar y consiste principalmente de parafinas pesadas y cicloparafinas (Speight, 1999) y su densidad oscila entre 800 y 900Kg/m³. Los compuestos aromáticos consisten de uno o varios anillos aromáticos y pueden tener algunos anillos nafténicos unidos al anillo aromático y su densidad se encuentra entre 950 y 1050 Kg/m³. Las resinas consisten de compuestos aromáticos con núcleos grandes, tienen mayor polaridad, su

densidad aproximada es de 1050Kg/m^3 . Los asfáltenos son la fracción más polar en los crudos y más aromático que las resinas, heteroátomos como azufre, nitrógeno y oxígeno también se encuentran en la fracción de asfáltenos (Welch, 1998)

En la caracterización de fracciones pesadas, es usado normalmente el análisis SARA. La fortaleza de éste método radica en que analiza la muestra entera, permitiendo la comparación entre hidrocarburos de la misma naturaleza en base a un estándar consistente. Sin embargo, existen algunas limitaciones en éste análisis; entre ellas, las pérdidas significativas de compuestos volátiles en muestras con puntos de ebullición inicial por debajo de 260°C y los prolongados tiempos de respuesta (Lamus, Guzmán, Murcia, Cabanzo, & Mejía-Ospino, 2011), (Merdrignac & Espinat, 2007).

Existen diversos procedimientos diseñados para conocer y cuantificar las fracciones de saturados, aromáticos, resinas y asfáltenos presentes en el crudo entre los cuales encontramos, la cromatografía de adsorción basada en el método ASTM-D2007,D4124 (American Society Test and Materials (ASTM), 2009) (ASTM D2007-93 Anual Book of ASTM Standards, 1998), este método requiere una muestra grande, largo tiempo para el proceso, y grandes cantidades de disolventes. También existen otros métodos mejorados como la cromatografía líquida de alta presión (HPLC) y la cromatografía de capa delgada (TLC)

En la primera parte de la prueba se diluye la muestra del crudo en n-alcano tales como pentano o heptano, luego esta mezcla se filtra y los sólidos retenidos en el filtro constituyen los asfáltenos. La mezcla obtenida después del proceso de filtrado se conoce como maltenos; este contiene las tres fracciones, saturados, aromáticos y resinas. Estas tres fracciones posteriormente son separadas usando la técnica de cromatografía líquida (Barman & Cebolla, 2000).

Diferentes metodologías estándar, basadas en cromatografía líquida, se han desarrollado para análisis de diferentes tipos de hidrocarburos, fraccionamiento de crudos y de sus derivados (ASTM D2007-93 Anual Book of ASTM Standards, 1998). Las metodologías estándar ASTM-D4124, D2007 son usadas para el análisis de destilados pesados de hidrocarburos. Este método involucra el uso de n-pentano para la elución de saturados a través de una doble columna que contiene arcilla en la sección superior y arcilla y gel de sílice en la sección inferior. Durante la elución con n-pentano, los compuestos polares son retenidos en la arcilla y los aromáticos en la sílice. Posteriormente los polares son desorbidos de la arcilla usando una mezcla 1:1 (v/v) de acetona-tolueno. Los saturados y polares son determinados gravimétricamente después de completarse la evaporación de los solventes. La cantidad de aromáticos puede ser calculado por diferencia o puede ser recuperado de la sílice por extracción soxhlet usando tolueno (ASTM D2007-93 Anual Book of ASTM Standards, 1998) (Barman & Cebolla, 2000).

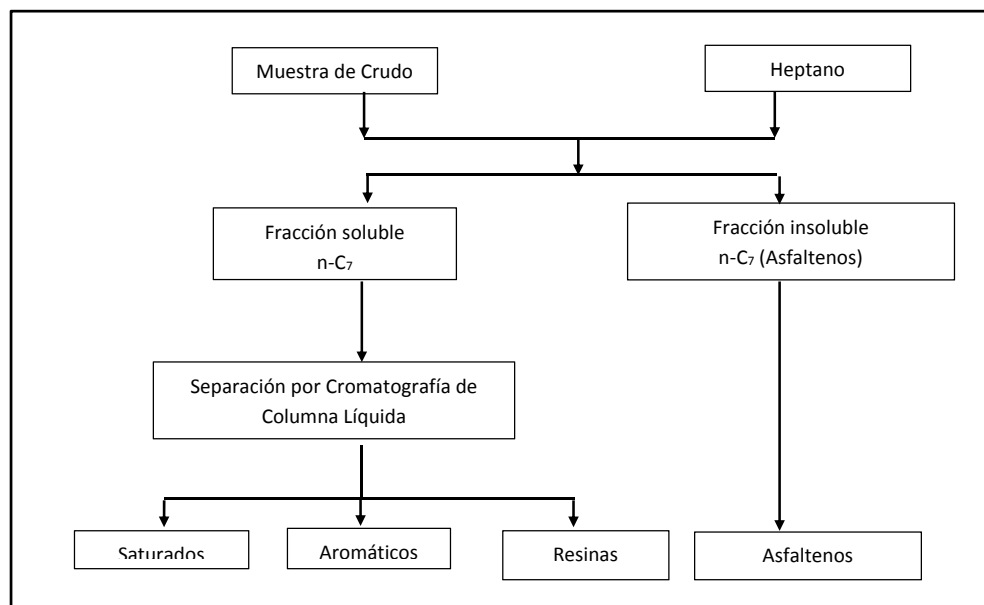


Figura 1. Esquema análisis SARA

La metodología estándar ha sido utilizada para el análisis de derivados del petróleo con puntos de ebullición por encima de 260°C, como también en crudos. Sin embargo, para la aplicación del método se requiere que la fracción correspondiente a los asfaltenos (insolubles en n-pentano) sea removida de la muestra a analizar. Las imprecisiones en los datos de la prueba ASTM en su mayor parte se originan a partir de efectos de sobre carga, causando contaminación cruzada de tipos de hidrocarburos, y se ha encontrado que esta es muy alta para destilados de alto punto de ebullición (Barman & Cebolla, 2000).

1.3 Espectrometría de masas

La espectrometría de masas (MS) está basada en la obtención de iones a partir de moléculas orgánicas en fase gaseosa con la ayuda de una fuente de ionización. Una vez obtenidos los iones, estos se separan de acuerdo a su relación masa carga (m/z) y finalmente se detectan con la ayuda de un dispositivo adecuado. La información presentada en el espectro representa la abundancia de los diferentes tipos de iones en función de la relación masa/carga de cada uno de ellos. Este espectro tiene relación con la estructura química del compuesto objeto de análisis.

Las ventajas de la MS se pueden resumir en los siguientes aspectos: proporciona una insuperable especificidad en la determinación del peso molecular debido a la posibilidad de medir con exactitud su masa molecular, así como obtener información a partir de los fragmentos iónicos de un analito. Su sensibilidad es elevada y en teoría la MS permite detectar una molécula; es muy versátil ya que es aplicable a todo tipo de muestras, polares, no polares, volátiles, no volátiles, sólidos, líquidos y gases, y es una de las técnicas más cualificada para analizar muestras complejas (Martinez Gomez & Ballesteros Gonzalez, 2009).

En un espectrómetro de masas se dan fundamentalmente cuatro procesos: el primero consiste en la vaporización de la sustancia, luego se debe dar la generación de iones a partir de las moléculas neutras en fase gaseosa, posteriormente una vez generados los iones se deben separar en función de su relación masa carga (m/z) y finalmente se detectan los iones y se registra la información de su relación m/z y su abundancia. El ion molecular (M^+) brinda la información más valiosa en el espectro; su masa y composición elemental indican las fronteras dentro de las cuales se deben encontrar los fragmentos que se detecten en el espectro (Martinez Gomez & Ballesteros Gonzalez, 2009). Para que en el espectrómetro se lleven a cabo los procesos mencionados, éste debe contar con un sistema de introducción de muestras, una fuente de iones, un analizador en donde se realice la separación de los iones y un sistema de detección y registro (fig. 2)

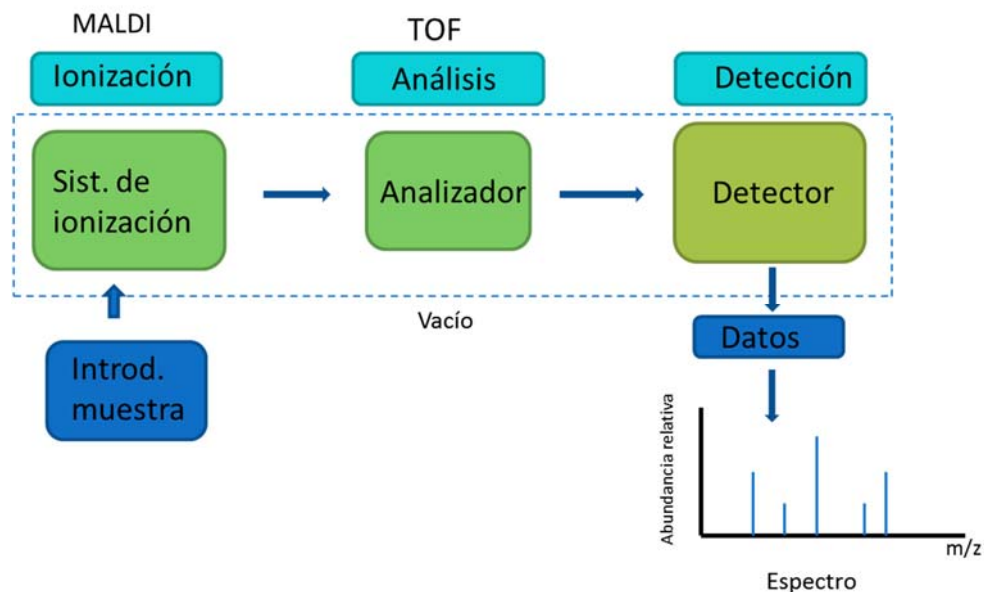


Figura 2. Esquema de un espectrómetro de masas

De acuerdo con la naturaleza de la muestra se determina el método de introducción de muestra, así como el método de ionización. Dentro de los sistemas de introducción de la muestra se encuentran: sistemas de entrada indirectos en los cuales la muestra se volatiliza externamente y se introduce en el sistema de ionización, sistema por sonda directa el cual se utiliza para introducir muestras sólidas y líquidas no volátiles y sistemas especiales para introducción de muestras acoplados a técnicas de separación como cromatografía de gases (GC), cromatografía líquida (GL) entre otros (Barman & Cebolla, 2000). Existe una amplia variedad de métodos de ionización, y la elección del método está determinada por la naturaleza de la muestra y la información que de ella se necesite, ya que la información presentada en el espectro de masas depende en gran medida del método de ionización utilizado. El factor principal que permitió la utilización de la MS en el análisis de moléculas de gran tamaño fue la aparición de técnicas de ionización suave, ya que las otras técnicas generaban degradaciones debido al exceso de energía depositada en moléculas (Abian, Carrascal, & Gay, 2008)

ESI-MS (Electrospray ionization mass spectrometry) y MALDI, son las dos técnicas de ionización utilizadas de forma general en el análisis de moléculas de alto peso molecular, conocidas como técnicas de ionización suaves, ya que el analito retiene muy poca energía residual. ESI-MS es una técnica de ionización a presión atmosférica, ideal para el análisis de moléculas de gran tamaño, no volátiles, que pueden ser analizadas en fase líquida (Enke, 2001).

MALDI, ionización/desorción láser asistida con matriz, se realiza mediante una radiación láser y en presencia de una matriz adecuada. En la ionización MALDI el analito se mezcla con una sustancia matriz que se encuentra en mucha mayor proporción que la muestra, la mezcla se deposita en un dispositivo porta muestras especialmente diseñado para este sistema de

ionización. Después de la evaporación del solvente, el analito y la matriz se co-cristalizan para posteriormente ser irradiada por un láser pulsado de alta potencia (10^6 Wcm^{-2}) durante algunos nanosegundos. Durante la incidencia del láser sobre el porta muestras se desorben las moléculas de la matriz y de la muestra y posteriormente se ionizan las moléculas de la muestra, manteniéndolas en fase gaseosa (fig. 3). La matriz es fundamental en el proceso de ionización ya que esta debe ser capaz de absorber del láser la energía suficiente y así causar indirectamente la vaporización del analito. Como el analito en sí no absorbe la energía del láser directamente, el método se considera una técnica de ionización suave (Marvin, Roberts, & Fay, 2003), permitiendo el análisis de muestras complejas de miles de Dalton. La matriz también juega un papel muy importante ya que favorece la ionización del analito ya sea siendo dador o receptor de un protón (Lewis, Wei, & Siuzdak, 2000).

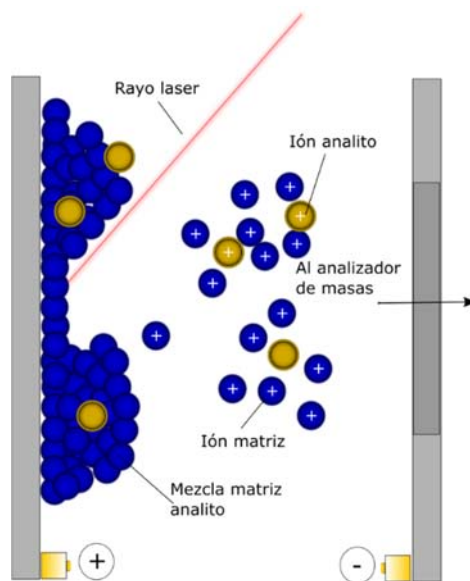


Figura 3. MALDI, ionización/desorción láser asistida con matriz

En éste proceso de ionización se llevan a cabo procesos complejos que involucran fenómenos ópticos, termodinámicos y fisicoquímicos. Distintas teorías se han desarrollado para explicar la desorción de moléculas de gran tamaño por MALDI. El origen de los iones en el proceso MALDI no está aun completamente definido, se cree que la ionización de las moléculas del analito se produce gracias a un fenómeno de transferencia de protones de las moléculas de la matriz al analito una vez sean desorbidas (Marvin, Roberts, & Fay, 2003), esto deja claro que el proceso de ionización depende críticamente de la combinación matriz-analito.

De los cuatro grandes grupos en los que se pueden dividir los analizadores de masas están, analizadores de sectores (eléctricos o magnéticos), de cuadrupolo, de tiempo de vuelo (TOF) y de confinamiento de iones (trampa de iones), resonancia de ion ciclotrón, con transformada de Fourier (FT-ICR).

Dentro de los analizadores usados más comúnmente está el analizador tiempo de vuelo (TOF- Time of flight), trabajado en modo lineal o reflectron, el cual posee excelentes ventajas como el registro completo de los iones procedentes de la fuente, sin limitaciones de masa de los iones y análisis extremadamente corto.

El analizador TOF se basa en el hecho de que los iones generados en la fuente se aceleran acelerados por un campo eléctrico, lo que proporciona la misma energía cinética, E_k , a todos los iones,

$$E_k = eV \quad (1)$$

al entrar al tubo (libre de cualquier campo eléctrico o magnético), adquieren distintas velocidades según el valor de su relación m/z y por tanto tardan distinto tiempo en recorrer una determinada

distancia (fig. 4). Debido a que todos los iones que entran en el tubo tienen idealmente la misma energía cinética.

$$\frac{1}{2} m v^2 = eV \quad (2)$$

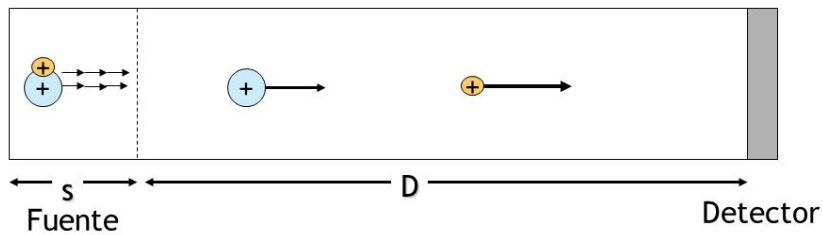


Figura 4. Esquema analizador TOF

Sus velocidades en el tubo deberán variar de modo inverso a sus masas, y por tanto las partículas ligeras llegarán antes al detector que las pesadas,

$$v = \sqrt{\frac{2eV}{m}} \quad (3)$$

La medida del tiempo de vuelo permite determinar de forma muy precisa la masa de cada uno de estos iones. Para una longitud L de analizador dada, el tiempo que tarda un ion en atravesar la distancia L es:

$$L = vt \quad (4)$$

$$t = \frac{L}{\sqrt{2V}} \sqrt{\frac{m}{e}} \quad (5)$$

$$t = K \sqrt{\frac{m}{e}} \quad (6)$$

En donde K está definido por los parámetros de diseño del analizador

A diferencia de los sistemas de cuadrupolo o de sectores, los cuales filtran en cada instante grupos de iones dentro de un pequeño rango de valores m/z desechando el resto de la población de iones (Marshall & Hendrickson, 2008), el analizador TOF separa y detecta en una escala de tiempo (tiempo de vuelo) el paquete completo de iones procedente de la fuente. El sistema trabaja por este motivo en régimen discontinuo por lo que es un detector indicado para técnicas de ionización de carácter pulsante como el MALDI. El TOF es un analizador relativamente simple, de relativo bajo costo, con un ciclo de barrido rápido y con un rango de masas en teoría ilimitado (Marshall & Hendrickson, 2008).

Mejoras sustanciales en la resolución de los espectros de masa se han logrado, mediante la implementación de una segunda etapa o reflectron, el cual combina la tecnología TOF con un analizador electrostático (Lewis, Wei, & Siuzdak, 2000). El reflectron situado en el extremo del tubo se utiliza para compensar la diferencia en el tiempo de vuelo de los iones de igual relación m/z , debido a sus diferentes energías cinéticas a la entrada del analizador (fig. 5). En el reflectron se incrementa el tiempo t , que los iones necesitan para alcanzar el detector mientras reduce la diferencia entre las velocidades de los iones con la misma masa, reduciendo el Δt , que conduce a un aumento en la resolución (Xian & Hendrickson, 2012).

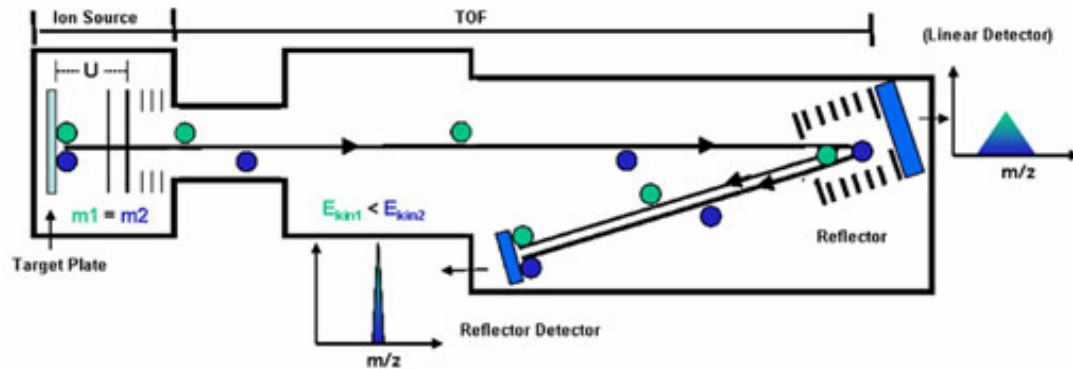


Figura 5. Analizador en modo reflector

1.4 Espectroscopia Raman

La espectroscopia Raman no está basada en un proceso de absorción, sino en un efecto de dispersión de luz, llamado más tarde por el nombre del científico Indio que lo observó en 1928, Chandasekhara Venkata Raman. Este científico dió nombre al fenómeno inelástico de dispersión de la luz que permite el estudio de rotaciones y vibraciones moleculares. El estudio sobre este fenómeno se basó en los trabajos realizados por Rayleigh.

Si se hace incidir un haz de fotones de energía $E = h\nu_0$, y frecuencia ν_0 (h es la constante de Planck) sobre una muestra cuyas características moleculares se desean determinar, la mayoría de los fotones que componen la radiación serán dispersados sin cambiar su energía, este proceso constituye una dispersión elástica o dispersión Rayleigh, y no aporta información sobre la composición molecular de la muestra irradiada. Por otro lado, una muy pequeña fracción de los fotones al interactuar con la molécula se dispersan con una frecuencia diferente, es la que proporciona información de la molécula y es llamada dispersión inelástica o Raman (Brown, 2003) (Hollas, 2002). La intensidad de la radiación dispersada varía fuertemente con la longitud de onda (Hollas, 2002):

$$I_s \propto \lambda^{-4} \quad (7)$$

La energía interna de una molécula (rotacional y vibracional) está cuantizada, cada uno de los movimientos vibracionales y rotacionales de la molécula corresponderá a un valor determinado de energía. Las variaciones de frecuencia que se observan en la dispersión Raman, equivalen a variaciones de energía. Esta dispersión puede ser interpretada como sigue: un primer fotón incidente no tiene energía suficiente para excitar la molécula a un estado electrónico de mayor energía y lleva a la molécula transitoriamente a un “estado virtual”, que no es un verdadero estado cuántico de la molécula (Brown, 2003) (R.L. McCreery, 2000); este estado virtual es abandonado rápidamente para pasar a un estado de energía permitido, emitiendo un fotón de una frecuencia que dependerá del salto energético realizado por la molécula (fig. 6).

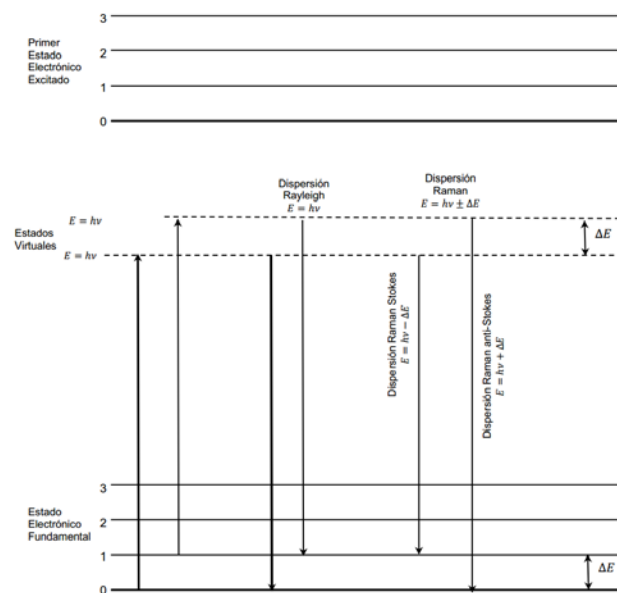


Figura 6. Efecto Raman

Si el fotón dispersado tiene una frecuencia menor que la del incidente, la molécula después de pasar al estado virtual de energía vuelve a uno de mayor energía que el que tenía

inicialmente antes de interactuar con el fotón; el fotón es dispersado con energía $h(\nu_0 - \nu_1)$, correspondiente a la energía vibracional cuantizada, tomada por la molécula y es llamada dispersión Raman Stokes (Brown, 2003). Por otra parte, si el fotón dispersado tiene una frecuencia mayor que la frecuencia del fotón incidente, la molécula transfiere energía al fotón; esto significa que la molécula no se encontraba en su estado vibracional (o rotacional) fundamental, sino en uno de mayor energía y después de la interacción, el fotón es dispersado con frecuencia $\nu_0 + \nu_1$ y se produce la dispersión Raman anti-Stokes.

¿Cuál es el papel que juegan estos dos fotones en el nivel molecular? Una interpretación clásica del efecto Raman es útil considerando el caso de una molécula diatómica. El primer fotón expone la molécula a un campo eléctrico oscilante, que genera oscilación en los electrones de la molécula,

$$E = E_0 \cos 2\pi\nu_0 t \quad (8)$$

en donde E_0 es la amplitud de la onda electromagnética incidente y ν_0 es su frecuencia. Como resultado se induce en la molécula un momento dipolar eléctrico dependiente del tiempo. De acuerdo a la física clásica este momento dipolar oscilante a su vez emite radiación electromagnética (esto explica porque el fotón es emitido), dejando a la molécula en un nivel rotacional o vibracional diferente de acuerdo a las reglas de selección. Para baja intensidad de luz, la magnitud del momento dipolar inducido P es proporcional al campo eléctrico, la constante de proporcionalidad es conocida como polarizabilidad α de la molécula, que se puede visualizar como un cambio en la forma de la nube electrónica.

$$P = \alpha E = \alpha E_0 \cos 2\pi\nu_0 t \quad (9)$$

Para una amplitud de vibración pequeña, la polarizabilidad α puede ser expresada en términos de la separación internuclear (Brown, 2003):

$$\alpha = \alpha_0 + \left(\frac{\partial \alpha}{\partial Q}\right)_0 Q_0 + \dots \quad (10)$$

en donde α_0 es la polarizabilidad en la posición de equilibrio Q_0 . La distancia Q de la molécula que vibra con frecuencia ν_m :

$$Q = Q_0 \cos 2\pi\nu_m t \quad (11)$$

Reemplazando en la expresión del momento dipolar se obtiene:

$$P = \left(\alpha_0 + \left(\frac{\partial \alpha}{\partial Q}\right)_0 Q\right) E_0 \cos 2\pi\nu_0 t \quad (12)$$

$$P = \alpha_0 E_0 \cos 2\pi\nu_0 t + \left(\frac{\partial \alpha}{\partial Q}\right)_0 Q_0 \cos 2\pi\nu_m t E_0 \cos 2\pi\nu_0 t \quad (13)$$

$$P = \alpha_0 E_0 \cos 2\pi\nu_0 t + \frac{1}{2} \left(\frac{\partial \alpha}{\partial Q}\right)_0 Q_0 E_0 [\cos 2\pi t(\nu_m + \nu_0) + \cos 2\pi t(\nu_0 - \nu_m)] \quad (14)$$

En donde el primer término corresponde a un dipolo oscilante que emite radiación con una frecuencia igual a la incidente ν_0 (dispersión Raleigh), el segundo corresponde a la dispersión Raman anti-Stokes de frecuencia $\nu_m + \nu_0$ y dispersión Raman Stokes de frecuencia $\nu_0 - \nu_m$, respectivamente. La temperatura es una variable importante en los resultados ya que a temperatura ambiente la población de moléculas en estados vibracionales bajos es más alta que la población de moléculas con estados altos, es decir habrá más moléculas realizando la transición

tipo Stokes reflejando una intensidad más alta en el espectro. La población de moléculas está determinada por la distribución de Boltzman (Brown, 2003):

$$\frac{N_i}{N_0} = e^{-\frac{\Delta E}{kT}} \quad (15)$$

La energía intercambiada entre la molécula y la radiación en el efecto Raman involucra la polarizabilidad y el campo eléctrico (Brown, 2003),

$$W = \frac{1}{2} \propto E^2 \quad (16)$$

1.5 Quimiometría

El desarrollo de la Quimiometría como disciplina está fuertemente relacionada por el uso de los computadores en química. Actualmente los laboratorios cuentan con equipo sofisticado que generan un elevado volumen de información, que deben ser procesados e interpretados adecuadamente para que representen una información útil para el usuario (Otto, 2007)

En los últimos años las técnicas espectroscópicas junto con métodos matemáticos de tratamientos de datos han dado origen a una nueva área de la química analítica llamado Quimiometría (Filzmoser, 2009).

La Quimiometría está actualmente definida como: disciplina química que usa métodos matemáticos y estadísticos para diseñar o seleccionar experimentos y procedimientos de medidas

óptimos y suministrar la máxima información química por medio del análisis de los datos químicos (Otto, 2007)

Entre las áreas que contempla la Quimiometría se destaca el diseño de experimentos, el reconocimiento de patrones y la calibración multivariable. En el área de diseño experimental se buscan las variables que más afectan a determinado proceso, así como la interacción entre ellos. En reconocimiento de patrones, a partir de una amplia gama de información (medidas espectrales, por ejemplo) en una serie de objetos, se pretende encontrar grupos de muestras (objetos) que son similares entre sí y por lo tanto detectar las tendencias de datos. En la calibración multivariable, se busca establecer un modelo que relaciona una serie de medidas realizadas sobre muestras con una cierta propiedad (Otto, 2007). En la calibración multivariable el objetivo es el desarrollo de modelos predictivos de propiedades fisicoquímicas con base en propiedades medidas del sistema químico (tales como datos espectrales), que arrojen los mismos resultados que los encontrados con los procedimientos estándar para dichas propiedades, con alta confiabilidad, pero con menores tiempos de análisis y costos (Filgueiras P. R., 2016).

A partir de la correlación del espectro Raman y el espectro de masas con las propiedades obtenidas por métodos estandarizados de referencia (análisis SARA, contenido de carbón Conradson), se pueden construir modelos mediante la regresión de mínimos cuadrados parciales (PLSR) que puedan predecir en el futuro estas propiedades.

1.5.1 Calibración multivariable. El análisis de muestras complejas constituye un reto importante de la química analítica. El desarrollo rápido de la Quimiometría en química analítica se debe fundamentalmente a la existencia de herramientas informáticas acopladas a la moderna instrumentación. Las medidas con gran contenido de información, por ejemplo, medidas espectrales (multivariable), se pueden transformar matemáticamente a la forma más adecuada para la obtención de un modelo de calibración, que permita predecir el valor de la propiedad fisicoquímica estudiada a partir de la relación entre la matriz de datos X , que contiene los datos espectroscópicos y la matriz de datos Y que representan los valores de las propiedades de las muestras. Una vez obtenido el modelo de calibración la propiedad se puede predecir por medio de la ecuación:

$$Y = X B + E \quad (17)$$

Donde X es la matriz de datos espectroscópicos, B es la matriz de regresores que se obtiene en el proceso de calibración del modelo y E es la matriz de residuales aleatorios de la propiedad fisicoquímica.

Dentro de las técnicas más usadas tenemos la regresión por componentes principales(PCR), regresión por mínimos cuadrados parciales(PLSR) y la regresión lineal múltiple(MLR).

El método PCR presenta inconvenientes si la selección de los PC no son los apropiados para predicción de propiedades de los analitos de estudio; por esta razón la técnica de calibración por regresión parcial por mínimos cuadrados (PLSR) se presenta como una excelente opción.

1.5.1.1 Regresión parcial por mínimos cuadrados (PLSR). La técnica PLSR busca hacer la descomposición de la matriz X independientemente de la matriz Y (figuras 7 y 8). Usa la información contenida en la matriz X (datos espectroscópicos) y la información en la matriz Y (propiedad a estimar), y obtiene unas variables auxiliares llamadas variables latentes, parecidas a los componentes principales. Cada dirección de la variable latente de la matriz X es modificada de manera que la covarianza entre esta y la matriz Y es maximizada. Las matrices X y Y se descomponen simultáneamente en matrices más pequeñas de acuerdo al esquema (Otto, 2007)

$$X = TP^T + E = \sum_{i=1}^d t_i p_i^T + E \quad (18)$$

$$T = XP^T + E$$

Figura 7. Notación matricial para la descomposición de la matriz de regresores

$$Y = UQ^T + F = \sum_{i=1}^d u_i q_i^T + F \quad (19)$$

$$\begin{array}{c} m \\ \boxed{Y} \\ n \end{array} = \begin{array}{c} d \\ \boxed{U} \\ n \end{array} \begin{array}{c} m \\ \boxed{Q} \\ d \end{array} + \begin{array}{c} m \\ \boxed{F} \\ n \end{array}$$

Figura 8. Notación matricial para la descomposición de la matriz de predictores

Donde $T_{n \times d}$ es la matriz de scores, $P_{d \times p}$ es la matriz de pesos, y E los residuales para la matriz de datos X ; $U_{n \times d}$ es la matriz de scores, $Q_{d \times m}$ de pesos y F de residuales, para la matriz de la propiedad a determinar Y . De la misma manera n es el número de muestras, d es el número de componentes, p son las variables originales (espectros) y m las variables dependientes (propiedades). En este caso los pesos no coinciden con los PCA, ya que están corregidos para conseguir la máxima capacidad predictiva para la matriz Y

La descomposición de las matrices se realiza simultáneamente, estableciéndose una relación lineal interna entre los scores de las matrices X e Y , de manera que se cumpla para cada componente p :

$$u_d = b_d t_d \quad (20)$$

Donde b_d es el coeficiente de regresión para cada componente, el cual considera la contribución de cada variable al modelo PLS. Se calcula entonces el valor de Y

$$Y = TBQ^T + F \quad (21)$$

Donde B es la matriz de los regresores b_d , agrupados en una matriz diagonal B y F la matriz de los residuales de Y .

Cuando ya se ha conseguido el modelo de calibración se puede realizar la predicción de la propiedad modelada en un nuevo conjunto de muestras,

$$y_i = bX_i + e_i \quad (22)$$

Siendo X_i vector que contiene la señal de la muestra, y_i el vector de propiedades a determinar y b un vector que permite determinar la predicción de la propiedad de una muestra sin necesidad de descomponer su matriz en matrices T y P .

1.6 Antecedentes de la espectrometría de masas, espectroscopia Raman y Quimiometría en el análisis de crudos.

Debido a la gran aplicación de la MS en el análisis del petróleo se ha generado un área de investigación llamada petroleómica (*petroleomics*) enfocada en obtener una comprensión detallada de la relación entre la composición química del petróleo y sus propiedades fisicoquímicas con el objetivo de ayudar en la solución de problemas de producción y procesamiento (Hsu, Hendricson, Rodgers, McKenna, & Marshall, 2011) . En las primeras aplicaciones de la MS al análisis del petróleo, el método de ionización más práctico fue el bombardeo por electrones de alta energía (~70 eV), que generaba una alta fragmentación, la cual era perjudicial, pues la generación de más de una señal por ion analito complicaba enormemente el espectro de masas; esto limitó el análisis por MS a compuestos volátiles y semivolátiles para GC/MS (Rogers, Shaub, & Marshall, 2005) (Merdrignac & Espinat, 2007).

La GC/MS ofreció detallado análisis de la composición del petróleo más liviano y de sus destilados (Eckert & Roach, 2012) (Rogers, Shaub, & Marshall, 2005), sin embargo, hasta hace poco se sabe acerca de la composición de los crudos más pesados. El factor principal que permitió la utilización de la MS en el análisis del crudo pesado fue la aparición de técnicas de ionización suave, ya que las otras técnicas generaban degradaciones y alta fragmentación debido al exceso de energía depositada en dichas moléculas. ESI y MALDI, son las dos técnicas de ionización suave utilizadas de forma general en la ionización de moléculas de alto peso molecular.

La generación de iones cuasi moleculares ($[M + H]^+$ y $[M-H]^-$), por ionización química, ESI, MALDI ha permitido la caracterización detallada de especies de baja volatilidad. La ionización por ESI se hace sobre las especies polares del crudo y la fragmentación durante la ionización es mínima. El análisis de crudo por ESI FT-ICR MS ha hecho posible obtener espectros de masa de muy alta resolución ($m/\Delta m_{50\%}) > 300000$) en aproximadamente 1 segundo, por lo que es posible identificar alrededor de 17000 especies polares en una muestra de crudo (10% en peso). De la misma manera se han realizado comparaciones de crudos de origen geográfico similar que han sufrido diferentes grados de biodegradación (Rogers & Marshall, 2004).

MALDI –TOF/MS se ha utilizado para el análisis de la composición de muestras de crudo, así como de los asfáltenos extraídos de éstas muestras, confirmando que el análisis MALDI-TOF es un proceso eficiente, con gran potencial para la estimación de propiedades fisicoquímicas de asfáltenos (Araujo Pantoja, Mendes, & Oller Nascimento, 2003). De la misma manera el sistema MALDI/TOF ha sido adaptado para compuestos polares y aromáticos y se

han reportado el uso de matrices no polares, para el análisis de fracciones de crudo de alto punto de ebullición, las cuales son muy eficaces para la caracterización de fracciones pesadas del crudo, proporcionando espectros de masas de alta calidad (Robins & Limbach, 2003) (Intech, 2012); los resultados obtenidos fueron comparados con otros obtenidos utilizando matrices polares así como también con la técnica (LDI-TOF)(19). Por otra parte, se han realizado estudios de asfáltenos utilizando la desorción/ionización láser sin matriz (LDI-TOF), para analizar fracciones pesadas del crudo de pesos moleculares de hasta 30.000 Da y más (Araujo Pantoja, Mendes, & Oller Nascimento, 2003). En estudios recientes los datos de espectros obtenidos con y sin matriz se han comparado, indicando que el uso de una matriz para muestras de crudo podría ser necesario, pero por otra parte también se demostró que para muestras de asfáltenos los resultados con o sin matriz no presentan diferencias significativas (Araujo Pantoja, Mendes, & Oller Nascimento, 2003), pudiéndose obtener espectros de masas, correlacionándolos con sus propiedades físicas (por ej. Viscosidad) (Araujo Pantoja, Mendes, & Oller Nascimento, 2003). Una de las ventajas más prometedoras de la caracterización del crudo por espectrometría de masas combinado con métodos estadísticos quimiométricos es el potencial para rápidamente analizar y distinguir los componentes y el origen de las muestras con un mínimo o ningún tratamiento previo de la muestra. La técnica también ayuda a reducir modificaciones inesperadas a la composición de las muestras que pueden ocurrir durante su pretratamiento.

Por otro lado, el efecto Raman ocurre cuando hay un cambio en la polarización molecular, y proporciona información complementaria a la espectroscopía de infrarrojo, de hecho, estas técnicas son similares en términos de los resultados y se pueden utilizar en combinación para obtener información sobre el análisis de muestras de petróleo (Silva, Silva, Ribeiro, Martins, Da Silva, & Silva, 2011). Aunque no se considera la mejor opción para el análisis de crudos debido

a que la señal obtenida en el espectro Raman es débil y presenta alta fluorescencia, posee algunas ventajas como mínima preparación de la muestra y rapidez en la obtención del espectro. La espectroscopia Raman se utiliza en el análisis de compuestos aromáticos en mezclas de hidrocarburos, en elementos ligeros como la gasolina y diésel. También la espectroscopia Raman puede combinarse con la transformada de Fourier (FT- Raman), en la determinación del número de octano (RON) (Silva, Silva, Ribeiro, Martins, Da Silva, & Silva, 2011). A pesar de la complejidad de una muestra de petróleo, los espectros Raman de estas muestras pueden proporcionar alguna información química con solamente una inspección superficial (Britain cooper, 1999). Aunque la composición química de un combustible de petróleo está contenida en un espectro Raman, la no linealidad de la escala de octano, así como la compleja dependencia del índice de octano con los cientos de posibles especies existentes en la muestra sugiere que una correlación simple no es posible. El análisis de componentes principales (PCA), para la reducción de los datos espectrales y posteriormente la utilización de algunos algoritmos como la regresión por mínimos cuadrados parciales (PLSR) son útiles para construir un modelo de clasificación que permita correlacionar los espectros Raman de los combustibles de petróleo con los números de octano (Britain cooper, 1999) (Li & Dao, 2012).

Algunos trabajos reportados han usado (FT-IR/ATR) en combinación con calibración multivariable basada en regresión de soporte vectorial (SVR) y PLSR para predecir propiedades de petróleo crudo como gravedad API (*American petroleum institute*), viscosidad y contenido de agua (Filgueiras & M.S., 2014). Por otra parte, en la predicción de análisis SARA para petróleos colombianos se han desarrollado trabajos usando las técnicas espectroscópicas ATR-FTIR y NIR en combinación con métodos quimiométricos. (Lamus, Guzmán, Murcia, Cabanzo, & Mejía-Ospino, 2011) y (Melendez, Lache, Orrego, Pachon, & Mejia Ospino, 2012).

2. Resultados Experimentales

El presente trabajo se desarrolló en el laboratorio de espectroscopía atómico molecular (LEAM) de la Universidad Industrial de Santander, en donde se han venido realizando trabajos de investigación enfocados a la caracterización y análisis de crudos, sus fracciones y productos de su proceso de refinación.

2.1 Selección de muestras

En este trabajo se estudiaron las muestras de 67 crudos con el fin de determinar su análisis SARA y contenido de carbón Conradson, información muy importante para el proceso de refinación y preparación del crudo para estudios posteriores. Las muestras fueron suministradas por el Instituto Colombiano de Petróleo (ICP) y tienen su origen en diferentes regiones de Colombia.

2.1.1 Caracterización de las muestras. Todas las muestras contaron con su respectivo análisis SARA (Saturados, aromáticos, resinas y asfaltenos) y contenido de carbón Conradson, información obtenida a través de la metodología estándar ASTM D-4124 para el análisis SARA y ASTM 4530 para el análisis del contenido de carbón Conradson. Esta información fue suministrada por el laboratorio de espectroscopia (LE) del Instituto Colombiano del petróleo ICP en el marco del convenio UIS-ICP 8806.

Tabla 1. Análisis SARA y contenido de carbón Conradson de las muestras de crudos

Propiedad	Valor Max.	Valor Mín.	Valor Promedio
	(wt %)	(wt %)	(wt %)
Saturados	41.86	12.44	27.37
Aromáticos	47.15	8.31	31.19
Resinas	37.06	1.96	13.48
Asfáltenos	17.37	0.38	6.24
Carbón	20.30	0.74	8.84

En la tabla 1 se muestran los rangos de referencia del contenido de saturados, aromáticos, resinas, asfáltenos y carbón Conradson obtenidos por las metodologías estándar, para los 67 crudos. En el anexo A se encuentra la caracterización detallada de cada una de las muestras utilizadas en este estudio. La etiqueta corresponde al consecutivo del LEAM.

2.1.2 Preparación de las muestras. Las 67 muestras que hacen parte de este estudio se prepararon de dos maneras diferentes, con el objetivo de explorar la calidad de los espectros y así comparar los resultados obtenidos.

2.1.2.1 Método directo. Las muestras de crudo sin disolver fueron utilizadas en las pruebas preliminares de adquisición de espectros Raman. Cada muestra de aproximadamente de 1 gramo fue depositada con la ayuda de una espátula en un porta muestras y posteriormente colocada en el plato del microscopio Raman.

2.1.2.2 Método Indirecto. La muestra de crudo de aproximadamente 1 mg se disolvió en 1 ml de diclorometano [1:1]. Estas muestras fueron utilizadas para la obtención de los espectros Raman y los espectros de masa (fig. 9)

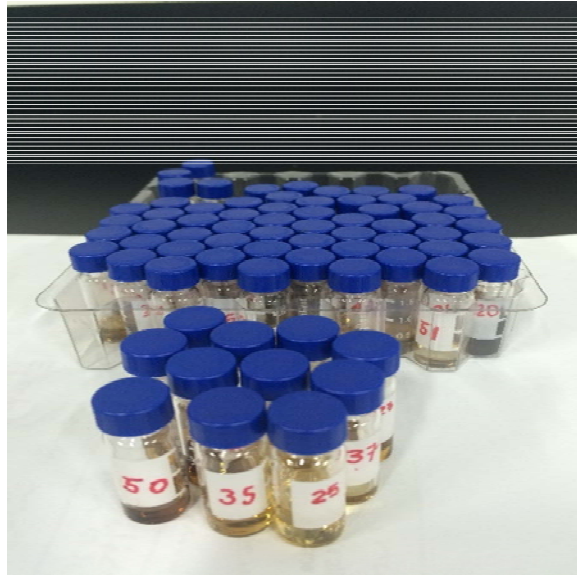


Figura 9. Muestras diluidas en diclorometano

2.2 Instrumentación

2.2.1 Espectroscopía Raman. La adquisición de los espectros Raman se realizó en un microscopio Raman LabRam HR Evolution. HORIBA Scientific (JOBIN YVON), con tres láseres (473, 532 y 785 nm) como fuentes de excitación. Este sistema permite la observación de los espectros en la región comprendida entre 50 cm^{-1} y 4000 cm^{-1} (fig. 10). Este instrumento se encuentra ubicado en el Edificio de Investigaciones del Parque Tecnológico de Guatiguará, Universidad Industrial de Santander. Para determinar las condiciones apropiadas de toma de espectros, se realizaron medidas preliminares modificando los siguientes parámetros: longitud de onda de excitación, Intensidad del láser, Apertura, Enfoque, tiempo de integración, entre otros.



Figura 10. Microscopio Raman LabRam HR Evolution. HORIBA.

2.2.1.1 Prueba de la reproducibilidad en la obtención de los espectros Raman. El desempeño del microscopio Raman se verificó a partir del análisis de ocho espectros adquiridos bajo los mismos parámetros experimentales del crudo 45 (fig. 11). Esta prueba consistió en analizar la repetibilidad en la intensidad para un número de onda

determinado (tabla 2) y el desplazamiento en el número de onda para el valor máximo de intensidad de cada espectro (tabla 3).

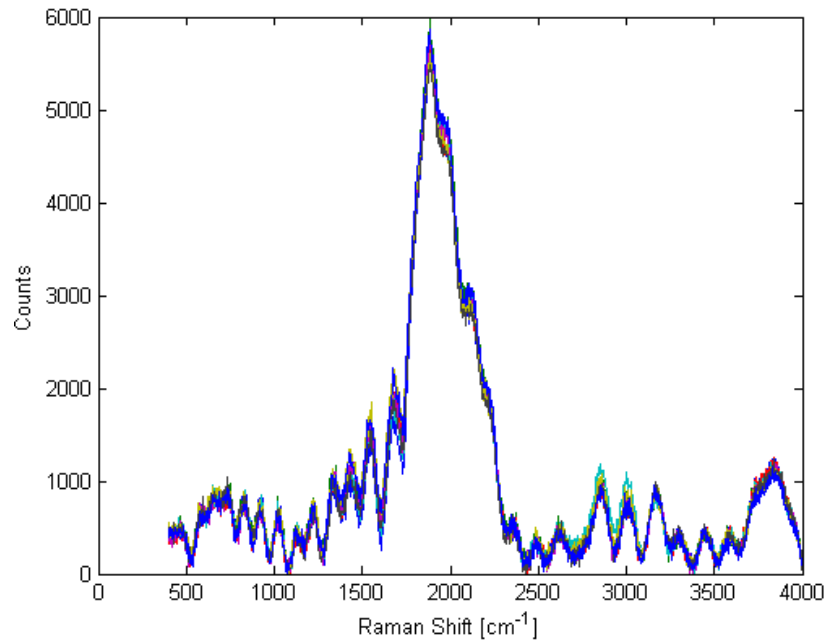


Figura 11. Espectros Raman crudo 45

La repetibilidad se evaluó a partir del cálculo de la desviación estándar de las intensidades de los espectros para 1884.2cm^{-1} , en donde el espectro exhibió la máxima intensidad. Se encontró que la desviación estándar fue de 159.28 y el porcentaje de desviación obtenido para la intensidad fue de 2.89%, que representa un porcentaje muy bajo considerando el ruido que acompaña a la señal espectral. Por otra parte, la desviación estándar del número de onda para la intensidad máxima de los espectros fue de 1.91 y el porcentaje de desviación 0.10%, lo que indica que la repetibilidad en la adquisición de los espectros es alta, así como el desempeño del microscopio Raman.

Tabla 2. Intensidad de espectros Raman crudo 45 a 1884.2 cm⁻¹

Espectro	Counts a 1884.2cm⁻¹
1	5427.3
2	5737.2
3	5462.5
4	5566.2
5	5624.3
6	5372.4
7	5279.4
8	5680.6
Media	5518.7
Desviación estándar	159.28
% desviación estándar	2.89

Tabla 3. Máxima intensidad de los espectros Raman del crudo Santos 65 deshidratado a su respectivo número de onda

Espectro	Máxima Intensidad	Raman shift(cm⁻¹)
1	5427.3	1884.2
2	5737.2	1884.2
3	5462.5	1881.2
4	5566.2	1881.2
5	5624.3	1887.2
6	5372.4	1884.2
7	5279.4	1884.2
8	5680.6	1884.2
Media		1883.8
Desviación estándar		1.91
% desviación estándar		0.10

2.2.2 Espectrometría de masas. Los espectros de masas MALDI/LDI fueron obtenidos con un espectrómetro de masas MALDI TOF/TOF UltrafleXtreme, *Bruker*TM, de alta resolución (40000) y que permite observar masas desde 100 hasta 200 kDa, en dos modos de trabajo; modo lineal para grandes masas (mayores a 10000 Da) y modo reflectron para masa menores (menores a 10000 Da), este último con mayor resolución. Este instrumento se encuentra ubicado en el Edificio de Investigaciones del Parque Tecnológico de Guatiguará, de la Universidad Industrial de Santander (fig. 12).



Figura 12. Espectrómetro de masas MALDI.TOF

2.3 Parámetros de adquisición de los espectros

2.3.1 Espectroscopía Raman. Las condiciones utilizadas para la adquisición de los espectros fueron determinadas con una exploración previa de espectros adquiridos por el

método directo e indirecto. En esta prueba preliminar se detectaron alteraciones en las muestras una vez se obtenía el espectro por el método directo (quemaduras), como resultado de la alta potencia de salida de la fuente de excitación. Debido a esto se ajustaron los diferentes parámetros de adquisición de los espectros, como, potencia de la salida del diodo láser, tiempo de integración y número de acumulaciones.

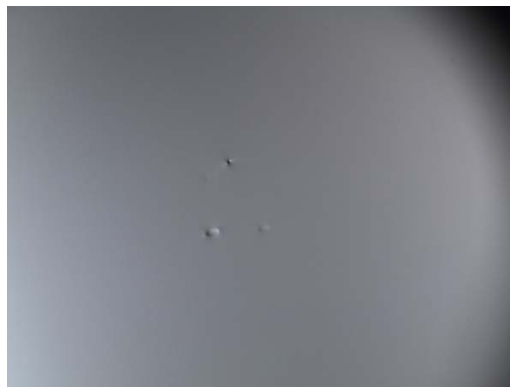


Figura 13. Imagen de la muestra posterior a la adquisición del espectro preliminar.

Todos los espectros se adquirieron usando la opción *flat correction*, la cual eliminó el efecto de la fluorescencia presentada en los espectros de los crudos, aplicando sobre el espectro un procesamiento que corresponde a una sustracción de línea base. Por otra parte, para la adquisición de los espectros se ajustó el factor *binning* de acuerdo al tipo de muestra, el cual mejoró la apariencia del espectro (relación señal/ruido). Cuando el «*Binning Factor*» se establece en 1, se utiliza el número total de puntos de datos para crear el espectro. El número exacto depende de la configuración específica del sistema Raman. Su valor predeterminado es 1. Si se ajusta un valor mayor que 1 el número de datos utilizados para crear un espectro es reducida, por la adición de co-píxeles

adyacentes en el detector. El aumento de “Binning Factor” genera espectros que tienen menos puntos de datos, los archivos de los espectros serán de menor tamaño y la calidad espectral (por ejemplo, relación señal a ruido) se incrementará.

La siguiente tabla muestra las diferentes condiciones utilizadas para la adquisición de los espectros Raman.

Tabla 4. Parámetros de adquisición de los espectros Raman.

Parámetros		Método directo (espectros)	Método directo (condiciones)	Método indirecto
Rango	de	400 – 4000 cm ⁻¹	400 – 4000 cm ⁻¹	400 – 4000 cm ⁻¹
Tiempo	de	1s	3s	1s
Acumulaciones		100	100	10
Láser		532nm	532nm	532nm
Objetivo		10 X	10 X	Celda de cuarzo de
Factor Binning		1	3	2
Potencia del láser		0.1% - 10%	0.1%	10%

2.3.2 Espectrometría de masas. Para determinar las condiciones apropiadas de toma de espectros, se realizaron medidas preliminares modificando los siguientes parámetros: Rango de masas, Intensidad del Láser, Modo del analizador (lineal, reflectron).

El rango de número de masas (m/z) utilizado para la adquisición de los espectros estuvo entre 250Da y 1000Da. El modo reflectron fue el seleccionado para la operación del analizador, pues de acuerdo al rango (m/z) utilizado el espectrómetro adquiere los espectros con mejor resolución [15]. La intensidad del láser fue del 60% de su energía

total ya que a mayor intensidad se observa una distribución adicional que podría deberse a procesos de ablación láser.

2.4 Adquisición de los espectros

2.4.1 Espectroscopía Raman. La adquisición de los espectros Raman se realizó mediante el software LabSpec6 del microscopio Raman LabRam HR Evolution HORIBA Scientific (JOBIN YVON). Para la adquisición de los espectros por el método indirecto se utilizó una celda de cuarzo de paso 10 mm en donde se depositaron 2 ml de la solución crudo-diclorometano. Una vez se adquirió el espectro, la celda de cuarzo fue lavada con diclorometano; posteriormente se adquirió un el espectro de diclorometano para verificar que la celda no contenía restos de muestra. Este procedimiento se repitió para la adquisición de todos los espectros.

De la misma manera se obtuvieron los espectros Raman por el método directo, es decir colocando una muestra de crudo sin diluir de aproximadamente 1 g sobre el plato portamuestras del microscopio Raman.

En una primera inspección de los espectros y de acuerdo al resultado obtenido con los modelos de predicción, se encontró que el mejor desempeño se obtuvo para los modelos desarrollados a partir de los espectros tomados por el método indirecto; debido a esto el desarrollo de los modelos de predicción del análisis SARA y el porcentaje del contenido de carbón Conradson se construyeron a partir de los espectros obtenidos por el método indirecto. En la figura 14 se observan los espectros obtenidos por el método indirecto.

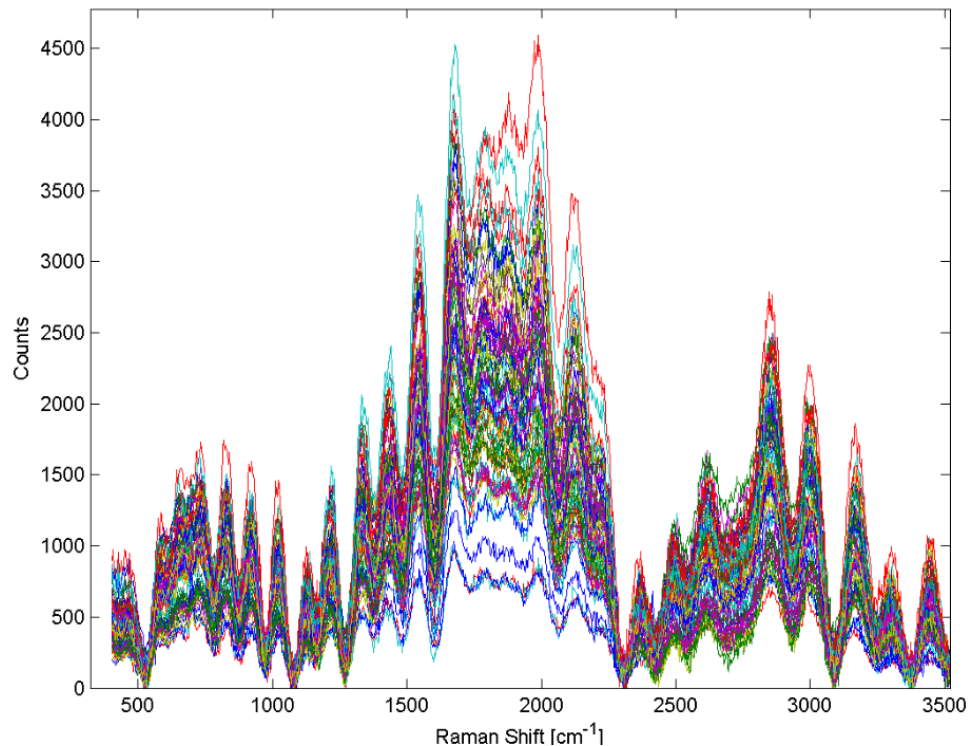


Figura 14. Espectros obtenidos para las 67 muestras de crudo por método indirecto

2.4.2 Espectrometría de masas. Un volumen de 1 μl de muestra preparada por método indirecto, fue depositada sobre el spot de una placa de acero inoxidable (target). Una vez el solvente fue evaporado, las muestras se introdujeron al espectrómetro para la adquisición de los espectros.

La adquisición de los espectros se realizó mediante el software de análisis FlexControl, el cual permite el ajuste de parámetros como: intensidad del láser, método de adquisición de los espectros, rango de masas, voltaje de extracción de las placas, etc.

La figura 15, muestra los espectros de masas adquiridos con los parámetros establecidos. Si se enfoca una pequeña parte del espectro, aproximadamente entre 313.5m/z y 314m/z se puede observar el corrimiento entre todos los espectros alrededor de un mismo m/z.

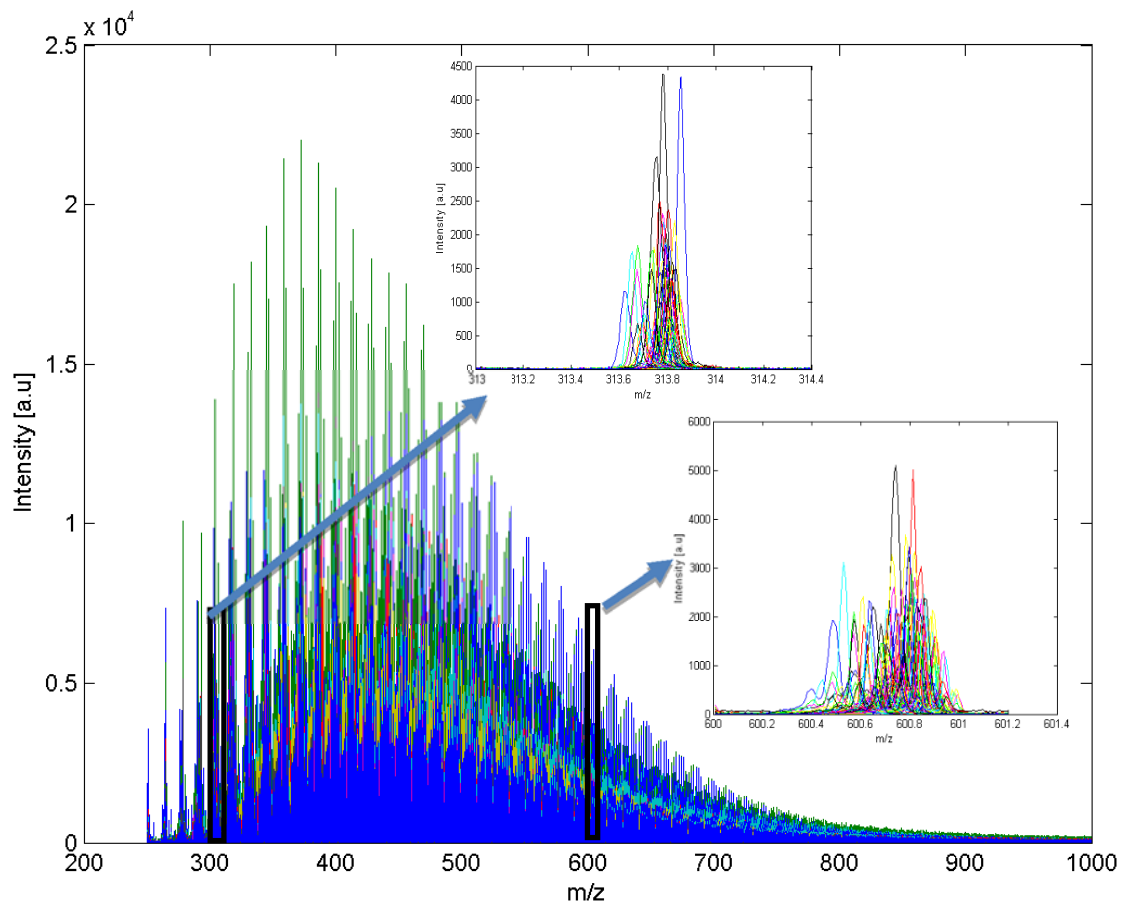


Figura 15. Espectros Raman obtenidos para las 62 muestras de crudo

2.5 PRE-PROCESAMIENTO DE DATOS

El pretratamiento de la señal es uno de los primeros pasos que se lleva a cabo en el análisis de datos multivariantes. Consiste en manipulaciones matemáticas que se aplican antes de cualquier otro tipo de análisis y tratan de anular o, al menos reducir, fuentes de variabilidad en la señal, ya

sea de carácter aleatorio (como el ruido) o de carácter sistemático (variaciones en la línea base, etc.), que no están relacionadas con el analito o propiedad de interés. Entre los procedimientos más utilizados se encuentran, el suavizado que consiste en reducir matemáticamente el ruido aleatorio que acompaña a la señal analítica, el proceso de normalización, que busca generar una nueva escala estándar en donde todos los valores de las variables de los espectros de las diferentes muestras conserven sus proporciones originales y la corrección de línea base que selecciona determinados puntos dentro del espectro, luego los ajusta a un polinomio y la función resultante es restada del espectro.

2.5.1 Espectroscopía Raman. Como se mencionó anteriormente, en todos los espectros Raman se eliminó la fluorescencia de fondo utilizando la opción “flat correction”, algoritmo que garantizó una línea base plana, aunque en este proceso se generaron algunos valores negativos en los espectros. Este resultado requirió realizar una corrección de línea base adicional que consistió en sustraer de cada espectro el valor más negativo de intensidad, posteriormente los espectros se normalizaron, dividiendo cada espectro en su respectiva área bajo la curva espectral. Todo el procesamiento mencionado anteriormente fue realizado desarrollando un código en *Matlab R2012 (The MathWorks Inc.)*. En la figura 16, se muestran los espectros preprocesados.

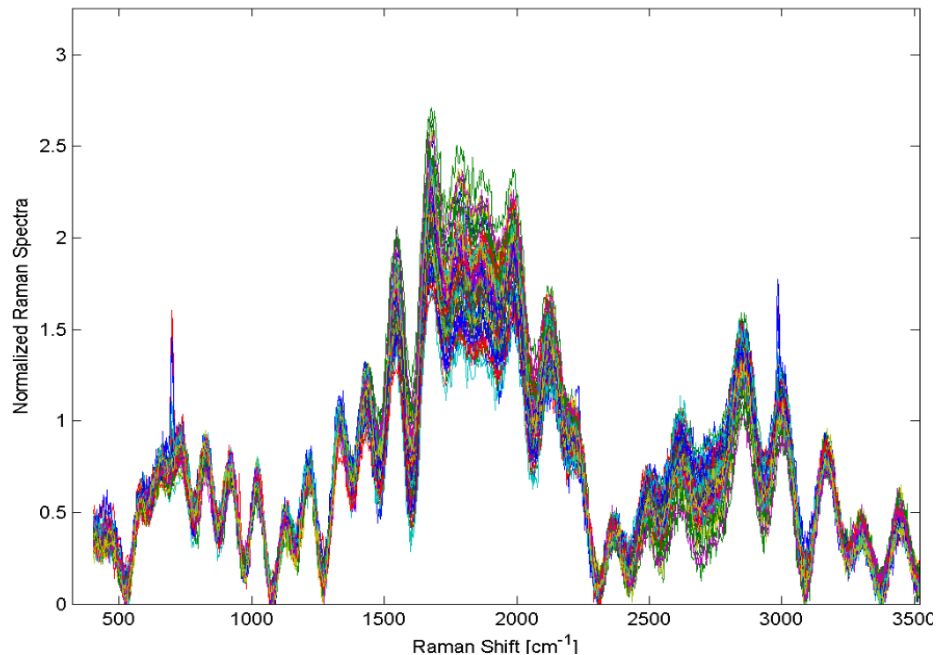


Figura 16. Espectros Raman para las muestras de crudo obtenidos por método indirecto después del preprocesamiento

2.5.2 Espectrometría de masas. Una vez obtenidos los espectros de masas se les aplicó un proceso de normalización en donde cada espectro fue dividido en su respectiva área bajo la curva espectral. Debido a errores instrumentales se genera un desplazamiento entre los espectros para un mismo valor de (m/z) (fig. 15). Este desplazamiento es originado debido a las pequeñas diferencias que pueden ocurrir en las diferencias de potencial generando una energía cinética diferente para sistemas moleculares iguales, originando un desplazamiento a valores (m/z) iguales para diferentes espectros.

La exigencia de una alta colinealidad entre las variables del espectro (m/z) para el desarrollo de los modelos de calibración PLS requirió la corrección del desplazamiento entre los espectros. Este proceso se realizó desarrollando un código en Matlab y consistió en realizar la correlación cruzada entre un espectro de referencia y cada uno de los demás espectros. El origen respecto al cual se midieron los corrimientos fue la posición del

máximo de la auto correlación del primer espectro. Posteriormente a través de la correlación cruzada se realizó el ajuste del corrimiento sobre cada espectro, logrando la superposición de todos los espectros de masas. El resultado de la corrección de los desplazamientos de los espectros se observa en la figura 17.

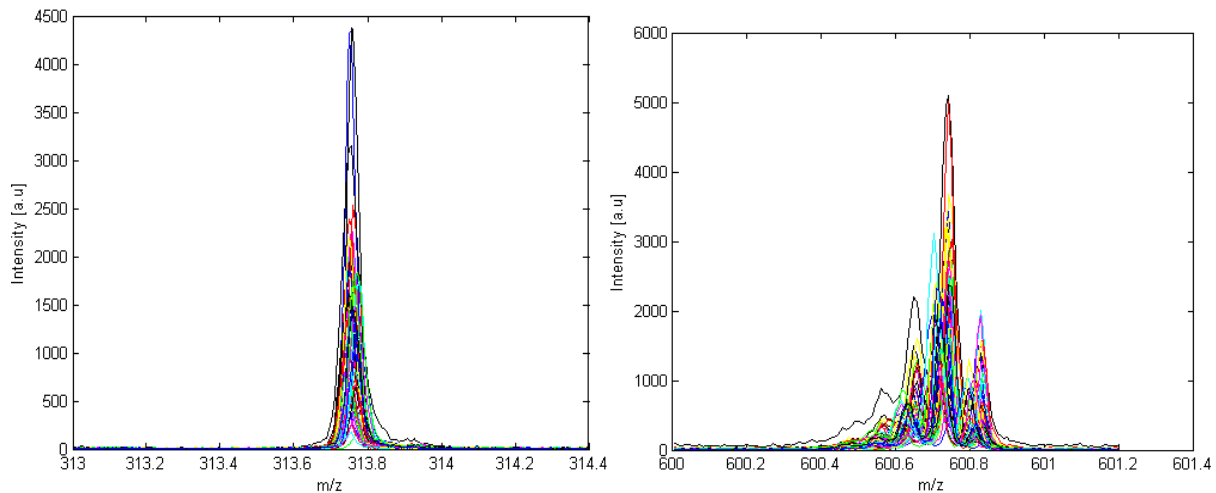


Figura 17. Espectros de masas pre-procesados (corrección del corrimiento)

2.6 Análisis Quimiométrico de datos.

2.6.1 Análisis de residuales. Una vez fue realizado el pre-procesamiento a los espectros Raman y de masas se llevó a cabo el análisis quimiométrico de estos espectros que consistió en realizar la calibración multivariable por regresión de mínimos cuadrados parciales (PLS) con el fin de desarrollar los modelos de predicción de contenidos de saturados, aromáticos, resinas, asfaltenos y carbón Conradson.

Los residuales miden que tan bien las muestras se ajustan al modelo determinado por las componentes. Hotelling, mide la distancia desde la muestra proyectada (es decir su

modelo de aproximación) al centro (punto medio). Las muestras con un alto residual no están bien descritas por el modelo, sin embargo, no son extremas en el modelo, es decir no influyen considerablemente en el modelo. Las muestras con bajo residual y bajo Hotelling están muy bien descritas por el modelo, mientras que las muestras con alto Hotelling tienen mayor influencia en el modelo que otras muestras, pueden ser o no, valores atípicos o Outlier. Una muestra es atípica con alto Hotelling y alto residual, es el peor de los casos y puede ser detectado usando un gráfico de influencia. De acuerdo a las gráficas de influencia de las muestras (fig. 18 y fig. 19), construidas a partir de los espectros Raman y espectros de masas se observa que no hay muestras que exhiban alto Hotelling y alto residual, las cuales podrían ser muestras atípicas y deberían ser retiradas de la población de estudio. En la figura 18 se observa la muestra 65 que presenta un considerable valor de Hotelling. Una vez se desarrolló el modelo a partir de los espectros Raman, se exploró la influencia de esta muestra en el modelo encontrándose que no representaba ser una muestra atípica y no era necesario retirarla de la población, pues los parámetros estadísticos que muestran el desempeño del modelo no mostraron cambios importantes cuando esta muestra fue retirada de la población. Por otra parte para el desarrollo de los modelos a partir de los espectros de masas la población inicial de muestras no tuvo ninguna modificación.

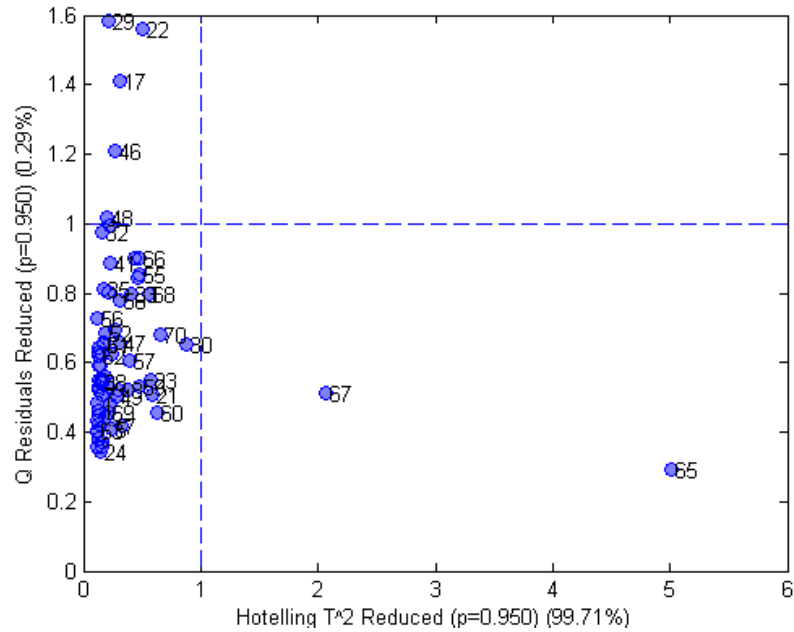


Figura 18. Gráfica de influencia muestras en el modelo. Espectroscopía Raman.

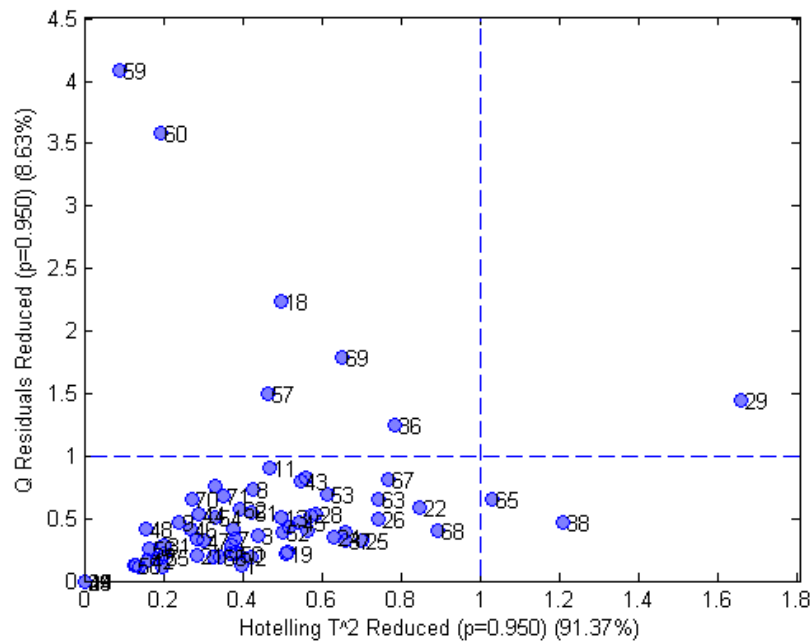


Figura 19. Gráfica de influencia de muestras en el modelo. Espectrometría de masas

2.6.2 Desarrollo del modelo PLS. En ésta etapa se desarrollaron los modelos PLS para las fracciones de saturados, aromáticos, resinas, asfaltenos y contenido de carbón Conradson, a partir de los espectros obtenidos por espectroscopía Raman y espectrometría de masas.

Utilizando el algoritmo matemático de Regresión de Mínimos Cuadrados Parciales (PLS) y el programa MatLab se evaluó la capacidad de la espectroscopía Raman y la espectrometría de masas para predecir el análisis SARA y el contenido de carbón Conradson de los crudos colombianos. PLSR es una herramienta matemática ampliamente utilizada en análisis quimiométrico de datos espectrales con alta colinealidad. El algoritmo implementado en Matlab para tal fin fue el NIPALS (Non-linear Iterative Partial Least Squares).

La técnica PLS usa la información contenida en los datos espectroscópicos y la información de la propiedad que se quiere predecir y obtiene unas variables auxiliares llamadas variables latentes o componentes que reducen la dimensionalidad de los datos. Estas nuevas variables contienen los datos en X (datos espectroscópicos) que mejor predicen la propiedad (Otto, 2007).

El desarrollo del modelo se realizó en dos etapas, calibración y validación. Para esto el grupo total de muestras se separan en dos grupos; el grupo de calibración que está compuesto por las muestras usadas para construir el modelo, sus variables X y sus propiedades de referencia Y, y el grupo de validación que contiene a las muestras restantes para las cuales los valores de las variables X se introducen al modelo y se obtienen los valores predichos Y.

La técnica de validación cruzada de K iteraciones o K-fold (K= 10) fue usada sobre la población de calibración para determinar el número de variables latentes y consiste en dividir aleatoriamente el grupo de muestras en k subconjuntos. Los K-1 subconjuntos se utilizan como datos de entrenamiento o calibración y el subconjunto restante se utiliza como datos de validación. El proceso de validación cruzada es repetido durante k iteraciones, con cada uno de los posibles subconjuntos de datos de prueba. Para encontrar los mejores grupos de entrenamiento y validación, se ejecuta el K-fold varias veces (multi cross-validation) hasta que se obtiene el mejor modelo con un número óptimo de variables latentes, de acuerdo con el mejor valor de R^2 y el RMSECV (error cuadrático medio de validación cruzada) más bajo, calculado por la ecuación 28,

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{N_v} (y_i - \hat{y}_i)^2}{N_v}} \quad (23)$$

En la ecuación (23), N_v es el número de muestras usadas en la validación, y_i es la propiedad medida y \hat{y}_i es el valor de la propiedad predicha por el modelo.

El valor de la propiedad predicha por el modelo \hat{y}_i , para un espectro X de una muestra determinada es dada por:

$$\hat{y}_i = bX + e \quad (24)$$

En la ecuación 24, X es el espectro de la muestra a la cual se le quiere predecir la propiedad, e es el residual y b es vector de regresión que considera la contribución de cada variable al modelo PLS, esto significa que entre más alto sea el valor de b más importe es la variable para el modelo de calibración.

Una vez obtenido el modelo de calibración se evalúa el desempeño del modelo para predecir la propiedad con el grupo de validación. Los resultados obtenidos se comparan con los observados produciendo un residuo de predicción que se puede utilizar para calcular un RMSEP (Root mean Square Error of prediction), que da información del desempeño del modelo en la predicción de la propiedad para muestras que no fueron utilizadas en la construcción del modelo, calculado por la ecuación 30.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{N_p} (y_i - \hat{y}_i)^2}{N_p}} \quad (25)$$

En donde N_p es el número de muestras de predicción, y_i es la propiedad medida y \hat{y}_i es el valor de la propiedad predicha por el modelo.

El modelo de calibración se evaluó de acuerdo a los parámetros estadísticos: coeficiente de determinación R^2 , raíz del error cuadrático medio de validación cruzada (RMSECV) y el error cuadrático medio de predicción (RMSEP).

2.6.1.1 calibración y validación de modelos PLS a partir de espectroscopía raman. A continuación, se describe el procedimiento empleado para el desarrollo y evaluación de los modelos de predicción de contenido de saturados, aromáticos, resinas, asfáltenos y carbón conradson, a partir de los espectros raman de las muestras preparadas por el método indirecto utilizando el algoritmo nipals programado en matlab

2.6.1.1.1 Calibración y validación del modelo PLS para la fracción de saturados. Una vez obtenido el mejor modelo PLS de calibración para la predicción de la fracción de saturados que presentaba el menor valor de RMSECV y el mayor R^2 , se encontró que los parámetros estadísticos RMSECV, RMSEP y R^2 variaban con el número de componentes (variables latentes).

Se observa como disminuye el comportamiento de RMSECV a medida que aumenta el número de componentes, obteniéndose el valor más bajo, 1,75w%, para 8 componentes (fig. 20), Este parámetro estadístico da información de la medida de la dispersión de las muestras de validación alrededor de la línea de regresión cuando es usada la validación cruzada.

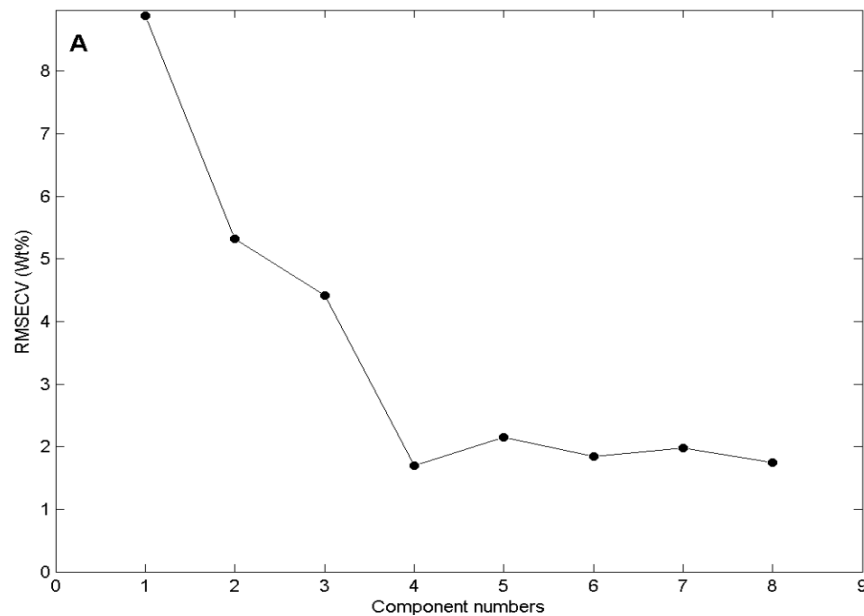


Figura 20. Raíz del error cuadrático medio de validación cruzada (RMSECV) en función del número de componentes (variables latentes) para el modelo de predicción de la fracción de saturados.

Sin embargo, la selección del número de componentes dependió además del valor del coeficiente de determinación R^2 , el cual evolucionó igualmente con el número de

componentes obteniéndose su mejor valor de 0,954 para 8 componentes, determinando de esta manera que 8 es el número de componentes óptimo para construir el modelo PLSR de predicción del contenido de saturados mostrando que la recta de regresión ajusta bien los datos de calibración, comportamiento que se confirma con el valor del RMSEC.

La figura 21 muestra la evolución de la varianza explicada acumulativa de validación con el número de componentes. Se encontró que con 8 componentes es posible explicar el 99.82% de la variabilidad de los datos de contenido de saturados para el grupo de validación, que es una estimación de la capacidad del modelo de predecir la propiedad para nuevas muestras.

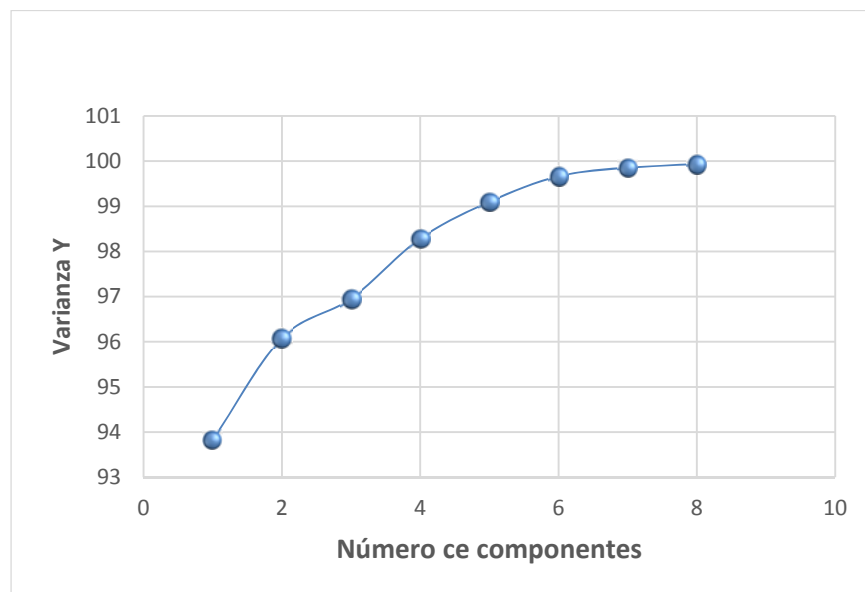


Figura 21. Varianza explicada acumulativa de contenido de saturados

Los coeficientes de regresión muestran cómo se pondera cada variable X al predecir una respuesta particular Y. Resume la relación entre los predictores X y una respuesta dada y pueden calcularse para cualquier número de componentes (Otto, 2007). Los predictores

con coeficientes cercanos a cero significan que no aportan a la predicción de la propiedad y podrían ser retirados de la matriz X. En la figura 22, se muestra claramente un aporte uniforme en todas las regiones del espectro, tanto negativa como positivamente a la respuesta Y (porcentaje en peso de saturados), de manera que no es conveniente retirar variables espectrales.

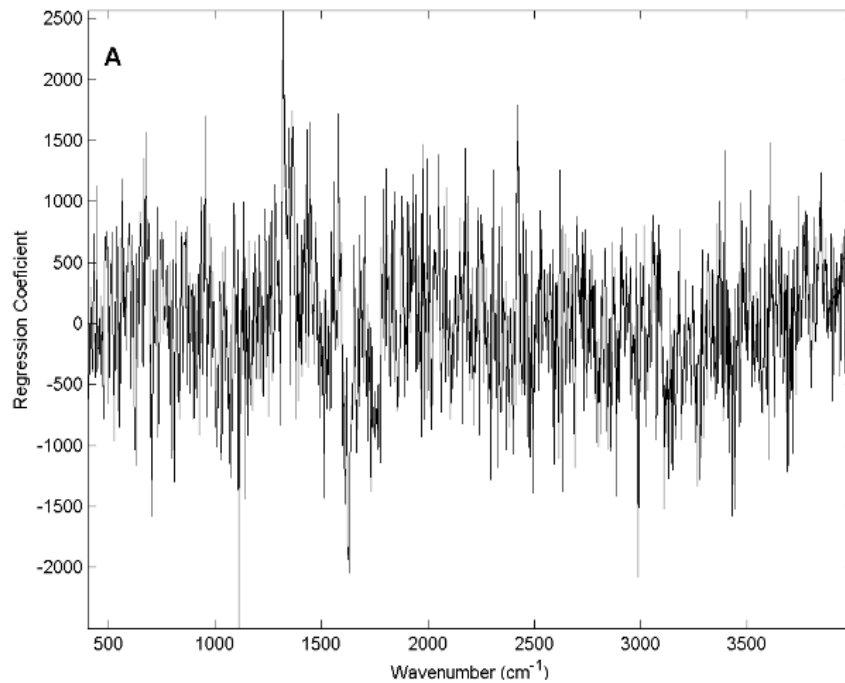


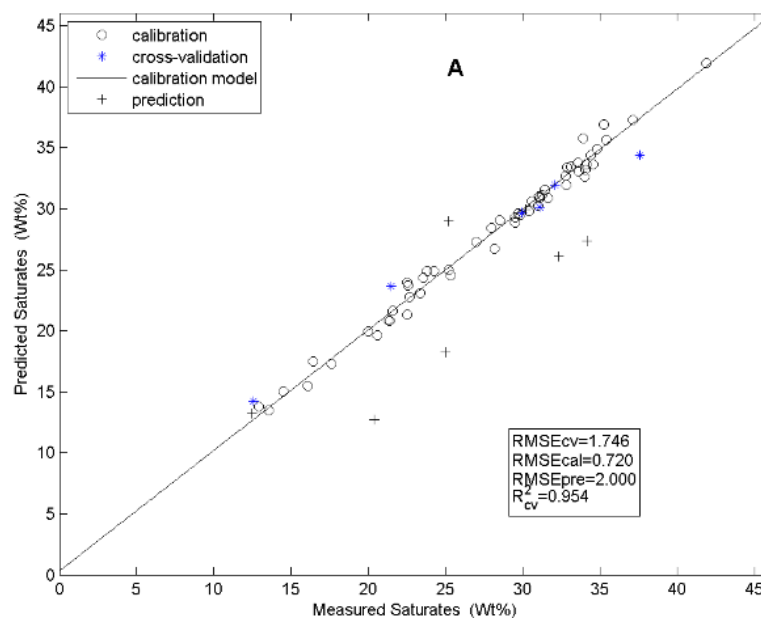
Figura 22. Coeficientes de regresión para el modelo de predicción PLSR de contenido de saturados

Los parámetros estadísticos obtenidos para el modelo PLSR de predicción de contenido de saturados se muestran en la siguiente tabla:

Tabla 5. Parámetros estadísticos del modelo de calibración para el contenido de saturados

propiedad	Número de component es	<i>RMSEC</i> <i>V</i> (wt %)	<i>RMSE</i> <i>C</i> (wt %)	<i>RMSEP</i> (wt %)	R_{cv}^2	Varianza explicada (%)
Saturados	8	1.75	0.72	2.00	0.95	99.81

La figura 23 muestra la curva de predicción del contenido de saturados contra valores medidos por la metodología estándar ASTM-D4124 para las muestras de calibración y de predicción de crudos colombianos.

**Figura 23.** Curva de valores predichos contra valores medidos del contenido de saturados

Un modelo está concebido para hacer predicciones futuras, es por esta razón que la validación de dicho modelo es muy importante. La validación de un modelo basado en

datos experimentales busca estimar la incertidumbre de predicciones futuras, es decir evalúa el desempeño del modelo con nuevos datos de la misma clase de los que se usaron para su construcción. Si la incertidumbre es razonablemente baja el modelo es válido (Cooper, 1999).

El desempeño del modelo PLSR para predecir el contenido de saturados se evaluó de dos formas. La primera consiste en una validación cruzada (K-fold), con múltiples iteraciones que se realizó sobre el grupo de muestras de calibración resultando un error RMSECV de 1.746wt%. Por otra parte, ya que el modelo de regresión está concebido para hacer predicciones futuras, se realizó una validación con muestras que no participaron en la construcción del modelo obteniéndose un error RMSEP de 2.00wt% (tabla 5).

En la figura 24 se observan los residuales (diferencia entre los valores de referencia y los valores predichos), que pueden ser negativos o positivos y se distribuyen con alta probabilidad alrededor de cero, permaneciendo aproximadamente constantes.

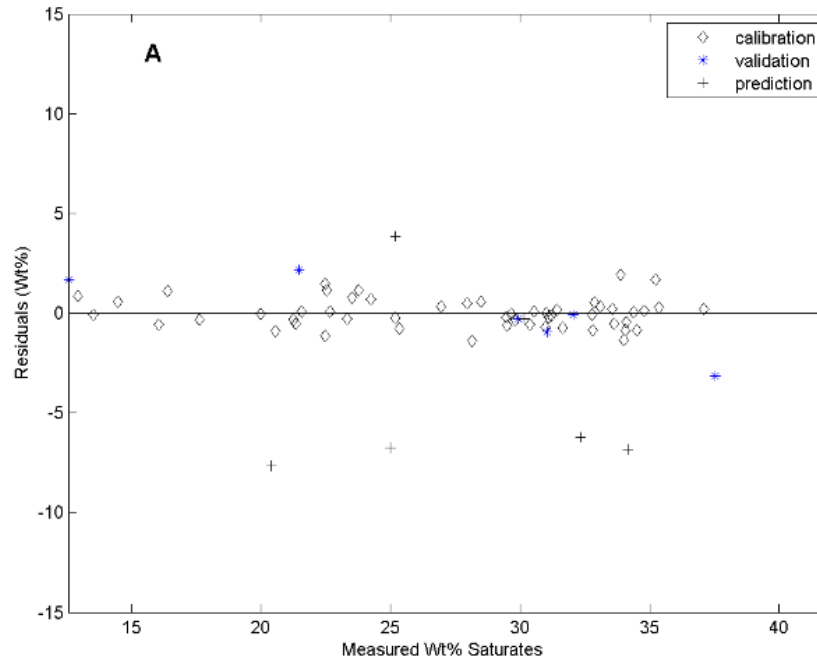


Figura 24. Gráfica de residuales para el porcentaje del contenido de Saturados. A continuación, se presentan los resultados más importantes obtenidos en el desarrollo de los modelos de predicción para las fracciones de aromáticos, resinas, asfaltenos y contenido de carbón Conradson, siguiendo la misma metodología expuesta anteriormente.

2.6.1.1.2 Calibración y validación del modelo PLS para la predicción del contenido de aromáticos. Siguiendo el mismo procedimiento para el desarrollo del modelo de predicción del contenido de saturados, se obtuvo el mejor modelo PLS de calibración para la predicción de la fracción de aromáticos que presentaba los mejores valores para RMSECV y R^2 , encontrándose que los parámetros estadísticos RMSECV, RMSEP y R^2 variaban con el número de componentes (variables latentes).

En la figura 26 se observa como disminuye el comportamiento de RMSECV a medida que aumenta el número de componentes. Para este caso en particular los valores óptimos de RMSECV y R^2 se obtuvieron para un valor de 8 componentes, permaneciendo constantes a partir de este número. De la misma manera la varianza explicada de

validación alcanza su valor máximo con 8 componentes, es decir se explica el 99.82% de la varianza de los datos del contenido de aromáticos para el grupo de validación (fig. 25).

Análogamente al modelo de predicción del contenido de saturados, la gráfica de los coeficientes de regresión (figura 26) muestra la contribución de cada variable X al modelo obtenido para predecir el contenido de aromáticos.

En la figura 26 se observa la curva de valores predichos del contenido de aromáticos contra valores medidos, con sus parámetros estadísticos RMSECV de validación cruzada, RMSEC de calibración, R^2 de validación cruzada y el error de predicción RMSEP.

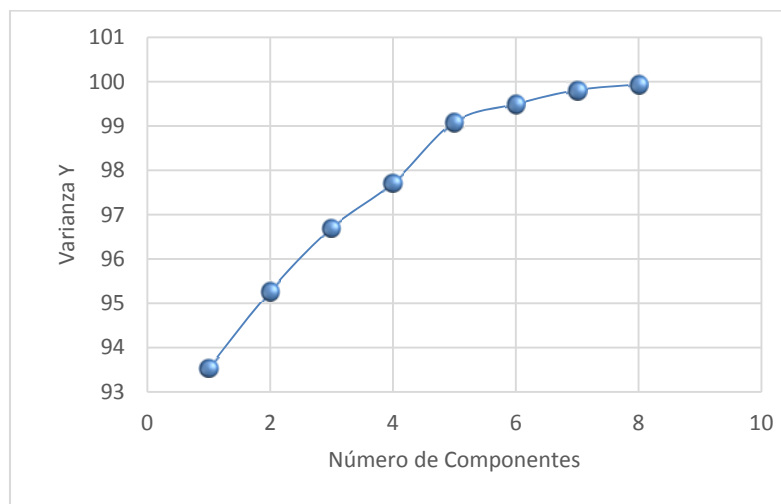


Figura 25. Varianza explicada acumulativa de contenido de aromáticos

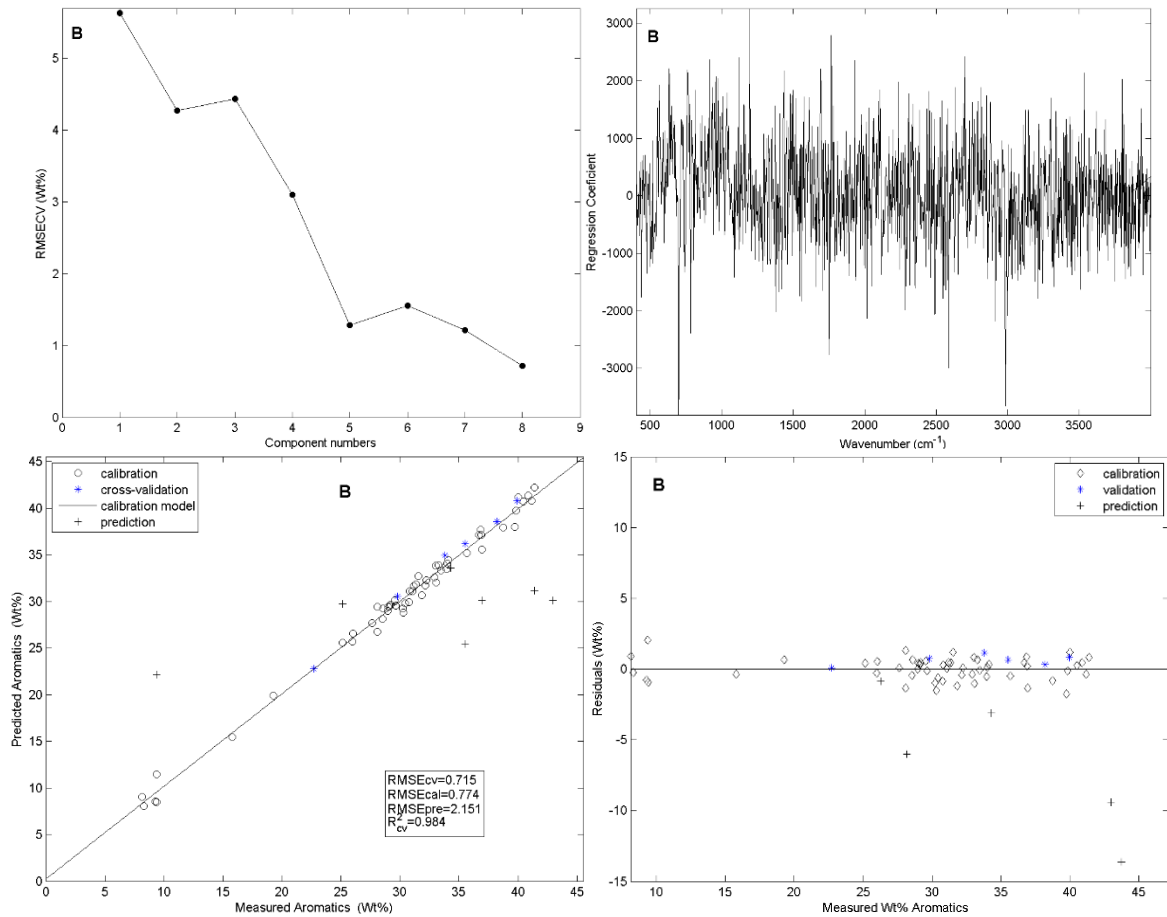


Figura 26. RMSECV, gráfica de coeficientes, modelo de calibración y gráfica de residuos para contenido de aromáticos

Los parámetros estadísticos que caracterizan el modelo obtenido se encuentran en la tabla 6.

Tabla 6. Parámetros estadísticos para el modelo de calibración del contenido de aromáticos

propiedad	Número de component es	$RMSEC$ V (wt %)	$RMSE$ C (wt %)	$RMSEP$ (wt %)	R_{CV}^2	Varianza explicada (%)
Aromáticos	8	0.72	0.77	2.15	0.98	99.82

El RMSECV obtenido para el grupo de validación es 0.72w%. El modelo exhibe un alto coeficiente de determinación R_{CV}^2 de 0.98 que indica la alta tendencia lineal de los datos,

es decir las muestras de calibración se ajustan bien al modelo, lo cual es confirmado con un bajo error RMSEC, de 0.77w%. Aunque hay una relación lineal entre el porcentaje del contenido de aromáticos predicho y el valor real se encontró que para las muestras usadas en la validación externa con los valores más elevados de contenido de aromáticos los residuales estuvieron por encima del 5wt%. El error de predicción RMSEP del contenido de aromáticos es de 2,15 wt %, el cual es consistente con la gráfica de residuales (fig. 26). Por otra parte, se observa que para las muestras de calibración los residuales presentan un comportamiento homocedástico alrededor del cero, en un rango entre -2 wt % y 2 wt %

2.6.1.1.3 Calibración y validación del modelo PLS para la predicción del contenido de resinas. El modelo desarrollado para predecir el contenido de resinas a partir del espectro Raman con 6 componentes explica el 99,78% de la varianza de los datos con respecto al contenido de resinas en las muestras de crudo (fig. 27). De la misma manera, se obtuvo un RMSECV y el R^2 con el número de componentes obteniéndose los mejores valores de estos parámetros para 6 componentes, permaneciendo constantes a partir de las siguientes componentes. Los coeficientes de regresión se exhiben en la figura 28.

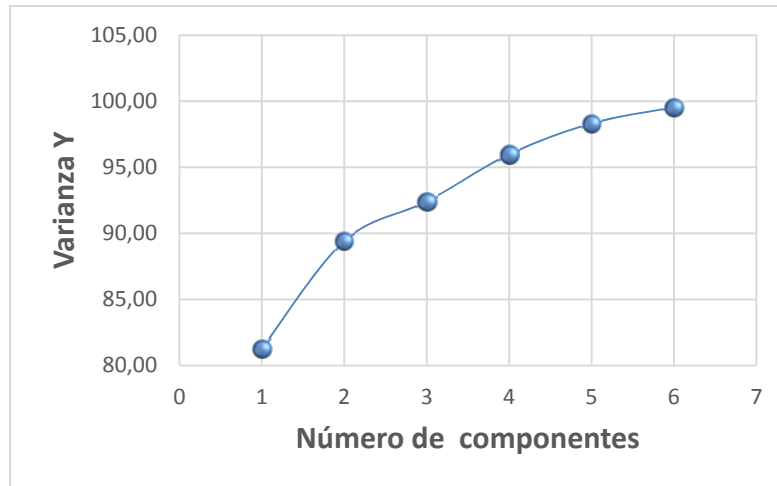


Figura 27. Varianza explicada acumulativa de contenido de resinas

En la figura 28 se muestra la curva de valores predichos contra valores medidos del contenido de resinas con los parámetros estadísticos que permiten evaluar el modelo, los cuales se presentan en la tabla 7.

Tabla 7. Parámetros estadísticos del modelo de calibración para el contenido de resinas

propiedad	Número de component es	<i>RMSEC</i> <i>V</i> (wt %)	<i>RMSE</i> <i>C</i> (wt %)	<i>RMSEP</i> (wt %)	R_{CV}^2	Varianza explicada (%)
Resinas	6	0.69	0.72	2.66	0.98	99.78

De acuerdo a la tabla 1, el contenido de resinas está en el rango entre 1.96 wt % y 37.06 wt %, para el número total de muestras y un valor promedio de 13.48 wt %. El valor de $RMSECV$ es de 0.69wt% y el coeficiente de determinación de validación cruzada R_{CV}^2 indica el buen ajuste de los datos al modelo de calibración, confirmado con un bajo error $RMSEC$, de 0.72w%. Aunque hay una relación lineal entre el porcentaje del contenido de resinas predicho y el valor real, se encontró que el error de predicción

RMSEP es 2.66 wt %. La figura 28, muestra que los residuales se encuentran distribuidos aleatoriamente alrededor del cero con valores en un rango entre -5wt% a 5wt% de residual para las muestras de validación externa, mientras que para las muestras de calibración este intervalo se encuentra aproximadamente entre -2.5wt% y 2.5wt%.

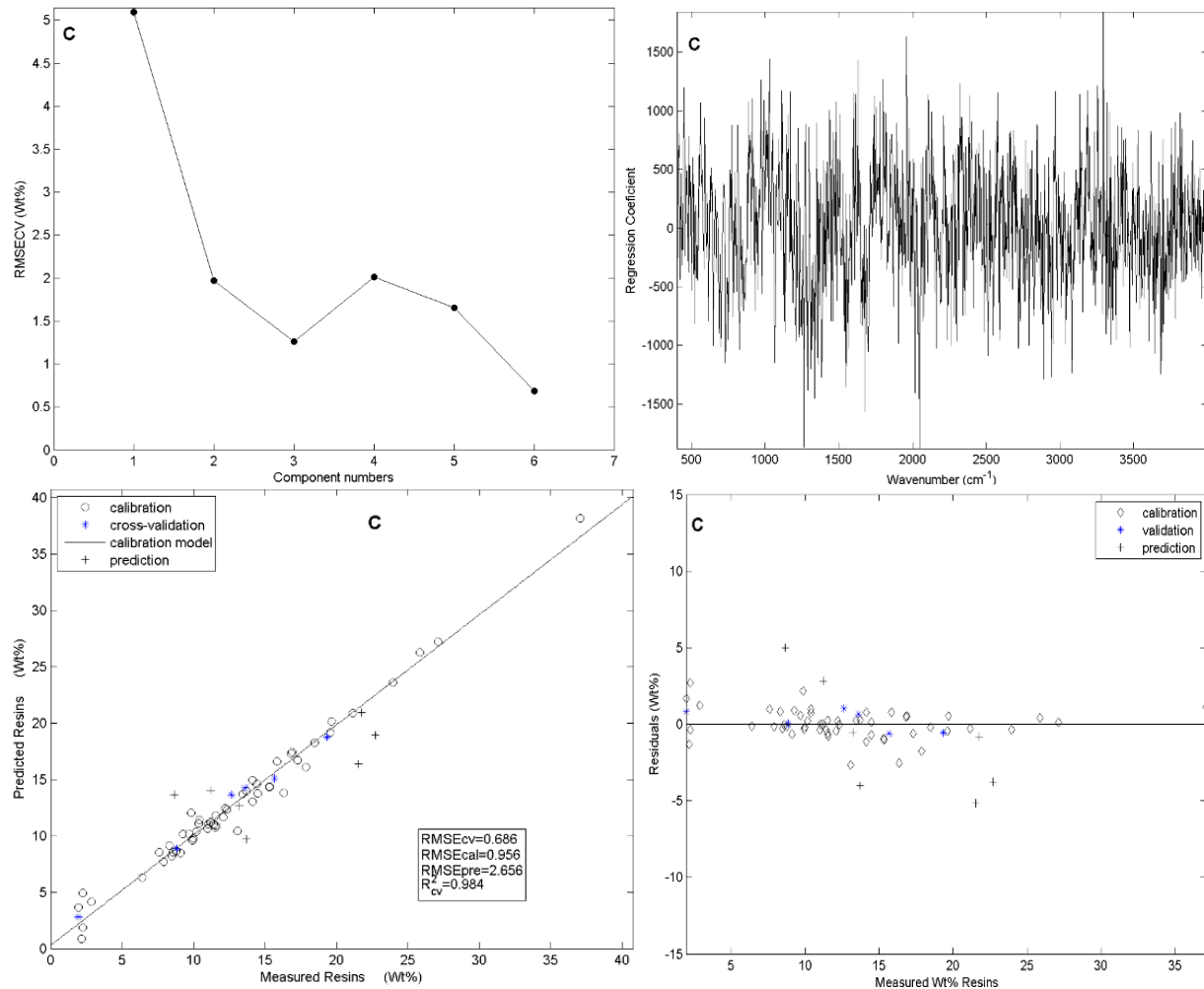


Figura 28. RMSECV, gráfica de coeficientes, modelo de calibración y gráfica de residuales para contenido de resinas

2.6.1.1.4 Calibración y validación del modelo PLS para la predicción del contenido de asfáltenos. El modelo desarrollado para predecir el contenido de asfáltenos mostró bajos errores RMSECV de 1.01wt%, RMSEC de 0.57wt% y RMSEP de 1.59wt% con 7 componentes. El valor de R^2 de validación cruzada para esta situación fue de 0.94 y la varianza explicada de validación para el contenido de asfáltenos fue 99.81% (Fig. 29). La gráfica de los coeficientes muestra una participación de todas las variables espectroscópicas con contribuciones positivas y negativas al modelo.

En la tabla 8 se presentan los parámetros estadísticos que caracterizan el modelo PLS de predicción de contenido de asfáltenos.

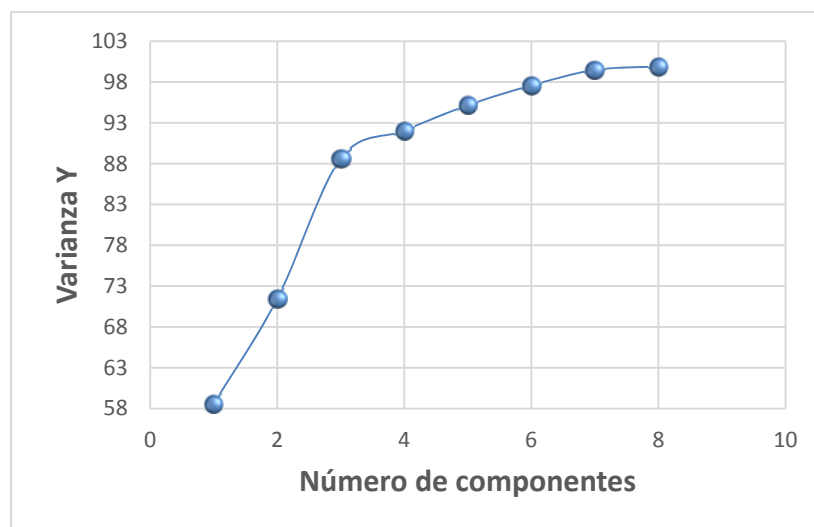
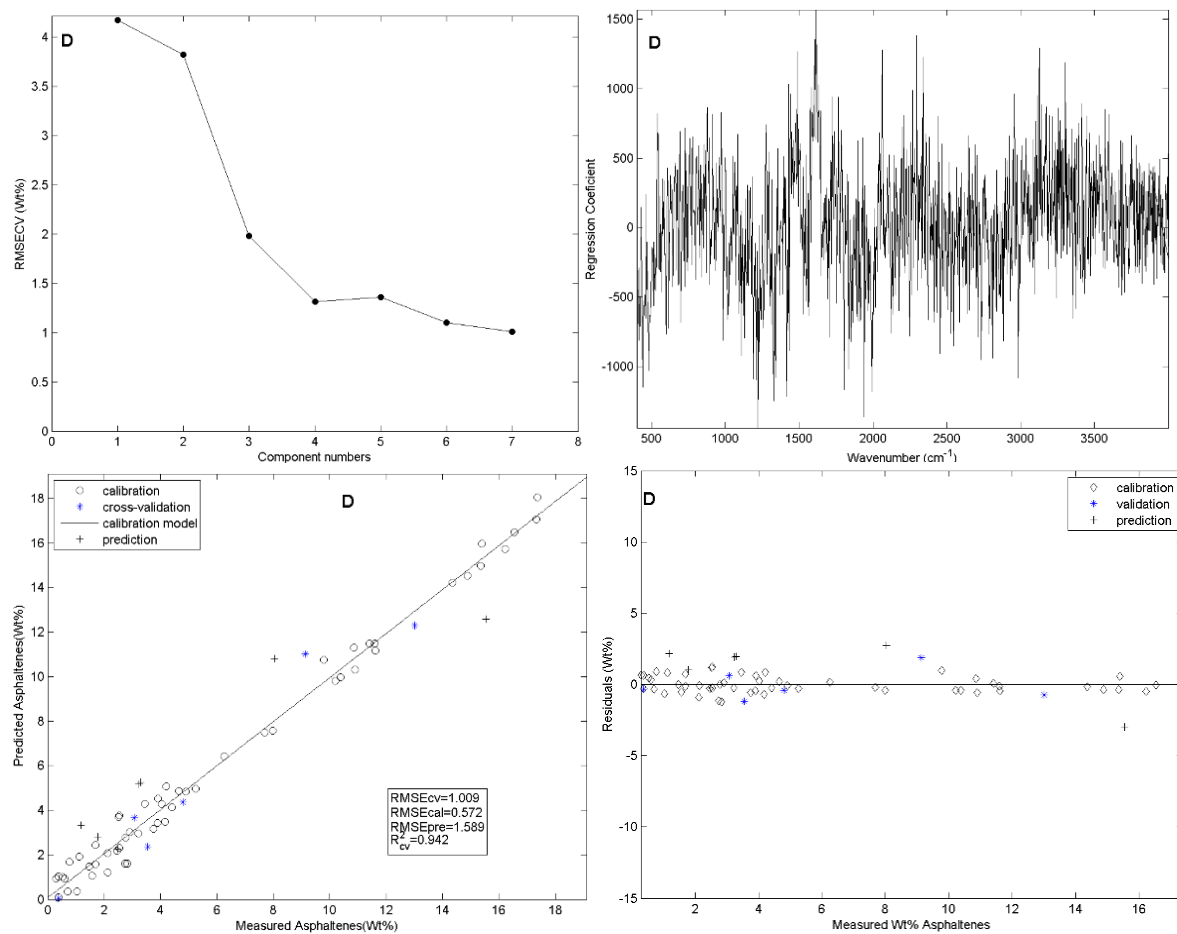


Figura 29. Varianza explicada acumulativa de contenido de asfáltenos

Como se mencionó anteriormente el RMSECV (1.01wt%) (ver tabla 8). El bajo error de predicción RMSEP de 1.59wt%, muestra un modelo con un buen desempeño para predecir el contenido de asfáltenos para muestras de alto porcentaje de contenido de asfáltenos. Para muestras con bajos contenidos de asfáltenos el desempeño se reduce drásticamente. Este resultado se corrobora en la gráfica de residuales (fig. 30).

Tabla 8. Parámetros estadísticos del modelo de calibración para el contenido de asfaltenos

propiedad	Número de component es	$RMSEC$ V (wt %)	$RMSE$ C (wt %)	$RMSEP$ (wt %)	R_{CV}^2	Varianza explicada (%)
asfaltenos	7	1.01	0.57	1.59	0.94	99.81

**Figura 30.** RMSECV, gráfica de coeficientes, modelo de calibración y gráfica de residuales para contenido de asfaltenos.

2.6.1.1.5 Calibración y validación del modelo PLS para la predicción del contenido de carbón Conradson. El modelo de calibración obtenido para predecir el contenido de carbón Conradson a partir de espectros Raman modelo exhibió al igual que los modelos anteriores un comportamiento decreciente del error RMSECV con el número de componentes, logrando minimizar el error para 7 componentes. En esta condición también se obtuvo el mejor valor para el coeficiente de determinación R_{CV}^2 , que indica el ajuste lineal de los datos de calibración al modelo figura 32 y una varianza acumulativa explicada de 99.80% (fig. 31). El modelo exhibe un buen desempeño en la predicción de altos contenidos de carbón Conradson y su desempeño se deteriora para muestras de bajo contenido de éste. La gráfica de los coeficientes de regresión (fig. 32), muestra que el aporte de las variables al modelo es aproximadamente uniforme de forma que no es práctico prescindir de variables pues todas aportan al modelo.

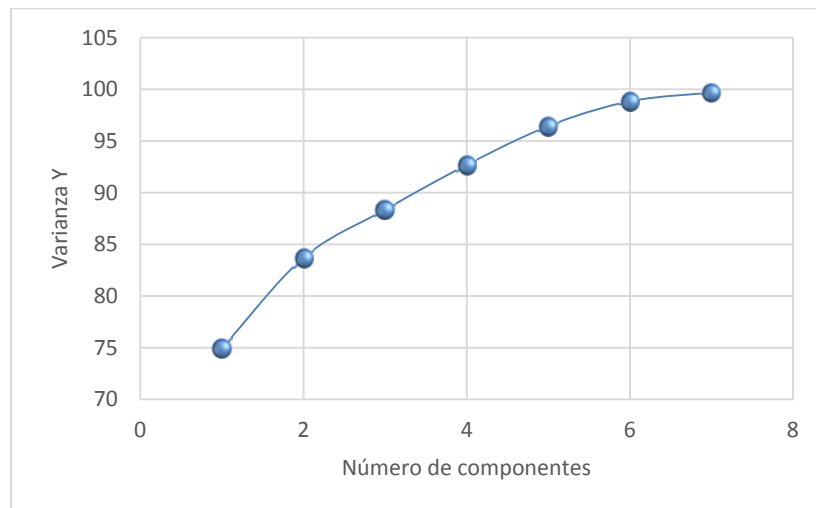
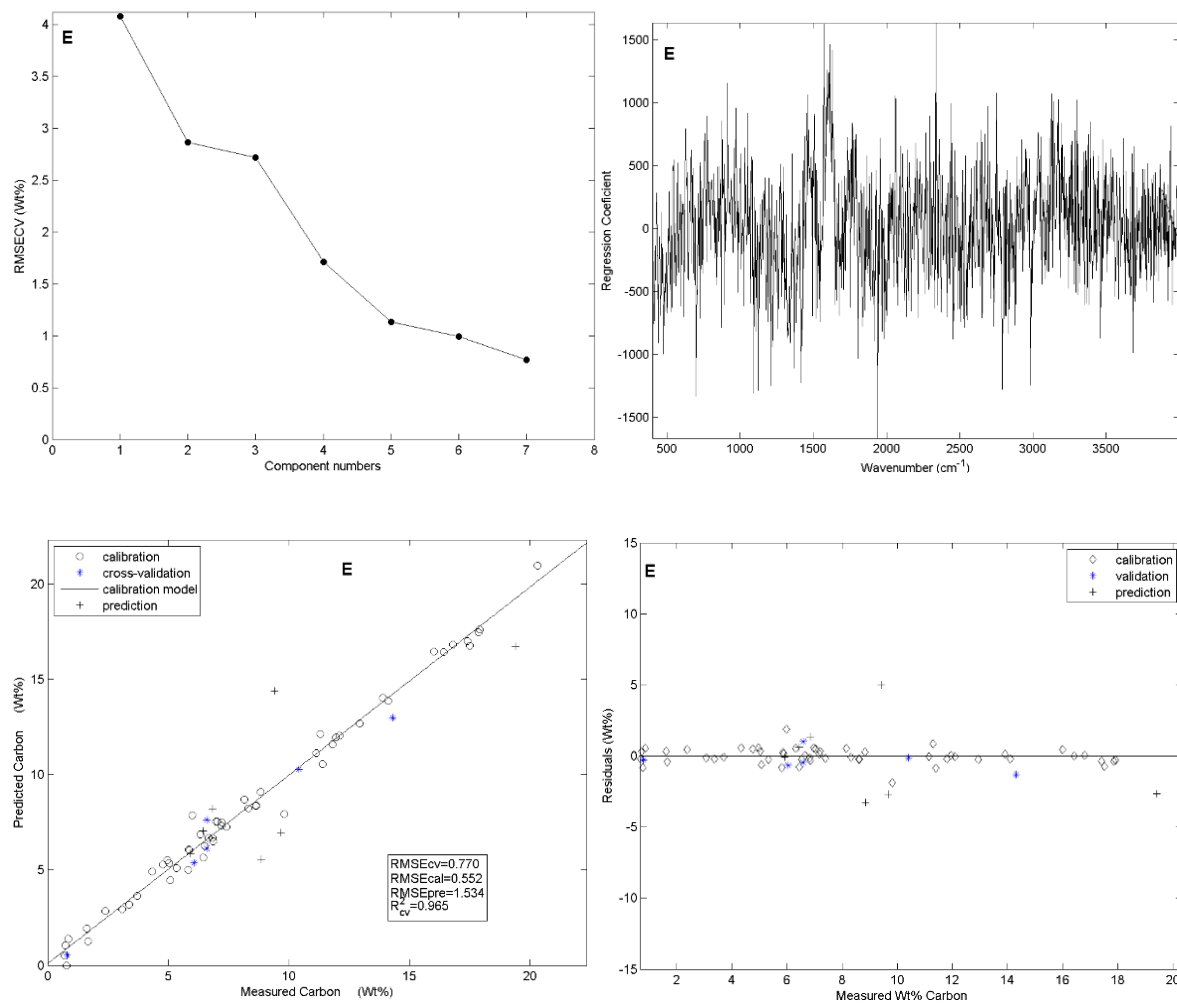


Figura 31. Varianza explicada acumulativa de contenido de carbón Conradson

Tabla 9. Parámetros estadísticos del modelo de calibración para el contenido de carbón Conradson

propiedad	Número de componentes	RMSEC V (wt %)	RMSE C (wt %)	RMSEP (wt %)	R_{CV}^2	Varianza explicada (%)
Carbón Conradson	7	0.77	0.55	1.53	0.97	99.80

**Figura 32.** RMSECV, gráfica de coeficientes, modelo de calibración y gráfica de residuales para contenido de carbón Conradson

2.6.1.2 Calibración y validación de modelos PLS a partir de Espectrometría de masas.

A continuación, se describe el procedimiento empleado para el desarrollo y evaluación de los modelos de predicción de contenido de saturados, aromáticos, resinas, asfaltenos y carbón Conradson, a partir de los espectros de masas de las muestras preparadas por el método indirecto obtenidos entre el rango de 250 a 1000 m/z y utilizando el algoritmo NIPALS programado en Matlab.

Las 62 muestras que correspondieron a la población total se dividieron en dos grupos. Un grupo de muestras de calibración (50) y las muestras restantes se usaron para la validación del modelo.

2.6.1.2.1 Calibración y validación del modelo PLS para la fracción de saturados. Una vez obtenido el mejor modelo PLS de calibración para la predicción de la fracción de saturados que presentó el menor valor de RMSECV y el mejor R^2 , se encontró que los parámetros estadísticos RMSECV, RMSEP y R^2 variaban con el número de componentes (variables latentes).

Se observa como disminuye el comportamiento de RMSECV a medida que aumenta el número de componentes, que para el modelo de saturados obtuvo su valor más bajo en 12 componentes (fig. 33), Este parámetro estadístico da información de la medida de la dispersión de las muestras de validación alrededor de la línea de regresión cuando es usada la validación cruzada.

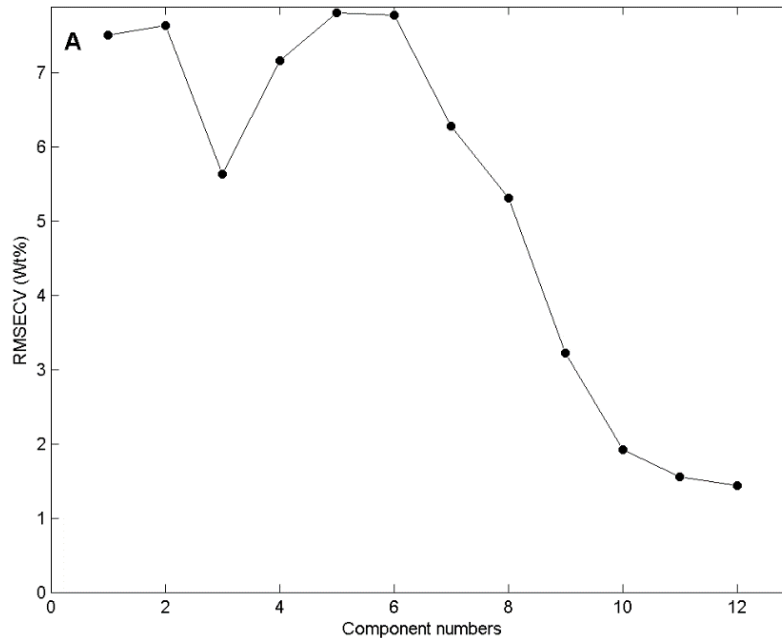


Figura 33. Raíz del error cuadrático medio de validación cruzada (RMSECV) en función del número de componentes (variables latentes) para el modelo de predicción de la fracción de saturados.

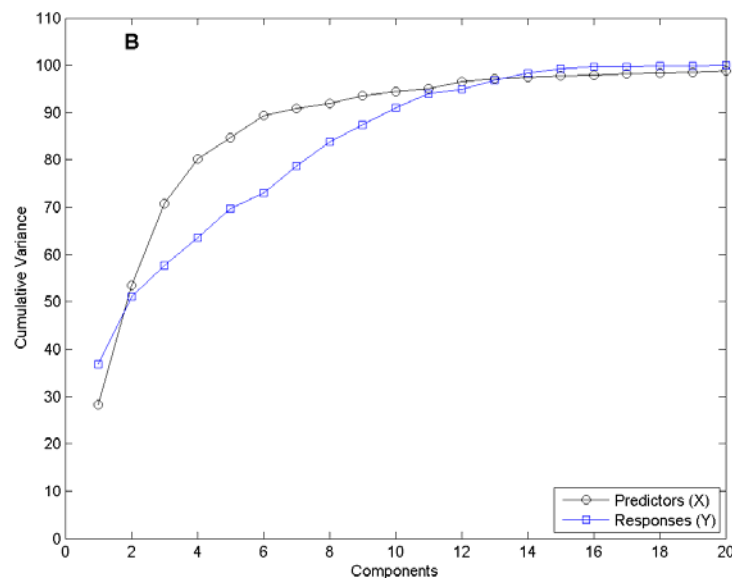
Sin embargo, la selección del número de componentes dependió además del valor del coeficiente de determinación R^2 de validación cruzada, el cual evolucionó igualmente con el número de componentes obteniéndose su mejor valor de 0.97 para 12 componentes, determinando así que 12 es el número de componentes óptimo para construir el modelo PLSR de predicción del contenido de saturados, mostrando que la recta de regresión ajusta bien los datos de calibración, comportamiento que se confirma con el valor del RMSEC.

Los parámetros estadísticos obtenidos para el modelo PLSR de predicción de contenido de saturados se muestran en la siguiente tabla:

Tabla 10. Parámetros estadísticos del modelo de calibración para el contenido de saturados

propiedad	Número de componentes	$RMSEC$ V (wt %)	$RMSE$ C (wt %)	$RMSEP$ (wt %)	R_{CV}^2	Varianza explicada (%)
Saturados	12	1.44	1.22	2.37	0.97	99.93

La figura 34 muestra la evolución típica de la varianza explicada acumulativa con el número de componentes, para los modelos de calibración a partir de los espectros de masas. Para este caso particular se encontró que con 12 componentes es posible explicar el 99.93% de la varianza de los datos de contenido de saturados, que es una estimación de la capacidad del modelo de predecir la propiedad para nuevas muestras.

**Figura 34.** Varianza explicada acumulativa de predictores y de contenido de saturados

Los coeficientes de regresión muestran cómo se pondera cada variable X al predecir una respuesta particular Y. Resume la relación entre los predictores X y una respuesta dada y pueden calcularse para cualquier número de componentes. Los predictores con

coeficientes cercanos a cero significan que no aportan a la predicción de la propiedad y pueden ser retirados de la matriz X de predictores. En la gráfica 35 se muestra un comportamiento típico de los coeficientes de regresión. Claramente se puede ver que en la región comprendida entre 750m/z y 1000m/z aproximadamente, los coeficientes de regresión cercanos a cero reflejan la casi nula contribución de estas variables al modelo, indicando que estas variables pueden ser retiradas del conjunto de predictores reduciendo el tiempo de cálculo en el desarrollo del modelo. Por otra parte la región entre 250m/z a 750m/z poseen coeficientes que indican un aporte de las variables de este rango al modelo y contribuyen tanto negativamente como positivamente a la respuesta Y. Sin embargo, la máxima contribución se observa en el rango de variables comprendido entre 250m/z y 500m/z.

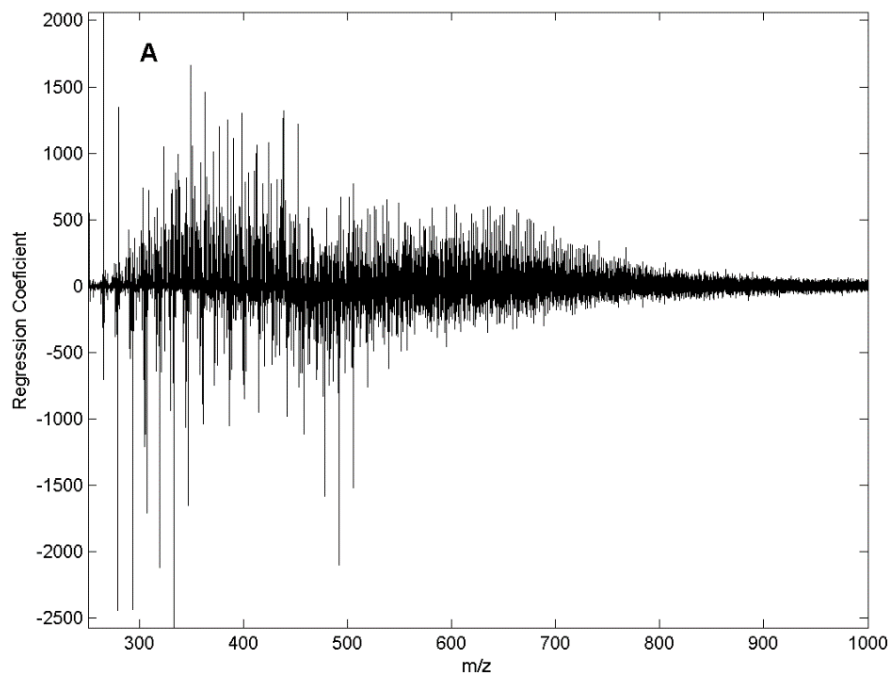


Figura 35. Coeficientes de regresión para el modelo de predicción PLSR de contenido de saturados

La figura 36 muestra la curva de predicción del contenido de saturados contra valores medidos para las muestras de calibración y de predicción.

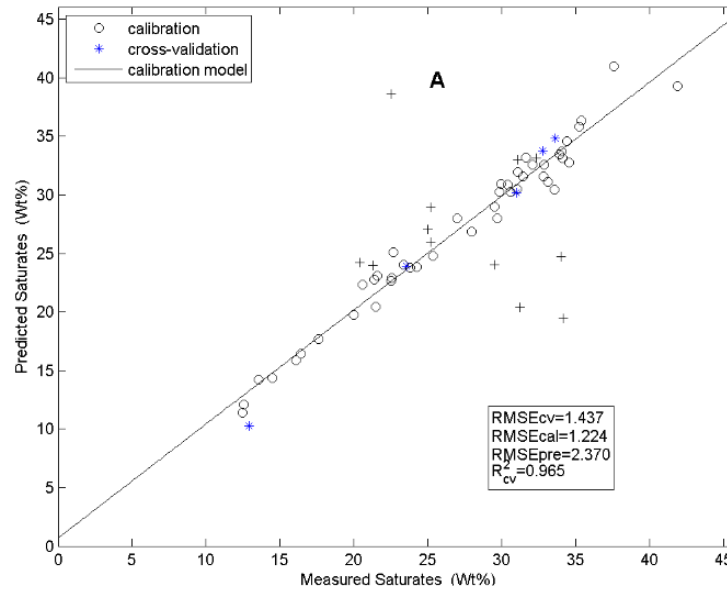


Figura 36. Curva de valores predichos contra valores medidos del contenido de saturados

Un modelo está concebido para hacer predicciones futuras, es por ésta razón que la validación de dicho modelo es muy importante. La validación de un modelo basado en datos experimentales busca estimar la incertidumbre de predicciones futuras, es decir evalúa el desempeño del modelo con nuevos datos de la misma clase de los que se usaron para su construcción. Si la incertidumbre es razonablemente baja el modelo es válido.

Para la validación cruzada (K-fold), con múltiples iteraciones que se realizó sobre el grupo de muestras de calibración se encontró un error RMSECV de 1.437%. Por otra parte, ya que el modelo de regresión está concebido para hacer predicciones futuras, se realizó una validación con muestras que no participaron en la construcción del modelo

obteniéndose un error RMSEP de 2.37wt% del peso del contenido de saturados (Ver tabla 10)

En la figura 37 se observan los residuales (diferencia entre los valores de referencia y los valores predichos), que pueden ser negativos o positivos y se distribuyen aleatoriamente alrededor del cero, permaneciendo aproximadamente constantes entre -4wt% a 4wt%, excepto algunos residuales superiores para algunas muestras de validación externa con mayores porcentajes de contenido de saturados.

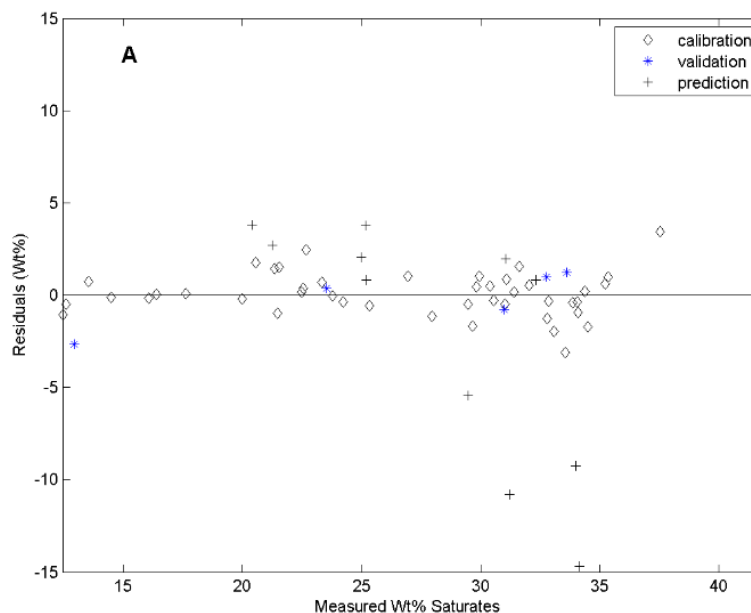


Figura 37. Gráfica de residuales para el porcentaje del contenido de Saturados

2.6.1.2.2 Calibración y validación del modelo PLS para las fracciones de aromáticos, resinas, asfáltenos y contenido de carbón Conradson. Análogamente a lo observado en el modelo para la predicción de saturados, se observa como disminuye el comportamiento de RMSECV a medida que aumenta el número de componentes, parámetro estadístico que da información de la medida de la dispersión de las muestras de validación alrededor

de la línea de regresión cuando es usada la validación cruzada. De la misma manera se observa el comportamiento de los coeficientes de regresión obtenidos para los modelos de predicción del contenido de aromáticos, resinas, asfaltenos y carbón Conradson (ver anexo B).

A continuación, se presentan los resultados más importantes obtenidos en el desarrollo de los modelos de predicción para las fracciones de aromáticos, resinas, asfaltenos y contenido de carbón Conradson a partir de los espectros de masas, siguiendo la misma metodología expuesta anteriormente.

Tabla 11. Parámetros estadísticos del modelo de calibración para el contenido de saturados, aromáticos, resinas, asfaltenos y carbón Conradson.

propiedad	Número de componentes	RMSECV (wt %)	RMSEC (wt %)	RMSEP (wt %)	R_{CV}^2	Varianza explicada (%)
Saturados	12	1.44	1.22	2.37	0.97	99.93
Aromáticos	9	1.16	3.40	6.59	0.96	99.90
Resinas	14	1.43	1.12	2.17	0.96	99.96
Asfaltenos	13	0.58	0.69	1.33	0.99	99.98
Carbón	14	1.03	0.83	1.61	0.97	99.96

La tabla 11, muestra algunos parámetros estadísticos PLSR relacionados con los modelos predictivos para el análisis SARA y contenido de carbón Conradson. Para los modelos de calibración para el contenido de aromáticos, resinas, asfaltenos y carbón Conradson se observa que el número óptimo de componentes o variables latentes de acuerdo al valor más bajo de error RMSECV oscila entre 9 y 14 componentes (ver anexo B), el cual asegura una alta varianza explicada acumulada entre 99.90% y 99.98%. Para el contenido de aromáticos el RMSECV fue de 1.16wt% con 9 componentes que. Igualmente, el

RMSECV para las fracciones de resinas, asfaltenos y contenido de carbón Conradson es respectivamente 1.43wt%, 0.58wt% y 1.03wt%, ver tabla 11.

En la figura 38(B a E) se muestran los modelos de calibración. Estos modelos exhiben altos valores de coeficientes de determinación R_{CV}^2 , los cuales se encuentran en el rango entre 0.96 y 0.99, que demuestran que la línea de regresión ajusta muy bien los datos de calibración (modelo de calibración). Este resultado se confirma con los bajos valores de calibración RMSEC.

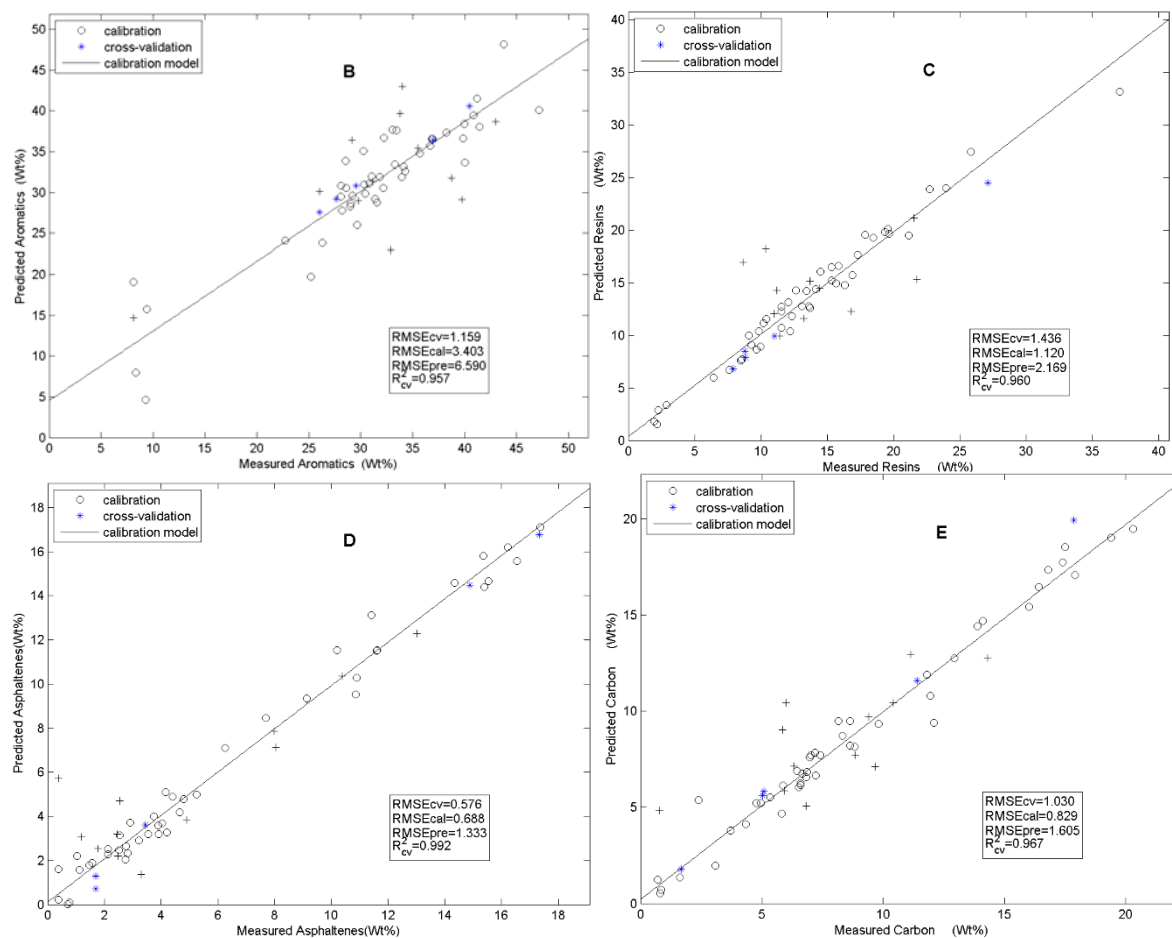


Figura 38. Curva de valores predichos contra valores medidos del contenido de aromáticos (B), resinas(C), asfaltenos (D) y carbón Conradson (E)

El error RMSEP para las muestras que no fueron utilizadas en el grupo de calibración para construir los modelos se encuentran en la tabla 11. El rango para este error de predicción se encuentra entre 1.33wt% y 6.59wt% encontrándose el valor más bajo para el porcentaje del contenido de asfaltenos, seguido por el porcentaje de contenido de carbón Conradson. Estos resultados se confirman en la figura 39 (B a E) de residuales, en las cuales se observa un comportamiento homocedástico, exhibiendo resultados para los modelos de predicción del contenido de asfaltenos, entre -2wt% y 2wt%.

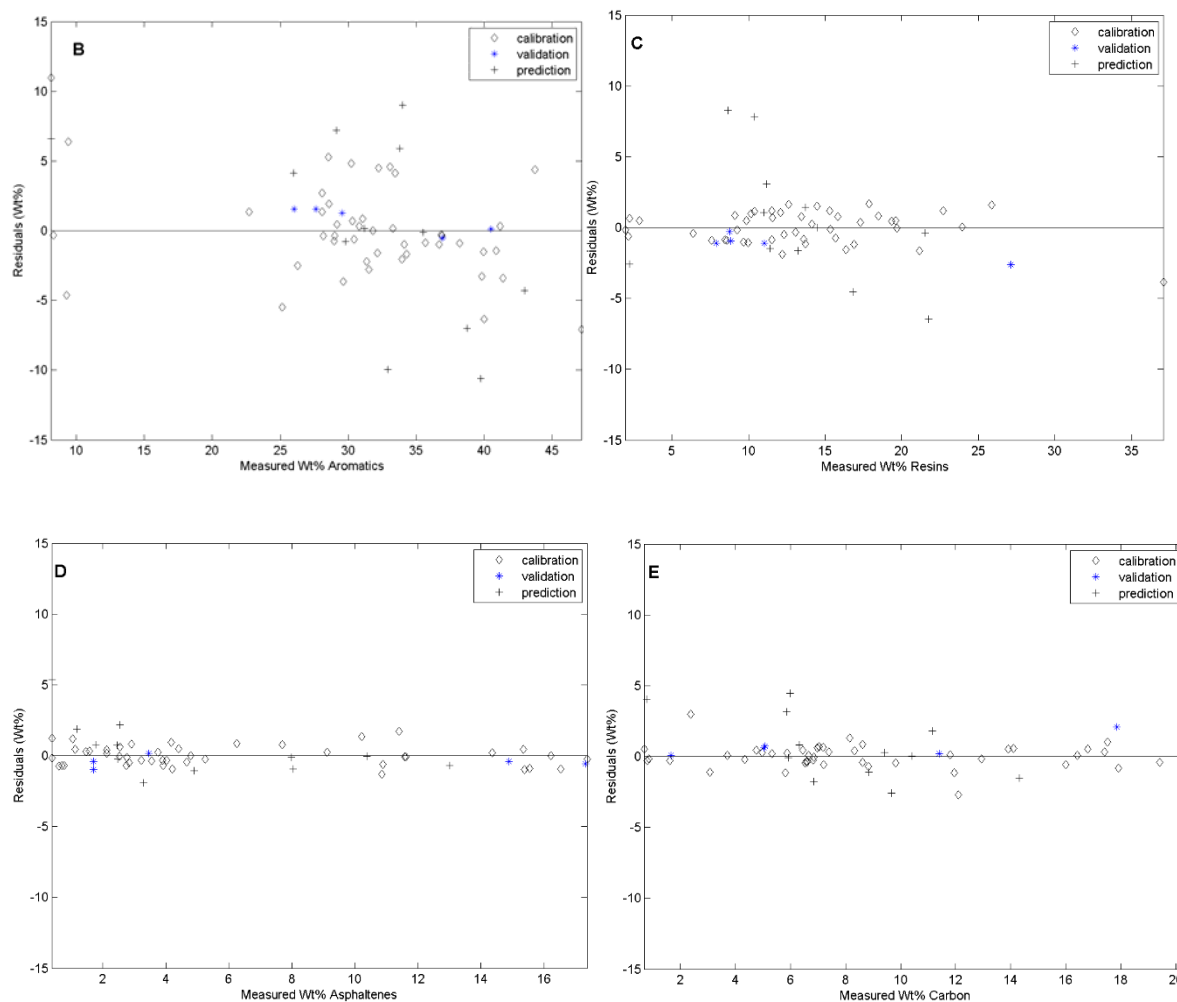


Figura 39. Grafica de residuales del contenido de aromáticos (B), resinas(C), asfaltenos (D) y carbón Conradson (E)

3. CONCLUSIONES

Las técnicas espectroscopía Raman y calibración multivariable como PLSR han demostrado ser alternativas para determinar el porcentaje del contenido de saturados, aromáticos, resinas y asfaltenos (Análisis SARA) así como el contenido de carbón Conradson, mostrando buenos desempeños.

La calibración multivariable (PLSR) en conjunto con la espectrometría de masas generan modelos capaces de predecir el porcentaje del contenido de carbón Conradson y el análisis SARA de los crudos colombianos con parámetros estadísticos que demuestran desempeños de los modelos especialmente interesantes para predecir el contenido de asfaltenos y contenido de carbón Conradson.

Los modelos de predicción obtenidos en este estudio son útiles para predecir el análisis SARA y el porcentaje de contenido de carbón Conradson para muestras externas que tengan un análisis SARA y contenido de carbón Conradson en el rango comprendido entre el valor máximo y mínimo reportado para las muestras usadas para la calibración de los modelos. Así mismo los espectros Raman deben ser adquiridos bajo los mismos parámetros y condiciones experimentales reportados.

Debido a que los modelos fueron desarrollados con la ayuda de un algoritmo propio programado en Matlab, permite flexibilidad para variar los parámetros de diseño del modelo de manera que se pueden buscar las mejores condiciones que una herramienta tradicional en esta tarea no permite.

Se logró desarrollar un algoritmo que permitió realizar la corrección del desplazamiento de los espectros de masas de una manera automática y rápida con el fin de obtener la máxima colinealidad posible requerida para el desarrollo de los modelos de calibración.

Técnicas como la espectroscopía Raman, espectrometría de masas y calibración multivariable (PLS) han permitido desarrollar modelos de predicción del análisis SARA y contenido de carbón Conradson con ventajas muy importantes sobre las metodologías estándar como el corto tiempo de análisis y la reducción drástica en el uso de solventes altamente contaminantes.

4. RECOMENDACIÓN

Los modelos propuestos podrían ser mejorados incluyendo muestras en el grupo de calibración con mayor variabilidad en los porcentajes de contenido de saturados, aromáticos, resinas, asfaltenos y carbón Conradson. Esto podría contribuir a obtener mejores y más robustos modelos de predicción, permitiendo estimar rápidamente el análisis SARA y el contenido de carbón Conradson.

REFERENCIAS BIBLIOGRAFICAS

- Abian, J., Carrascal, M., & Gay, M. (2008). Introducción a la espectrometría de masas para la caracterización de péptidos y proteínas. *Proteómica*, 16-33.
- American Society Test and Materials (ASTM). (2009). Standard Test Method for Separation of asphalt into four fractions. D4124-01. En *ASTM Standards* (págs. 1-8). United States: ASTM International.
- Araujo Pantoja, P., Mendes, M., & Oller Nascimento, C. (2003). Contribution of mass spectrometry in assessing quality of petroleum fractions. The use of mass spectrometry for assessing asphaltenes. *Journal of petroleum science and engineering*, 198-205.
- ASTM D2007-93 Annual Book of ASTM Standards. (1998). American Society for testing Materials. 1998, 655-661.
- Barman, B., & Cebolla, V. (2000). Chromatographic Techniques for petroleum and related products. *Critical reviews in analytical chemistry*, 6-10.
- Brittain Cooper, J. (1999). Chemometric analysis of Raman spectroscopic data for process control applications. *Chemometrics and intelligent laboratory systems*, 231-247.
- Brown, J. M. (2003). *Molecular Spectroscopy*. New York, United States: Oxford Science Publications.
- Cooper, J. B. (1999). Chemometrics Analysis of Raman Spectroscopic data for process control applications. Modeling physical properties of petroleum fuels: determination of octane numbers. *Chemometrics and Intelligent Laboratory Systems*, 46, 231-247.
- Council, T. i. (2011). *Introducción a la refinación del petróleo*. TT Math Pro.

- Eckert, P., & Roach, P. (2012). Chemical Characterization of crude Petroleum using Nanospray desorption electrospray Ionization coupled with High-resolution Mass Spectrometry. *Analytical Chemistry*, 84, 1517-1525.
- Enke, N. B. (2001). Practical Implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrometry Reviews*, 363-373.
- Filgueiras, P. R. (2014). Determination of API gravity, Kinematic viscosity and water content in petroleum by ATR-FTIR spectroscopy and multivariate calibration. (Elsevier, Ed.) *Fuel*, 116, 123-130.
- Filgueiras, P. R. (2016). Determination of saturates, Aromatics, and Polars in Crude Oil by ¹³C NMR and support Vector Regression with variable Selection by Genetic Algorithm. (A. publications, Ed.) *Energy & Fuels*, 30, 1972-1978.
- Filgueiras, P., & M.S., C. (2014). Determination of API gravity, Kinematic viscosity and water content in petroleum by ATR-FTIR spectroscopy and multivariate calibration. *Fuel*, 122-129.
- Filzmoser, K. V. (2009). *Introduction Multivariable Statistical Analysis in Chemometrics*. Boca Ratón: CRC Press.
- G. Marshall, A., & Ryan P., R. (2004). Petroleomics: The next grand challenge for chemical analysis. *Acc Chem*, 37, 53-59.
- Hollas, M. (2002). *Basic Atomic and Molecular Spectroscopy*. New York, United States: Wiley Interscience. Royal Society of Chemistry.
- Hsu, C., Hendricson, C., Rodgers, R., McKenna, A., & Marshall, A. (2011). Petroleomics: advanced molecular probe for petroleum heavy ends. *Journal Mass Spectrometry*, 46, 337-343.
- Ingvar Eide, K. Z. (2005). A Novel Method for Chemical Fingerprinting of Oil and Petroleum Products Based on Electrospray Mass Spectrometry and Chemometrics. *Energy and Fuels*, 19, 964-967.

- Intech. (2012). Analysis of polar components in crude oil by ambient Mass Spectrometry. En M. El-Sayed (Ed.), *Crude Oil Emulsions composition stability and characterization* (pág. 110).
- Lamus, C., Guzmán, A., Murcia, B., Cabanzo, R., & Mejía-Ospino, E. (2011). Uso de Análisis Multivariado en la determinación SARA de crudos por Espectroscopía NIR. *Revista Colombiana de Física*, 43(3), 635-642.
- Lewis, J., Wei, J., & Siuzdak, G. (2000). Matrix-assisted Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis. En *Encyclopedia of Analytical Chemistry* (págs. 5880-5894). Chichester: John Wiley & Sons Ltd.
- Li, S., & Dao, L. (2012). Classification of gasoline brand and origin by Raman spectroscopy and novel R-weighted LSSVM algorithm. *Fuel* 96, 146-152.
- Marshall, A. G., & Hendrickson, C. (2008). High resolution mass spectrometers. *Annu. Rev. Chem*, 1-21.
- Martinez Gomez, M., & Ballesteros Gonzalez, M. (2009). Espectrometría de masas y análisis de biomarcadores. *Universidad de Alcalá. Madrid*, 118.
- Marvin, L., Roberts, M., & Fay, L. (2003). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in clinical chemistry. *Clinica Chimica Acta*, 337, 11-21.
- Melendez, L., Lache, A., Orrego, J., Pachon, Z., & Mejia Ospino, E. (2012). Prediction of the SARA analysis of colombian crude oils using ATR-FTIR spectroscopy and chemometric methods. *Journal of petroleum science and engineering*, 90-91, 56-60.
- Merdrignac, I., & Espinat, D. (2007). Physicochemical characterization of petroleum fractions: the state of the Art. *Oil and Gas science and technology*, 7-32.
- Ministerio de minas y energía, 2013. (s.f.). *Cadena de petróleo*.
- Otto, M. (2007). *Chemometrics* (2a Edición ed.). Germany: Wiley-Vch.

- Pudenzi, M. (2012). Petroleum Thermal Evolution Assessment Using MALDI-MS and chemometric Analysis of Asphaltenes. *LATIN AMERICAN CONGRESS ON ORGANIC GEOCHEMISTRY*.
- R.L. McCreery. (2000). *Raman Spectroscopy for Chemical Analysis*. New York: John Wiley & Sons Inc.
- Robins, C., & Limbach, P. (2003). The use of nonpolar matrices for matrix-assisted laser desorption/ionization mass spectrometric analysis of high boiling crude oil fractions. *Rapid Commun. Mass Spectrom.*, 17, 2839-2845.
- Rogers, R., & Marshall, A. (2004). . characterization of heavy, biodegraded crude oils by high resolution ESI FT-ICR mass spectrometry. *Am. Chem Soc. Fuel Chem*, 49(29), 484.
- Rogers, R., Shaub, T., & Marshall, A. (2005). Petroleomics: MS returns to its roots. *Analytical Chemistry*, 21-27.
- Silva, S., Silva, A., Ribeiro, J., Martins, F., Da Silva, F., & Silva, C. (2011). Chromatographic and spectroscopic analysis of heavy crude oil mixtures with emphasis in nuclear magnetic resonance spectroscopy: A review. *Analytica Chimica Acta*, 18-37.
- Speight, J. (1999). *The chemistry and technology of petroleum* (3 ed.). New York: Marcel Dekker Inc.
- Welch, B. K. (1998). Determination of asphaltenes in petroleum Crude Oils by Fourier Transform Infrared Spectroscopy. *Energy & Fuels*, 12, 1008-1012.
- Xian, F., & Hendrickson, C. (2012). High Resolution Mass Spectrometry. *Analytical Chemistry*, 84,708-719.

ANEXO A

Análisis SARA y contenido de carbón conradson de referencia de las muestras de crudo empleadas en este trabajo

Propiedad					
Etiqueta (LEAM)	Saturados (%wt)	Aromáticos (%wt)	Resinas (%wt)	Asfáltenos (%wt)	Carbón Conradson (%wt)
1	31.05	29.15	11.2	4.91	6.32
3	34.39	39.84	17.86	3.22	6.67
4	35.24	33.27	9.1	7.69	11.94
5	29.48	25.99	16.8	10.39	11.14
6	30.55	26.04	9.97	5.25	6.53
7	35.37	33.96	7.9	10.9	11.8
8	34	33.98	10.38	2.54	5.98
9	33.88	34.12	8.78	11.59	12.94
10	33.09	30.29	10.17	6.25	8.15
11	17.61	25.16	15.69	2.83	9.81
12	22.48	8.13	2.24	0.38	5.88
13	31.21	32.9	11.41	2.46	5.85
14	29.81	28.54	12.21	2.12	5.82
16	31.63	28.97	11.04	3.91	6.6

17	22.48	8.13	2.24	0.38	0.79
18	26.96	9.29	2.16	0.38	0.81
19	34.1	27.62	11.55	4.21	6.86
20	34.06	31.36	15.33	3.45	7.21
21	12.92	40.87	19.35	17.33	17.84
22	23.77	8.31	1.96	0.7	0.71
23	28.49	9.38	1.97	0.29	0.74
24	32.32	29.78	13.22	3.29	6.84
25	30.98	30.41	16.34	2.13	5.03
26	31.41	29.01	11.54	1.7	0.86
27	31.09	29.18	12.33	3.55	6.82
28	32.04	28.1	13.08	3.75	6.44
29	32.87	28.09	6.4	0.61	1.63
30	16.06	38.21	21.16	14.89	17.9
31	21.46	36.9	15.3	14.36	16.41
33	28.13	30.78	14.14	3.07	6.06
34	30.17	19.28	11.22	3.24	6.44
35	34.79	15.8	8.31	0.53	3.38
36	31.01	35.66	13.44	0.78	4.34
37	25.17	38.74	13.71	1.78	5.93
38	23.34	36.69	12.63	4.17	8.32
39	25.34	40	16.9	4.79	8.83

40	27.96	30.24	12.09	11.41	13.9
41	13.52	33.06	19.59	17.37	20.3
42	19.98	36.83	17.31	16.54	16.8
43	33.61	28.57	11.53	2.54	5.08
44	30.39	22.7	9.25	4.66	5.34
45	29.49	29.65	13.71	3.89	6.6
46	37.53	32.14	8.84	10.86	12.1
47	20.39	31.17	21.75	8.04	9.41
48	41.86	31.83	8.57	9.13	2.38
49	37.09	33.02	9.95	9.78	11.3
50	29.66	30.81	13.61	4.41	7.4
51	12.44	39.94	22.71	15.55	19.4
52	24.99	39.73	21.53	2.47	9.67
53	29.94	31.54	9.86	2.77	4.77
54	21.55	29.53	10.41	4.03	6.97
55	12.57	32.22	37.06	15.35	17.5
56	25.19	35.52	10.99	13.01	14.3
57	34.51	9.39	2.89	1.04	1.67
58	14.49	41.39	18.46	16.22	17.4
59	32.76	26.29	8.47	1.47	3.71
60	32.81	28.18	7.59	1.57	3.08
61	20.55	34.29	15.83	15.39	16

62	23.53	31.05	9.67	11.62	11.4
63	24.23	36.92	14.15	2.52	4.97
64	34.16	42.96	8.67	1.18	8.85
65	33.55	33.45	27.11	1.12	7.03
67	22.66	43.74	23.95	2.74	8.62
68	21.34	47.15	25.86	2.91	8.63
69	22.54	41.16	14.49	1.7	7.19
70	16.39	40.48	19.67	10.21	14.1
71	21.25	33.78	14.46	7.99	10.4

ANEXO B.**RMSECV y coeficientes de regresión de los modelos de predicción del contenido de aromáticos (B), resinas(C), asfaltenos (D) y carbón Conradson (E)**