

**PREDICCIÓN DEL ANÁLISIS ELEMENTAL POR ESPECTROSCOPIA
FOTOACÚSTICA EN LA REGIÓN DEL INFRARROJO MEDIO Y MÉTODOS
QUIMIOMÉTRICOS A LAS FRACCIONES DE LOS FONDOS DE VACÍO**

CRISTIAN FERNEY LEGUIZAMO CASTELLANOS

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE QUÍMICA
LABORATORIO DE ESPECTROSCOPIA ATÓMICA Y MOLECULAR (LEAM)
BUCARAMANGA
2012**

**PREDICCIÓN DEL ANÁLISIS ELEMENTAL POR ESPECTROSCOPIA
FOTOACÚSTICA EN LA REGIÓN DEL INFRARROJO MEDIO Y MÉTODOS
QUIMIOMÉTRICOS A LAS FRACCIONES DE LOS FONDOS DE VACÍO**

CRISTIAN FERNEY LEGUIZAMO CASTELLANOS

**Trabajo de grado para optar el título de
QUÍMICO**

**DIRECTOR:
ENRIQUE MEJÍA OSPINO
Químico, Ph. D.**

**COORDIRECTOR
JORGE ARMANDO ORREGO
Químico, M. Sc.**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE QUÍMICA
LABORATORIO DE ESPECTROSCOPIA ATÓMICA Y MOLECULAR (LEAM)
BUCARAMANGA
2012**

DEDICATORIA

A Dios, por ser la luz en mi camino.

A mi mamá Smith Leguizamo Castellanos, que desde el cielo siempre estuvo presente para alcanzar este gran logro.

A mi Hijo, Juan Esteban, por ser mi mayor motivación para salir adelante y triunfar en la vida

Mi familia por estar conmigo incondicionalmente, y apoyarme en todo momento.

A mis amigos que siempre me han acompañado en el transcurso de mi carrera universitaria...muchas GRACIAS

A todos ellos les dedico esta Tesis...

AGRADECIMIENTOS

Al Profesor Enrique Mejía Ospino (Director) y al Profesor Rafael Cabanzo por brindarme la oportunidad de trabajar en Laboratorio de Espectroscopia Atómica y Molecular (LEAM) y por su apoyo incondicional en cada paso de la realización de este trabajo.

Al Msc. Jorge Armando Orrego (Codirector) por el aporte de sus conocimientos que ayudaron a culminar este proyecto.

A mis Familiares por ser el eje que sostiene mi vida, por su constante apoyo y amor.

A mis amigos Oscar Castañeda, Fabián Camargo, Oscar Dallos, Diego Valencia, Mario Llinas, Johan Gallo, Sergio Garrido, Mireya Moreno, Fay zuly, Roika...por su sincera amistad y hacer de los momentos vividos más alegres. Gracias por su compañía, cariño y ayuda en todo momento.

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

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

CONTENIDO

Pág.

INTRODUCCIÓN	15
1. CONSIDERACIONES TEORICAS	17
1.1 PETROLEO GENERALIDADES	17
1.2 CLASIFICACION DEL PETROLEO	18
1.2.1 Composición química	18
1.2.2 Densidad	19
1.2.3 Contenido Azufre.....	21
1.3 ANALISIS COMPOSICIONAL.....	22
1.3.1 Análisis elemental.....	24
1.3.1.1 Contenido de Azufre	25
1.3.1.2 Contenido de Nitrógeno	27
1.3.1.3 Contenido de Oxígeno	29
1.4 ESPECTROSCOPIA INFRARROJA	30
1.4.1 Historia de la técnica	30
1.4.2 Regiones espectrales	31
1.4.3 Análisis infrarrojo	32
1.4.3.1 Espectroscopia Fotoacústica	34
1.5 QUIMIOMETRIA	39
1.5.1 Calibración multivariable	40
1.5.1.1 Clasificación de los modelos de calibración multivariable	41
1.5.2 Construcción de modelos de calibración multivariable	43
1.5.3 Pretratamiento de los datos registrados	46
1.5.3.1 Corrección de línea base	47
1.5.3.2 Normalización	47
1.5.3.3 Suavizado espectral	48
1.5.3.4 Centrado	48
1.5.3.5 Derivadas	49
1.5.4 Análisis por componentes principales (PCA).....	49
1.5.5 Regresión por mínimos cuadrados parciales (PLS)	52
2. METODOLOGIA	55
2.1 ESQUEMA EXPERIMENTAL.....	55
2.2 SELECCIÓN DE MUESTRAS.....	56
2.3 MONTAJE EXPERIMENTAL	57

2.3.1 Instrumentación.....	57
2.4 DETERMINACION DE LOS PARAMETROS	59
2.4.1 Verificación del espectrómetro	62
2.5 ADQUISICION Y ANALISIS DE ESPECTROS PAS.....	64
2.5.1 Vibraciones en los fondos de vacío	65
2.6 PROCESAMIENTO DE DATOS.....	70
2.6.1 Pretratamiento de datos	70
2.6.2 Análisis por componentes principales (PCA).....	73
2.6.3 Regresión por mínimos cuadrados parciales	73
3. RESULTADOS Y ANALISIS	74
3.1 ANALISIS DE COMPONENTES PRINCIPALES (PCA)	74
3.2 ANALISIS DE REGRESION POR MINIMOS CUADRADOS PARACIALES (PLS-REGRESION).....	80
3.2.1 Modelo PLS para la predicción del % p/p Carbono	80
3.2.1.1 Calibración del modelo PLS.....	81
3.2.1.2 Validación cruzada del modelo PLS.....	83
3.2.1.3 Análisis de muestras excluidas	85
3.2.2 Modelo PLS para la predicción del % p/p Hidrógeno	87
3.2.2.1 Calibración del modelo PLS.....	87
3.2.2.2 Validación Cruzada del modelo PLS.....	90
3.2.2.3 Análisis de muestras excluidas	91
3.2.3 Modelo PLS para la predicción del % p/p Nitrógeno	93
3.2.3.1 Calibración del modelo PLS.....	93
3.2.3.2 Validación Cruzada del modelo PLS.....	95
3.2.3.3 Análisis de muestras excluidas	97
3.2.4 Modelo PLS para la predicción del % p/p Azufre.....	99
3.2.4.1 Calibración del modelo PLS.....	99
3.2.4.2 Validación Cruzada del modelo PLS.....	101
3.2.4.3 Análisis de muestras excluidas	103
4. CONCLUSIONES	106
5. RECOMENDACIONES.....	108
6. REFERENCIA BIBLIOGRAFICA	109

LISTA DE TABLAS

	Pág.
Tabla 1. Clasificación de Crudos según densidad y gravedad API.....	21
Tabla 2. Composición elemental del petróleo	25
Tabla 3. Análisis elemental de las muestras.....	56
Tabla 4. Datos de absorción y desviación estándar para la prueba de repetibilidad	63
Tabla 5. Bandas vibracionales en el infrarrojo medio presentes en los fondos de vacío	66
Tabla 6. Vibraciones de flexión δ (C-H) aromático.....	69
Tabla 7. Calculo del CV para tres números de onda característicos aplicando diferentes pretratamientos	71
Tabla 8. Varianza de los 7 componentes principales de los 20 fondos de vacío ...	74
Tabla 9. Parámetros del modelo seleccionado para la predicción % p/p Carbono	81
Tabla 10. Validación cruzada del modelo de predicción % p/p Carbono	84
Tabla 11. Distribución % p/p Carbono en las muestras	85
Tabla 12. Parámetros del modelo seleccionado para la predicción de % p/p Hidrógeno	87
Tabla 13. Validación cruzada del modelo de predicción % p/p Hidrógeno.....	90
Tabla 14. Parámetros estadísticos del modelo desarrollado para la predicción de % p/p Nitrógeno	93
Tabla 15. Validación cruzada del modelo de predicción % p/p Nitrógeno.....	95
Tabla 16. Parámetros estadísticos del modelo desarrollado para la predicción % p/p Azufre.....	99
Tabla 17. Validación cruzada del modelo de predicción % p/p Azufre.....	102

LISTA DE FIGURAS

	Pág.
Figura 1. Nomenclatura y tipos de compuestos orgánicos de Azufre presentes en el petróleo	26
Figura 2. Nomenclatura y tipos de compuestos orgánicos de Nitrógeno presentes en el petróleo	28
Figura 3. Región infrarroja del espectro electromagnético.....	31
Figura 4. Espectro FTIR-PAS de un fondo de vacío	34
Figura 5. Esquema de una celda fotoacústica	35
Figura 6. Generación de la señal fotoacústica en un sólido.....	38
Figura 7. Interpretación geométrica de un PCA.....	51
Figura 8. Descripción gráfica del modelo de regresión PLS	53
Figura 9. Esquema metodológico	55
Figura 10. Sistema PAS acoplado al equipo IR Tensor 27 marca Bruker.....	57
Figura 11. Celda PAS modelo MTEC 200.....	58
Figura 12. Efecto de la relación señal ruido en un espectro de una muestra.....	60
Figura 13. Curva normal de error.....	61
Figura 14. Prueba de repetibilidad del espectrómetro FTIR-PAS	63
Figura 15. Espectros PAS de todas las muestras.....	64
Figura 16. Clasificación posibles vibraciones presentadas por los grupos funcionales en el infrarrojo.....	66
Figura 17. Espectros normalizados por máximos	72
Figura 18. Gráfica de Scores para PC1 Vs PC2 de los Fondos de vacío	75
Figura 19. Descomposición por componentes principales de la información espectral para PC1 y PC2. Gráfica de Score con Hotelling T^2	76
Figura 20. Gráfico de X-Loadings de PC1 y PC2.....	77
Figura 21. Gráfica de Influencia para PC1	79
Figura 22. Grafica de Leverage para PC2 para el modelo PCA	80

Figura 23. Varianza explicada en el modelo PLS para % p/p Carbono	81
Figura 24. Error calculado en función del número de componentes para el modelo PLS % p/p Carbono	82
Figura 25. Coeficiente de regresión para el primer componente principal PLS % p/p Carbono	83
Figura 26. Curva de calibración y validación del modelo de predicción % p/p Carbono	84
Figura 27. Varianza residual de la muestra para el modelo % p/p Carbono	86
Figura 28. Determinación de valores atípicos para el modelo % p/p Carbono.....	87
Figura 29. Varianza explicada en el modelo PLS para % W Hidrógeno	88
Figura 30. Error calculado en función del número de componentes para el modelo PLS % W Hidrógeno	88
Figura 31. Coeficiente de regresión para el primer componente principal PLS % W Hidrógeno	89
Figura 32. Curva de calibración y validación del modelo de predicción % p/p Hidrógeno	91
Figura 33. Gráfica de Influencia para modelo predicción % p/p Hidrógeno	92
Figura 34. Varianza explicada de la muestra para el modelo % p/p Hidrógeno.....	92
Figura 35. Varianza explicada en el modelo PLS para % W Nitrógeno	94
Figura 36. Error calculado en función del número de componentes para el modelo PLS % W Nitrógeno	94
Figura 37. Coeficiente de regresión para el primer componente principal PLS % p/p Nitrógeno	95
Figura 38. Curva de calibración y validación del modelo de predicción % p/p Nitrógeno	96
Figura 39. Varianza explicada de la muestra para el modelo % p/p Nitrógeno.....	97
Figura 40. Muestras atípicos para el desarrollo del modelo % p/p Nitrógeno	98
Figura 41. Gráfica de Influencia para modelo predicción % p/p Nitrógeno	98
Figura 42. Varianza explicada en el modelo PLS para % p/p Azufre.....	100
Figura 43. Error calculado en función del número de componentes para el modelo PLS % p/p Azufre.....	100

Figura 44. Coeficiente de regresión para el primer componente principal PLS % p/p Azufre.....	101
Figura 45. Curva de calibración y validación del modelo de predicción % p/p Azufre	103
Figura 46. Varianza explicada de la muestra para el modelo % p/p Azufre.....	104
Figura 47. Muestras atípicos para el desarrollo del modelo % p/p Azufre	104
Figura 48. Varianza explicada de la muestra modelo sin FV 8, 10, 12, 16 % p/p Azufre.....	105

RESUMEN

TÍTULO

PREDICCIÓN DEL ANÁLISIS ELEMENTAL POR ESPECTROSCOPIA FOTOACÚSTICA EN LA REGIÓN DEL INFRARROJO MEDIO Y MÉTODOS QUIMIOMÉTRICOS A LAS FRACCIONES DE LOS FONDOS DE VACÍO.*

AUTOR

CRISTIAN FERNEY LEGUIZAMO CASTELLANOS. **

PALABRAS CLAVES

FOTOACÚSTICA, FONDOS DE VACÍO, ANÁLISIS ELEMENTAL, PCA, PLS.

DESCRIPCIÓN

En el presente trabajo se desarrollaron cuatro modelos matemático – estadístico que permiten determinar el análisis elemental (% p/p Carbono, Hidrógeno, Nitrógeno y Azufre) en los fondos de vacío y sus fracciones, empleando la espectroscopia fotoacústica (FTIR-PAS) en la región del infrarrojo medio. La obtención de veinte muestras de se llevó a cabo utilizando una celda fotoacústica MTEC 200 que se acopló a un espectrómetro IR 27 marca Bruker. Como parámetros del equipo para la toma de los espectros se utilizó una resolución de 8cm^{-1} , ganancia del amplificador de 1, velocidad del interferómetro de 0.0583 cm/s y 256 barridos.

El trabajo consistió en construir modelos matemáticos a partir de los espectros fotoacústicos de las muestras de fondos de vacío para predecir al análisis elemental de los mismos. Se utilizó análisis de componentes principales (PCA) para reducir la dimensionalidad de los datos espectrales y luego regresión de mínimos cuadrados parciales (PLS) para generar los modelos predictivos del análisis elemental por medio de correlaciones lineales de las matrices de los datos espectroscópicos y los valores teóricos del análisis elemental.

La validación mostró resultados satisfactorios para la predicción del análisis elemental de los fondos de vacío. Para cada elemento se obtuvieron errores estándar de predicción (SEP) de 0.16, 0.28, 0.04 y 0.07. En todos los casos el coeficiente de correlación (R^2) entre los valores de referencia y predichos por los modelos fue superior a 0.96.

* Trabajo de Grado

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

SUMARY

TITLE

PREDICTION OF ELEMENTAL ANALYSIS BY PHOTOACOUSTIC SPECTROSCOPY IN THE REGION OF MIDDLE INFRARED AND CHEMOMETRIC METHODS FRACTIONS OF A VACUUM FUNDS. *

AUTHOR

CRISTIAN FERNEY LEGUIZAMO CASTELLANOS. **

KEY WORDS

PHOTOACOUSTIC, VACUUM FUNDS, ELEMENTAL ANALYSIS, PCA, PLS.

DESCRIPTION

In the present work four mathematical models were developed - allowing statistical determine the elemental analysis (% p/p Carbono, hydrogen, nitrogen and sulfur) in the bottom of innovation and its fractions, using photoacoustic spectroscopy (FTIR-PAS) in the mid-infrared region. The collection of twenty samples was carried out using a MTEC 200 photoacoustic cell is coupled to an IR spectrometer Bruker 27 mark. The parameters of the equipment for making the spectra we used a resolution of 8 cm⁻¹, amplifier gain of 1, interferometer speed 0.0583 cm / s and 256 scans.

The work consisted of constructing mathematical models based on the photoacoustic spectra of samples of vacuum funds for predicting the elemental analysis thereof. We used principal components analysis (PCA) to reduce the dimensionality of the spectral data and then partial least squares regression (PLS) to generate predictive models of elemental analysis by linear correlations of the spectroscopic data matrices and the values theoretical elemental analysis.

The validation showed satisfactory results for predicting the elemental analysis of vacuum funds. For each element were obtained standard errors of prediction (SEP) of 0.16, 0.28, 0.04 and 0.07. In all cases the correlation coefficient (R^2) between the values of reference and predicted by the models was higher than 0.96.

* Work Degree

** Sciences Faculty. Chemistry School. Directreess: Enrique Mejía Ospino. Codirectreess: Jorge Armando Orrego.

INTRODUCCIÓN

En la actualidad las reservas de petróleo en el mundo están conformadas por un 30% de crudos livianos y 70% de crudos pesados, extrapesados y bitumen^[1]. Las reservas de crudos han venido disminuyendo en proporciones elevadas debido al aumento en la demanda de productos obtenidos por destilación atmosférica y al vacío, los cuales representan un menor costo de refinación frente a los crudos pesados y las fracciones pesadas, además estos poseen mayores concentraciones de metales y estructuras más condensadas y complejas que vuelven más denso y viscoso (densidad API < 22.3^o [2]) el crudo lo que genera un aumento de costos en la producción y refinación.

En Colombia se identificó que los crudos livianos han tenido una participación promedio en la producción total de petróleo del país del 25%. Es necesario destacar que los crudos livianos han venido disminuyendo su participación en la canasta de crudos extraídos en el país; en el 2006 la participación fue del 29,1%, mientras que en el 2009 esta cifra fue del 20,7%. Por ende los crudos pesados y los residuos de las destilaciones tales como los fondos de vacío (FV) han comenzado a ser estudiados en los últimos años por parte de la industria petrolera la cual ha puesto su interés, conocimiento y tecnología hacia las grandes reservas de crudo pesado (incluidos los fondos de vacío), para esto ha realizado investigaciones sobre técnicas que favorezcan su producción, disminuyan los costos y además nos permitan caracterizarlos y facilitar su refinación. Los estudios se han realizado por diferentes técnicas analíticas convencionales, las cuales han generado buenos resultados en los análisis, pero no contribuyen en otros aspectos como: los elevados tiempos de respuesta, altos costos de operación, y uso de solventes contaminantes.

Dentro de las estrategias de caracterización de este tipo de muestras pesadas a parte de la viscosidad, la gravedad API y el análisis SARA^[3], se tiene el análisis elemental (Carbono, Hidrógeno, Nitrógeno, Azufre y Oxígeno), este análisis se lleva a cabo mediante procedimientos estándar regidos por la ASTM (American Society Test and Materials).

Para evitar los inconvenientes generados por las normas ASTM se han implementado técnicas espectroscópicas de análisis tales como la espectroscopia fotoacústica en el infrarrojo medio (FTIR-PAS), la espectroscopía de fluorescencia inducida por láser (LIF), espectroscopia de plasma inducido por láser (LIBS) y la Resonancia Magnética Nuclear (RMN-H)^[4,5] entre otras, que brinden la información necesaria de las muestras de una manera satisfactoria.

Dentro de las técnicas de análisis químico instrumental empleadas en la industria, la espectroscopia fotoacústica en el infrarrojo medio (FTIR-PAS) ha probado ser muy eficaz para el estudio estructural, mostrando superioridad sobre otras técnicas analíticas convencionales, debido a que la muestra no es destruida durante el análisis y ésta no requiere tratamiento previo (disminuyendo los costos al no utilizar solventes), además los espectros FTIR-PAS se obtienen directamente, por lo que las dificultades presentadas con los anteriores métodos instrumentales se ven significativamente superados^[6,7]. Sin embargo la alta complejidad de los datos obtenidos por las técnicas espectroscópicas conlleva a la utilización de análisis multivariable, el cual permite avanzar en la determinación de las propiedades fisicoquímicas de las muestras. El análisis por componentes principales (PCA) y la regresión por mínimos cuadrados parciales (PLS) son técnicas empleadas en la quimiometría que permiten desarrollar modelos para la predicción de propiedades a partir de datos espectroscópicos^[8].

1. CONSIDERACIONES TEORICAS

1.1 PETROLEO GENERALIDADES

El Petróleo (o crudo, como también se le conoce) es una mezcla compleja de hidrocarburos líquidos insolubles en agua, compuesto en mayor medida de Carbono e Hidrógeno, que pueden incluir desde 1 hasta 60 átomos de Carbono, el número de átomos de Carbono y la forma en que están colocados dentro de las moléculas de los diferentes compuestos proporciona al petróleo diferentes propiedades físicas y químicas. Así tenemos que los hidrocarburos compuestos por uno a cuatro átomos de Carbono son gaseosos, los que contienen de 5 a 20 son líquidos, y los de más de 20 son sólidos a la temperatura ambiente. Además pueden contener en su estructura Azufre, Nitrógeno, Oxígeno y algunos elementos metálicos como níquel, Vanadio, Hierro y cobre en proporciones más pequeñas. Esta compleja mezcla es formada por la descomposición y transformación de restos animales y plantas que han estado enterrados a grandes profundidades durante varios siglos, sometidos a grandes temperaturas y presiones. El petróleo crudo ha sido uno de los grandes protagonistas de la revolución industrial del planeta, convirtiéndose en la base económica de diversos sectores como la industria y el transporte, los cuales dependen directamente de los productos que de él se derivan ^[9,10,11].

De los productos derivados del petróleo se obtienen un sin número de materias primas y sustancias diversas que son de usos muy variados en nuestra vida cotidiana. Todas ellas se extraen de los productos de la destilación del petróleo por medio de procesos químicos muy elaborados. En términos muy generales a estos procesos se les denomina la química del petróleo o petroquímica. De hecho, uno de los grandes dilemas que se le presentan a la humanidad es decidir si

seguir quemando el petróleo como combustible o utilizarlo para la fabricación de materias petroquímicas, las cuales tienen una utilidad práctica enorme.

El petróleo presenta gran variación en diversos parámetros como color, densidad, capacidad calorífica, viscosidad, número de octano, tensión superficial, curvas de destilación entre otras; estas variaciones se deben a la diversidad de concentraciones de los hidrocarburos que componen la mezcla ^[12].

Se considera que un crudo promedio contiene aproximadamente de 84 a 87 % de Carbono, 11 a 14 % de Hidrógeno, entre 0 a 5 % de Azufre y menos del 1 % de Nitrógeno, Oxígeno, trazas de metales y sales ^[13]. Estas diferencias en composición influyen de manera apreciable en los aspectos de diseño y localización de las plantas de refinación, en la determinación de los procesos de conversión y tratamiento requeridos y en la producción de los derivados de mayor demanda e importancia económica.

1.2 CLASIFICACION DEL PETROLEO

La presencia en diversas cantidades de cada uno de los elementos químicos (orgánicos e inorgánicos) que componen el petróleo, determinan sus características particulares las que nos permiten clasificarlo de diferentes maneras. Los tipos de petróleo pueden ser determinados de distintos modos en función al criterio que se desee considerar como predominante, siendo los más comunes:

1.2.1 Composición química

Este tipo de clasificación se basa en un factor desarrollado por la Universal Oil Products (UOP), conocido como factor UOP (identificado con la letra k), el cual se

determina por medio de la relación $K=TB/S$ donde TB es el punto de ebullición molar promedio en grados Rankine ($^{\circ}F$ absoluto) y S es el peso específico a $60^{\circ}F$.

El cual para los crudos parafínicos fluctúa entre 12.15 y 12.9, los mixtos se sitúan entre 11.5 y 12.1. Finalmente los nafténicos presentan un rango entre 10.5 y 11.5. Esta clasificación depende estrictamente de la presencia de ciertos componentes químicos en el petróleo, así como de la unión de éstos en elementos más complejos. Su importancia radica en las características particulares que cada uno de estos elementos le añade al petróleo. Así tenemos que se puede clasificar en:

Parafínico: formado principalmente por parafinas (cadenas lineales de Carbono asociadas a Hidrógeno, las cuales pueden ser lineales o ramificadas, en si son hidrocarburos saturados). Son muy fluidos y de color claro. Al ser incorporados al proceso de destilación producen abundante parafina y poco asfalto. Proporcionan una mayor cantidad de nafta y aceites lubricantes.

Nafténicos: siendo sus componentes principales los naftenos (moléculas en las que se forman ciclos de Carbono) y los hidrocarburos aromáticos. Son petróleos muy viscosos y de coloración oscura. Generan una gran cantidad de residuo de asfalto tras el proceso de refinación.

Mixtos: De composición de bases intermedias, formados por toda clase de hidrocarburos: Saturados, no saturados (etilénicos y acetilénicos) y cíclicos (ciclánicos o nafténicos y bencénicos o aromáticos). La mayoría de los yacimientos mundiales son de este tipo.

1.2.2 Densidad

La clasificación más sencilla, pero no menos importante en cuanto a los resultados económicos, es la clasificación en crudos pesados y ligeros. Al estar formado

principalmente por moléculas hidrocarbonadas, la densidad de un crudo será tanto menor cuanto mayor sea la relación atómica C/H. La densidad de los crudos puede oscilar entre 0,7 y 1, expresándose con mucha frecuencia en grados API (American Petroleum Institute, ecuación 1) cuyo valor varía entre 70 y 5; esta variabilidad de la densidad es consecuencia de composiciones en familias químicas muy diferentes ^[14].

La densidad es la relación entre el peso de un determinado volumen de muestra a una temperatura t y el peso del mismo volumen de agua a una temperatura determinada. El grado A.P.I., es definido como una función hiperbólica de la densidad. Esta unidad corresponde al antiguo grado Baumé. Los grados API se basan en la comparación de la densidad del petróleo con la densidad del agua, es decir, se busca determinar si el petróleo es más liviano o pesado que ésta última. La clasificación propuesta por el Instituto de Petróleo Americano (API) indica que a una mayor gravedad API el petróleo será más liviano, como se puede ver en la Tabla 1. Cabe indicar que los petróleos ligeros son también los más apetecidos en el mercado, y al mismo tiempo los de mayor precio, debido a que tienen menor viscosidad, escasa tendencia aditiva y alta tendencia a emulsificarse, por tanto los costos de extracción como de refinación son menores en comparación con petróleos pesados. Así, se da una relación directa entre la gravedad API y la calidad del petróleo, petróleos más ligeros requieren de menores costos para ser aprovechados que aquellos más pesados. Así, un crudo con una baja gravedad específica presentará un alto valor comercial debido al mayor porcentaje de fracciones livianas, que de él pueden extraerse, entre las cuales se encuentran las gasolinas ^[11].

$$^{\circ} API = \frac{141.5}{Gravedad\ Especifica} - 131.5 \quad Ecuación\ (1)$$

Los crudos están constituidos por mezclas de un número muy elevado de componentes puros, aumentando la dificultad de la descripción de las distintas fracciones. En términos de la densidad API, los crudos ligeros o de baja gravedad

específica presentan alto dicho valor. Los crudos con bajo contenido de Carbono, alto contenido de Hidrógeno y alta densidad API son generalmente ricos en hidrocarburos parafínicos y tienden a producir mayores cantidades de gasolina y productos ligeros; aquellos crudos con alto contenido de Carbono, bajo contenido de Hidrógeno y baja densidad API son ricos en hidrocarburos nafténicos y aromáticos. De acuerdo al contenido de compuestos con Azufre, el crudo puede clasificarse como agrio, si presenta cantidades apreciables de estos compuestos o como dulce, si presenta cantidades muy pequeñas.

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

Tabla 1. Clasificación de Crudos según densidad y gravedad API

Aceite crudo	Densidad (g/cm ³)	Gravedad API
Extrapesado	> 1.0	1
Pesado	1.0 - 0.92	10.0
Mediano	0.92 - 0.87	22.3 - 31.1
Ligero	0.87 - 0.83	31.1 – 39
Superligero	< 0.83	> 39

Fuente: WAUQUIER, J. P. El refinado del petróleo. Petróleo crudo, productos petrolíferos, esquemas de fabricación ^[15].

1.2.3 Contenido Azufre

El Azufre es uno de los componentes que están presentes en los hidrocarburos (así sea en pequeños porcentajes). Pero su presencia en los hidrocarburos implica la necesidad de mayores procesos de refinamiento, y por ende un mayor costo final, razón por la cual la presencia de Azufre es también un determinante del valor comercial del petróleo. Así, tenemos que el petróleo puede clasificarse de 2 formas dependiendo la cantidad de Azufre:

- **Petróleo Dulce (Sweet Crude Oil)**, es aquel que contiene menos de 0.5% de contenido sulfuroso, es decir, con presencia de Azufre. Es un petróleo de alta calidad y es ampliamente usado para ser procesado como gasolina.
- **Petróleo Agrio (Sour Crude Oil)**, es aquel que contiene al menos 1% de contenido sulfuroso en su composición. Debido a la mayor presencia de Azufre su costo de refinamiento es mayor, razón por la cual es usado mayormente en productos destilados como el diesel, dado su menor costo de tratamiento.

Las tres formas generales de categorización mencionadas anteriormente permiten establecer criterios básicos para determinar la calidad del petróleo, los cuales influirán finalmente en la determinación de los precios de cada uno de ellos.

1.3 ANALISIS COMPOSICIONAL

Una forma de definir un crudo es conociendo la composición y estructura de todos los compuestos presentes. Sin embargo, debido a la gran variedad y al número elevado de componentes que le forman, conocer la composición exacta del crudo es casi imposible.

Un número de técnicas de ASTM se utilizan comúnmente para evaluar la calidad de los carburantes, entre dichas técnicas se encuentra la clasificación total (Análisis Composición) del tipo de crudo. El análisis de mezclas de hidrocarburos es importante en la determinación de las condiciones de operación de la refinería. Además de estos datos es la composición de hidrocarburos la que determina la selección de catalizadores adecuados y las operaciones de fusión. Los grupos de hidrocarburos principales presentes en la nafta y gasolina son las parafinas, olefinas, naftenos y aromáticos (PONA) ^[16].

Parafinas: Nombre común para un grupo de hidrocarburos alcanos normales. La molécula más simple de parafina es el metano (gas a temperatura ambiente). Más pesados miembros de la serie, incluyen octano (líquido a temperatura ambiente) y la cera de parafina (sólidas a temperatura ambiente).

Aromáticos: Un compuesto aromático es aquel que contiene uno o más anillos de benceno. Comúnmente los compuestos aromáticos de benceno incluyen tolueno y naftaleno.

Naftenos: También llamados cicloalcanos, que son similares a los alcanos en sus propiedades físicas, pero tienen mayores puntos de ebullición y densidades. Los cicloalcanos se clasifican en cicloalcanos pequeños (ciclopropano y ciclobutano), normal (ciclopentano a cicloheptano) y grandes con más de 20 átomos de Carbono los cuales se suelen llamar cicloparafinas.

Olefinas: También llamados alquenos. Poseen en su estructura dobles enlaces, la forma más simple de olefina es el etileno (también llamado eteno) se caracterizan por su mayor reactividad frente a las parafinas.

El análisis composicional es una de las características más importantes de las fracciones. Basados en la naturaleza del crudo existen varias maneras de expresar la composición de un crudo. Algunas de las más importantes son ^[17]:

- PONA (parafinas, olefinas, naftenos y aromáticos)
- PNA (parafinas, naftenos y aromáticos)
- PIONA (parafinas, isoparafinas, olefinas, naftenos y aromáticos)
- SARA (saturados, aromáticos, resinas y asfáltenos)
- Análisis elemental (C, H, N, O, S)

Ya que la mayoría de las fracciones son libres de olefinas la composición puede ser expresada en términos del análisis PNA. Este tipo de análisis es útil en cargas livianas con rangos de ebullición estrechos; mientras que el análisis SARA es útil para fracciones pesadas y residuos que contienen un alto contenido de resinas y asfaltenos^[18].

1.3.1 Análisis elemental

El porcentaje de Carbono, Hidrógeno, Nitrógeno, Oxígeno y Azufre es quizá el primer paso para examinar la naturaleza general y la calidad de una carga. Las relaciones atómicas de varios elementos con respecto al Carbono (H/C, N/C, O/C y S/C) son usadas con frecuencia para este propósito, así como también la concentración de elementos traza como Hierro, Vanadio y Níquel. Los métodos propuestos para el análisis elemental son:

Contenido de Hidrógeno: ASTM D1018, ASTM D3178, ASTM D3343, ASTM D3701, y ASTM E777

Contenido de Nitrógeno: ASTM D3179, ASTM D3228, ASTM D3431, ASTM E148, ASTM E258, and ASTM E778

Contenido de Azufre: ASTM D124, ASTM D1266, ASTM D1552, ASTM D1757, ASTM D2662, ASTM D3177, ASTM D4045 y ASTM D4294^[18,19].

La Norma ASTM D5291, Esta es la primera norma de la ASTM que cubre la determinación simultánea de Carbono, Hidrógeno y Nitrógeno en los productos derivados del petróleo y lubricantes.

Los análisis de Carbono, Hidrógeno y Nitrógeno son útiles para determinar la naturaleza compleja de los tipos de muestra. Los resultados CHN se pueden

utilizar para estimar el potencial de procesamiento, refinación y producción en la industria petroquímica.

Con pocas excepciones, las proporciones de los elementos (Carbono, Hidrógeno, Nitrógeno, Oxígeno, Azufre y metales) en el petróleo (sin importar la fuente) varían en un estrecho límite. Ver tabla 2.

Tabla 2. Composición elemental del petróleo

Componente	Presencia
Carbono	83,0 a 87,0 % W
Hidrógeno	10,0 a 14,0 % W
Nitrógeno	0.1 a 2.0 % W
Oxígeno	0.05 a 1.5 % W
Azufre	0.005 a 6.0 % W
Metales (Ni y V)	< 1000 ppm

Fuente: SPEIGH, J. The Chemistry and Technology of Petroleum ^[19].

1.3.1.1 Contenido de Azufre

El Azufre elemental es un componente común de los sedimentos y, si están presentes en las rocas de los yacimientos, se disuelve en el petróleo crudo y reacciona lentamente para producir compuestos de Azufre o varios sulfuro de Hidrógeno, que puede reaccionar más fácilmente con ciertos componentes del petróleo. Estas reacciones se presume que son en gran parte las responsable por el contenido de Azufre de un petróleo, dichas reacciones están acompañadas por un oscurecimiento del aceite y un aumento significativo en el peso específico y viscosidad de los crudos ^[19].



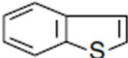
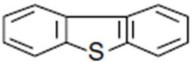
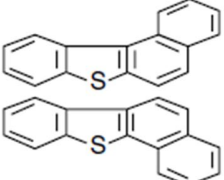
Todos los petróleos tienen compuestos de Azufre, desde 0,03% en los crudos de Pennsylvania hasta 5,0% en los crudos mexicanos. Los compuestos de Azufre

determinan la corrosividad potencial del crudo. En las naftas presentan dificultades en cuanto a corrosión, olor y explosiones deficientes de las mismas.

Han sido aislados una gran cantidad de derivados de Azufre incluyendo Azufre elemental, sulfuro de Hidrógeno, mercaptanos, tioéteres disulfuros y tiofenos. Los compuestos de Azufre son complejos y generalmente térmicamente inestables. Los cíclicos como el tiofeno son más estables. Se eliminan craqueándolos durante el proceso de refinación en el cual se forma SH_2 y compuestos orgánicos simples de Azufre.

Ciertos compuestos de Azufre que son corrosivos por su acidez como SH_2 y mercaptanos, se eliminan con tratamientos químicos. Los compuestos de Azufre no ácidos se transforman en SH_2 durante el tratamiento.

Figura 1. Nomenclatura y tipos de compuestos orgánicos de Azufre presentes en el petróleo

RSH	Thiols (Mercaptans)
RSR'	Sulfides
	Cyclic Sulfides
RSSR'	Disulfides
	Thiophene
	Benzothiophene
	Dibenzothiophene
	Naphtobenzothiophene

Fuente: SPEIGH, J. the Chemistry and Technology of Petroleum ^[19].


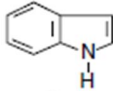
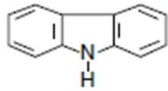
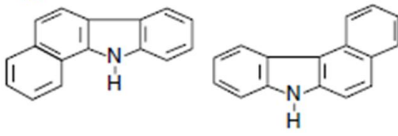

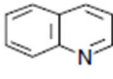
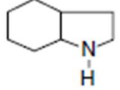
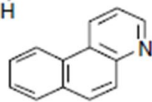
1.3.1.2 Contenido de Nitrógeno

El contenido de Nitrógeno en la mayoría de los crudos existentes, es relativamente bajo, generalmente menor que 0,1% en peso aumentando en las fracciones de alto punto de ebullición.

Hay varios tipos de hidrocarburos nitrogenados y son considerablemente más complejos que los compuestos hidrocarbonados de Azufre. Son completamente estables térmicamente y por esta razón en los cortes livianos solo aparecen trazas de Nitrógeno. Se cree que son derivados de las proteínas presentes en los materiales a partir de los cuales se formó el crudo. Una pequeña cantidad presente en el gasoil de la alimentación para cracking catalítico tiene un efecto adverso sobre la selectividad de la catálisis.

La separación de los diferentes tipos de compuestos es dificultosa y los mismos son susceptibles de alteración y pérdidas por manipulación. Según sus características se los clasifica en básicos y no básicos. Para reducir su presencia en la carga a procesos catalíticos se efectúa hidrotratamiento. En los petróleos crudos no existen compuestos básicos de N_2 , estos se forman por descomposición en la destilación.

Figura 2. Nomenclatura y tipos de compuestos orgánicos de Nitrógeno presentes en el petróleo

<i>Nonbasic</i>		
Pyrrole	C_4H_5N	
Indole	C_8H_7N	
Carbazole	$C_{12}H_9N$	
Benzo(a)carbazole	$C_{16}H_{11}N$	
<i>Basic</i>		
Pyridine	C_5H_5N	
Quinoline	C_9H_7N	
Indoline	C_8H_9N	
Benzo(f)quinoline	$C_{13}H_9N$	

Fuente: SPEIGH, J. the Chemistry and Technology of Petroleum ^[19].

La concentración de Nitrógeno en los fondos de vacío es una medida de la presencia de aditivos que contienen Nitrógeno. El conocimiento de su concentración puede ser utilizado para predecir el rendimiento. Algunos productos derivados del petróleo también contienen Nitrógeno de origen natural.

Conocimiento del contenido de Hidrógeno en las muestras es útil en el tratamiento de sus características de rendimiento. Relación Hidrógeno-Carbono es útil para evaluar el desempeño de los procesos de modernización.

1.3.1.3 Contenido de Oxígeno

Al igual que los compuestos de Nitrógeno, estos son generalmente más complejos que los de Azufre. En su mayor parte son ácidos carboxílicos, fenoles y cresoles, con menores cantidades de sustancias no ácidas como ésteres, amidas, cetonas y benzofuranos.

El total de ácido contenido en el petróleo varía desde 0.03 % (petróleo iraquí y egipcio) a 3% (algunos petróleos de Estados Unidos). El contenido total de Oxígeno del crudo es por lo general menos del 2 % en peso, cuando el contenido de Oxígeno es extraordinariamente elevado o se encuentra muy por encima de los valores estándar, puede ser que el petróleo ha sufrido una exposición prolongada a la atmósfera, ya sea durante o después de la producción^[19].

Los ácidos nafténicos que representan la acidez primaria del crudo, son de considerable importancia petroquímica. Los que se encuentran en nafta pesada, kerosene y gas oil, se destruyen en el cracking. Los que se encuentran en la fracción gas oil son cadenas de alquil- cicloparafinas, se eliminan como naftenato de sodio por extracción con soda caustica.

La presencia de compuestos de Oxígeno en la alimentación a procesos catalíticos, no presenta mayores problemas. Los asfaltos contienen algunos compuestos altamente oxigenados que son fácilmente separados del crudo y sus fracciones.

1.4 ESPECTROSCOPIA INFRARROJA

1.4.1 Historia de la técnica

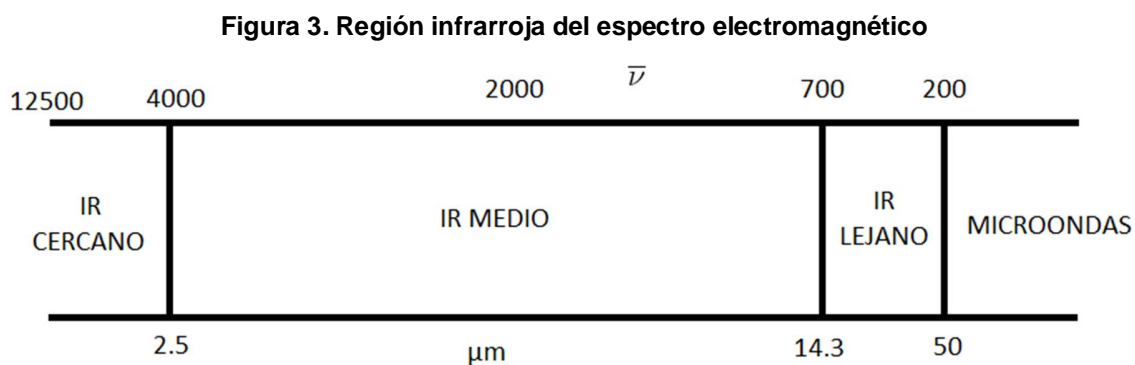
La radiación Infra Roja (IR) fue descubierta por William Herschel a comienzos del siglo XIX. Las dificultades en construir un detector IR adecuado impidieron cualquier aplicación. Solo hasta llegar a los años 30 del siglo XX fue cuando se vio el potencial de la espectroscopía IR en el campo de la química, debido a la construcción de los primeros equipos IR. Las necesidades analíticas asociadas con la producción comercial de goma sintética estimularon el desarrollo comercial de la espectrometría IR y estudios básicos en este campo. Hasta los años 60 se utilizaron equipos dispersivos que utilizaban monocromadores de prisma o red. Un avance considerable se alcanzó con la introducción de la técnica de transformada de Fourier (FT) [20].

Este procedimiento está basado en el interferómetro de Michelson (desarrollado inicialmente para determinar con exactitud la velocidad de la luz) y en el método del matemático francés Fourier que permite convertir la información obtenida (interferograma) en un espectro. El método de la transformada de Fourier fue utilizado en un principio por los astrónomos (década del 50) para aislar señales débiles, procedentes de estrellas distantes, del ruido de fondo ambiental. El rápido desarrollo de las tecnologías de los láseres y la computación son responsables del predominio actual de los equipos FT, rápidos y sensibles, especialmente exitosos en su combinación con las técnicas cromatográficas. La micro-espectroscopía IR resurge con fuerza mediante la tecnología FT-IR a partir de 1983, convirtiéndose en un procedimiento eficaz en varios campos de la química [18, 20].

La espectroscopia infrarroja la cual es una técnica muy empleada en la industria por parte de los químicos debido a su fácil manejo, bajo mantenimiento de equipo, poco tiempo empleado en el análisis y evaluación de la muestra [21].

1.4.2 Regiones espectrales

La región infrarroja del espectro electromagnético se extiende entre la zona del visible y la de las microondas, tal como se muestra en la Figura 3.



Fuente:[(en línea)
http://www.fq.uh.cu/dpto/qf/docencia/pregrado/estruc_2/ir/descargas/1_ir_tecnica.pdf [citado el 10 de diciembre de 2010][20].

La región de mayor utilidad práctica de la extensa región IR es la que se extiende entre 4000 y 700 cm^{-1} denominada región infrarroja media (MIR), donde tienen lugar las vibraciones fundamentales; es una de las técnicas analíticas disponibles más importantes para conseguir información sobre aspectos cualitativos y cuantitativos de analitos en tiempo real. La utilización de la región IR lejana (Far Infrared, FIR), entre 700 y 200 cm^{-1} , se producen las absorciones debido a cambios rotacionales, por ende se ha ampliado considerablemente en los últimos tiempos, sobre todo para el estudio de compuestos órgano-metálicos o inorgánicos (átomos pesados, enlaces débiles), debido a las vibraciones de

tensión y flexión de los enlaces entre átomos metálicos y ligandos inorgánicos u orgánicos tienen lugar generalmente, a frecuencias inferiores a 650 cm^{-1} . La región IR cercana (Near Infrared, NIR), entre 12500 y 4000 cm^{-1} , accesible a la óptica de cuarzo, donde se presentan las bandas armónicas, en esta región se producen absorciones debidas a sobretonos y combinaciones de las bandas vibracionales de tensión fundamental que se producen en la región de 3000 a 1700 cm^{-1} , los enlaces implicados son por lo general C-H, N-H, O-H; ha sido utilizada para determinaciones cuantitativas de rutina de especies tales como agua, proteínas, hidrocarburos de bajo peso molecular, grasas en productos agrícolas, petróleo e industria química ^[17,20].

1.4.3 Análisis infrarrojo

El análisis infrarrojo es una de las herramientas más importantes para observar espectros vibracionales. Un espectro IR proporciona los primeros datos para interpretar la estructura molecular de una muestra ya que proporciona información de la aromaticidad, de la presencia de unidades estructurales básicas y grupos funcionales, así como información sobre el contenido y la naturaleza del material mineral presente ^[22].

Aunque el espectro IR no caracteriza a cada compuesto, si brinda información de que ciertos enlaces atómicos que dan lugar siempre a bandas en un determinado intervalo de frecuencias, independiente de la naturaleza del resto de la molécula. La presencia de estas bandas, características de grupos funcionales, permite una amplia utilización de la espectroscopía IR en la determinación estructural. Sin embargo pueden presentarse problemas con la reproducibilidad de los espectros por errores aleatorios que están ligados a la preparación de las muestras, el operario, etc ^[20].

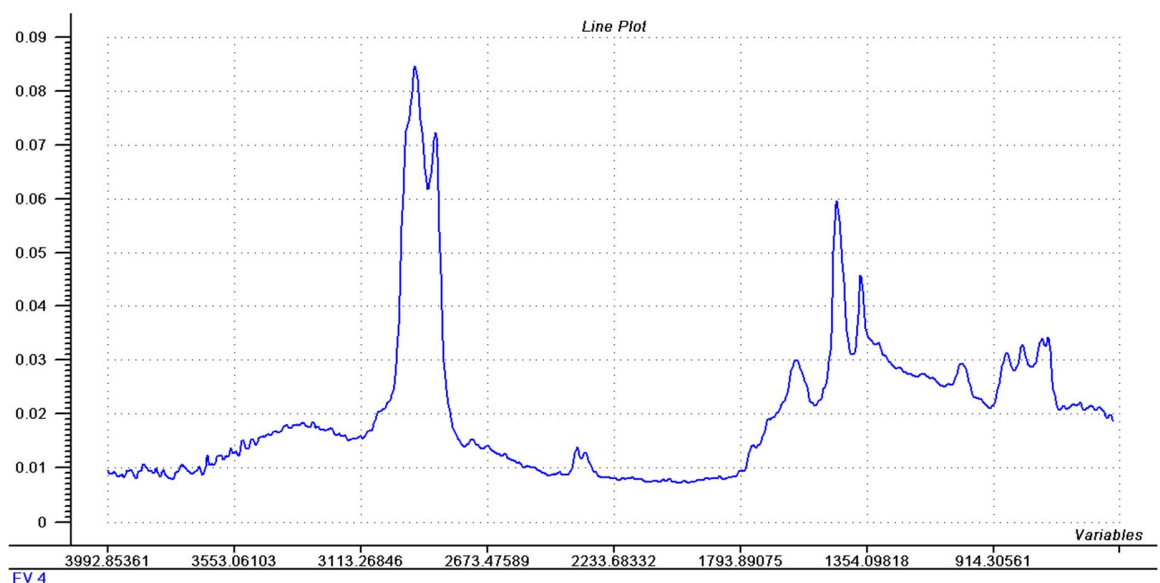
La espectroscopía infrarroja juega un papel importante en el análisis de combustibles para caracterizar cualitativamente sus diversos componentes. Con el desarrollo de la espectroscopía IR con transformada de Fourier (FT-IR), se crea la posibilidad de desarrollar cuantitativamente estos métodos de análisis que son útiles para muestras complejas; debido a los avances en las técnicas de manipulación de muestras y la obtención de mayor información del analito conlleva al uso de nuevos métodos quimiométricos para facilitar el análisis cuantitativo ^[23]. Por tal motivo, es indispensable la aplicación de la transformada de Fourier (FT), que como se menciona anteriormente fue desarrollada para mitigar las limitaciones encontradas con los instrumentos dispersivos. De este modo, la técnica FT ha conseguido para la espectroscopía infrarroja una significativa ventaja práctica, además ha hecho posible el desarrollo de nuevas técnicas, como FTIR-CA (de camino abierto), FTIR-PAS (espectroscopía fotoacústica) y FTIR-ATR (de reflectancia total atenuada) que fueron diseñadas para solucionar los desafíos que eran planteados a la técnica ^[18].

La espectroscopia infrarroja (IR) puede ser utilizada para estudiar una amplia variedad de sustancias. Pero presenta como desventaja el tratamiento previo que se le debe realizar a la muestra antes de ser tomado los espectros (preparación de las pastilla de KBr en caso de que las muestras sean solidas), con la preparación de las pastillas de KBr se disminuye la densidad óptica de la muestra de modo que la radiación no es absorbida totalmente y puede detectarse la radiación transmitida^[22], este fenómeno introducen errores en la medida debido a la absorción de agua por parte del KBr, afectando la región espectral donde aparecen las señales; para poder solucionar estos inconvenientes se utiliza la espectroscopia infrarroja acoplada con la PAS (FTIR-PAS), donde la muestra no necesita ningún tratamiento y el análisis se realiza con escasa o nula manipulación de la misma, constituyendo de este modo un método confiable.

1.4.3.1 Espectroscopia Fotoacústica

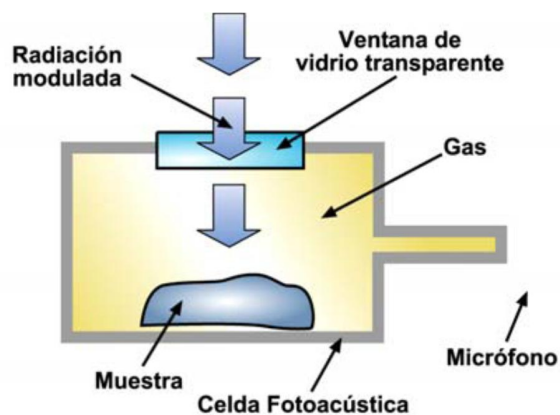
El efecto fotoacústico fue descubierto por Alexander Graham Bell hacia 1880, quien observó que se producían sonidos audibles en muestras ópticamente absorbentes, cuando estas se exponían a la luz solar modulada [22]. El efecto fotoacústico (PA), consiste en la generación de ondas acústicas en una muestra, debido a la absorbancia de radiación electromagnética pulsada o modulada en la fracción no radiactiva (producción de calor). Este efecto es causado por el calentamiento de la muestra después de haber absorbido energía óptica (efecto fototérmico) correspondiente a la energía liberada por el decaimiento desde un estado excitado al estado fundamental y esto da lugar a variaciones de presión que son detectadas por un transductor acústico (micrófono o cristal piezoeléctrico) [24], esta técnica proporciona espectros similares a los de absorción óptica en cualquier tipo de material condensado o gases figura 4. Además, puesto que sólo la luz absorbida se convierte en sonido, la luz dispersada no genera interferencias en las señales.

Figura 4. Espectro FTIR-PAS de un fondo de vacío



La detección de la señal puede ser directa, cuando la onda acústica se produce y detecta en la muestra; o indirecta cuando la onda acústica se genera en un medio de acoplamiento (que puede ser un gas o líquido) adyacente a la muestra, y se debe a la transferencia de calor entre la muestra y el medio de acoplamiento, y a la transmisión acústica desde la muestra figura 5.

Figura 5. Esquema de una celda fotoacústica



Aunque el efecto FA en sólidos ganó el interés de algunos investigadores, como Röntgen, Tyndall y (Lord) Rayleigh, permaneció como una curiosidad científica por casi medio siglo hasta que, gracias en gran medida al desarrollo del micrófono, que sustituiría al tubo de escucha en el montaje experimental de Bell, a la fabricación de fuentes luminosas intensas (fuentes de radiación laser de gran potencia) y al desarrollo de detectores acústicos de alta sensibilidad (amplificadores sensitivos a fase o sincrónicos), además la capacidad de procesar datos, comenzaron a gestarse las primeras aplicaciones prácticas^[25].

La diferencia fundamental entre la PAS y la espectroscopia de absorción convencional, es la forma en que se detecta la absorción. En la espectroscopia convencional esta medición es indirecta, en realidad se detecta la radiación transmitida y posteriormente se realiza la conversión a unidades de absorbancia.

Las muestras con bajo coeficientes de absorción, β , difícilmente pueden ser analizadas por esta espectroscopia porque la magnitud de la señal transmitida es cercana a la señal de referencia, y es difícil discriminar entre las dos señales. Lo contrario ocurre con muestras con alto valor de β , porque la magnitud de la señal transmitida es muy pequeña comparada con la señal de referencia y fácilmente se pierde en el ruido de fondo. El rango de detección lineal de este tipo de experimentos, está limitado a las muestras con un rango medio de β .

Mientras que, la absorción de la radiación electromagnética en espectroscopia fotoacústica se detecta con un transductor acústico (micrófono o cristal piezoeléctrico), que produce una señal eléctrica cuya amplitud pico a pico, es directamente proporcional a la magnitud de los cambios de presión en la muestra. Por lo tanto, el rango de detección lineal de este tipo de experimento es muy bueno para las muestras con valores altos de β , debido a que el cambio inducido de presión es directamente proporcional a la energía absorbida.

El modelo normal que describe la señal PA de una muestra sólida, se propuso por Rosencwaig y Gersho ^[26]. Estos autores, pudieron demostrar que la generación de la señal acústica se produce por el flujo de calor periódico entre la muestra y el gas de la cámara cerrada. Cuando la energía radiante incide sobre una muestra sólida y opaca esta absorbe energía y parte de ella es transformada en calor. Si la fuente de energía es modulada, una variación periódica de temperatura se producirá en la muestra.

Rosencwaig y Gersho ^[26], se basaron en la opacidad, la difusividad térmica y el grosor total de la muestra, para explicar el fenómeno acústico para muestras ópticamente opacas, la magnitud de la señal fotoacústica, q , y la fase \emptyset , se pueden expresar como se muestra en las ecuaciones 2 a 4.

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

$$\phi = \tan^{-1} \left(1 + \frac{2}{\beta \mu_s} \right) \quad \text{Ecuación 3}$$

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

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

La señal fotoacústica es proporcional a la intensidad de la radiación, inversamente proporcional a la temperatura y volumen de la celda, y dependiente de las propiedades termofísicas de la muestra y del gas circundante. El coeficiente de absorción óptica, β , propio de cada material, se define como el inverso de la profundidad en donde el 63% de la radiación incidente es absorbida. Mientras que, μ , es la profundidad a la cual la energía absorbida constituye un 63% de la señal fotoacústica. Esto está dado por la ecuación 5

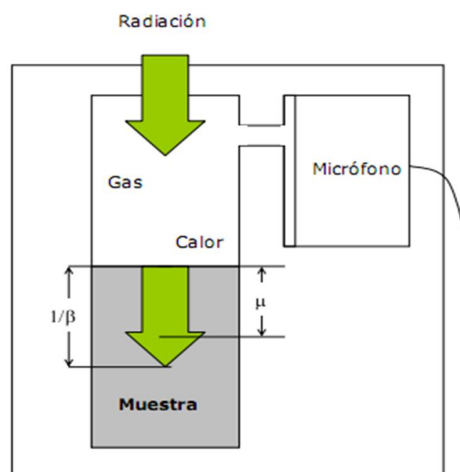
$$\mu = \left(\frac{\alpha}{\pi f} \right)^{1/2} \quad \text{Ecuación 5}$$

Donde α ($\alpha = \beta^{-1}$) es la difusividad térmica y está dada por la ecuación 6.

$$\alpha = \left(\frac{k}{\rho C_p} \right) \quad \text{Ecuación 6}$$

La señal fotoacústica se conduce desde μ hasta la superficie donde calienta una capa delgada del gas sobre la superficie de la muestra. La cantidad de calor transportada de este modo es dependiente en las características termofísicas de la muestra y del gas. La capa calentada del gas actúa como pistón acústico que se amplía y contrae contra el gas restante en la celda. Este proceso causa una oscilación periódica de la presión (onda acústica), que se propaga a través del volumen del gas. Un micrófono situado en el espacio del gas detecta esta señal y la retransmite a un amplificador. Este toma la señal y la amplifica rechazando las señales de altas frecuencias que normalmente corresponden al ruido electrónico. Un sistema de adquisición de datos recoge y registra la señal de salida del amplificador figura 6 [18, 22, 24].

Figura 6. Generación de la señal fotoacústica en un sólido



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

1.5 QUIMIOMETRIA

La quimiometría es un método que fue empezado a usar a finales de los años 70 para obtener información útil especialmente en la química analítica, donde se realizan diversos procesos que transforman datos complejos en información que se pueda comprender. La automatización y computarización de los laboratorios ha llevado consigo diversas consecuencias. Una de ellas es la rápida adquisición de gran cantidad de datos. Ahora bien, la posesión de dichos datos dista, muchas veces, de proporcionar respuestas adecuadas y significativas. La obtención de datos no es sinónimo de poseer información; debemos interpretarlos y colocarlos en el contexto adecuado para convertirlos en información útil para el usuario. La quimiometría es la disciplina que tiene esta finalidad. El desarrollo rápido de la quimiometría en química analítica se debe fundamentalmente a la utilización masiva de los ordenadores acoplados a la moderna instrumentación. Se ha pasado de hacer medidas puntuales a una longitud de onda (univariadas), a medidas espectrales (multivariadas) con gran contenido de información donde los métodos tradicionales de análisis y procesado de datos resultan totalmente insuficientes. La quimiometría presenta diferentes formas de análisis como son los métodos matemáticos, estadísticos y logísticos.

Los diferentes análisis químicos que se pueden realizar por estos métodos, son determinaciones cuantitativas y cualitativas, requiriendo diferentes herramientas para cada uno de los análisis.

Los métodos espectroscópicos pueden proporcionar datos analíticos sobre muchos componentes de una única muestra. Situaciones como estas, en que se miden varias variables para cada muestra, proporcionan datos multivariantes. En la actualidad, la calibración multivariable se ha empleado en el análisis de sistemas químicos complejos, incluyendo aguas residuales, Análisis de los

componentes principales de los alimentos por espectroscopia fotoacústica en infrarrojo medio y cercano^[26], maderas, formulaciones de medicamentos, aceites y combustibles, caracterización de materiales industriales, con énfasis especial en la producción de hidrocarburos^[27], teniendo la ventaja de no necesita la separación de los solutos antes de su análisis, como en el caso de los análisis por cromatografía. La calibración multivariable es un método de calibración efectivo en el cual la información química (absorción, emisión, transmisión, etc) de un conjunto de mezclas patrón, registrados con las diferentes variables (números de onda) están relacionados con la concentración de los compuestos químicos presentes en las mezclas. La información contenida en las variables de una señal puede ser reducida a un número menor de variables sin que haya pérdida de información relevante. Para esto, se requiere el uso de herramientas matemáticas y estadísticas de análisis multivariante, en el que el proceso de calibración se realiza, no sobre los datos originales sino sobre estas nuevas variables, simplificando el modelo y la interpretación de los resultados.

Por ende, uno de los campos de trabajo fundamental de la quimiometría es el desarrollo y aplicación de los nuevos métodos de calibración multivariable, que permiten extraer la máxima información química del análisis de muestras a partir de señales o respuestas instrumentales con muy poca selectividad^[28].

1.5.1 Calibración multivariable

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

La calibración multivariante es la encargada de predecir una propiedad de interés utilizando respuestas instrumentales no selectivas, es decir, la búsqueda de relaciones entre una o más variables de respuesta “Y” y una matriz de variables predictoras “X” como por ejemplo el espectro de una muestra, de manera que se cumpla la ecuación 7^[29].

$$Y = G(X) \quad \text{Ecuación 7}$$

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

El tipo de modelo multivariante más común es el lineal: la concentración del analito se predice multiplicando cada absorbancia del espectro por un coeficiente y sumando como se muestra en la ecuación 8:

$$C = b_0 + b_1 r \lambda_1 + b_2 r \lambda_2 + \dots + b_j r \lambda_j \quad \text{Ecuación 8}$$

1.5.1.1 Clasificación de los modelos de calibración multivariable

Se han propuesto sistemas de clasificación según Martens & Neas^[30]; basados en los métodos de calibración y conceptos fundamentales no tan matemáticos como:

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

- *Calibración lineal* descrita por un modelo lineal,

$$y = b_0 + \sum_{k=1}^K b_k X_k \quad \text{Ecuación 9}$$

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

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

$$y = \alpha x^\beta \quad \text{Ecuación 10}$$

Donde α y β son los parámetros a determinar.

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

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

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

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

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

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

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

1.5.2 Construcción de modelos de calibración multivariable

El objetivo de los métodos de calibración multivariable es construir modelos de calibración que sean capaces de predecir el valor de la propiedad estudiada para nuevas muestras, a las que se ha registrado la señal analítica. La fiabilidad de las predicciones depende de los modelos obtenidos, dichos modelos deben seguir las etapas que se describen a continuación para obtener modelos válidos:

Las etapas descritas a continuación, son las usadas en la construcción de modelos multivariados a partir de la espectroscopia infrarroja, siguiendo los parámetros de la Norma ASTM E-1655^[31].

1) Selección de muestras para la calibración: Muestras altamente representativas del sistema, se deben considerar tanto las variaciones químicas como las físicas. El conjunto de muestras debe ser representativo del sistema en estudio (amplio intervalo de concentraciones) y debe permitir realizar predicciones futuras sobre un conjunto nuevo de muestras.

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

3) Toma de espectros infrarrojos: Incluye la selección de condiciones experimentales óptimas de adquisición espectral. En este trabajo la información es espectrofotométrica, obtenida a partir de los espectros de FTIR-PAS, en la región del Infrarrojo medio.

4) *Pretratamiento de los datos*. En esta etapa, se minimizan las contribuciones no deseadas presentes en la señal analítica que disminuyen la reproducibilidad y pueden provocar que el sistema presente comportamientos que darían lugar a

estimaciones erróneas de los parámetros deseados debido a errores de origen aleatorio o sistemático. Los más comunes son el suavizado para reducir el ruido en los datos, la normalización para lograr que los datos estén aproximadamente en la misma escala, el centrado para evitar que ciertos puntos tengan más peso que otros en el modelo y las derivadas de diferente orden para extraer información detallada que no puede ser observada en el espectro normal.

5) Cálculo del modelo matemático: después de encontrar el pretratamiento óptimo de la señal, se emplean técnicas de regresión sobre los datos espectrales

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

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

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

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

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

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

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

- Validación mediante cross-validación:

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

$$RMSECV = \sqrt{\frac{\sum_{i=1}^c (y_i - \hat{y}_{CV,i})^2}{n_c}} \quad \text{Ecuación 12}$$

- Validación mediante conjunto de validación externa:

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

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n_p} (y_i - \hat{y}_i)^2}{n_p}} \quad \text{Ecuación 13}$$

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

8) Implementación del modelo al análisis de muestras desconocidas: Con el modelo construido y validado, se calcula la propiedad estudiada en nuevas muestras a las que, previamente, se ha registrado la señal analítica, el modelo se instala como herramienta de análisis de rutina y se realizan chequeos periódicos para evaluar su desempeño [32,33,33,18].

1.5.3 Pretratamiento de los datos registrados

Los pretratamientos espectrales son correcciones del espectro basadas en tratamientos matemáticos con la finalidad de aumentar la calidad de la señal instrumental, reduciendo al máximo la información no asociada a la propiedad estudiada, como pueden ser desplazamientos de la línea base, ruido espectral, etc. Se denomina ruido al componente espectral no relacionado con ninguna

propiedad física o química de la muestra como puede ser el ruido instrumental o de iluminación. Además, otras fuentes de variación como la temperatura y humedad pueden aportar interferencias a la señal analítica.

Existen bastantes tratamientos matemáticos que se pueden aplicar antes de intentar hallar la relación señal-concentración para mejorar la calidad de los datos, entre las más utilizadas están [33]:

1.5.3.1 Corrección de línea base

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

1.5.3.2 Normalización

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

- Media de la normalización (Mean Normalization): Es el tipo de normalización más clásica. Consiste en dividir cada fila de una matriz de datos por su media, neutralizando así la influencia de los factores ocultos.
- Máxima normalización (Maximum normalization): Esta es una alternativa a la normalización clásica. Divide cada fila por su valor máximo absoluto en lugar de la media.

- Normalización por rangos (Range normalization): En esta transformación se normaliza un espectro X_i calculando el área bajo la curva del espectro. Se intenta corregir el espectro de longitud de la trayectoria indeterminada cuando no hay forma de medirla, o aislar a un grupo de un componente constante.

1.5.3.3 Suavizado espectral

Mediante el suavizado se intenta reducir el ruido aleatorio existente en la señal instrumental. La técnica más utilizada es la propuesta por Savitzky y Golay^[34]. El principio en el que se basa este método es que, para un pequeño intervalo de longitudes de onda, se puede ajustar un polinomio del grado adecuado. Los nuevos valores tras el ajuste son una mejor estima que los valores medidos ya que se ha eliminado parte del ruido que los afectaba. El suavizado espectral se aplica en aquellos casos en que la relación señal / ruido es pequeña.

1.5.3.4 Centrado

El centrado evita que ciertos puntos tengan más peso que otros en el modelo. Consiste en calcular el valor medio de cada variable (X_m) del conjunto de calibración (de cada columna de la matriz), y restar este valor a cada punto ($X_{i,m}$).

$$x'_{i,m} = x_{i,m} - \bar{x}_m \quad \text{Ecuación 14}$$

Siendo $x'_{i,m}$ el dato centrado, $x_{i,m}$ el dato de la fila i y la columna m antes del centrado, \bar{x}_m la media de la columna m ($\bar{x}_m = \sum x_{i,m} / I$). La propiedad fundamental de los datos centrados es que el valor medio de cada una de las Fundamentos

teóricos variables es igual a cero. Este pretratamiento no modifica la varianza de los datos

1.5.3.5 Derivadas

La diferenciación o cálculo de derivadas permite acentuar las diferencias existentes en los datos espectrales. Tanto la primera como la segunda derivada se utilizan a menudo para el tratamiento de los datos. La segunda derivada elimina el ruido de fondo lineal y constante. Los dos principales algoritmos de diferenciación son el de Savitzky-Golay y el de Norris ^[35,36]. El primero, permite calcular derivadas de primer orden o mayor incluyendo un factor de suavizado que determina el número de variables adyacentes que se usarán en la estimación de la aproximación polinómica utilizada en la derivación. El algoritmo de Norris, a diferencia del anterior, solo permite el cálculo de derivadas de primer orden.

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

1.5.4 Análisis por componentes principales (PCA)

El análisis en componentes principales (PCA) se aplica en química con diferentes finalidades [Jackson 1991; Wold, 1987]. Entre ellas cabe destacar la clasificación

de muestras y la reducción en la dimensionalidad de los datos, siendo éste el primer paso de la mayoría de métodos multivariantes de análisis de datos.

La técnica multivariante PCA consiste en reducir la dimensión (número de variables) a un conjunto de datos con muchas variables (como lo son los datos espectroscópicos) reduciéndolo a un menor número perdiendo la menor cantidad de información posible, eliminando la información redundante o reduciendo la contribución del ruido. El PCA construye una transformación lineal que escoge un nuevo sistema de coordenadas para el conjunto original de datos denominados componentes principales (PCs) el cual explica la máxima variabilidad de las muestras; el segundo PC se escoge de forma que sea ortogonal al primero y que explique la máxima variabilidad de las muestras una vez restada la explicada por el primer PC.

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

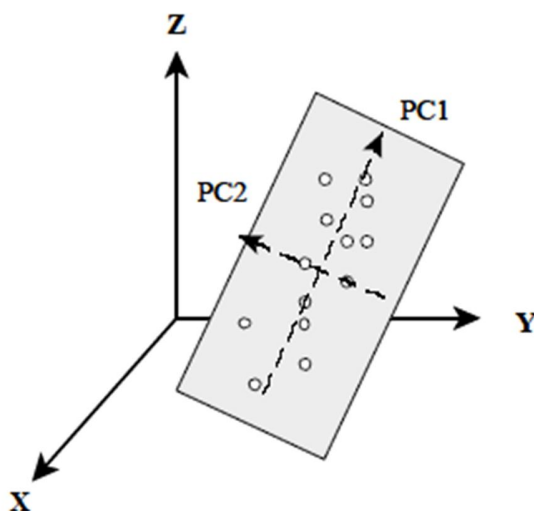
La ASTM E131, es la encargada de definir el análisis por componentes principales, como un procedimiento matemático para resolver conjuntos de datos hallando componentes o vectores ortogonales entre sí, llamados componentes principales (PC). Tales componentes modelan la variación estadísticamente significativa en el conjunto de datos así como el error aleatorio presente, de manera que se pueden detectar fenómenos que a primera vista no parecen evidentes.

Geoméricamente, el PCA es un cambio en la dirección y dimensionalidad de los ejes (figura 7). Busca las direcciones ortogonales que explican la máxima variabilidad de las muestras y las utiliza como nuevos ejes de coordenadas

llamados componentes principales (PCs) para representarlas. El primer componente principal es la dirección que explica la máxima variabilidad de la matriz X ; el segundo se escoge de tal forma que sea perpendicular al primero y que explique la máxima variabilidad de la matriz X la cual no se encuentra explicada en el primer componente, y así sucesivamente.

Para poder definir estos nuevos ejes se utiliza la grafica de pesos (*loadings*) que son los cosenos de los ángulos que forman con los ejes antiguos y contiene los coeficientes de las transformaciones lineales de las variables originales. Por su parte el grafico de puntuaciones (*scores*) representa las coordenadas de las muestras en estos nuevos ejes^[17,24].

Figura 7. Interpretación geométrica de un PCA



Fuente: GEMPERLINE, P. Practical guide to chemometrics. Segunda edición. Editorial Taylor y Francis Grupo. 2006. 541 p.^[33]

Matemáticamente, la matriz de datos X (datos espectrales) se descompone en el producto de dos matrices, T (matriz de *scores*) y P (matriz de *loadings*), más una matriz E de residuales de X

$$X = TP^T + E \quad \text{Ecuación 15}$$

Los diferentes componentes principales no contienen la misma información. Los primeros describen la fuente de variación más importante de los datos, que se puede asociar a la información más relevante. Para tener una reproducción perfecta de la matriz \mathbf{X} se necesitan k vectores de *loadings* y *scores*, sin embargo, es posible representar la matriz original con un número $a < k$ de vectores

$$X = t_1p_1^T + t_2p_2^T + \dots + t_ap_a^T + E \quad \text{Ecuación 16}$$

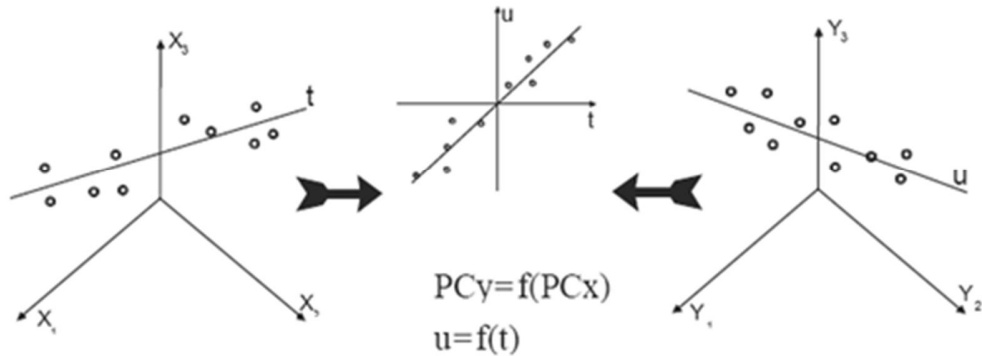
Puesto que la información relevante está contenida en los a primeros componentes, mientras que los últimos sólo describen variaciones debidas al ruido. Esto permite una reducción importante del número de variables.

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

1.5.5 Regresión por mínimos cuadrados parciales (PLS)

El método de Regresión Parcial por Mínimos Cuadrados (PLS, *Partial Least-Squares Regression*) fue desarrollado por H. Wold en 1975 [Wold, H.; 1975]. Este tipo de regresión es probablemente, la más utilizada en su aplicación al tratamiento de datos multivariados, este método pretende que los primeros componentes contengan la mayor información para la predicción de la matriz \mathbf{Y} . Para ello, durante la etapa de calibración, el algoritmo PLS utiliza tanto la información contenida en la matriz de datos (matriz \mathbf{X} , Datos espectroscópicos) como la información contenida en la matriz de la propiedad a determinar (matriz \mathbf{Y} , composición elemental de los FV), obteniéndose unas variables auxiliares llamadas variables latentes. ^[17,37] En la Figura 8 se muestra la descripción del PLS.

Figura 8. Descripción gráfica del modelo de regresión PLS



Como en el caso del análisis por componentes principales, estas nuevas variables en X y Y se pueden representar como un producto de matrices según se muestra en las ecuaciones 17 y 18.

Donde:

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

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

E y F son los residuos.

Las variables originales X y Y se pueden relacionar mediante los scores de cada una de las nuevas variables latentes como lo muestra la ecuación 19.

Donde b_a es el coeficiente de regresión para cada variable latente obtenido a través de la ecuación 20.

Los coeficientes \mathbf{b}_a hallados para cada componente son agrupados en una matriz diagonal \mathbf{B} , que contiene los coeficientes de regresión de los scores \mathbf{U} y \mathbf{T} de las matrices \mathbf{Y} y \mathbf{X} respectivamente. De tal manera, la matriz \mathbf{Y} puede ser calculada por medio de la ecuación 21.

$$Y = TBQ^T + F \quad \text{Ecuación 21}$$

Una de las ventajas que presenta la regresión por mínimos cuadrados parciales frente a otros métodos de regresión es que, al estar basada en variables latentes, se puede utilizar para el tratamiento de datos altamente correlacionados como son los datos espectroscópicos.

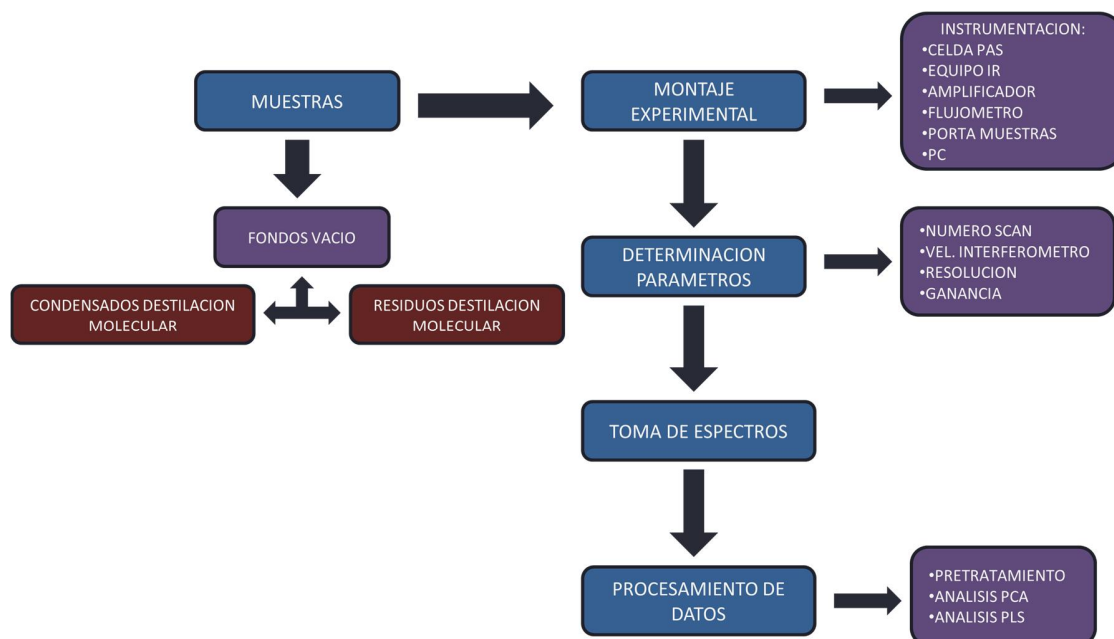
2. METODOLOGIA

En el Laboratorio de Espectroscopia Atómica y Molecular, LEAM de la Universidad Industrial de Santander, se han venido desarrollando en los últimos años investigaciones, que van encaminadas a la caracterización del petróleo Colombiano y sus residuos, generando aportes científicos y tecnológicos en esta área.

2.1 ESQUEMA EXPERIMENTAL

En la figura 9 se muestra un esquema de los pasos que se realizaron para el desarrollo del proyecto.

Figura 9. Esquema metodológico



2.2 SELECCIÓN DE MUESTRAS

En este trabajo se estudiaron 20 muestras conformadas por fondos de vacío de diferentes crudos colombianos, los cuales se separaron por destilación molecular obteniendo residuos y condensados. Las muestras fueron suministradas por el Instituto Colombiano del Petróleo (ICP), así como los datos de análisis elemental. (Tabla 3).

Tabla 3. Análisis elemental de las muestras

	Nombre de las Muestras	% C	% H	% S	% N
FV 1	FONDO DE VACIO GALAN	85.9800	10.8800	1.7700	0.7024
FV 2	GALAN CORTE 1 RESIDUOS	86.2200	11.7300	1.9100	0.7475
FV 3	GALAN CORTE 1 CONDESADOS	85.7100	11.7600	1.4100	0.3601
FV 4	GALAN CORTE 2 RESIDUOS	85.5000	10.7200	2.0000	0.8343
FV 5	GALAN CORTE 2 CONDENSADOS	85.8300	11.8900	1.3900	0.3964
FV 6	GALAN CORTE 3 RESIDUOS	85.7000	10.1800	2.0700	0.9324
FV 7	GALAN CORTE 3 CONDENSADOS	86.0300	12.1600	1.4100	0.3953
FV 8	FONDO DE VACIO TIBU ZULIA	86.6200	11.1100	2.2000	0.4500
FV 9	TIBU ZULIA COLECTOR 2 CONDENSADOS	85.9300	12.4300	1.4700	0.1181
FV 10	TIBU ZULIA COLECTRO 4 RESIDUOS	86.1000	11.2500	2.4700	0.5786
FV 11	TIBU ZULIA COLECTRO 4 CONDENSADOS	85.7000	12.1600	1.5400	0.1339
FV 12	TIBU ZULIA COLECTRO 6 RESIDUOS	85.4700	10.1500	2.7300	0.7007
FV 13	TIBU ZULIA COLECTRO 6 CONDENSADOS	86.1000	12.5200	1.5300	0.1852
FV 14	FONDO DE VACIO RUBIALES	87.3400	9.5200	2.1100	0.7639
FV 15	RUBIALES COLECTOR 2 RESIDUOS	86.5200	9.0200	2.1500	0.7635
FV 16	RUBIALES COLECTOR 2 CONDENSADOS	86.9600	11.0400	1.4900	0.3522
FV 17	RUBIALES COLECTOR 4 RESIDUOS	86.9800	9.0400	2.2300	0.7986
FV 18	RUBIALES COLECTOR 4 CONDENSADOS	86.7600	11.6100	1.5200	0.3439
FV 19	RUBIALES COLECTOR 6 RESIDUOS	86.7900	8.6000	2.2900	0.8484
FV 20	RUBIALES COLECTOR 6 CONDENSADOS	86.5200	11.8100	1.4500	0.3551

2.3 MONTAJE EXPERIMENTAL

2.3.1 Instrumentación

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

Figura 10. Sistema PAS acoplado al equipo IR Tensor 27 marca Bruker



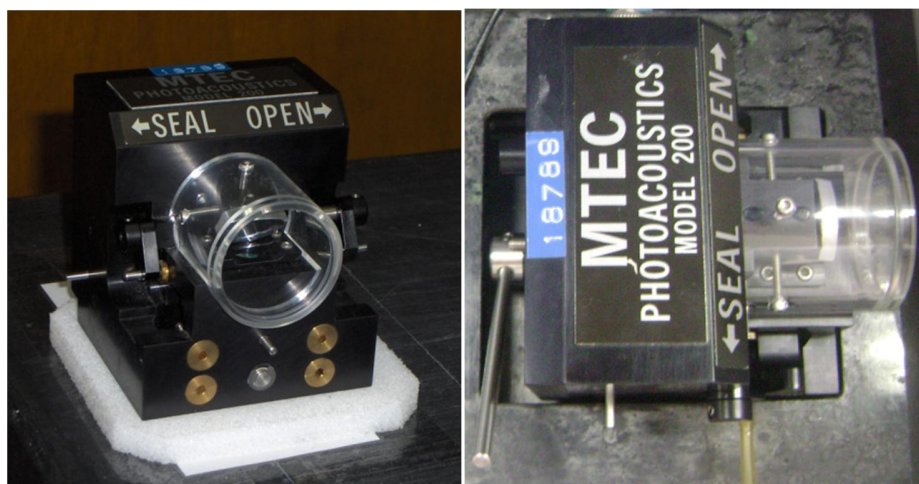
Fuentes: ORREGO Jorge. Estudio de la Estructura de Cinco Carbones Colombianos por Espectroscopía Fotoacústica en el Infrarrojo Medio ^[22].

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

1. *Celda PAS* (figura 10a). Está formada principalmente por un compartimiento para la muestra (portamuestra), por una cámara que mantiene confinado el gas inerte y garantiza que esté en contacto con la muestra mientras se adquiere cada espectro, y por un micrófono que detecta las variaciones de presión del gas. La

figura 11 muestra el exterior de la celda PAS. En el interior de la celda la cámara es sellada para garantizar que el gas no se escape a la atmósfera durante la toma de cada espectro.

Figura 11. Celda PAS modelo MTEC 200



2. *Amplificador* (figura 10b) usado para amplificar la señal y controlar remotamente la ganancia del pre-amplificador localizado dentro del micrófono. La ganancia puede modificarse desde 1 hasta 128. La magnitud de la ganancia se determina de acuerdo a la velocidad del espejo del interferómetro.

3. *Interface* (figura 10c) usada para llevar la señal fotoacústica hasta el procesador de la señal del equipo FTIR.

4. *Flujómetro* (Figura 10d) se encarga de controlar el flujo de gas He (figura 10e) que es introducido a la celda. El gas He además de servir como el portador de la señal, se usa para purgar la celda de los gases atmosféricos (vapor de H₂O y CO₂). El flujo de gas debe ser controlado para minimizar el riesgo de daño del micrófono.

2.4 DETERMINACION DE LOS PARAMETROS

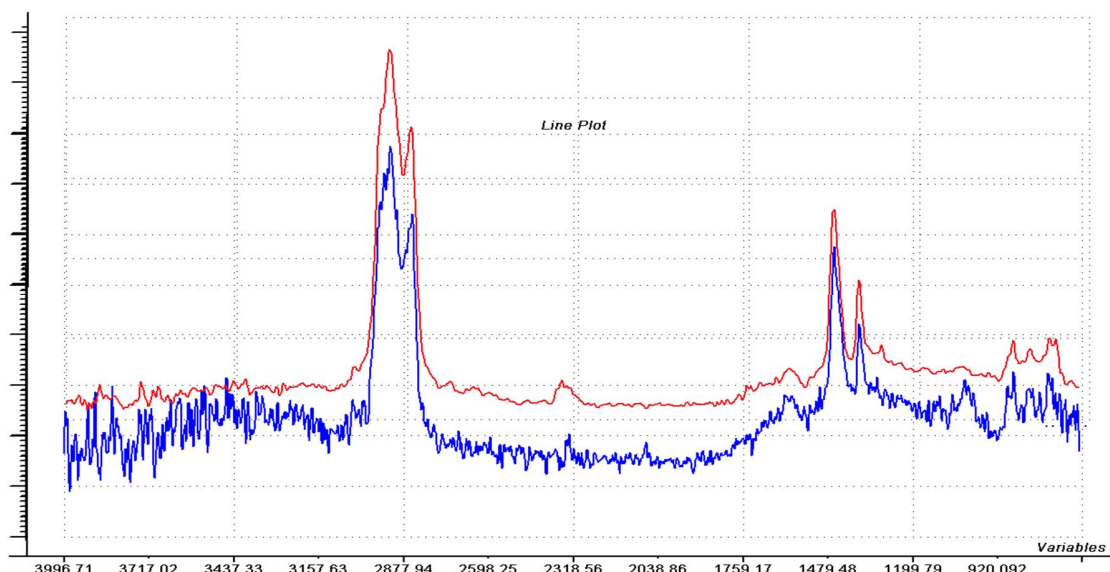
Luego de acoplar la celda se variaron parámetros del sistema PAS y del equipo IR para obtener espectros con relación señal ruido aceptable. La relación señal-ruido (S/N) es una medida útil para describir la calidad de un método analítico o el funcionamiento de un instrumento. La señal lleva la información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

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

$$\frac{S}{N} = \frac{\text{media}}{\text{desviación estándar}} = \frac{\mu}{\sigma} \quad \text{Ecuación 22}$$

Esta relación se determinó a partir de la ecuación 23. La figura 12 muestra el aspecto de dos espectros de las muestras con diferente relación S/N.

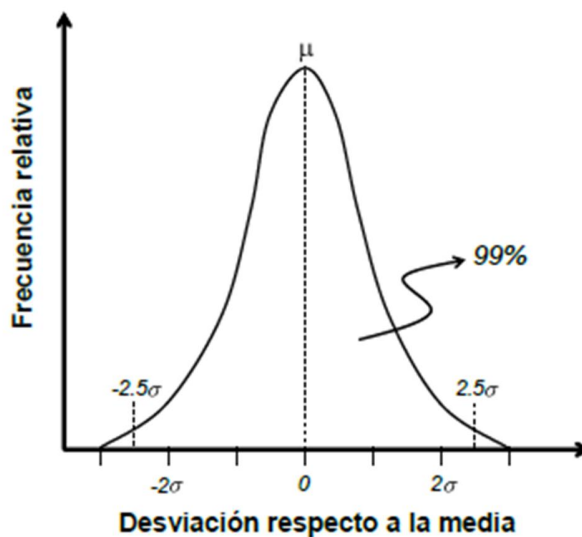
Figura 12. Efecto de la relación señal ruido en un espectro de una muestra



La desviación estándar (ruido) puede estimarse para un nivel de confianza del 99% dividiendo por cinco la diferencia entre la señal máxima y la mínima en una región que contenga el fondo espectral (ecuación 23), teniendo en cuenta que el 99% de los datos bajo la curva normal de error caen dentro de ± 2.33 de la media (figura 13) [22].

En el presente trabajo se determinó la relación señal-ruido en la región comprendida entre 2000 y 1800 cm^{-1} , en donde no aparecen señales asignables a grupos funcionales (fondo espectral).

Figura 13. Curva normal de error



Fuente: SKOOG, D., A. Principios de análisis instrumental^[38].

Los parámetros que se variaron para obtener la menor relación señal/ruido fueron:

- Numero de Scan o barridos
- Resolución
- Ganancia
- Velocidad de espejo

La relación señal ruido se determinó a partir de la ecuación 23, donde S_{max} y S_{min} representan las señales máxima y mínima comprendidas entre 2000 y 1800 cm^{-1} , y μ el promedio de las mediciones en este rango.

La señal fotoacústica, como se mostró en las ecuaciones 2 y 4, depende drásticamente de la frecuencia de modulación que a su vez es función de la velocidad del espejo del interferómetro ($f = V/\lambda$ donde V es la velocidad del espejo y λ es el número de onda). El equipo IR Bruker Tensor 27 donde se acopló el sistema PAS, permite trabajar a tres únicas velocidades de espejo del interferómetro: 0.0583, 0.633 y 1.27 cm/s. La relación existente entre la velocidad

del espejo y la señal fotoacústica es la siguiente: Una velocidad alta implica que el tiempo de interacción de la radiación con la muestra sea corto y por lo tanto la intensidad de la señal fotoacústica es baja. A medida que la velocidad del espejo disminuye la intensidad de la señal aumenta. Un aumento de la señal podría causar una saturación de la señal.

La selección de las mejores condiciones para la toma de espectros se basó en las relaciones señal ruido que se calcularon variando los parámetros, encontrando que a partir de una velocidad de espejo de 0.0583 cm/s, una resolución de 8 cm⁻¹, 256 barridos y una ganancia de 1, la relación alcanzaba un valor 45.11, por lo cual se establecieron estas condiciones como los parámetros para la toma de todos los espectros de las muestras.

2.4.1 Verificación del espectrómetro

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

Según la ASTM en sus normas E177, Práctica para el uso de los términos precisión y desviación en los métodos de prueba ASTM y E456, Terminología relacionada con la calidad y las estadísticas.

La repetibilidad nos indica el grado de concordancia entre resultados mutuamente independientes de un ensayo o prueba, obtenidos utilizando el mismo método, en idénticos materiales, en un mismo laboratorio, y tomados por el mismo operario, usando el mismo equipo en un intervalo de tiempo corto. La determinación de la repetibilidad es una condición necesaria para un análisis más profundo de la realidad de un sistema (espectrómetro). Cuando se hace un experimento por

segunda vez, éste se repite en algunos aspectos pero no en todos; puesto que siendo lo suficientemente analítico, se puede considerar variaciones, por ejemplo, las condiciones externas e internas del experimento no son exactamente las mismas y que, por tanto, el experimento es diferente para cada evento.

Para el presente caso, se calcularon las desviaciones estándar para las absorbancias medidas a 2924,23 y 1450,54 cm^{-1} en 10 espectros adquiridos sobre una misma muestra (Figura 14). Las desviaciones encontradas para cada uno de los espectros fueron inferiores a 0.002 (Tabla 4), asegurando de esta manera la repetibilidad del espectrómetro.

Figura 14. Prueba de repetibilidad del espectrómetro FTIR-PAS

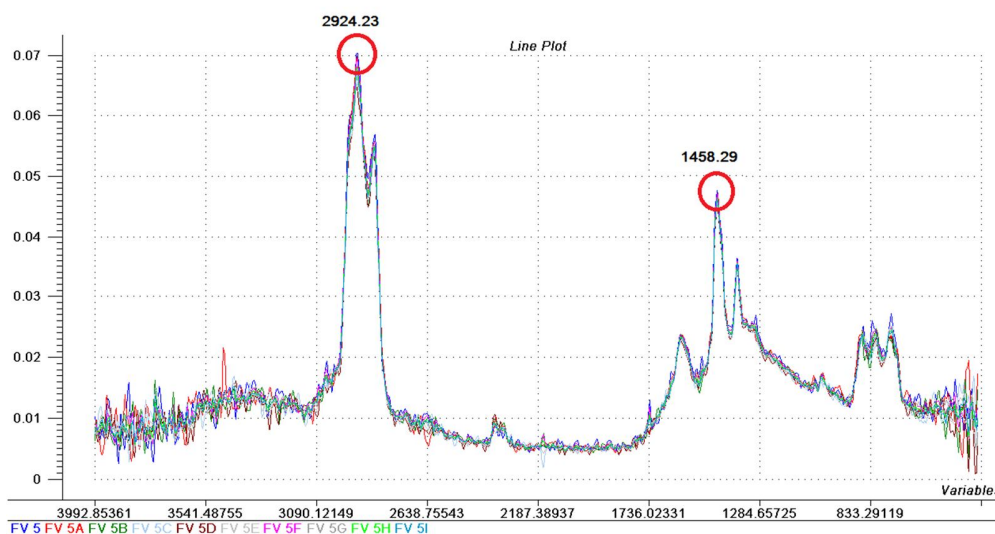


Tabla 4. Datos de absorción y desviación estándar para la prueba de repetibilidad

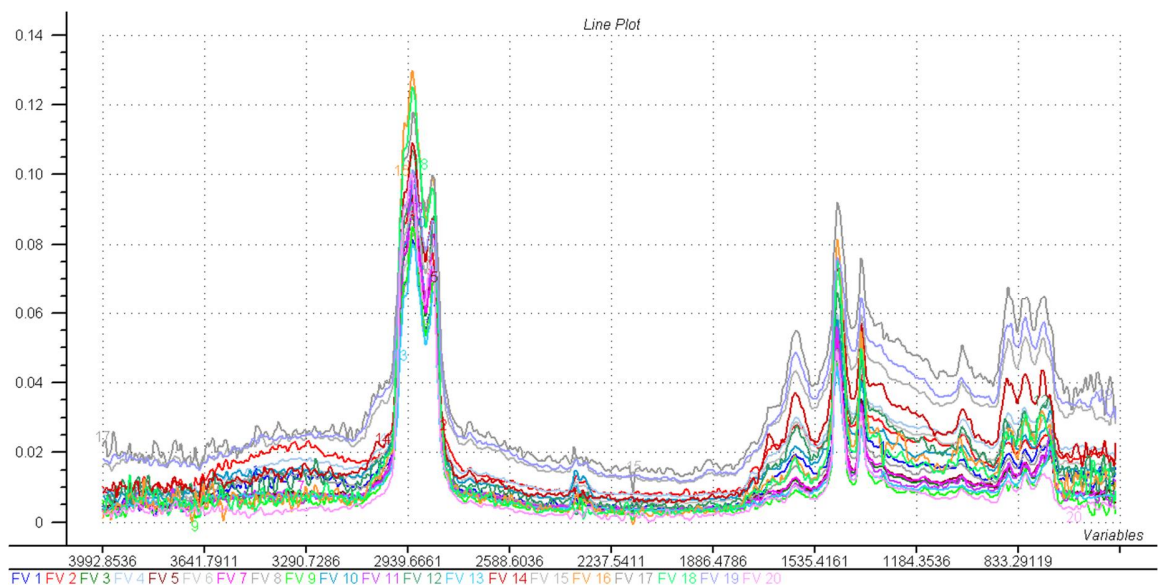
Espectro	A 2924.23 cm^{-1}	A 1458,29 cm^{-1}
FV 5	0.07081	0.04763
FV 5 ^a	0.06978	0.04719
FV 5B	0.06820	0.04664
FV 5C	0.06581	0.04595
FV 5D	0.06473	0.04486
FV 5E	0.06527	0.04541

FV 5F	0.06951	0.04714
FV 5G	0.06777	0.04625
FV 5H	0.06809	0.04656
FV 5I	0.06726	0.04628
Media	0.06772	0.04639
Desviación	0.00200	0.00084

2.5 ADQUISICION Y ANALISIS DE ESPECTROS PAS

Los espectros FTIR-PAS de cada muestra, se tomaron en un intervalo espectral comprendido entre 4000 y 500 cm^{-1} , con 256 Scan, una velocidad de espejo de 0.0583 cm/s y una resolución de 8 cm^{-1} . Para el background se usó negro de humo y helio como gas encargado de transmitir la onda acústica de cada espectro. La figura 15 muestra los espectros PAS para las 20 muestras.

Figura 15. Espectros PAS de todas las muestras



2.5.1 Vibraciones en los fondos de vacío

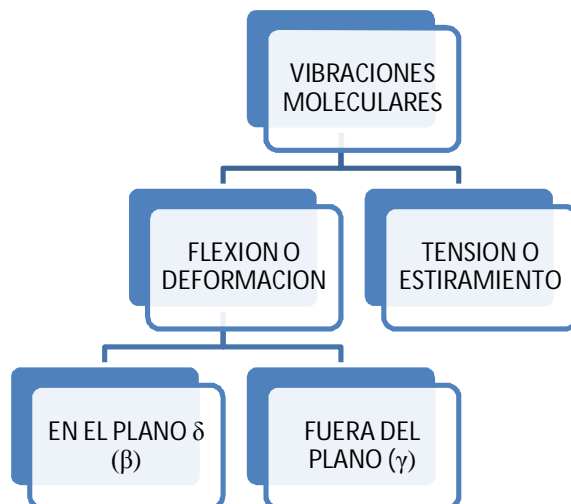
En general los espectros de los 20 fondos de vacío presentan las mismas bandas típicas para hidrocarburos, ya que presentan grupos funcionales similares:

En las moléculas poliatómicas podemos clasificar a las vibraciones, de forma aproximada, de acuerdo a los parámetros geométricos que cambian en los grupos funcionales. Las vibraciones que corresponden esencialmente a variación de las distancias internucleares, movimientos a lo largo de los ejes de los enlaces, reciben el nombre de vibraciones de tensión o estiramiento. El otro tipo de vibraciones corresponde a la modificación de los ángulos entre dos enlaces y reciben el nombre de vibraciones de deformación o flexión. En el caso particular de un grupo o molécula con sus átomos en un plano, las vibraciones de flexión pueden a su vez clasificarse como flexiones en el plano o flexiones fuera del plano, de acuerdo a las direcciones de desplazamiento de los átomos. La simbología utilizada para representar las vibraciones es la siguiente:

- Vibraciones de tensión o estiramiento: ν
- Vibraciones de doblaje en general o doblajes en el plano: δ (β)
- Vibraciones de doblaje fuera del plano: γ

Si se presentan acoplamientos entre dos movimientos de igual naturaleza, las vibraciones resultantes se clasifican como simétricas o antisimétricas.

Figura 16. Clasificación posibles vibraciones presentadas por los grupos funcionales en el infrarrojo.



En la tabla 5 se muestran las principales asignaciones a las vibraciones de los grupos funcionales presentes en las muestras.

Tabla 5. Bandas vibracionales en el infrarrojo medio presentes en los fondos de vacío

Posición de las bandas (cm ⁻¹)	Grupos funcionales
3700 – 3300	Vibraciones de tensión con Hidrógeno
3400-3300	Vibraciones de tensión con H del Aril-NH-R
3090 – 3030	Vibraciones de tensión (C-H) _{ar}
2978 – 2960	R-CH ₃ estiramiento asimétrico
2925 – 2918	R ₂ -CH ₂ estiramiento asimétrico
2860	Estiramiento del CH alifático
2854	R-CH ₂ estiramiento simétrico
2600 – 2540	RS-H vibraciones de tensión (S-H)
1690 – 1520	Vibraciones de tensión de alifáticos
1650 – 1615	Vibraciones de tensión de aromáticos
1660	Enlace aislado de C=C
1654	-C=O conjugado
1650 – 1570	(C-H) _{ar} sistema aromático polinuclear
1500 – 1450	(C-C) _{ar} estiramiento
1460 – 1450	Flexión C-H alifático
1500 – 1100	R ₂ N(C=S)NR ₂ vibraciones de tensión

1400-1200	Aril-NH-R vibraciones de tensión ν_{Aril-N}
1380	Deformación simétrica $-CH_2-$ (flexión)
1300-1000	Deformación C-O fenolica (estiramiento)
1275 – 1030	$R_2C=S$ tensión en las tioacetonas $\nu_{C=S}$
1215	$R(C=S)SH$ estiramiento $\nu_{C=S}$ en ditiocidos
1140 – 1090	$R(C=S)NR_2$ tensión $\nu_{C=S}$ en las tioamidas
925 – 670	Enlaces aromáticos ; flexión fuera del plano del $(C-H)_{ar}$
873	Anillo bencénico sustituido con H "aislado"
758	Anillo bencénico o-m-sustituido sistema de anillos aromáticos
720	Banda de balanceo de los CH_2

Fuente: SKOOG, D., A. Principios de análisis instrumental^[38].

Vibraciones asociadas o los grupos alifáticos alifáticas:

Las flexiones asociadas a la cadena (δ_{C-C}) son vibraciones de baja energía, fuertemente acopladas y fuera del rango del IR medio. Las vibraciones de tensión (ν_{C-C}) son en general débiles (debido a la baja polaridad de los enlaces) y fuertemente acopladas, observándose en la zona de la huella dactilar de la molécula (1600 y 1300 cm^{-1}). Las vibraciones de estiramiento (ν_{C-H}) se presentan en la zona de 3000 – 2800 cm^{-1} , su posición es muy estable y corresponden esencialmente a los movimientos acoplados de los Hidrógenos unidos a un mismo Carbono.

- Vibraciones asociadas al grupo metilo:

Las vibraciones (ν_{C-H}) del grupo metilo corresponden a movimientos acoplados. Se observan 2 bandas: la vibración de estiramiento en el plano de los tres enlaces CH_3 a 2870 cm^{-1} y a la vibración de estiramiento asimétrico fuera del plano a 2960 cm^{-1} . Las vibraciones de flexión mas relevantes de los grupos metilo son las correspondientes a la flexión de deformación en el plano $\delta(\beta)_{CH_3}$ a 1378 cm^{-1} , esta banda permite la detección de ramificaciones; y la flexión de deformación asimétrica fuera del plano $\delta(\gamma)_{CH_3}$ a 1463 cm^{-1} . La presencia de más de un grupo

metilo sobre un mismo Carbono produce pequeños acoplamientos entre estas vibraciones, que aparecen como dobletes tal como se muestra en la figura X.

- Vibraciones asociadas al grupo metileno:

Al igual que en el grupo metilo, las vibraciones de estiramiento (ν_{C-H}) de los grupos metileno tiene carácter acoplado. Se observan dos bandas que corresponden a la vibración de estiramiento en el plano ν_{CH_2} a 2815 cm^{-1} , y a la vibración de estiramiento fuera del plano ν_{CH_2} a 2926 cm^{-1} . La vibración de flexión en el plano más interesante es la de tijera $\delta(\beta)_{CH_2}$ a 1463 cm^{-1} , que usualmente aparece superpuesta con la banda de flexión fuera del plano $\delta(\gamma)_{CH_3}$ de los grupos metilo. La banda de balanceo τ_{CH_2} de los grupos CH_2 posee cierta utilidad diagnóstica, ya que cuando está presente una cadena Carbonada de cuatro o más grupos metilenos contiguos aparece como una banda débil a 720 cm^{-1} .

Vibraciones asociadas a los grupos aromáticos:

Los hidrocarburos aromáticos presentan numerosas bandas de absorción en el MIR. Las vibraciones de tensión ν_{C-H} causadas por la hibridación sp^2 de los Carbonos aromáticos aparecen entre $3000 - 3100\text{ cm}^{-1}$, generalmente son varias bandas cercanas, en ocasiones no resueltas: $3096 - 3048$, $3039 - 3025$ y $3020 - 3000\text{ cm}^{-1}$.

A 1604 , 1585 , 1500 y 1450 cm^{-1} se presentan varias bandas producto del acoplamiento de las vibraciones de los enlaces (C=C) del anillo aromático $\nu_{C=C}$. Dependiendo de la simetría y el patrón de sustitución algunas de estas bandas pueden que no se observen.

Las vibraciones de flexión δ_{C-H} dan lugar a bandas intensas en la zona de bajas frecuencias y dependen del número de Hidrógenos vecinos en el anillo. En la tabla

6 se muestra la posición de las bandas δ_{C-H} de acuerdo al número de Hidrógenos vecinos.

Tabla 6. Vibraciones de flexión δ (C-H) aromático

Hidrógenos vecinos	Posición bandas (cm^{-1})
1	900 – 860
2	860 – 800
3	810 – 750
4	770 – 735
5	770 – 730

- Vibraciones de tensión con el Hidrógeno:

En las vibraciones donde varía la distancia internuclear (vibraciones de estiramiento o tensión) de un enlace donde participa un átomo de Hidrógeno, las correspondientes frecuencias tienen un valor elevado. Este tipo de vibración ν_{XH} presenta en general débiles acoplamientos y aparece en IR a frecuencias altas (en la región entre $3700 - 2500 \text{ cm}^{-1}$) caracterizando a las agrupaciones hidrogenadas ($\nu_{CH}, \nu_{OH}, \nu_{NH}, \nu_{SH}$). Cuando un mismo átomo forma dos enlaces con Hidrógeno, las vibraciones γ_{XH} se acoplan entre sí. Las frecuencias de las vibraciones acopladas pueden evaluarse a partir de la siguiente ecuación:

$$\nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu} \left(1 + \frac{\mu}{M} \cos \theta \right)} \quad \text{Ecuación 24}$$

Donde:

θ = Angulo HXH

μ = masa reducida de la vibración XH

M = masa del átomo central X

k - constante de fuerza del enlace XH

Véase que en este caso el acoplamiento es moderado y la diferencia entre las frecuencias acopladas es discreta, dado que $\mu/M \ll 1$.

Así por ejemplo tenemos para el agua H₂O:

$$\nu_{H_2O}^{as} = 3756 \text{ cm}^{-1} \quad \nu_{H_2O}^s = 3652 \text{ cm}^{-1} \quad \text{Ecuación 25}$$

2.6 PROCESAMIENTO DE DATOS

El pretratamiento y análisis de los datos espectroscópicos se realizó utilizando las herramientas del software The Unscrambler® versión 9.7 [38] licenciado por Ecopetrol S.A.

Inicialmente se realizó una inspección visual de los espectros PAS de la figura 15 con el fin de descartar muestras atípicas o la influencia de posibles errores instrumentales y humanos durante la toma de muestras y posterior adquisición de espectros.

La región comprendida desde 3100 hasta 600 cm^{-1} es la que registra la señal correspondiente a las vibraciones existentes en las moléculas que componen las muestras; por lo tanto las regiones restantes en un rango espectral de 4000 cm^{-1} a 400 cm^{-1} simplemente no registran información de utilidad.

2.6.1 Pretratamiento de datos

El pretratamiento de datos se realizó utilizando algunas herramientas disponibles en el software, con la finalidad de eliminar la variabilidad presente en las muestras

debido a efectos del ruido, corrimiento de la línea base, diferencias en escala entre otras; haciendo uso solamente de aquellas que permitieran de alguna manera obtener la máxima y más útil información de los datos.

En esta parte del trabajo se determinó una medida de dispersión empleada en estadística la cual da una idea de en qué medida los datos están más o menos juntos (concentrados) o más o menos dispersos. Se determinó el coeficiente de variación (CV) que permite comparar la dispersión de dos o más distribuciones; a menor coeficiente de variación menor dispersión de los datos (o mayor concentración)^[22]. Las bandas escogidas para el análisis fueron 3 que son de las mas representativas de los espectros (2924,23481; 1548,29958; y 744,56111 cm^{-1}) para determinar el pretratamiento indicado para realizarle a las muestras, ya que el software presenta diferentes opciones para el pretratamiento de datos espectrales. Los datos se muestran en la tabla 7.

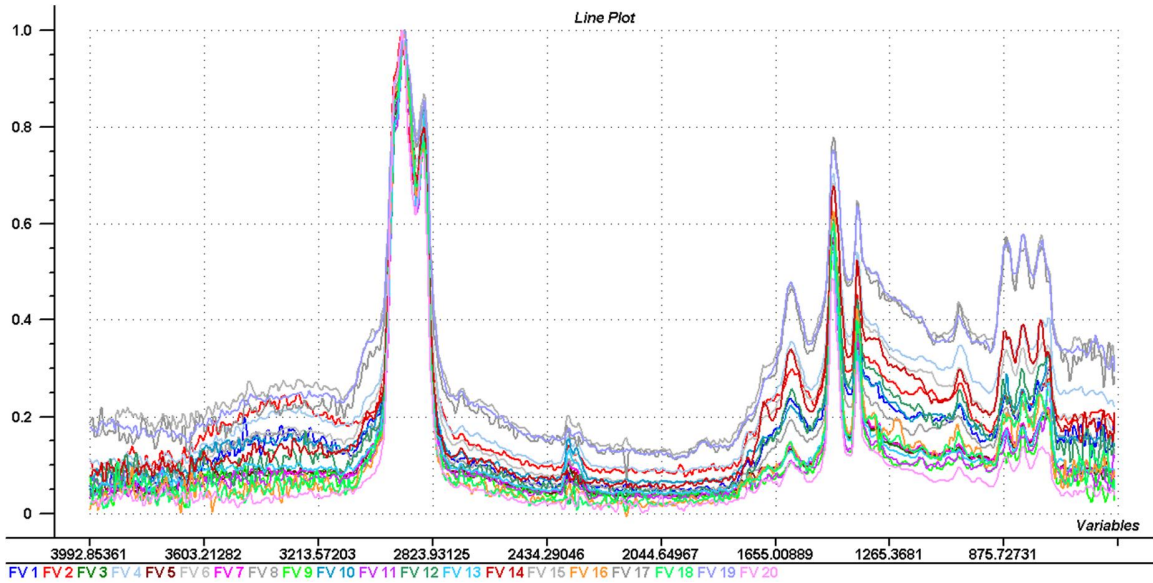
Tabla 7. Calculo del CV para tres números de onda característicos aplicando diferentes pretratamientos

Pretratamiento	Numero de onda (cm^{-1})		
	2924,23481	1458,29958	744,56111
Datos originales	14,67	21,54	49,16
Línea base (LB)	16,33	20,82	47,66
Smooth (MA)	14,68	21,83	48,88
Smooth (SG)	14,67	21,70	49,07
Normalización media	34,14	24,45	11,36
Normalización rango	3,830	16,52	48,28
Normalización máximo	0.70	13.08	44.13
Primera derivada	25,86	911,36	198,77
Normalización vector unitario	20,19	10,43	19,45
Centrado	31,03	43,61	92,81
LB-Smooth	16,33	21,03	47,29
LB-Normalización media	27,31	19,55	15,06
LB-Centrado-Normalización media	51,18	35,89	40,65

Smooth-LB-normalización media	28,19	18,21	13,10
-------------------------------	-------	-------	-------

La herramienta utilizada como pretratamiento de datos fue la Normalización por máximos, que proporciona una transformación orientada a cada número de onda (pico), es decir, la información de un pico en los datos espectroscópicos puede que este influenciado por los números de onda vecinos, en general, la normalización por máximos (figura 17) se escogió como pretratamiento ya que generó los menores valores de CV para las bandas evaluadas (Tabla 7). La normalización por máximos divide cada fila por su valor máximo absoluto en lugar de la media, neutralizando la influencia de factores ocultos, además permite obtener todos los datos en aproximadamente la misma escala, con una distribución mucho más equitativa de las variaciones y los valores medios. Con este pretratamiento se pudo obtener la mayor información de los datos, lo cual se evidencia más adelante en el análisis por componentes principales.

Figura 17. Espectros normalizados por máximos



2.6.2 Análisis por componentes principales (PCA)

Se realizó una descomposición en componentes principales (PC_s) a los 20 espectros de los fondos de vacío, corregidos aplicando normalización por máximos pretratamiento realizado anteriormente, para determinar los componentes (direcciones) que explican la máxima variabilidad de las muestras y usarlas como nuevos ejes de coordenadas, llamados componentes principales (PC_s), de esta forma se reduce la dimensionalidad.

2.6.3 Regresión por mínimos cuadrados parciales

Empleando la regresión PLS, disponible en el software unscrambler 9.7 se evaluó el potencial de la espectroscopia FTIR-PAS para predecir la composición en %W del análisis elemental (Carbono, Hidrógeno, Nitrógeno, Azufre) presente en los fondos de vacío.

El desarrollo de los modelos de predicción comprende dos etapas: calibración y validación interna

Calibración y validación cruzada:

Teniendo en cuenta el número óptimo de componentes principales y las regiones espectrales más representativas, se desarrollaron modelos de predicción variando el número de muestras de calibración en función del análisis PCA. La selección del modelo de predicción se realizará evaluando parámetros estadísticos como el RMSEP, RMSEC, R^2 y la varianza explicada para determinar el número óptimo de variables latentes.

3. RESULTADOS Y ANALISIS

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

3.1 ANALISIS DE COMPONENTES PRINCIPALES (PCA)

Inicialmente se realizó una descomposición por componentes principales a las 20 muestras pretratadas anteriormente, en toda la región espectral, para encontrar el número de componentes óptimo en las posteriores etapas de calibración y validación del modelo.

Los datos de varianza explicada son muy importantes para saber cuántos componentes principales se van a utilizar en el análisis. Debido a que no existe una regla definida sobre el número que se debe utilizar, se buscó decidir en función del número de variables iniciales (recordando que el ideal de la técnica PCA es disminuir la dimensionalidad de los datos lo más posible) y de la proporción de varianza explicada acumulada (Tabla 8).

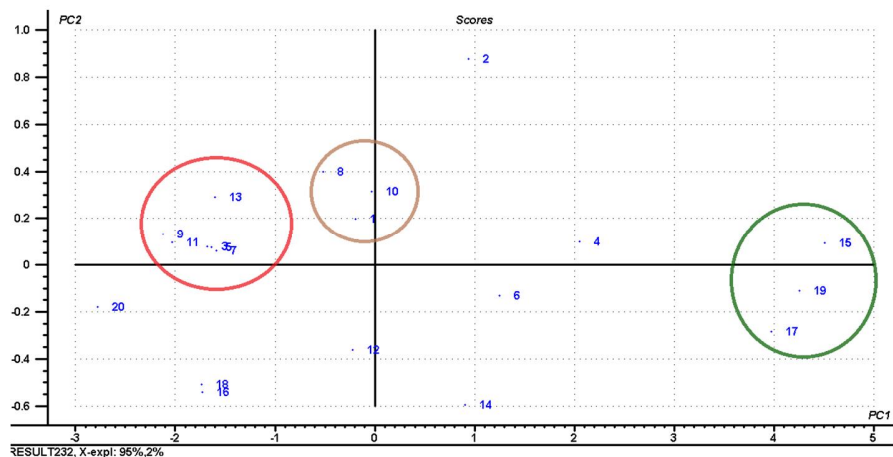
Tabla 8. Varianza de los 7 componentes principales de los 20 fondos de vacío

Componente principal	varianza explicada %	varianza explicada acumulada %
1	94.85	94.85
2	2.36	97.21
3	1.57	98.72
4	0.06	98.78
5	0.27	98.99
6	0.2	99.19
7	0.12	99.31

Teniendo en cuenta el porcentaje de varianza de los datos espectrales que explican los 7 primeros componentes principales se determinó que 3 es el número de componentes a utilizar en los modelos de predicción ya que a partir del PC4 el aumento de la varianza explicada acumulada es menor a 0.06 %.

La representación de los dos primeros componentes principales (PC1 y PC2) se muestran en la figura 18.

Figura 18. Gráfica de Scores para PC1 Vs PC2 de los Fondos de vacío



La figura de scores, es un gráfico de dispersión en dos dimensiones de los componentes principales PC1 y PC2 en los cuales se explica el 95 % y el 2 % de varianza respectivamente, lo que indica que estos dos componentes contienen el 97 % de la varianza total de las intensidades espectrales. Asimismo éste grafico permite establecer las relaciones entre las muestras según su cercanía en el plano PC1-PC2, ya que cuanto más cerca están las muestras al eje de coordenadas del grafico, más similares son con respecto a los dos componentes en cuestión.

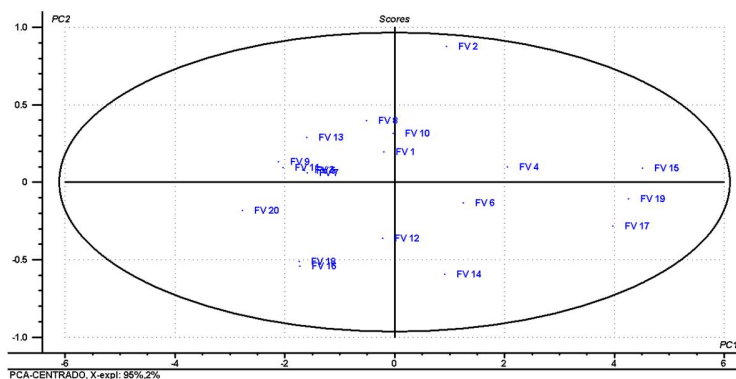
Comparando los grupos que se formaron de acuerdo con el análisis elemental se encontró que las muestras se agrupaban de acuerdo a su composición similar y su yacimiento (figura 18). Con lo anterior se pudo establecer que las muestras se pueden dividir en 3 grupos:

- Grupo verde: conformado por FV 15, FV17 y FV19, son los condensados del fondo de vacío rubiales en los colectores 2, 4 y 6, respectivamente.
- Grupo rojo: conformado por los FV 3, 7, 9, 11 y 13, representa a fondos del mismo yacimiento en el caso de los FV 9, 11 y 13, que son los condensados del fondo de vacío Tibu Zulia en los colectores 2, 4 y 6 respectivamente, sin embargo todos los fondos de este grupo presentan valores de % W de Azufre muy similar.
- Grupo café: conformado por los FV 1, 8 y 10 no presenta información contenida en el análisis elemental que permita describir el comportamiento singular de dichas muestras, la información se debe a otra propiedad fisicoquímica.

Para las demás muestras se observó distancias elevadas entre ellas, lo que impide la conformación de posibles grupos.

La aplicación de de la elipse de Hotelling T^2 , representa los dos primeros componentes principales (PC1 y PC2) los cuales contiene cierta información que permite la clasificación de las muestras a partir de su espectro FTIR-PAS (Figura 19).

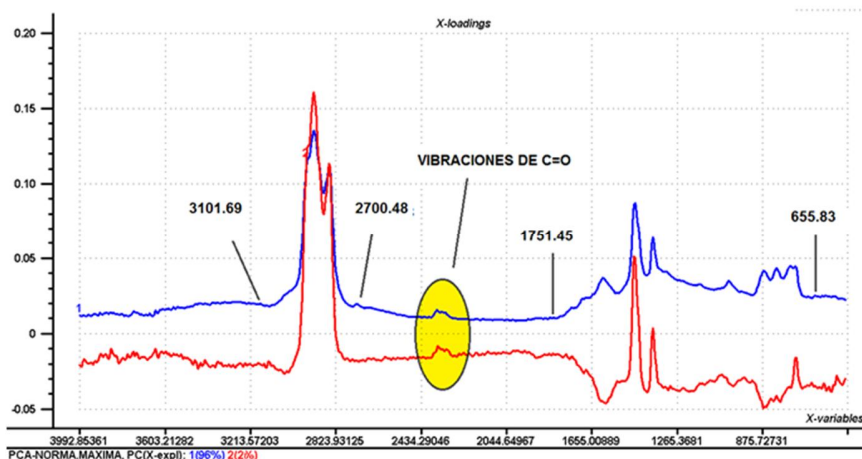
Figura 19. Descomposición por componentes principales de la información espectral para PC1 y PC2. Gráfica de Score con Hotelling T^2



Mediante la descomposición por componentes principales se identificó claramente que con el primer y segundo componente principal se puede identificar y clasificar las muestras en un grupo, además de representar el 97 % de la varianza total de las intensidades espectrales, además en la figura 19 se muestra como la posición de las muestras indica la descripción de variabilidad presentada por el componente principal 1 que corresponde a 95 % de la varianza total. Utilizando la aplicación de la elipse Hotelling T^2 , se buscaba encontrar outliers potenciales (puntos alejados que quedan fuera de la elipse) o muestras atípicas, se observó que el FV 2 se encuentra en el límite de la elipse, lo que podría describirla como un potencial outlier.

La gráfica de *X-loadings* (figura 20) se utilizó para determinar los números de onda de mayor importancia los cuales se encuentran aproximadamente en los intervalos de $3101,69 - 2700,48 \text{ cm}^{-1}$, $1751,45 - 655,83 \text{ cm}^{-1}$ del espectro PAS. Las señales presentes en $2700-1750 \text{ cm}^{-1}$ no se tuvieron en cuenta en el análisis debido a que estas regiones pertenecen usualmente a vibraciones del CO_2 presente.

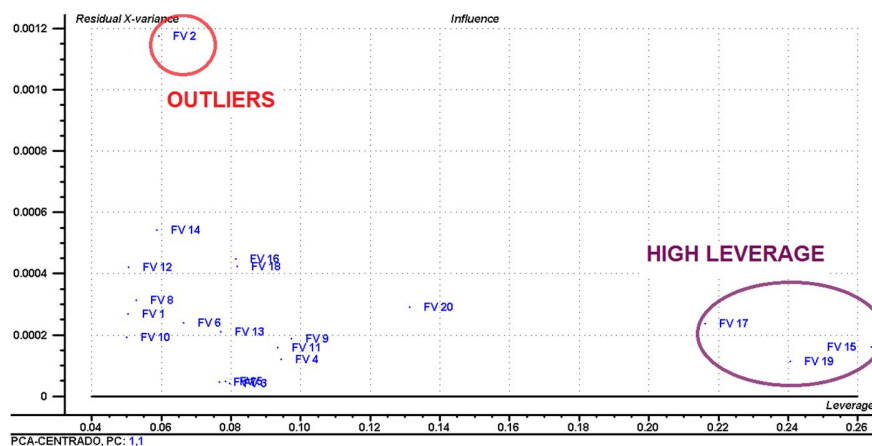
Figura 20. Gráfico de X-Loadings de PC1 y PC2



En esta figura también se observó la varianza explicada por los dos primeros PC según la información de los espectros, en el cual el 95 % de las variables contribuyen al primer componente, este componente describe en su mayoría el perfil del espectro inicial, representado por las regiones correspondientes a vibraciones de alifáticos ($3100-2700\text{ cm}^{-1}$), a la combinación de alifáticos y aromáticos y de los grupos con presencia de Azufre ($1700-900\text{ cm}^{-1}$) y en menor proporción a las vibraciones de grupos aromáticos ($900-680\text{ cm}^{-1}$). El otro 3 % de las variables están representadas por el segundo componente, sin embargo en la figura 20 se puede ver como el perfil del PC2 es muy similar al PC1 variando solo la contribución en la región de vibraciones de grupos aromáticos.

La figura 21 representa la grafica de influence, la cual permitió identificar con mucha más convicción aquellos muestras atípicas (outliers), muestras de alta influencia y posibles dangerous outliers (valores extremos peligrosos) de las muestras analizadas, que podrían de alguna manera distorsionar los modelos. Las muestras ubicadas en la parte superior izquierda de la grafica de influence (figura 21) presentan altos valores de varianza residual y pueden ser catalogadas como posibles outliers. Las muestras ubicadas en la parte inferior derecha de la grafica son catalogadas como muestras con alta influencia para los modelos denominadas muestras con alto leverage (apalancamiento), lo que significa que dichas muestras atraerán el modelo para que las describa mejor, las muestras con valores altos de leverage y varianza residual se encuentran en la parte superior derecha y se definen como dangerous outliers, lo que significa que no serán descritas por el modelo que describe la mayoría de las muestras, éstas distorsionan el modelo con el fin de estar mejor descritas.

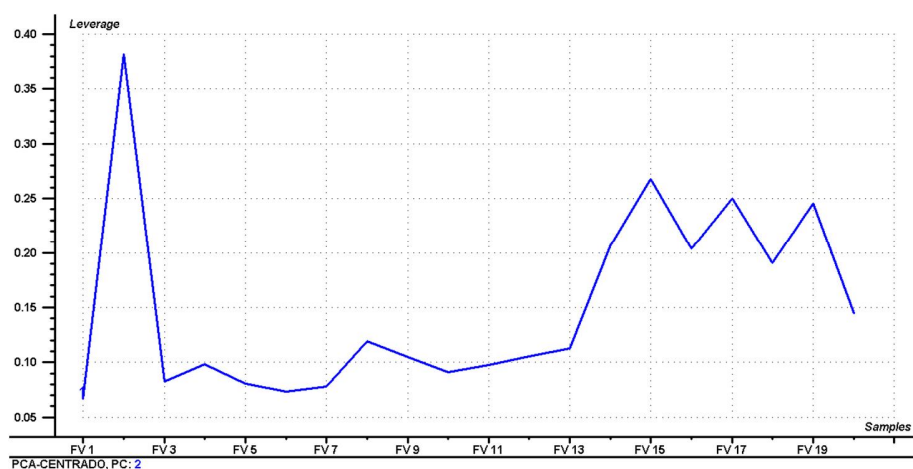
Figura 21. Gráfica de Influencia para PC1



Tomando como referencia la información suministrada por la grafica de influence se encuentra que el FV 2 presenta alta varianza residual, clasificando como posible outlier, Por ende el FV 2 debe considerarse como una posible muestra a excluir para el desarrollo del PLS. La comparación de los valores del análisis elemental no muestra indicios de la alta varianza de dicha muestra, teniendo como referencia esta propiedad fisicoquímica. Las muestras que presentan alto leverage como las muestras 15, 17 y 19 (figura 21), atraerán el modelo para que las modele mejor. La discriminación de las muestras 15, 17 y 19 pueden ser explicadas por los datos de la tabla 3, en el cual dichas muestras presentan valores más bajos del % p/p Hidrógeno y valores muy cercanos del % p/p Carbono.

En la grafica de leverage (figura 22), se evidencio más aun, el carácter negativo que representa el FV 2, presentando el valor más alto de apalancamiento, teniendo en cuenta que muestras con alto leverage perjudican los futuros modelos.

Figura 22. Grafica de Leverage para PC2 para el modelo PCA



3.2 ANALISIS DE REGRESION POR MINIMOS CUADRADOS PARACIALES (PLS-REGRESION)

Los resultados obtenidos en el análisis de regresión por mínimos cuadrados parciales, para la predicción del análisis elemental de las muestras, se muestran a continuación.

3.2.1 Modelo PLS para la predicción del % p/p Carbono

Mediante el algoritmo de regresión lineal PLS, se evaluó el potencial de la FTIR-PAS para predecir el análisis elemental en las muestras de fondos de vacío.

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

3.2.1.1 Calibración del modelo PLS

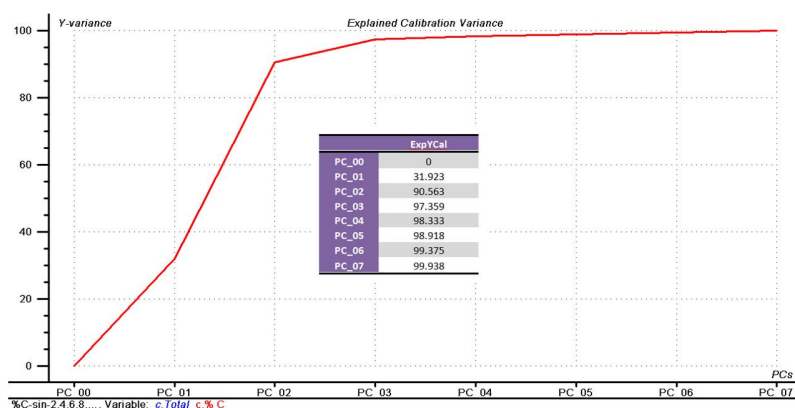
Se desarrolló el modelo de PLS para la predicción del contenido de Carbono expresado en % W, el rango de calibración en el que se desarrollaron los análisis fue de 85.7 – 86.98 % p/p Carbono. La tabla 9 muestra el error estándar de calibración (RMSEC), de validación (RMSEP), el R², la varianza explicada, las variables latentes (VL) y las muestras excluidas para el modelo de predicción teniendo como referencia los modelos obtenidos en el PCA (sección 3.1).

Tabla 9. Parámetros del modelo seleccionado para la predicción % p/p Carbono

Rangos usados (cm ⁻¹)	Muestras excluidas	VL	Varianza Explicada	RMSEC (%)	RMSEP (%)	R ²
1751.45 – 655.83 3101.69 – 2700.48	2, 4, 6, 8, 11, 12, 13, 14, 20	7	99.93 %	0.011	0.159	0.99

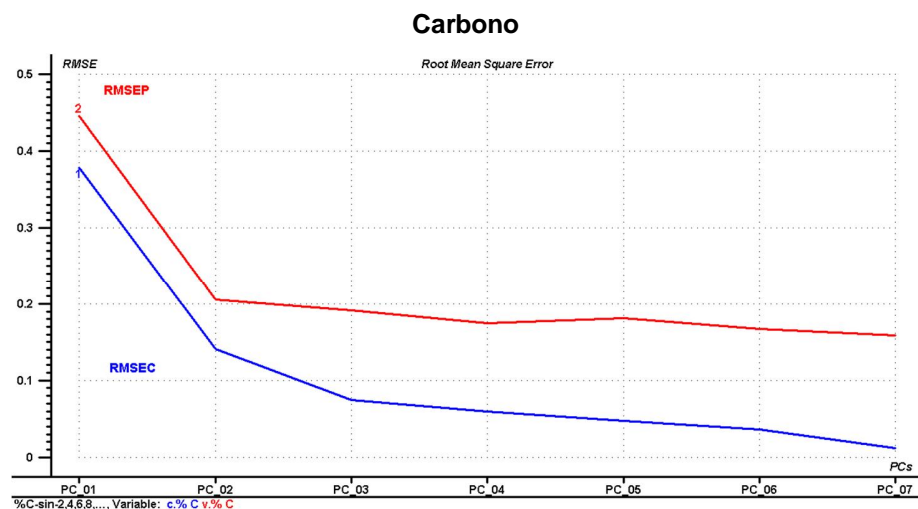
El modelo de predicción de Carbono explica: las regiones espectrales comprendidas entre 655 - 1750 cm⁻¹ y 2700 - 3101 cm⁻¹, el rango de calibración de 85.7 – 86.98 % p/p de Carbono, el 99 % en la variabilidad de los datos de referencia (variable dependiente Y) con siete variables latentes, a partir de los espectros FTIR-PAS en el MIR (Figura 23).

Figura 23. Varianza explicada en el modelo PLS para % p/p Carbono



La selección del número óptimo de componentes requirió evaluar los parámetros estadísticos RMSEC y RMSEP, los errores estándar de calibración y predicción respectivamente. El primero de estos parámetros es una medida de la desviación estándar de los residuales obtenidos por la diferencia entre los valores observados y predichos por el modelo para las muestras de calibración. El segundo parámetro se basa en un algoritmo iterativo que selecciona muestras dentro del grupo de calibración para desarrollar el modelo de predicción y posteriormente lo evalúa sobre las muestras restantes. (Figura 24).

Figura 24. Error calculado en función del número de componentes para el modelo PLS % p/p



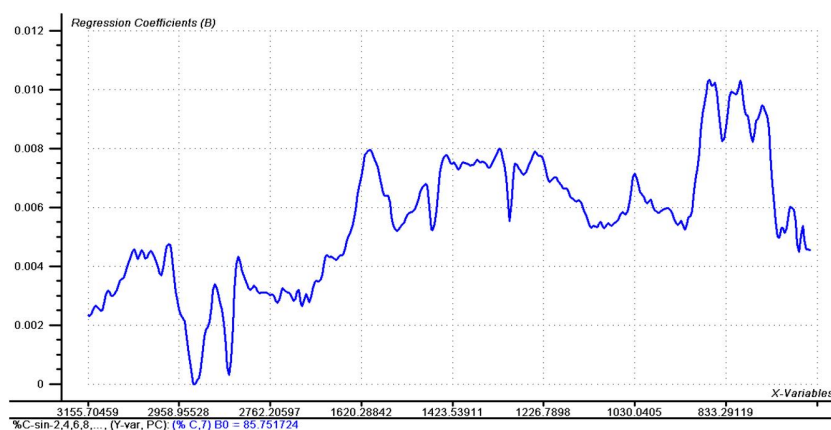
El modelo, mostró como el RMSEC disminuye de manera continua desde el segundo componente, además el RMSEP presenta un leve aumento del cuarto al quinto componente y de ahí en adelante disminuye hasta el séptimo componente (figura 24).

El error estándar en la etapa de calibración y validación disminuye apreciablemente desde el tercer componente, donde alcanza valores de 0.07 % y

0.19 % respectivamente hasta el séptimo componente donde los valores son 0.011 % y 0.159 % respectivamente.

La grafica de coeficientes de regresión (*Regression coefficients*, figura 25), muestra los coeficientes hallados para la séptima componente principal, ya que esta explica aproximadamente el 99 % de la variabilidad acumulada en los datos. Cada componente principal es una combinación lineal de las absorbancias medidas en las diferentes frecuencias del rango espectral, multiplicadas por un coeficiente de regresión que determina el peso o influencia de las frecuencias sobre el componente.

Figura 25. Coeficiente de regresión para el primer componente principal PLS % p/p Carbono



3.2.1.2 Validación cruzada del modelo PLS

La validación del modelo PLS para la predicción del % p/p Carbono en Fondos de vacío se desarrolló empleando el mismo grupo de muestras de fondos utilizadas para generar el modelo de calibración. Sobre los datos completos de los espectros FTIR-PAS con los pretratamientos escogidos en las regiones seleccionadas anteriormente gracias a los análisis de PCA, se aplicó el modelo descrito en la sección 3.2.1.1 (modelo predicción % p/p Carbono de los fondos de vacío).

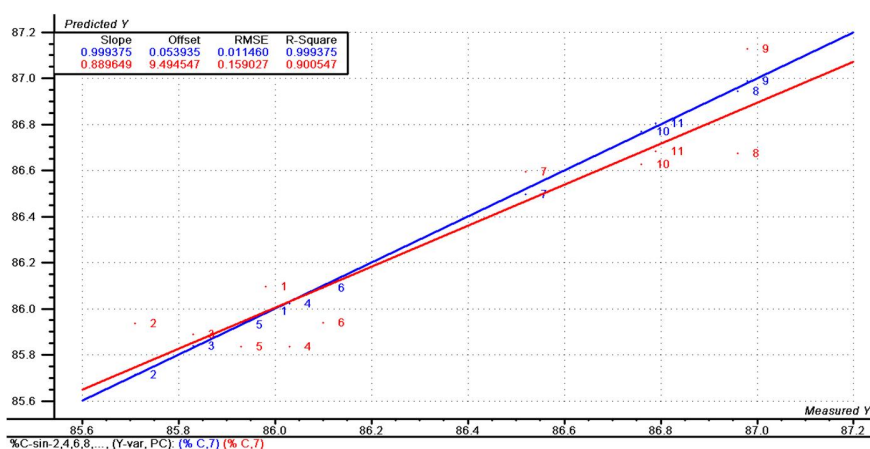
Los valores predichos por el modelo propuesto y los valores de referencia presentan valores muy similares entre sí, comprobando un óptimo desempeño por parte del modelo propuesto. (Tabla 10).

Tabla 10. Validación cruzada del modelo de predicción % p/p Carbono

Muestra	Referencia % C	Predicho % C	Residual	Error relativo
FV 1	85.98	86.097	0.117	0.136
FV 3	85.71	85.936	0.226	0.264
FV 5	85.83	85.89	0.06	0.069
FV 7	86.03	85.836	0.194	0.226
FV 9	85.93	85.836	0.094	0.109
FV 10	86.1	85.94	0.16	0.186
FV 15	86.52	86.595	0.075	0.087
FV 16	86.96	86.674	0.286	0.329
FV 17	86.98	87.127	0.147	0.169
FV 18	86.76	86.628	0.132	0.152
FV 19	86.79	86.683	0.107	0.123

El coeficiente de correlación entre los valores de referencia y de predicción hallado fue de 0.99 y 0.94 respectivamente, tanto para la curva de calibración como para la de validación. (Figura 26).

Figura 26. Curva de calibración y validación del modelo de predicción % p/p Carbono



El error estándar de predicción calculado para el conjunto de muestras de validación fue de 0.16.

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

3.2.1.3 Análisis de muestras excluidas

Se eliminaron del modelo nueve muestras (FV 2, 4, 6, 8, 11, 12, 13, 14 y 20). En el caso de los FV 6, 11, 13 y 20 porque presentaban igual valor en % p/p Carbono que otras muestras del modelo (tabla 11), lo que les daba más peso a la hora de predecir el modelo. El FV 4, 12 y 14 se eliminaron porque el contenido de Carbono en estos fondos es muy bajo en caso del FV 4 y 12 y demasiado alto en el FV 14 alejándolas del promedio de las demás muestras (86.23 % p/p Carbono). Mientras que las muestras presentan % p/p Carbono del 85.5, 85.47, 87.34 % p/p Carbono respectivamente (Tabla 11).

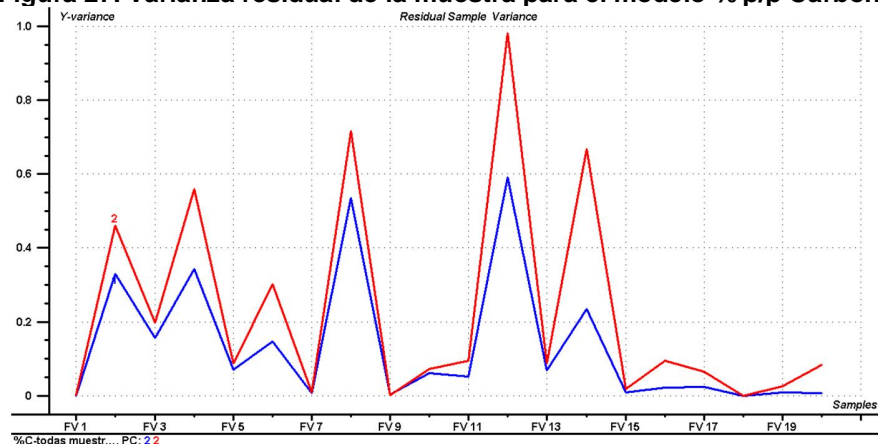
Tabla 11. Distribución % p/p Carbono en las muestras

% p/p Carbono	Muestra (FV)
85.47	12
85.5	4
85.7	6 – 11
85.71	3
85.83	5
85.93	9
85.98	1
86.03	7
86.1	10 – 13
86.22	2

86.52	15 – 20
86.62	8
86.76	18
86.79	19
86.98	17
87.34	12
86.23	Promedio

Sin embargo este no fue el único criterio que se tuvo en cuenta para la eliminación de estas variables, ya que son muestras que no explican el modelo, se consideraron atípicas como se muestra en el gráfico de la varianza residual de la muestra realizada como análisis exploratorio (*Residual sample Variance* figura 27).

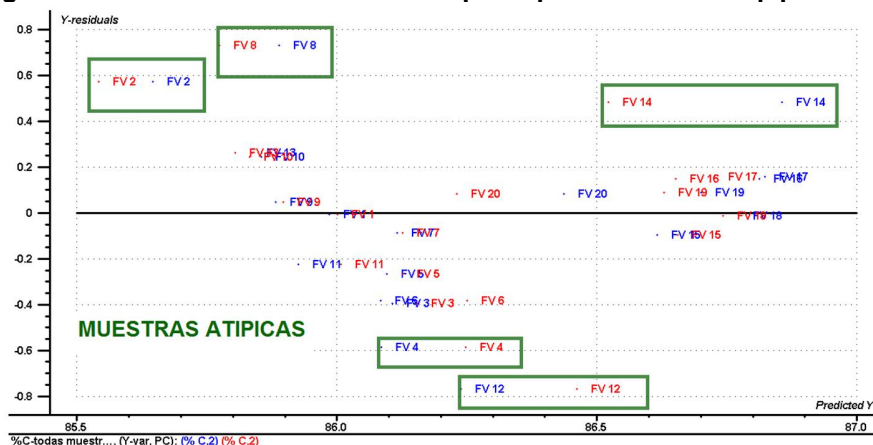
Figura 27. Varianza residual de la muestra para el modelo % p/p Carbono



En el caso del FV 2 presenta un valor de % p/p Carbono muy parecido al del promedio de las muestras (tabla 11), además esta muestra presentó indicios de ser un posible outlier (figura 21), Otro criterio empleado fue la figura 28 (gráfica Y residual Vs predited Y), esta gráfica permitió detectar valores atípicos, que conllevan a la falta de ajuste del modelo, debidos principalmente a errores sistemáticos. Como el caso del FV 8, el cual presenta inconsistencia con el modelo, dicha inadaptabilidad al modelo no es debida a la propiedad fisicoquímica

modelada en este caso el % p/p Carbono, sino debido a otras propiedades fisicoquímicas.

Figura 28. Determinación de valores atípicos para el modelo % p/p Carbono



3.2.2 Modelo PLS para la predicción del % p/p Hidrógeno

3.2.2.1 Calibración del modelo PLS

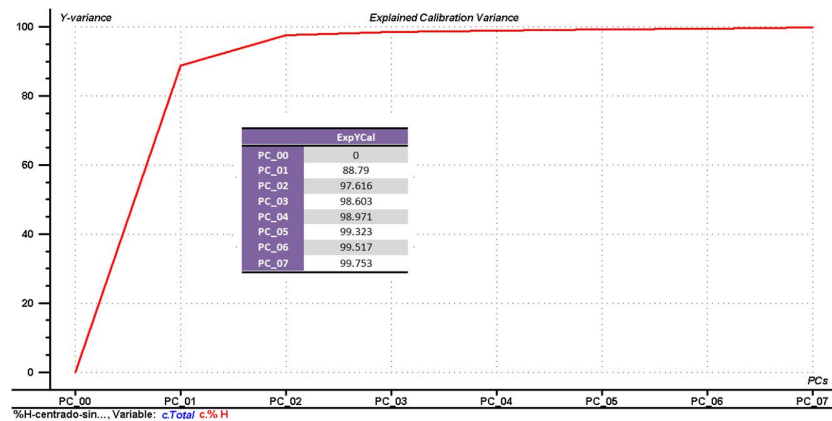
Se desarrolló el modelo de PLS para la predicción del contenido de Hidrógeno expresado en % p/p, el rango de calibración en el que se desarrollaron los análisis fue de 8.6 – 12.56 % p/p Hidrógeno. La tabla 12 muestra el error estándar de calibración (RMSEC), de validación (RMSEP), el R^2 , la varianza explicada y las muestras excluidas para el modelo de predicción.

Tabla 12. Parámetros del modelo seleccionado para la predicción de % p/p Hidrógeno

Rangos usados (cm^{-1})	Muestras excluidas	VL	Varianza Explicada	RMSEC (%)	RMSEP (%)	R^2
1751.45 – 655.83 3101.69 – 2700.48	1, 2, 4, 8	5	99.32 %	0.106	0.293	0.99

El modelo de predicción de Hidrógeno explica: las regiones espectrales comprendidas entre 655 - 1750 cm^{-1} y 2700 - 3101 cm^{-1} , el rango de calibración de 8.6 – 12.56 % p/p Hidrógeno, el 99 % en la variabilidad de los datos de referencia (variable dependiente Y) con un mínimo de cinco variables latentes, a partir de los espectros FTIR-PAS en el MIR (Figura 29).

Figura 29. Varianza explicada en el modelo PLS para % W Hidrógeno



Se determinó el número óptimo de componentes teniendo en cuenta la información de la Figura 30. La cual mostró las relaciones entre el RMSEC y el RMSEP

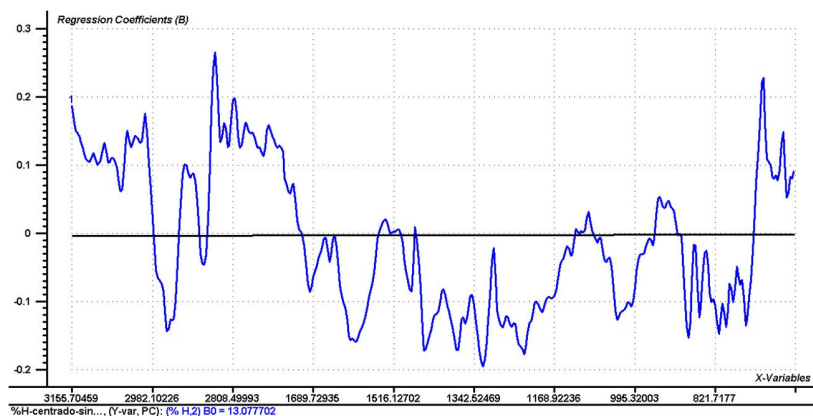
Figura 30. Error calculado en función del número de componentes para el modelo PLS % W Hidrógeno



El error estándar en la etapa de calibración y validación cruzada disminuye apreciablemente hasta el segundo componente, donde alcanza 0.19 % y 0.28 %, A partir del segundo componente el error de calibración sigue disminuyendo de forma constante, para el caso del error estándar de predicción se observó variaciones tanto de aumento como de disminución en el error estándar de predicción, sin embargo se determinó que el número de componentes óptimo era cinco, ya que el error estándar de calibración alcanza un valor 0.11 y el error estándar de predicción presenta valor similar al de la segunda componente 0.29 por tal motivo se decidió trabajar con 5 componentes (figura 30).

La grafica de coeficientes de regresión muestra como los coeficientes de regresión calculados para la segunda componente principal, explica la mayor variabilidad acumulada aproximadamente el 98 % de la variabilidad presente en los datos. Se determino que la región espectral donde se producen las vibraciones alifáticas del C-H, presenta un efecto positivo para la predicción del % p/p Hidrógeno, mientras que la región correspondiente a las vibraciones aromáticas presenta coeficientes de regresión negativos (figura 31),

Figura 31. Coeficiente de regresión para el primer componente principal PLS % W Hidrógeno



3.2.2.2 Validación Cruzada del modelo PLS

La validación del modelo PLS para la predicción del % p/p Hidrógeno en Fondos de vacío se desarrolló empleando el mismo grupo de muestras de fondos utilizadas para generar el modelo de calibración.

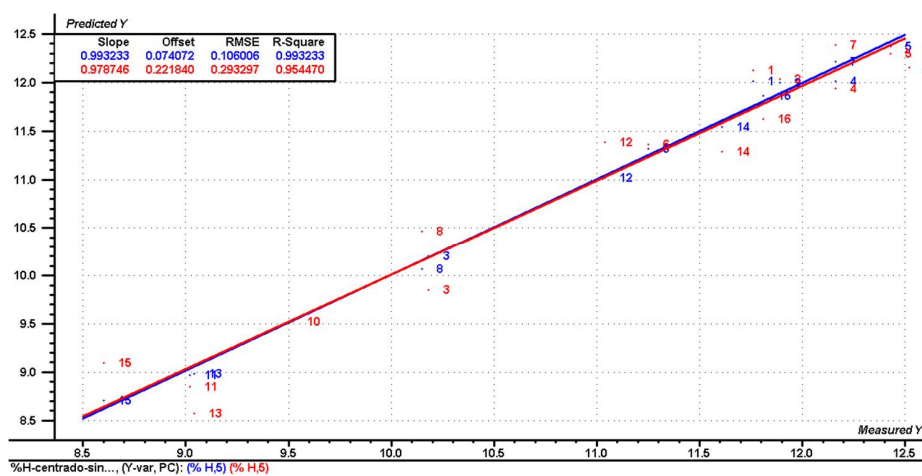
Los valores predichos por el modelo propuesto y los valores de referencia presentan valores muy similares entre sí, lo que comprueba un desempeño favorable por parte del modelo para predecir el % p/p Hidrógeno en los fondos de vacío. (Tabla 13).

Tabla 13. Validación cruzada del modelo de predicción % p/p Hidrógeno

Muestra	Referencia % H	Predicho % H	Residual	Error relativo
FV 3	11.76	12.123	0.363	3.087
FV 5	11.89	12.037	0.147	1.236
FV 6	10.18	9.852	0.328	3.222
FV 7	12.16	11.94	0.22	1.809
FV 9	12.43	12.301	0.129	1.037
FV 10	11.25	11.36	0.11	0.978
FV 11	12.16	12.391	0.231	1.899
FV 12	10.15	10.466	0.316	3.113
FV 13	12.52	12.16	0.36	2.875
FV 14	9.52	9.523	0.003	0.031
FV 15	9.02	8.848	0.172	1.907
FV 16	11.04	11.384	0.344	3.116
FV 17	9.04	8.569	0.471	5.210
FV 18	11.61	11.29	0.32	2.756
FV 19	8.6	9.098	0.498	5.791
FV 20	11.81	11.624	0.186	1.575

El coeficiente de correlación entre los valores de referencia y de predicción hallado fue superior a 0.97 tanto para la curva de calibración como para la de validación. (Figura 32). También se debe tener en cuenta que el elevado valor del porcentaje de error del FV 19 y 17 se debe a que están muy lejos del promedio de los datos se alejan demasiado de las otras muestras.

Figura 32. Curva de calibración y validación del modelo de predicción % p/p Hidrógeno

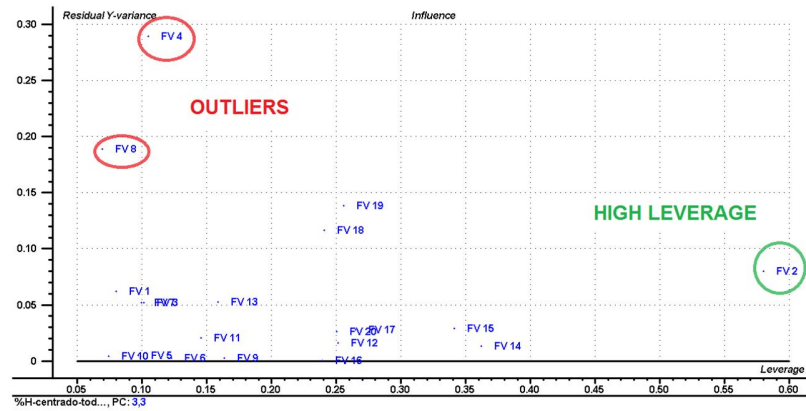


El error estándar de predicción calculado para el conjunto de muestras de validación fue de 0.28.

3.2.2.3 Análisis de muestras excluidas

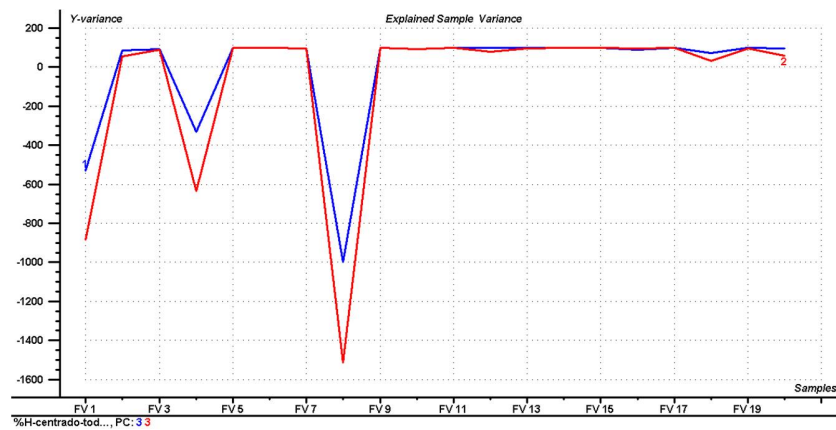
Se eliminaron del modelo nueve muestras (FV 1, 2, 4, 8). El FV 2, se eliminó, porque está descrito como posible outlier en el análisis PCA (figura 20), y además en la gráfica de influencia del modelo PLS con todas las muestras se corroboró el carácter negativo de este FV para el desarrollo del modelo ya que presentaba un alto leverage, lo que conllevaría, al modelo a predecirla de una mejor manera, pero afectaría la predicción del resto de muestras (figura 33).

Figura 33. Gráfica de Influencia para modelo predicción % p/p Hidrógeno



En la figura 33, se pudo ver que los FV 4 y 8, eran clasificados como outliers, para el modelo de predicción % p/p Hidrógeno. En el caso del FV 1 se empleó la gráfica de la varianza explicada de la muestra, la cual permitió encontrar muestras atípicas como el caso del FV 1, y ratificó que el FV 4 y 8 eran muestras que no explicaban el modelo (figura 34).

Figura 34. Varianza explicada de la muestra para el modelo % p/p Hidrógeno



3.2.3 Modelo PLS para la predicción del % p/p Nitrógeno

3.2.3.1 Calibración del modelo PLS

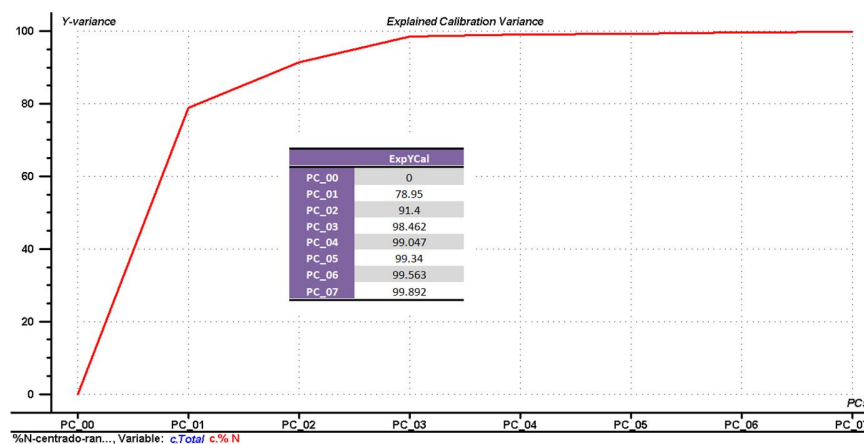
El rango de calibración en el que se desarrollaron los análisis fue de 0.344 – 0.932 % p/p Nitrógeno. La tabla 14 muestra el error estándar de calibración (RMSEC), de validación (RMSEP), el R^2 , la varianza explicada y las muestras para el modelo de predicción teniendo como referencia los modelos obtenidos en el PCA (sección 3.1).

Tabla 14. Parámetros estadísticos del modelo desarrollado para la predicción de % p/p Nitrógeno

Rangos usados (cm-1)	Muestras excluidas	VL	Varianza Explicada	RMSEC (%)	RMSEP (%)	R^2
1530 - 1200	1, 2, 9, 11, 13	3	98.46 %	0.026	0.039	0.98

El modelo de predicción de Nitrógeno explica: las regiones espectrales comprendidas entre 1200 - 1530 cm^{-1} , esta región se escogió debido a que es donde se presentan las vibraciones correspondientes al enlace Aril-N (tabla 5), el rango de calibración de 0.344 a 0.932 % p/p Nitrógeno, el 99 % en la variabilidad de los datos de referencia (variable dependiente Y) con un mínimo de tres variables latentes, a partir de los espectros FTIR-PAS en el MIR (Figura 35).

Figura 35. Varianza explicada en el modelo PLS para % W Nitrógeno



Se determinó el número óptimo de componentes teniendo en cuenta la información de la Figura 36. La cual mostró las relaciones entre el RMSEC y el RMSEP.

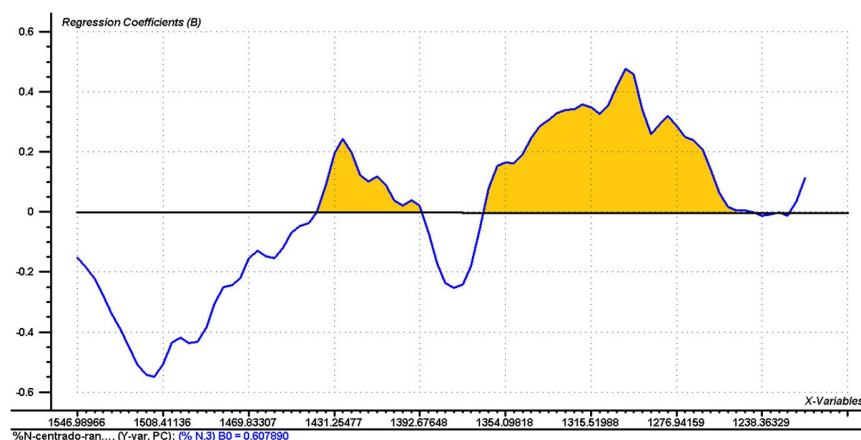
Figura 36. Error calculado en función del número de componentes para el modelo PLS % W Nitrógeno



El error estándar en la etapa de calibración y validación cruzada disminuye apreciablemente hasta el tercer componente donde alcanza los valores mínimos de 0.026 y 0.038 % respectivamente (figura 36).

La grafica de coeficientes de regresión muestra como los coeficientes de regresión calculados para la tercera componente principal, explica la mayor variabilidad acumulada aproximadamente el 99 % de la variabilidad presente en los datos. Se determino que toda la región espectral sombreada presenta un efecto positivo para la predicción del % p/p Nitrógeno (figura 37).

Figura 37. Coeficiente de regresión para el primer componente principal PLS % p/p Nitrógeno



3.2.3.2 Validación Cruzada del modelo PLS

La validación del modelo PLS para la predicción del % p/p Nitrógeno en Fondos de vacío se desarrolló empleando el mismo grupo de muestras de fondos utilizadas para generar el modelo de calibración.

Como se observa en la tabla 15, los valores predichos por el modelo propuesto y los valores de referencia presentan valores muy similares entre sí en el modelo para predecir el % p/p Nitrógeno en los fondos de vacío. (Tabla 15).

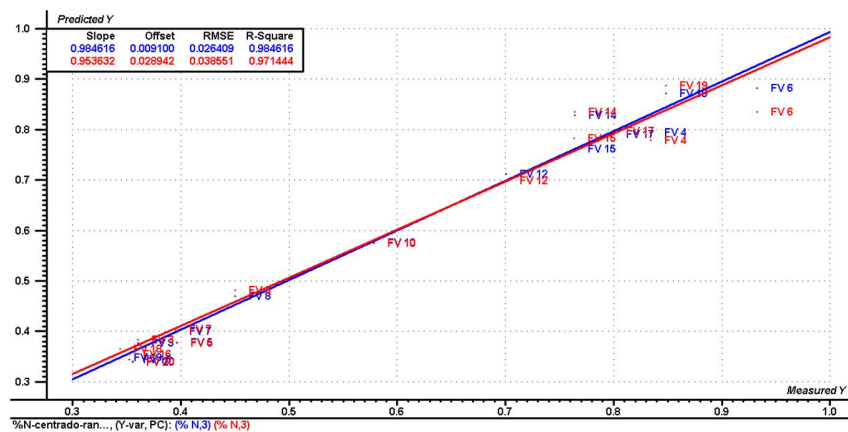
Tabla 15. Validación cruzada del modelo de predicción % p/p Nitrógeno

Muestra	Referencia % N	Predicho % N	Residual	Error relativo
FV 3	0.36	0.384	0.024	6.667

FV 4	0.834	0.779	0.055	6.595
FV 5	0.396	0.377	0.019	4.798
FV 6	0.932	0.835	0.097	10.408
FV 7	0.395	0.404	0.009	2.278
FV 8	0.45	0.482	0.032	7.111
FV 10	0.579	0.574	0.005	0.864
FV 12	0.701	0.699	0.002	0.285
FV 14	0.764	0.835	0.071	9.293
FV 15	0.763	0.783	0.02	2.621
FV 16	0.352	0.355	0.003	0.852
FV 17	0.799	0.797	0.002	0.250
FV 18	0.344	0.365	0.021	6.105
FV 19	0.848	0.888	0.04	4.717
FV 20	0.355	0.339	0.016	4.507

La validación cruzada del modelo mostró un desempeño favorable para la predicción de % p/p de Nitrógeno, ya que el coeficiente de correlación entre los valores de referencia y de predicción fue hallado superior al 0.99 para la curva de calibración y superior a 0.98 para la curva de validación (figura 38).

Figura 38. Curva de calibración y validación del modelo de predicción % p/p Nitrógeno

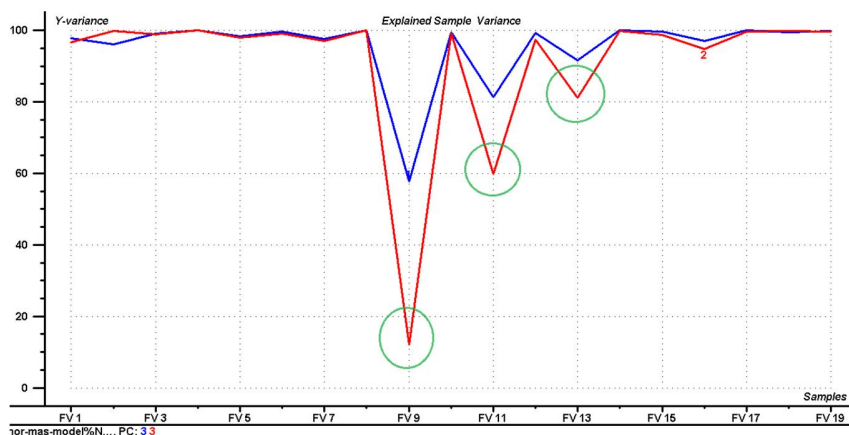


El error estándar de predicción calculado para el conjunto de muestras de validación fue de 0.039.

3.2.3.3 Análisis de muestras excluidas

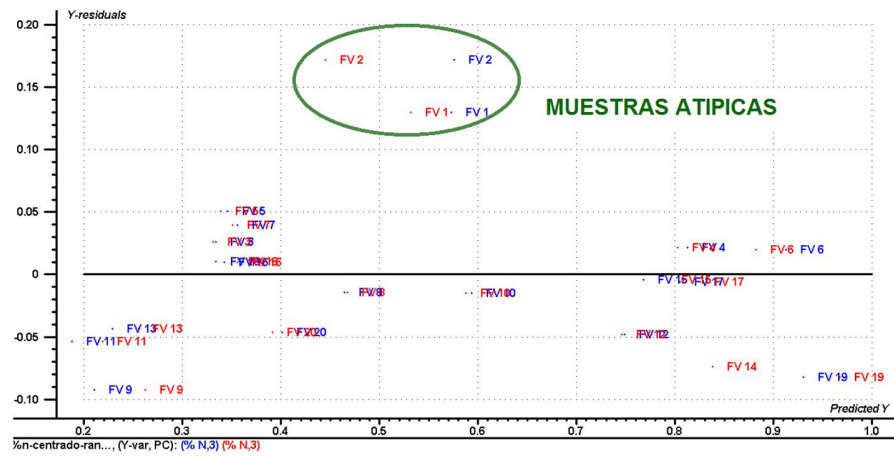
Se eliminaron del modelo cinco muestras (FV 1, 2, 9, 11, 13). En el caso de los FV 9, 11, 13 porque el contenido de Nitrógeno en estos fondos es el más bajos de las muestras y se encuentra muy por debajo del promedio del contenido de las demás muestras (0.548 % p/p Nitrógeno). Mientras que las muestras presentan % de Nitrógeno del 0.1181, 0.1339 y 0.1852 respectivamente. Igualmente estos tres fondos tiene en común la misma naturaleza, ya que los tres provienen del fondo de vacio Tibu Zulia, son sus fracciones condensadas obtenidas por la destilación molecular a dicho fondo del colector 2, 4, 6 respectivamente (tabla 3). Sin embargo este no fue el único criterio que se tuvo en cuenta para la eliminación de estas variables, ya que son muestras que no explican el modelo, se consideraron atípicas como se muestra en el grafico de la varianza explicada de la muestra realizada como análisis exploratorio (*Explained sample Variance* figura 39).

Figura 39. Varianza explicada de la muestra para el modelo % p/p Nitrógeno



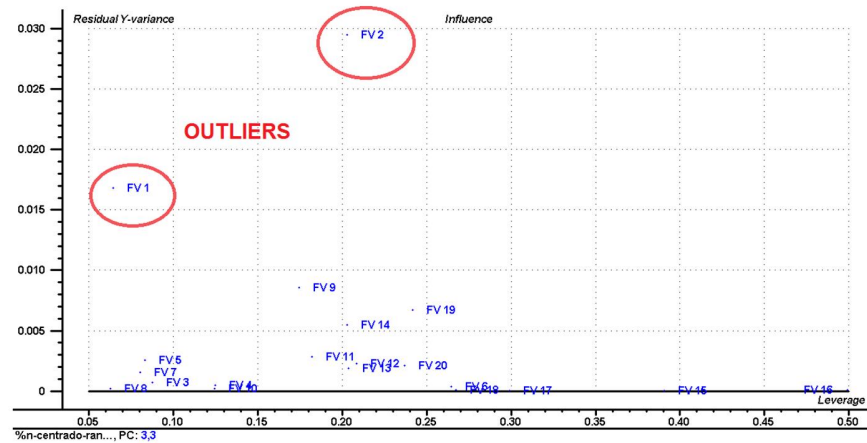
La figura 40, (grafica *Y residual Vs predited Y*), esta grafica permitió detectar valores atípicos, que conllevan a la falta de ajuste del modelo, debidos principalmente a errores sistemáticos. Entre los que se encuentra el FV 1 y 2.

Figura 40. Muestras atípicas para el desarrollo del modelo % p/p Nitrógeno



También se tuvo en cuenta la grafica de influencia (figura 41), que permitió corroborar que los FV 1 y 2, eliminados del modelo, eran clasificados como outliers, para el modelo de predicción % p/p Nitrógeno.

Figura 41. Gráfica de Influencia para modelo predicción % p/p Nitrógeno



3.2.4 Modelo PLS para la predicción del % p/p Azufre

3.2.4.1 Calibración del modelo PLS

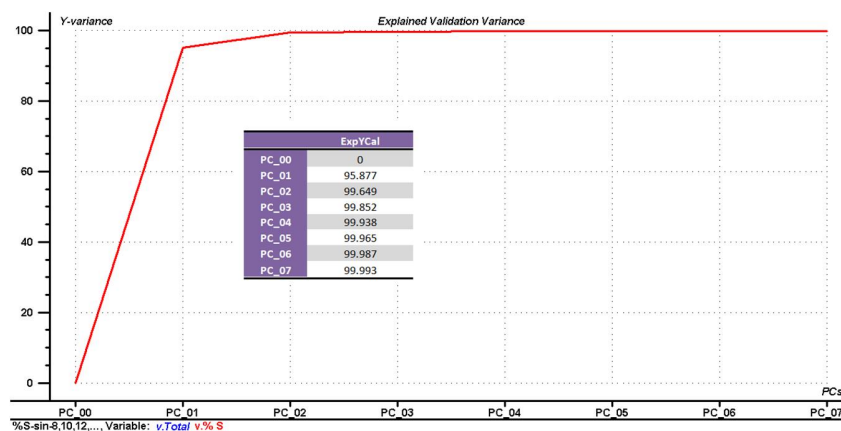
El rango de calibración en el que se desarrollaron los análisis fue de 1.39 – 2.29 % p/p de Azufre. La tabla 16 muestra el error estándar de calibración (RMSEC), de validación (RMSEP), el R^2 , la varianza explicada y las muestras excluidas para el modelo de predicción teniendo como referencia los modelos obtenidos en el PCA (sección 3.1).

Tabla 16. Parámetros estadísticos del modelo desarrollado para la predicción % p/p Azufre

Rangos usados (cm^{-1})	Muestras excluidas	VL	Varianza Explicada	RMSEC (%)	RMSEP (%)	R^2
1500,69 – 999,17	8, 10, 12, 14, 16	4	99.94 %	0.044	0.085	0.99

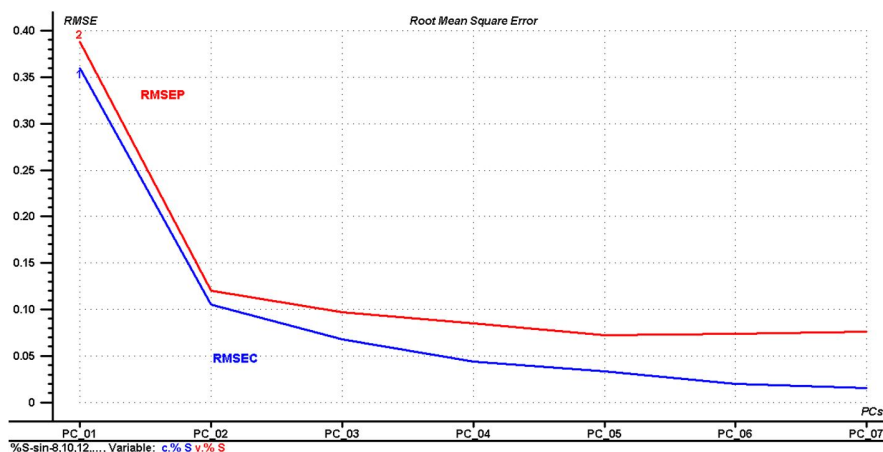
El modelo de predicción de Azufre explica: las regiones espectrales comprendidas entre 999 - 1500 cm^{-1} , la reducción del rango espectral se debió a que el Azufre presente en los fondos de vacío presenta vibraciones en el MIR en este rango de longitudes de onda (Tabla 5). Con este rango espectral se obtuvo un rango de calibración de 1.39 a 2.29 % p/p Azufre, en el cual el 99 % en la variabilidad de los datos de referencia (variable dependiente Y) con un mínimo de cuatro variables latentes, a partir de los espectros FTIR-PAS en el MIR (Figura 42).

Figura 42. Varianza explicada en el modelo PLS para % p/p Azufre



El número óptimo de componentes se determinó teniendo en cuenta la información de la figura 43. La cual mostró las relaciones entre el RMSEC y el RMSEP.

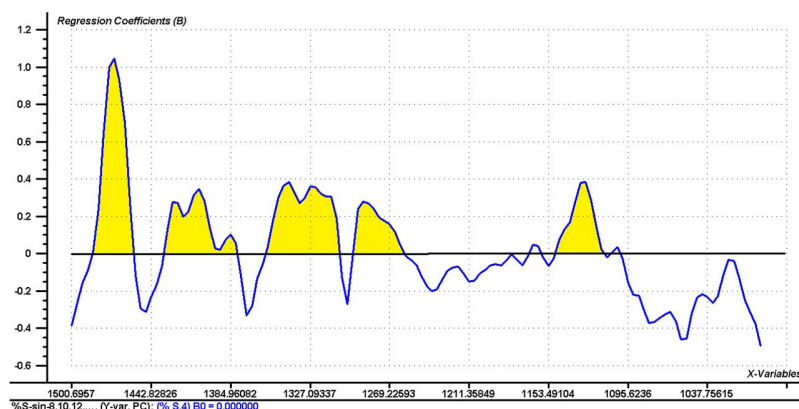
Figura 43. Error calculado en función del número de componentes para el modelo PLS % p/p Azufre



El error estándar en la etapa de calibración y validación cruzada disminuye apreciablemente hasta el tercer componente donde alcanza los valores mínimos de 0.044 % y 0.085 % respectivamente (figura 43).

La grafica de coeficientes de regresión mostró como dichos coeficientes calculados para la cuarta componente principal, explica la mayor variabilidad acumulada aproximadamente el 99 % de la variabilidad acumulada presente en los datos. Se determino que toda la región espectral sombreada presenta un efecto positivo para la predicción del % Azufre, ya que presenta las vibraciones mostradas anteriormente (tabla 5, vibraciones Azufre), que son las que aportan para el desarrollo del modelo, las vibraciones de los grupos funcionales presentes se encuentran sombreadas (figura 44).

Figura 44. Coeficiente de regresión para el primer componente principal PLS % p/p Azufre



3.2.4.2 Validación Cruzada del modelo PLS

La validación del modelo PLS para la predicción del % p/p de Azufre en Fondos de vacío se desarrolló empleando el mismo grupo de muestras de fondos utilizadas para generar el modelo de calibración.

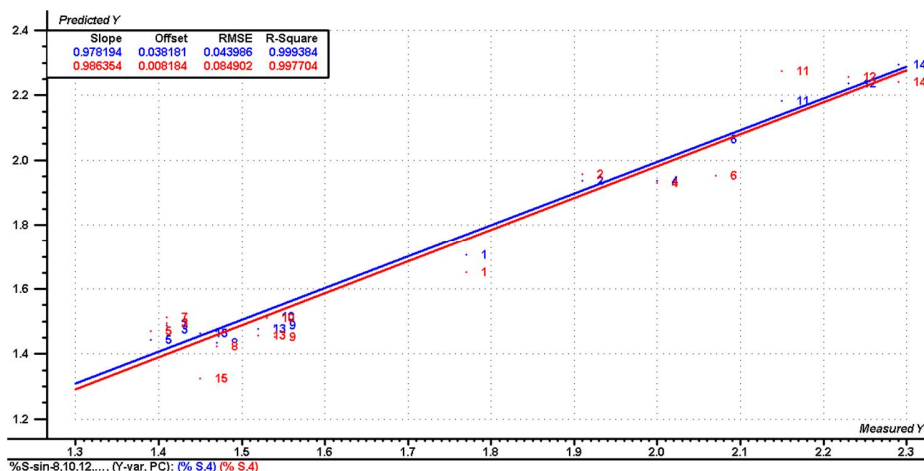
Los valores predichos por el modelo propuesto y los valores de referencia presentan valores muy similares entre sí, comprobando un óptimo desempeño por parte del modelo propuesto. (Tabla 17).

Tabla 17. Validación cruzada del modelo de predicción % p/p Azufre

Muestras	Referencia % S	Predicho % S	Residual	Error relativo
FV 1	1.77	1.673	0.097	5.480
FV 2	1.91	2.022	0.097	5.480
FV 3	1.41	1.447	0.112	5.864
FV 4	2	1.949	0.037	2.624
FV 5	1.39	1.402	0.051	2.550
FV 6	2.07	2.079	0.012	0.863
FV 7	1.41	1.465	0.009	0.435
FV 9	1.47	1.378	0.055	3.901
FV 11	1.54	1.499	0.092	6.259
FV 13	1.53	1.48	0.041	2.662
FV 15	2.15	2.265	0.05	3.268
FV 17	2.23	2.265	0.115	5.349
FV 18	1.52	1.434	0.035	1.569
FV 19	2.29	2.177	0.086	5.658
FV 20	1.45	1.37	0.113	4.934

La validación cruzada del modelo mostró un desempeño aceptable para la predicción de % p/p de Azufre, ya que el coeficiente de correlación entre los valores de referencia y de predicción fue hallado superior al 0.99 para la curva de calibración y superior a 0.98 para la curva de validación (figura 45).

Figura 45. Curva de calibración y validación del modelo de predicción % p/p Azufre

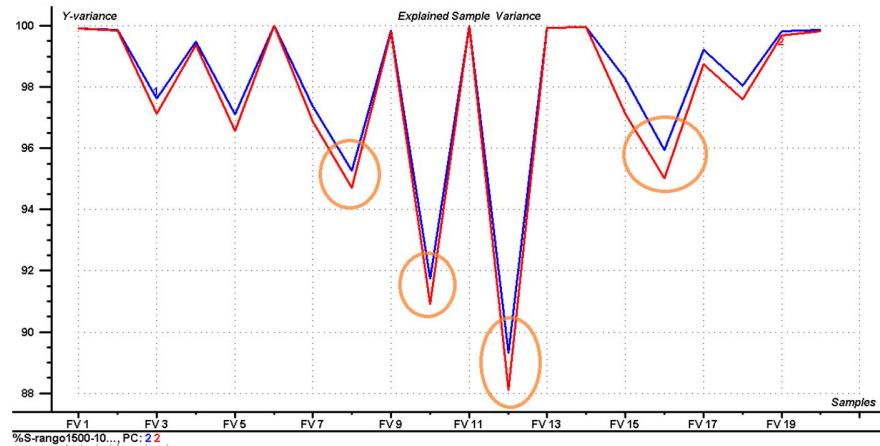


El error estándar de predicción calculado para el conjunto de muestras de validación fue de 0.086.

3.2.4.3 Análisis de muestras excluidas

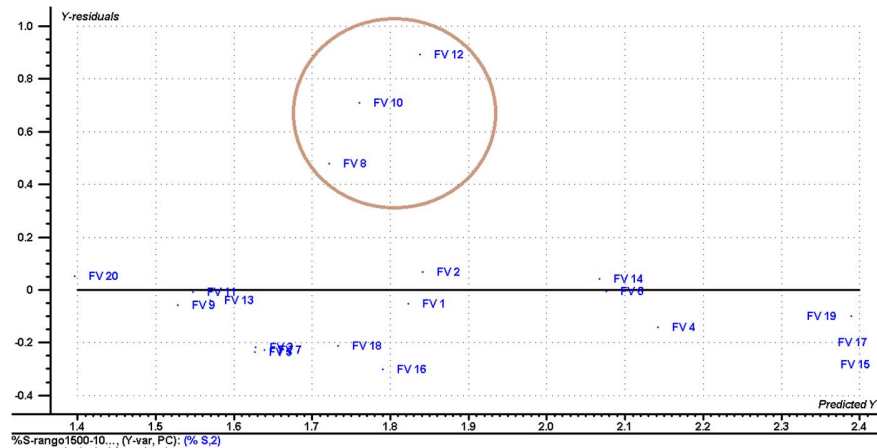
Se eliminaron del modelo cinco muestras (FV 8, 10, 12, 14, 16). En el caso de los FV 8, 10, 12 y 16. Se empleó como criterio la grafica de varianza explicada por cada muestra generado por el modelo PLS para la predicción de % Azufre sin eliminar ninguna Fondo en el rango 1500 a 1000 cm^{-1} , la grafica permite la detección de muestras atípicas que no aportan a la explicación de la variabilidad del modelo (figura 46).

Figura 46. Varianza explicada de la muestra para el modelo % p/p Azufre



Por demás los FV 8,10 y 12 se excluyeron ya que en la grafica de *Y residual Vs predited Y* (figura 47), fueron detectados como valores atípicos, que conllevan a la falta de ajuste del modelo, debidos principalmente a errores sistemáticos.

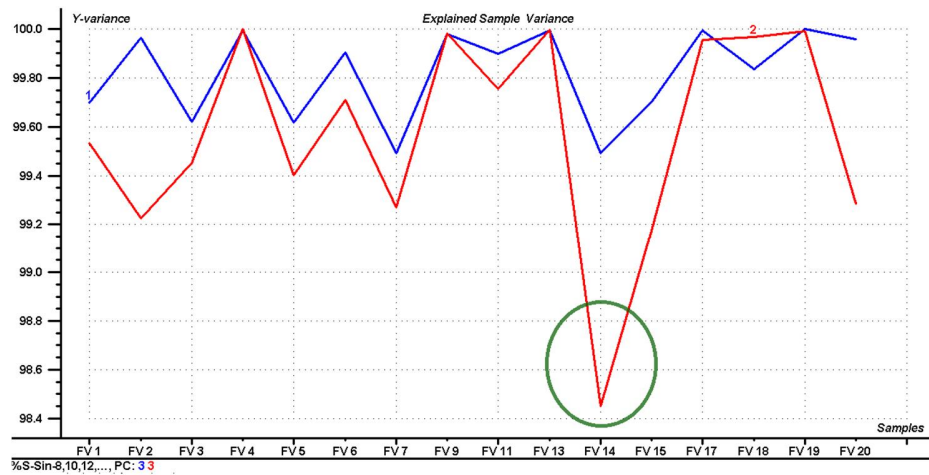
Figura 47. Muestras atípicas para el desarrollo del modelo % p/p Azufre



Se realizó un modelo tomando teniendo en cuenta la exclusión de las muestras atípicas (FV 8, 10, 12, 16), en la región de 1500 a 1000 cm^{-1} . Se empleó como criterio la grafica de varianza explicada por cada muestra generado por el modelo PLS para la predicción de % Azufre, la grafica permite la detección del FV 14

como muestra atípicas que no aportan a la explicación de la variabilidad del modelo (figura 48).

Figura 48. Varianza explicada de la muestra modelo sin FV 8, 10, 12, 16 % p/p Azufre



4. CONCLUSIONES

La espectroscopia FTIR-PAS combinada con técnicas de calibración multivariable como el PCA y PLS, permitió clasificar y analizar fondos de vacío para la obtención de modelos que pueden ser una alternativa, a la técnicas empleadas comúnmente (Normas ASTM); con la ventaja de evitar el pretratamiento de la muestra e incrementar la reproducibilidad en los resultados en un menor tiempo de análisis.

El análisis por componentes principales (PCA) permitió reducir la dimensionalidad de los datos con el uso de 3 componentes principales, las cuales explicaron aproximadamente el 100 % de la variabilidad de los datos. Además el PCA clasificó las muestras por el contenido de Azufre y por el yacimiento de procedencia, Se pudo determinar las regiones espectrales de mayor importancia, ubicadas en la región de 655 – 1750 y 2700 – 3100 cm^{-1} .

La espectroscopia fotoacústica en la región del infrarrojo medio se consolidó como alternativa de análisis de los fondos de vacío y sus fracciones, ya que permitió hacer un estudio de muestras sólidas y opacas, empleado un menor tiempo que las técnicas convencionales y ahorrando el uso de reactivos.

El modelo de regresión parcial por mínimos cuadrados (PLS) aplicado los espectros de FTIR-PAS mostró un desempeño satisfactorio en los modelos de predicción del análisis elemental expresado en % p/p de Hidrógeno, Nitrógeno y Azufre demostrando que los datos espectroscópicos pueden proporcionar información detallada de estas y otras propiedades de las muestras analizadas si se aplica un adecuado tratamiento quimiométrico.

El modelo para la predicción % p/p Carbono proporcionó los menores porcentajes de error, sin embargo fue el modelo con menor valor de correlación debido a que el rango de calibración fue muy pequeño.

Las graficas de coeficientes de regresión mostraron ser muy útiles para la escogencia de las regiones espectrales en el caso de los modelos de Nitrógeno y Azufre, ya que estos no fueron descritos muy bien por el rango espectral propuesto por el PCA, debido a que el aporte de estos elementos a la composición es mínimo.

Para el desarrollo de modelos más robustos es indispensable ampliar el rango de calibración de las muestras, ya que debido al estrecho margen de variabilidad de los valores teóricos de las muestras, conlleva a la eliminación de más de 3 muestras en el desarrollo de los modelos.

Para la posterior aplicación de los modelos desarrollados, se deben emplear muestras de la misma naturaleza. Adicionalmente, las señales espectroscópicas deben ser adquiridas en el espectrómetro empleado bajo las mismas condiciones y parámetros instrumentales.

5. RECOMENDACIONES

Conservando de manera adecuada las muestras empleadas en este estudio y aprovechando que se cuenta con su información espectral, es posible desarrollar modelos de predicción de nuevas propiedades de interés (True boilg point, densidad, entre otras). Además se pueden ser empleadas en el desarrollo de modelos de análisis elemental empleando otras técnicas.

Los resultados del estudio anteriormente presentados muestran el gran potencial de la espectroscopia PAS para la caracterización de fondos de vacío y muestras opacas. Sin embargo se recomienda aumentar los rangos de los datos para obtener modelos más robustos. También es recomendable alimentar los modelos con nuevas muestras, para asegurar que estos modelos siempre estén actualizados.

6. REFERENCIA BIBLIOGRAFICA

- [1]. CURTIS C., DECOSTER E., GUZMAN-GARCIA., HUGGINS C., Yacimientos de petróleo pesado. En: oilfield review 2002-2003. P32-35
- [2]. Comparing Venezuelan and Canadian Heavy Oil and Tar Sands. (2001-061: Calgary, Canada). Petroleum Society's Canadian International Petroleum Conference 2001, Calgary, Alberta, Canada.
- [3]. ASKE, N; KALLEVIK, H and SJOBLUM, J. Determination of saturate, aromatic, resin and asphaltenic (SARA) components in crude oils by means of infrared and Near-infrared spectroscopy. Energy and Fuels. 2001, 15, 1304-1312.
- [4]. CABRERA Edna, FRANCO Juan, MONDRAGON Fanor, y FERNÁNDEZ Jhon. Conversión de Fondos de Vacío de Petróleo a Semicoque. Universidad de Antioquia, Instituto de Química: Química de Recursos Energéticos y Medio Ambiente.
- [5]. COELHO Roberto, HOVELL Ian, de MELLO Marisa, MIDDEA Antonieta, and de SOUZA Andre. Characterisation of aliphatic chains in Vacuum Residues (VRs) of asphaltenes and Resins Using Molecular Modelling and FTIR Techniques. Fuel Processing Technology 87 (2006) 325-333.
- [6]. SANTIAGO G, PEURIOT A, SLEZAK V, and GONZÁLEZ M. Estudio de un Sistema de Detección Fotoacústica a Través de un Resonador doble y un Micrófono Diferencial. 62- Anales AFA vol. 18. 2006: ISSN 1850-1158
- [7]. YANG, Charles Q. Infrared spectroscopic analysis of Textile Materials Degradation Using Photoacoustic Detection. Ind. Eng. Chem. Res., 1992, 31 (2), 617-621.

- [8]. KALLEVIK, H. Characterization of crude oil and model oil emulsions by means of near infrared spectroscopy and multivariate analysis.
- [9]. JAPANWALA S, K. H. CHUNG, H. D. DETTMAN, M. R. GRAY. Quality of Distillates from Repeated Recycle of Residue, *Energy & Fuels* 16 (2002) 477–484.
- [10]. SPEIGHT J.G. Use of the Data: The Structure of Petroleum, *Handbook of Petroleum Analysis*, Wiley - Interscience, New York, 2001, pp. 403– 426.
- [11]. ALTGELT, K.H. AND BODUSZYNSKI, M. M. Composition and analysis of heavy petroleum fractions. Editorial Marcel Dekker. Segunda Edición. New York. 1994. P 495.
- [12]. HUSSEIN Alboudwarej, FELIX Joao, TAYLOR Shawn; La Importancia del Petróleo Pesado. *Oilfield Review*; Edmonton, Alberta, Canadá, otoño de 2006.
- [13]. VERSTRAETE, J.J., SCHNONGS., DULOT, H., HUDEBINE, D. Molecular Reconstruction of Heavy Petroleum Residue Fractions. *Chemical Engineering Science* (2009), doi:10.1016/j.ces.2009.08.033.
- [14]. CONAWAY, C. The petroleum industry: A Nontechnical al guide. Tulsa: Pennwell Publishing Co. 1999.
- [15]. WAUQUIER, J.P. El refino del petróleo. *Petróleo Crudo, Productos petrolíferos, Esquemas de fabricación*. En: *Composición de los petróleos crudos y de los productos petrolíferos*. España.: Editorial Diaz de Santos 2004. P 1-15.
- [16]. ASFAHA Iob, MOHAMMED A. Ali, BASSAM S. HydroCarbono group (PONA) analysis of reformato by FT-IR spectroscopy. *Fuel* 1996 volume 75 No 9
- [17]. APARICIO M, S. Metodologías analíticas basadas en espectroscopia de infrarrojo y calibración multivariante, aplicación a la industria petroquímica. Tarragona. Tesis doctoral, Universidad de Rovira y Virgili. Departamento de Química analítica y química orgánica. 2002. 20-29p.

- [18]. VARGAS, G. Predicción del análisis SARA de fondos de vacío colombianos utilizando espectroscopia fotoacústica en el infrarrojo medio y métodos quimiométricos. Bucaramanga 2011. Tesis de grado (Química). Universidad industrial de Santander. Facultad de ciencias. Escuela de Química.
- [19]. SPEIGH, J. The chemistry and Technology of Petroleum. Fourth edition. New York. CRC Press
- [20]. (en línea)
http://www.fq.uh.cu/dpto/qf/docencia/pregrado/estruc_2/ir/descargas/1_ir_tecnica.pdf [citado el 10 de diciembre de 2010]
- [21]. MICHAELIAN Kirk. Photoacoustic Infrared Spectroscopy of Industrial Materials. Review of Scientific Instruments (2003), Vol. 74, Nº 1, pp 659-662.
- [22]. ORREGO Jorge. Estudio de la Estructura de Cinco Carbonos Colombianos por Espectroscopía Fotoacústica en el Infrarrojo Medio. Bucaramanga 2008. Tesis de Maestría. Universidad industrial de Santander. Facultad de ciencias. Escuela de Química.
- [23]. WAUQUIER, J.P. El refinado del petróleo. Petróleo Crudo, Productos petrolíferos, Esquemas de fabricación. En: Composición de los petróleos crudos y de los productos petrolíferos. España.: Editorial Diaz de Santos 2004. P 1-15.
- [24]. MONROY, L. Predicción de la densidad en asfaltenos colombianos utilizando Espectroscopía Fotoacústica en la región del infrarrojo medio y métodos quimiométricos. Bucaramanga 2010. Tesis de grado (Química). Universidad industrial de Santander. Facultad de ciencias. Escuela de Química.
- [25]. MARIN, E. Escuchando la luz: breve historia y aplicaciones del efecto fotoacústico. Lat. Am. J. phys. Educ. Vol.2, No 2, May (2008). ISSN 1870-9095

- [26]. MCQUEEN, D; REGINALD, W. Near and mid-infrared photoacoustic analysis of principal components of foodstuffs. trends in analytical chemistry, vol. 14, no. 10, 1995.
- [27]. Kirk H. Michaelian. Photoacoustic infrared spectroscopy of industrial materials. REVIEW OF SCIENTIFIC INSTRUMENTS, Vol. 74, No. 1, January 2003.
- [28]. RUBINSON, K. A; RUBINSON, J. F. Análisis instrumental, Prentice Hall Hispanoamericana S.A. Madrid.: Editorial Pearson Educación, 2001.
- [29]. GELADI, P. Some recent trends in the calibration literature. En: Chemometrics and Intelligent Laboratory Systems. Vol. 60 (2002). p. 211 – 224.
- [30]. MARTENS, H; MARTENS, M. Multivariate analysis of quality an introduction. Segunda edición: Editorial John Wiley & Sons, 2001. P. 445.
- [31]. AMERICAN SOCIETY FOR TESTING AND MATERIALS. Standar practices for infrared multivariate quantitative analysis. Philadelphia: ASTM, 2005. 29h. (ASTM E 1655).
- [32]. MILLER, J. N. y MILLER, J. C. Estadística y quimiometría para química analítica. Cuarta edición. Editorial Prentice Hall. España. 2002. 296 p.
- [33]. GEMPERLINE, P. Practical guide to chemometrics. Segunda edición. Editorial Taylor y Francis Grupo. 2006. 541 p.
- [34]. FERNÁNDEZ, C. M. Quimiometría. Editorial Publicaciones Universidad de València. 2005. 423 p.
- [35]. SAVITZKY, A. and GOLAY, M.J.E. (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal. Chem. 36, 1627-1639.

[36]. NORRIS, K.H. and WILLIAMS, P.C. (1984) Optimization of mathematical treatments of raw near-infrared signal in the measurement of protein in hard red spring wheat. I. Influence of particle size. *Cereal Chemistry*, 61(2), 158-65.

[37]. PORCEL, G. M. Aplicación de técnicas quimiométricas para el desarrollo de nuevos métodos cinético-espectrofotométricos de análisis. Trabajo de grado Doctor en Ciencias Químicas. Bellaterra-España. Universitat Autònoma de Barcelona. Facultad de Ciencias. 2001. 5-26p

[38]. THE UNSCRAMBLER 9.7. METHODS. Software para diseño de experimentos y análisis multivariado. (En línea). <http://www.camo.com> [citado el 16 de enero de 2012].

[39]. SKOOG, D., A., HOLLER, F., J. y NIEMAN, T., A. Principios de análisis instrumental. 5ta Edición. Mc Graw Hill, Madrid, pp. 103-118. 2001