

**PREDICCIÓN DEL ANÁLISIS SARA DE FONDOS DE VACÍO COLOMBIANOS  
UTILIZANDO ESPECTROSCOPIA FOTOACÚSTICA EN EL INFRARROJO  
MEDIO Y MÉTODOS QUIMIOMÉTRICOS.**

**GLORIA ESTHER VARGAS PATIÑO**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE CIENCIAS  
ESCUELA DE QUÍMICA  
BUCARAMANGA  
2011**

**PREDICCIÓN DEL ANÁLISIS SARA DE FONDOS DE VACÍO COLOMBIANOS  
UTILIZANDO ESPECTROSCOPIA FOTOACÚSTICA EN EL INFRARROJO  
MEDIO Y MÉTODOS QUIMIOMÉTRICOS**

**GLORIA ESTHER VARGAS PATIÑO**

**Trabajo de grado para optar al título de  
QUÍMICA**

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

**CO-DIRECTOR: JORGE ARMANDO ORREGO RUIZ  
Qco, M.Sc.**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE CIENCIAS  
ESCUELA DE QUÍMICA  
BUCARAMANGA  
2011**

*Por todos los favores recibidos dedico  
este triunfo a mi Dios, en quien todo  
lo puedo*

## **AGRADECIMIENTOS**

**A toda mi gran familia por patrocinar mi carrera. A mis padres principalmente.**

**A Edwin Castro Jurado por sus ganas de seguir hacia adelante junto a mí.**

**Al profesor y maestro Enrique Mejía Ospino por su disposición para enseñarme en este proceso de aprendizaje. Así mismo le estoy muy agradecida a Jorge Orrego Ruiz porque hasta el final, también estuvo en todas las etapas de esta investigación.**

**A todas mis amigas y amigos, a la familia LEAM en cabeza del profesor Rafael Cabanzo quienes de una forma u otra me ayudaron a culminar este objetivo.**

**A la Universidad Industrial de Santander y al Instituto Colombiano del Petróleo por su apoyo económico.**

**Por último, a todas la personas de la Escuela de Química que me han colaborado.**

## CONTENIDO

|   | Pág. |
|---|------|
| INTRODUCCIÓN .....  | 14   |
| 1. CONSIDERACIONES TEÓRICAS.....  | 16   |
| 1.1. GENERALIDADES DE LOS FONDOS DE VACÍO .....                                       | 16   |
| 1.2. COMPOSICIÓN QUÍMICA DE LOS FONDOS DE VACÍO .....                                 | 16   |
| 1.2.1. Análisis SARA.....   | 17   |
| 1.2.2. Saturados.....   | 19   |
| 1.2.3. Aromáticos .....   | 19   |
| 1.2.4. Resinas .....  | 20   |
| 1.2.5. Asfaltenos .....   | 21   |
| 1.2.6. Análisis Composicional .....   | 22   |
| 1.3. ESPECTROSCOPIA INFRARROJA.....   | 24   |
| 1.3.1. Efecto Fotoacústico (PA) .....   | 29   |
| 1.4. QUIMIOMETRÍA .....   | 33   |
| 1.4.1. Calibración multivariable .....  | 35   |
| 1.4.2. Clasificación de los métodos de calibración multivariable .....                | 35   |
| 1.4.3. Construcción de modelos de calibración multivariable .....                     | 37   |
| 1.4.4. Técnicas de Pretratamiento de Datos .....                                      | 40   |
| 1.4.5. Métodos basados en reducción de variables .....                                | 43   |
| 1.4.6. Análisis Por Componentes Principales (PCA).....                                | 43   |
| 1.4.7. Regresión por Mínimos Cuadrados Parciales (PLS).....                           | 46   |
| 2. PARTE EXPERIMENTAL .....   | 49   |
| 2.1. ESQUEMA EXPERIMENTAL .....   | 49   |
| 2.2. SELECCIÓN DE MUESTRAS .....  | 50   |
| 2.3. MONTAJE EXPERIMENTAL .....   | 51   |
| 2.3.1. Instrumentación .....  | 51   |
| 2.4. DETERMINACIÓN DE PARAMETROS.....   | 53   |
| 2.4.1. Verificación del desempeño del Espectrofotómetro .....                         | 54   |
| 2.5. ADQUISICIÓN Y ANÁLISIS DE ESPECTROS PAS.....                                     | 55   |
| 2.5.1. Espectros MIR de fondos de vacío .....   | 55   |
| 2.6. PROCESAMIENTO DE DATOS .....   | 59   |
| 2.6.1. Pretratamiento de datos.....   | 59   |
| 2.6.2. Análisis de componentes principales (PCA) .....                                | 59   |
| 2.6.3. Regresión por mínimos cuadrados parciales (PLS).....                           | 60   |
| 3. RESULTADOS Y ANÁLISIS .....  | 61   |
| 3.1. ANÁLISIS ESPECTRAL DE FONDOS DE VACÍO .....                                      | 61   |
| 3.2. ANÁLISIS POR COMPONENTES PRINCIPALES (PCA).....                                  | 63   |
| 3.3. DESARROLLO DE LOS MODELOS PLS.....   | 69   |
| 3.3.1. Calibración y validación del modelo PLS para la fracción de saturados .....    | 69   |
| 3.3.2. Calibración y validación del modelo PLS para la predicción de Aromáticos ..... | 74   |
| 3.3.3. Calibración y validación del modelo PLS para la predicción de resinas .....    | 79   |
| Calibración y validación del modelo PLS para la predicción de asfaltenos.....         | 84   |

|  |    |
|--|----|
| 3.4. PRUEBAS T PARA MUESTRAS DE VALIDACIÓN. .... | 89 |
| 4. CONCLUSIONES .....                            | 92 |
| 5. RECOMENDACIONES.....                          | 94 |
| 6. REFERENCIAS BIBLIOGRÁFICAS.....               | 95 |

## LISTA DE FIGURAS

|   | Pág. |
|---|------|
| Figura 1. Esquema simplificado de la separación de los fondos de vacío en las fracciones SARA.....            | 18   |
| Figura 2. Estructura de una molécula promedio de saturados.....   | 19   |
| Figura 3. Estructura promedio de una estructura aromática.....  | 20   |
| Figura 4. Estructura de una molécula promedio de Resinas.....   | 20   |
| Figura 5. Estructura de una molécula promedio de Asfaltenos.....  | 21   |
| Figura 6. Procedimiento estándar para el análisis SARA según la norma ASTM: a. D- 2007 y b. D-4124.....       | 23   |
| Figura 7. Espectros MIR para el fondo de vacío 24.....  | 27   |
| Figura 8. Generación de la señal fotoacústica en un sólido.....   | 32   |
| Figura 9. Interpretación geométrica de un PCA.....  | 45   |
| Figura 10. Descripción gráfica del modelo de regresión PLS.....   | 47   |
| Figura 11. Esquema metodológico.....  | 49   |
| Figura 12. Sistema PAS acoplado al equipo IR Tensor 27 marca Bruker.....                                      | 51   |
| Figura 13. Celda PAS modelo MTEC 200.....   | 52   |
| Figura 14. Efecto de la relación señal ruido en un espectro de una muestra de fondo de vacío.....             | 53   |
| Figura 15. Prueba de repetibilidad del espectrofotómetro FTIR.....  | 55   |
| Figura 16. Deconvolución de la región entre 3000-2800 $\text{cm}^{-1}$ .....                                  | 56   |
| Figura 17. Bandas de vibración para el grupo $\text{CH}_3$ .....  | 57   |
| Figura 18. Espectros de Fondos de vacío.....  | 61   |
| Figura 19. Gráfico de <i>puntuaciones</i> : descomposición de la información espectral en el PC2 vs PC1.....  | 64   |
| Figura 20. Gráfico de <i>pesos</i> : <i>Importancia</i> de las diferentes variables para el PC1 y el PC2..... | 66   |
| Figura 21. Contribución de las variables en la formación PC1 y el PC2.....                                    | 67   |
| Figura 22. Gráfico de <i>puntuaciones</i> con el límite de Hotelling $T^2$ .....                              | 68   |
| Figura 23. Ubicación de las variables x en la región MIR para el modelo de saturados.....                     | 70   |
| Figura 24. Ubicación de los Coeficientes de regresión para el modelo de saturados.....                        | 71   |
| Figura 25. Curva de calibración del modelo seleccionado para Saturados.....                                   | 72   |
| Figura 26. Gráfico del valor predicho contra el valor de referencia para el modelo de saturados.....          | 73   |
| Figura 27. Gráfico del valor predicho contra el valor de referencia, eliminando una muestra.....              | 74   |
| Figura 28. Ubicación de las variables x en la región MIR para el modelo de aromáticos.....                    | 75   |
| Figura 29. Ubicación de los Coeficientes de regresión para 5 VL en el modelo de aromáticos.....               | 76   |
| Figura 30. Curva de calibración del modelo seleccionado para aromáticos.....                                  | 77   |

|   |    |
|---|----|
| Figura 31. Gráfico del valor predicho vs valor de referencia para aromáticos .....              | 78 |
| Figura 32. Ubicación de las variables x en la región MIR para resinas.....                      | 80 |
| Figura 33. Ubicación de los Coeficientes de regresión para el modelo de resinas                 | 81 |
| Figura 34. Curva de calibración del modelo seleccionado para resinas .....                      | 82 |
| Figura 35. Gráfico del valor predicho vs valor de referencia para el modelos de resinas .....   | 83 |
| Figura 36. Gráfico del valor predicho vs valor de referencia.....                               | 84 |
| Figura 37. Ubicación de las variables x en la región MIR para el modelo de asfaltenos .....     | 85 |
| Figura 38. Ubicación de los Coeficientes de regresión para el modelo de asfaltenos .....        | 86 |
| Figura 39. Curva de calibración del modelo seleccionado para asfaltenos .....                   | 87 |
| Figura 40. Gráfico del valor predicho vs valor de referencia para el modelo de asfaltenos ..... | 88 |

## LISTA DE TABLAS

|  | Pág. |
|--|------|
| Tabla 1. Regiones de absorción en el infrarrojo.....   | 26   |
| Tabla 2. Resumen de las bandas de absorción .....  | 27   |
| Tabla 3. Análisis SARA del grupo de calibración .....  | 50   |
| Tabla 4. Análisis SARA del grupo de validación .....   | 51   |
| Tabla 5. Vibraciones de flexión del C-H aromático .....  | 58   |
| Tabla 6. Cálculo de la RSD para tres números de onda característicos aplicando diferentes pretratamientos .....    | 59   |
| Tabla 7. Varianza explicada por los primeros 7 componentes principales de las muestras de calibración .....        | 63   |
| Tabla 8. Separación de las muestras según el contenido Asfaltenos .....  | 65   |
| Tabla 9. Parámetros del modelo seleccionado para la predicción del % de Saturados .....                            | 69   |
| Tabla 10. Valor de los coeficientes de regresión para 5 VL.....  | 71   |
| Tabla 11. Validación del modelo de predicción de %p de Saturados .....   | 72   |
| Tabla 12. Parámetros del modelo seleccionado para la predicción del %p de aromáticos .....                         | 74   |
| Tabla 13. Valor de los coeficientes de regresión para 7 VL en el modelo de aromáticos .....                        | 76   |
| Tabla 14. Validación del modelo de predicción de %p de Aromáticos.....   | 78   |
| Tabla 15. Parámetros del modelo seleccionado para la predicción del %p de Resinas.....                             | 79   |
| Tabla 16. Valores de los coeficientes de correlación .....   | 81   |
| Tabla 17. Validación del modelo de predicción de %p de resinas .....   | 83   |
| Tabla 18. Parámetros del modelo seleccionado para la predicción del %p de Asfaltenos .....                         | 84   |
| Tabla 19. Valores de los coeficientes de correlación para cada variable x.....                                     | 86   |
| Tabla 20. Validación del modelo de predicción de %p de Asfaltenos.....   | 88   |
| Tabla 21. Parámetros estadísticos usados para definir la calidad de las predicciones.....                          | 89   |
| Tabla 22. Parámetros estadísticos usados para evaluar la calidad de las predicciones a cuatro fondos de vacío..... | 90   |

## RESUMEN

### TÍTULO

PREDICCIÓN DEL ANÁLISIS SARA DE FONDOS DE VACÍO COLOMBIANOS UTILIZANDO ESPECTROSCOPIA FOTOACÚSTICA EN EL INFRARROJO MEDIO Y MÉTODOS QUIMIOMÉTRICOS.\*

### AUTOR

GLORIA ESTHER VARGAS PATIÑO. \*\*

### PALABRAS CLAVES

PAS-FTIR, FONDOS DE VACÍO, PLS, PCA, ANÁLISIS SARA

### DESCRIPCIÓN

28 muestras de fondos de vacío de crudos colombianos proporcionados por el Instituto Colombiano del Petróleo fueron analizadas por espectroscopia fotoacústica en la región del infrarrojo medio (PAS-FTIR) y métodos quimiométricos. Mediante el uso combinado de la espectroscopia PAS-FTIR y métodos de calibración multivariable, fue posible realizar el análisis de la composición química de fondos de vacío, encontrándose una relación directa entre las intensidades espectrales y el análisis SARA. El análisis por componentes principales (PCA) permitió clasificar las muestras de acuerdo al contenido de asfaltenos y determinar las variables espectrales de mayor importancia que ayudaron a la formación de los componentes principales. Estas variables se ubican en las regiones comprendidas entre 3000-2800, 1500-1350 y 950-700  $\text{cm}^{-1}$ . La regresión por mínimos cuadrados parciales (PLS-R) aplicada sobre los espectros PAS-FTIR, mostró un desempeño satisfactorio para predecir la composición en peso del análisis SARA, a ocho fondos de vacío que no fueron usados en la calibración de los modelos. Se comprobó que la combinación de la espectroscopia PAS-FTIR junto con el análisis quimiométrico puede ser un método alternativo de análisis que permite hacer un estudio de muestras sólidas, y opacas en un menor tiempo que las técnicas tradicionales.

---

\* Trabajo de Grado

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

## SUMMARY

### TITLE

SARA ANALYSIS PREDICTION FROM VACUUM RESIDUES COLOMBIANS USING PHOTOACOUSTIC SPECTROSCOPY IN THE MIDDLE INFRARED AND CHEMOMETRIC METHODS. \*

### AUTHOR

GLORIA ESTHER VARGAS PATIÑO. \*\*

### KEY WORDS

PAS-FTIR, VACUUM RESIDUES, PLS, PCA, SARA ANALYSIS.

### DESCRIPTION

28 samples of vacuum residues of colombian oils provided by the Colombian Petroleum Institute were analysed by photoacoustic spectroscopy in mid-infrared region (PAS-FTIR) and chemometric methods. Through the combined use of PAS-FTIR spectroscopy and multivariate calibration methods was possible analyzing chemical composition of vacuum residues, finding a direct relationship between spectral intensities and SARA analysis. Principal components analysis (PCA) allowed classifying samples according to the asphaltene content and determining the most important spectral variables wich helped to form the principal components. These variables are located in the regions between 3000-2800, 1500 - 1350 and 950 - 700  $\text{cm}^{-1}$ . Partial least squares regression (PLS-R) applied on the PAS-FTIR spectra, showed a satisfactory performance to predict SARA analysis of eight vacuum residues not used in the calibration of models. Combination of PAS-FTIR spectroscopy and chemometric analysis can be an alternative method of analysis that allows making a study of solid and opaque samples in less time respect to traditional techniques.

---

\* Work Degree

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

## INTRODUCCIÓN

Los crudos livianos en general presentan un costo de extracción y refinación menor al que presentan los crudos pesados, ya que estos hidrocarburos más densos y viscosos (densidad API < 22.3°<sup>[28]</sup>) tienden a poseer mayores concentraciones de metales y otros elementos, lo que exige mayor esfuerzo para su refinación y producción, aumentando de este modo su valor económico.

Actualmente el 70% de las reservas mundiales de petróleo se compone de petróleo pesado, extrapesado y bitumen<sup>[31]</sup>. Por tal motivo en los últimos años la industria petrolera ha tenido que desplazar su conocimiento y tecnología hacia los crudos pesados. Para esto realiza investigaciones sobre técnicas que favorezcan su producción y disminuyan los costos por los que anteriormente se habían dejado de lado. En trabajos previos<sup>[3]</sup> se ha dado a conocer, que los fondos de vacío pueden ser usados como derivados de gran interés industrial y no como subproductos de poco valor agregado, ya que contienen desde hidrocarburos parafínicos y nafténicos hasta hidrocarburos aromáticos.

Una de las estrategias de caracterización de este tipo de muestras pesadas es el análisis SARA (fracción de saturados, aromáticos, resinas y asfaltenos). Este análisis se lleva a cabo mediante procedimientos estándar regidos por la ASTM (American Society Test and Materials), en los cuales la muestra se somete a la acción de solventes como *n*-heptano y luego a cromatografía de elución con tolueno y diclorometano, cuyo resultado final tarda cerca de 2 días por muestra.

Sin embargo, esta metodología es costosa y poco amigable con el medio ambiente ya que hace uso de gran cantidad de solventes tóxicos. Para dar solución a estos inconvenientes se pueden aplicar técnicas quimiométricas de análisis multivariable en los cuales se relacionan los datos espectroscópicos y este análisis.

Para este propósito se usan técnicas como la espectroscopía infrarroja con transformada de Fourier (FT-IR), y la Resonancia Magnética Nuclear (RMN- $^1\text{H}$ )<sup>[3,27]</sup> que aunque han mostrado resultados satisfactorios requieren de preparación de las muestras, y en el caso de RMN altos costos de operación. La espectroscopía fotoacústica acoplada a la espectroscopía infrarroja con transformada de Fourier (PAS-FTIR) ha demostrado ser sensible y versátil para el estudio estructural de muestras opacas<sup>[15]</sup> mostrando superioridad sobre las técnicas de transmitancia convencionales, debido a que la muestra no es destruida durante el análisis y no requiere tratamiento previo, disminuyendo los costos al no utilizar solventes y la contaminación que estos causan.

De esta manera, el desarrollo el presente trabajo permitió clasificar y analizar fondos de vacío colombianos a partir de su espectro PAS-FTIR, mediante el uso de técnicas quimiométricas de análisis. La regresión por mínimos cuadrados parciales (PLS-R) aplicada sobre los espectros, mostró un desempeño satisfactorio para la predicción del análisis SARA (saturados, aromáticos, resinas y asfaltenos). Se comprobaron las ventajas que presenta la técnica PAS-FTIR para el estudio de la composición química de muestras complejas y opacas, ya que los espectros se obtuvieron en menos de 5 minutos, con buena relación señal/ruido y con una repetibilidad aceptable.

## **1. CONSIDERACIONES TEÓRICAS**

### **1.1. GENERALIDADES DE LOS FONDOS DE VACÍO**

En la refinación del petróleo crudo se obtienen gran cantidad de productos entre los que se encuentran los fondos de vacío (FV). Los fondos de vacío son fracciones pesadas que provienen de la destilación al vacío del petróleo, y pueden ser convertidos a fracciones que presentan un menor punto de ebullición por reducción de viscosidad, craqueo y procesos de hidrocrqueo. El hidrocrqueo de los fondos de vacío ofrece volúmenes adicionales, para combustibles de motor de alta calidad <sup>[1]</sup>.

Los fondos de vacío pueden ser utilizados para alimentar las unidades de craqueo térmico, donde se producen productos más ligeros, o para alimentar la unidad de producción de coque. Generalmente, los fondos son utilizados sin ningún tratamiento adicional como asfalto, aglutinante de los pavimentos en las carreteras, materiales para techos o como revestimientos de tubería.

Recientemente se ha publicado que el uso de los fondos de vacío es muy limitado, ya que considerablemente afectan el medio ambiente porque presentan en su estructura azufre y nitrógeno <sup>[2,4]</sup>. La emisión de azufre está relacionada con la formación de la lluvia ácida.

### **1.2. COMPOSICIÓN QUÍMICA DE LOS FONDOS DE VACÍO**

Los hidrocarburos presentes en esta fracción del petróleo concentran elevados puntos de ebullición (generalmente mas de 500 °C) <sup>[2]</sup> y moléculas complejas, ya que presentan una diversidad de estructuras químicas cuyos pesos moleculares varían en un amplio rango, comenzando con hidrocarburos parafínicos y nafténicos hasta hidrocarburos poliaromáticos. En menor cantidad contienen gran parte de las impurezas presentes en el crudo, estas impurezas incluyen azufre

(hasta 6% en peso en el residuo), nitrógeno (0.1-2% en peso) y oxígeno (0.005-1.5% en peso) <sup>[4]</sup>, los cuales se encuentran conformando grupos funcionales como ácidos carboxílicos, aldehídos, cetonas, tiofenos, anillos pirrólicos y bases nitrogenadas como piridina. A nivel de trazas contienen metales como níquel, vanadio, hierro, magnesio y calcio, presentes en forma de óxidos, sales o también ocupando el centro de estructuras organometálicas tipo porfirina <sup>[3]</sup>.

La razón por la cual los fondos de vacío presentan en su estructura moléculas complejas, se debe a que los hidrocarburos ligeros se eliminan previamente en el proceso de destilación.

### **1.2.1. Análisis SARA**

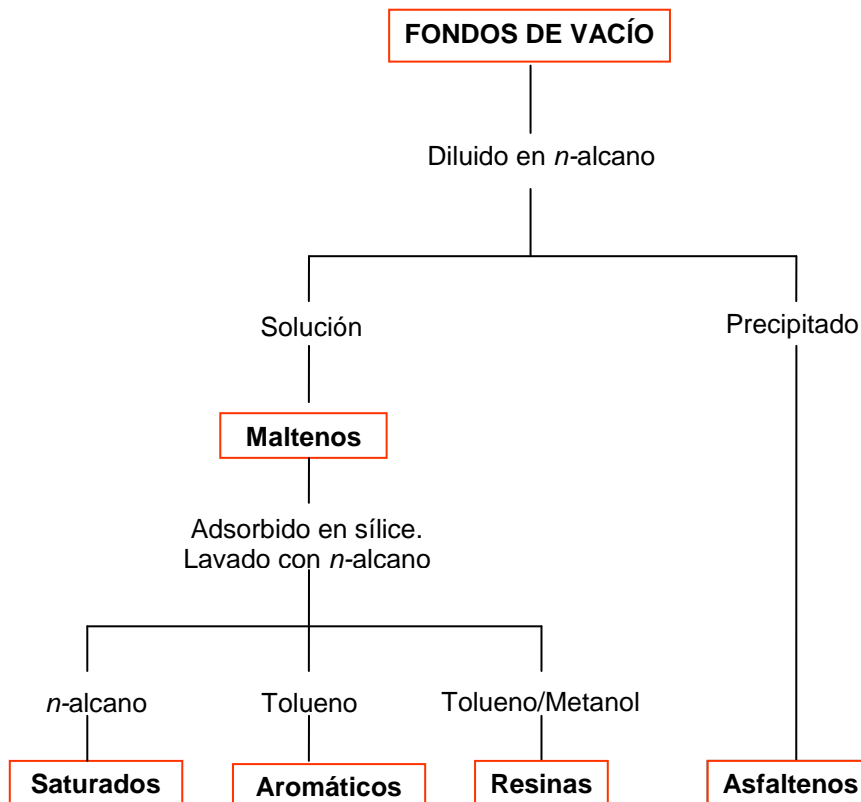
De las diversas técnicas de laboratorio que han sido desarrolladas para describir la composición química de los hidrocarburos, la más común es la cromatografía de gases (CG). Este tipo de análisis puede determinar la composición química de un petróleo convencional (petróleo liviano). No obstante el análisis por CG no puede diferenciar el alto número de componentes de los hidrocarburos pesados con suficiente grado de detalle como para ser utilizado en las operaciones de simulación.

Debido al gran número de compuestos que conforman los fondos de vacío es necesario separarlos en grupos más homogéneos, mediante técnicas de fraccionamiento para su identificación. Para la caracterización composicional de este tipo de fracciones pesadas, se pone en práctica técnicas de análisis adicionales que examinan en forma más exhaustiva estos fluidos de alta densidad y alta viscosidad. Entre estas técnicas se encuentra el análisis de las fracciones de saturados, aromáticos, resinas y asfaltenos (análisis SARA). El análisis SARA consiste en el fraccionamiento de la muestra en las cuatro fracciones mencionadas, expresándolas como porcentaje en peso. La fortaleza del método radica en que analiza la muestra entera, desde los compuestos livianos hasta los pesados, y de este modo permite comparar todos los petróleos con base en un

estándar consistente [8]. Estas técnicas requieren del uso de diferentes métodos para el pretratamiento de la muestra, así como de solventes para hacer posible su análisis. De esta manera, el estudio de las muestras pesadas de petróleo se convierte en un procedimiento tedioso, ya que requiere de un tiempo prolongado de análisis de entre 2 a 3 días por muestra.

El fraccionamiento SARA hace en primer lugar una separación mediante precipitación con hidrocarburos saturados de bajo peso molecular (*n*-heptano, o *n*-pentano), obteniéndose dos fracciones; a la fracción insoluble se le denomina *asfaltenos* y a la soluble *maltenos*; estos últimos a la vez contienen varios tipos de hidrocarburos que pueden dividirse en saturados, aromáticos y resinas [9] (figura 1).

**Figura 1. Esquema simplificado de la separación de los fondos de vacío en las fracciones SARA**

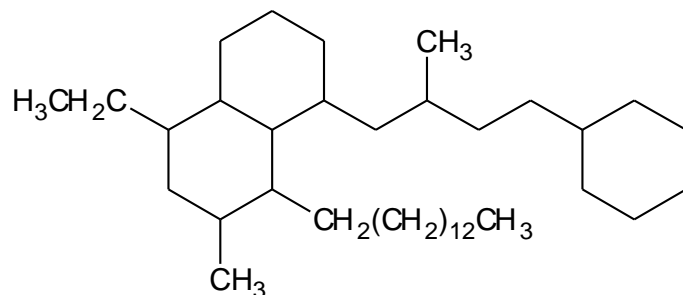


A continuación se muestra una breve descripción de cada grupo.

### 1.2.2. Saturados

El término saturado significa que la molécula contiene el número máximo de átomos de hidrógeno capaz de contener. Estos hidrocarburos son aceites claros, no polares constituidos por hidrocarburos alifáticos lineales o con cadenas laterales alifáticas y aromáticas. El rango de peso molecular medio está comprendido entre 300 y 2.000 Da <sup>[5]</sup>.

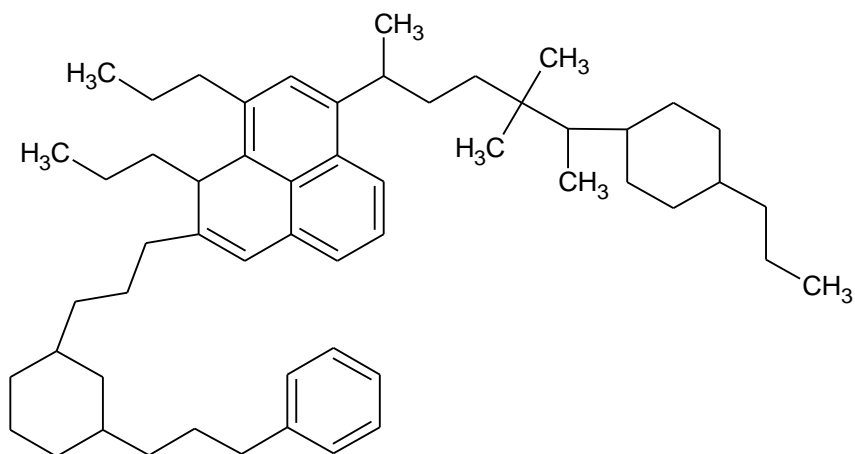
**Figura 2. Estructura de una molécula promedio de saturados**



### 1.2.3. Aromáticos

Son hidrocarburos cíclicos poliinsaturados que están presentes en gran proporción en el petróleo crudo, generalmente son líquidos viscosos de color marrón anaranjado. Sus usos son muy variados y dependen de la sustitución de los hidrógenos en el anillo aromático. Así, los compuestos como benceno, tolueno, y xileno son materias primas fundamentales de la petroquímica (además contribuyen igualmente a aumentar el número de octano de las gasolinas), pero a medida que aumenta la complejidad de las sustituciones, aumentan los problemas que estos compuestos causan tanto al medio ambiente, como al deterioro de la actividad de los catalizadores por su capacidad de formar coque <sup>[6]</sup>.

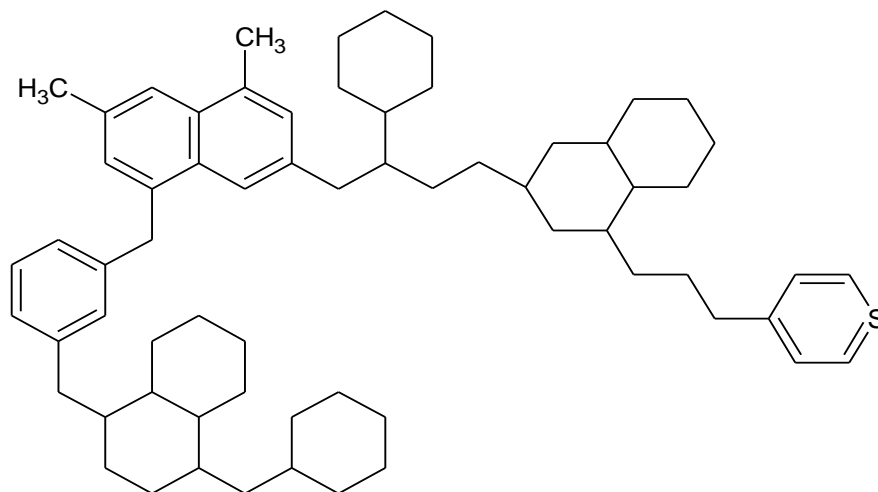
**Figura 3. Estructura promedio de una estructura aromática**



#### 1.2.4. Resinas

Son moléculas con un fuerte carácter aromático, al igual que los asfaltenos tienen una elevada proporción de hidrógeno y carbono, contienen pequeñas cantidades de oxígeno, azufre y nitrógeno. Son sólidos negros, brillantes, quebradizos y su naturaleza es muy polar. Las resinas constituyen el componente polar no volátil del petróleo, que es soluble en *n*-alcanos e insoluble en propano líquido [7].

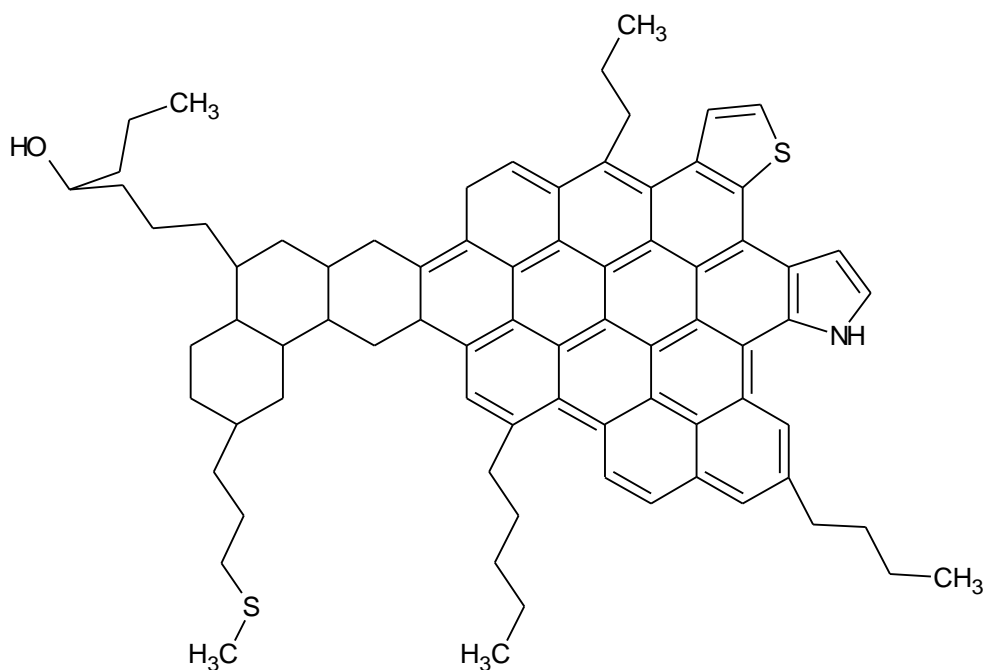
**Figura 4. Estructura de una molécula promedio de Resinas**



### 1.2.5. Asfaltenos

Dentro de los compuestos considerados, los asfaltenos son los que poseen las estructuras químicas más complejas y los más altos pesos moleculares; estos son principalmente aromáticos policíclicos condensados. Son sólidos amorfos de colores marrón oscuro o negros, insolubles en n-heptano y solubles en tolueno. Están constituidos por carbono, hidrógeno, nitrógeno, azufre y oxígeno, que dan lugar a ciclos tiofénicos y piridínicos que contienen elevada polaridad. La desalquilación y deshidrogenación de los asfaltenos se da como resultado del proceso de craqueo. Estas reacciones son favorables para el proceso de formación del coque [1].

**Figura 5. Estructura de una molécula promedio de Asfaltenos**



Por otra parte la operación industrial para la eliminación de asfaltenos en un aceite se realiza con propano o butano. El aceite separado de su fracción asfáltica se llama aceite desasfaltado (en inglés DAO: *DeAsphalted Oil*), y la parte precipitada se llama, por lo tanto, asfalto. Los heteroátomos se concentran en el asfalto (azufre, nitrógeno, níquel y vanadio), siendo su contenido muy variable,

dependiendo de la procedencia del crudo. Estos heteroátomos son la causa de numerosos problemas en la industria del petróleo <sup>[6]</sup>.

### **1.2.6. Análisis Composicional**

Un crudo estará bien definido si se conoce la composición y estructura de todos los compuestos presentes. Sin embargo, debido a la diversidad y al número elevado de constituyentes, conocer la composición exacta del crudo es casi imposible.

Después de los datos de destilación el análisis composicional es la característica más importante 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 <sup>[9,10]</sup>:

PONA (parafinas, olefinas, naftenos y aromáticos)

PNA (parafinas, naftenos y aromáticos)

PIONA (parafinas, isoparafinas, olefinas, naftenos y aromáticos)

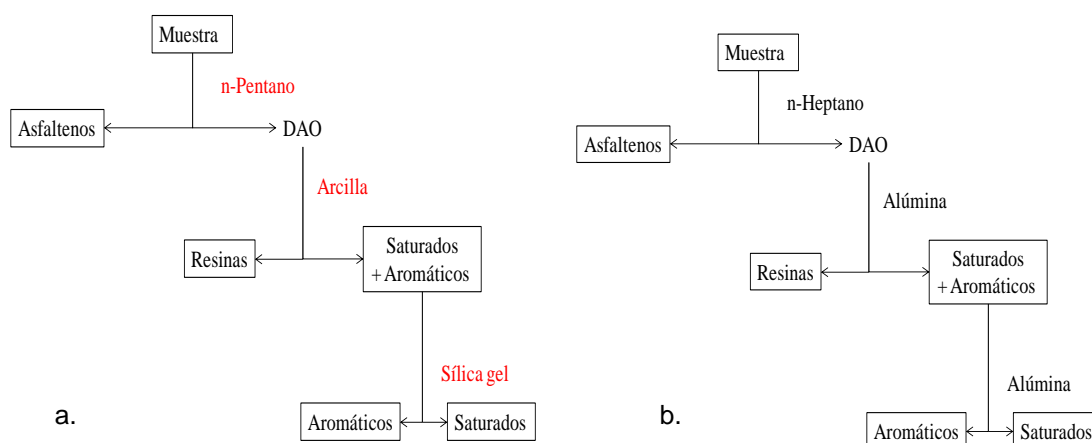
SARA (saturados, aromáticos, resinas y asfaltenos)

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.

La ASTM ofrece varios procedimientos para el análisis SARA: la norma ASTM D2007 y la ASTM D4124 mostrados en la figura 6, y la norma ASTM D2549 que usa cromatografía líquida para la separación de las fracciones de saturados y aromáticos.

**Figura 6. Procedimiento estándar para el análisis SARA según la norma ASTM: a. D-2007 y b. D-4124**



La porción de crudo que es soluble en pentano o heptano se conoce como maltenos. Esta porción puede a su vez, fraccionarse por percolación a través de cualquier material superficialmente activo como alúmina o arcilla. La fracción más fuertemente adsorbida y que se desorbe con piridina o una mezcla tolueno-metanol se conoce como resinas. La fracción que eluye por la acción del tolueno o el benceno se define como aromáticos, mientras que la fracción que lo hace por la acción del heptano se conoce como saturados.

Otro tipo de análisis usado para identificar grupos funcionales es la espectrometría de masas (EM) en la cual las moléculas son fragmentadas e ionizadas. Ese método presenta una alta reproducibilidad y ofrece información cualitativa y cuantitativa detallada de las moléculas orgánicas e inorgánicas que componen las muestras. Sin embargo, el uso de EM es limitado a compuestos que son estables hasta 300 °C.

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 oxígeno: ASTM E385

Contenido de azufre: ASTM D124, ASTM D1266, ASTM D1552, ASTM D1757, ASTM D2662, ASTM D3177, ASTM D4045 y ASTM D4294 <sup>[35]</sup>.

### **1.3. ESPECTROSCOPÍA INFRARROJA**

La radiación Infrarroja (IR) fue descubierta por William Herschel a comienzos del siglo XIX. Las dificultades en construir un detector IR adecuado impidieron cualquier aplicación. Solo hasta los años 30 del siglo XX fue cuando se desarrolló 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) <sup>[12]</sup>.

La técnica de transformada de Fourier está basada 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 <sup>[12]</sup>.

La aplicación de la Transformada de Fourier (FT) a la espectroscopía infrarroja extendió el uso de la técnica, ya que eliminó las limitaciones encontradas con los instrumentos dispersivos. De este modo, la FT ha conseguido para la espectroscopía infrarroja una significativa ventaja práctica, y ha hecho posible el desarrollo de nuevas técnicas, como FTIR-CA (de camino abierto) y FTIR-ATR (de reflectancia total atenuada).

### **Regiones Espectrales**

El espectro de infrarrojo se extiende desde 10 a 14300  $\text{cm}^{-1}$  pero desde un punto de vista funcional se divide en tres zonas. El IR lejano (FIR), donde se producen las absorciones debido a cambios rotacionales; se ha usado para el estudio de compuestos órgano-metálicos o inorgánicos (átomos pesados, enlaces débiles). El IR medio (MIR), donde tienen lugar las vibraciones fundamentales; es útil para conseguir información sobre aspectos cualitativos y cuantitativos de analitos en tiempo real. Finalmente en el IR cercano (NIR), donde se presentan las bandas armónicas, se producen absorciones debidas a sobretonos y combinaciones de las bandas fundamentales y ha sido utilizada para determinaciones cuantitativas y en menor proporción con fines estructurales <sup>[11,12]</sup>.

En la tabla 1 se muestra el rango del espectro electromagnético al que corresponde cada región IR.

**Tabla 1. Regiones de absorción en el infrarrojo**

| Región Infrarroja    | Longitud de Onda  | Número de Onda, (cm-1) | Transición característica  |
|----------------------|-------------------|------------------------|----------------------------|
| <b>Cercano (NIR)</b> | 780 a 2500 nm     | 12800 a 4000           | Sobretonos y Combinaciones |
| <b>Medio (MIR)</b>   | 2.5 a 50 $\mu$ m  | 4000 a 200             | Vibraciones fundamentales  |
| <b>Lejano (FIR)</b>  | 50 a 1000 $\mu$ m | 200 a 10               | Rotaciones                 |

### **Análisis Infrarrojo**

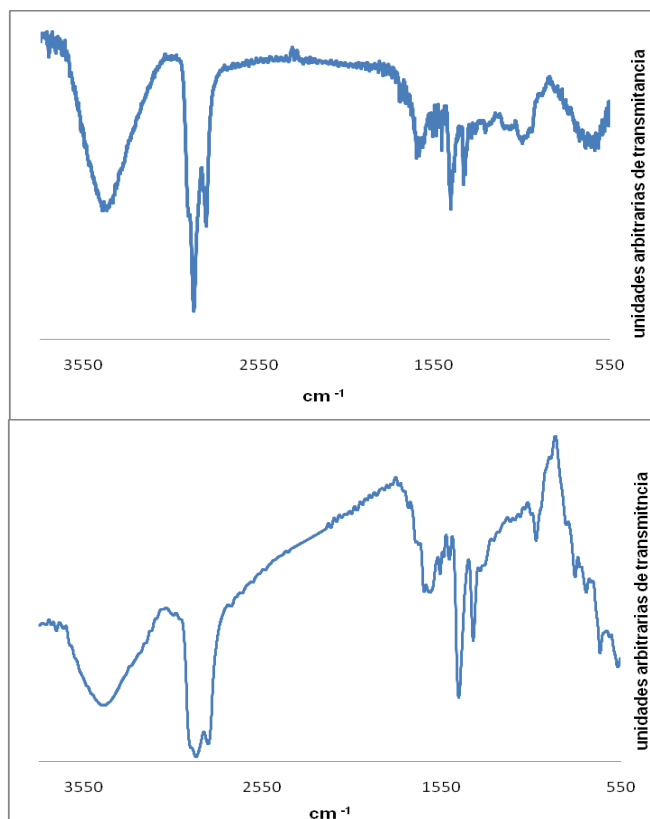
Un espectro IR proporciona los primeros datos para dilucidar la estructura molecular de una muestra ya que presenta la información sobre las unidades estructurales básicas y los grupos funcionales.

Aunque el espectro IR caracteriza a cada compuesto, ciertas agrupaciones atómicas dan lugar siempre a bandas en un determinado intervalo de frecuencias, independiente de la naturaleza del resto de la molécula. La existencia de estas bandas, permite una amplia utilización de la espectroscopía IR en la determinación estructural. Sin embargo pueden presentarse problemas con la reproducibilidad de los espectros por dificultades ligadas a la preparación de las muestras.

La espectroscopía infrarroja ha tenido siempre un lugar en el análisis de combustibles para caracterizar cualitativamente sus constituyentes. Con el advenimiento de la espectroscopía FT-IR y el avance en los métodos quimiométricos, se creó la posibilidad de desarrollar esta metodología para fines cuantitativos <sup>[13]</sup>. En el trabajo de Abbas y colaboradores <sup>[14]</sup> – muestran que, con la determinación del área de las bandas de absorción de cada uno de los espectros IR, fue posible clasificar el petróleo en términos de las etapas de biodegradación que sufrían los crudos analizados.

De esta manera, en nuestro trabajo se usó la espectroscopia FTIR para el análisis de fondos de vacío. En la figura 7 se muestra el espectro MIR de una muestra de fondo de vacío colombiano.

Figura 7. Espectros MIR para el fondo de vacío 24.



En la tabla 2 se muestran las asignaciones de algunas bandas del IR presentes en los fondos de vacío.

Tabla 2. Resumen de las bandas de absorción

| Posición de las bandas ( $\text{cm}^{-1}$ ) | Grupos funcionales  |
|---|---|
| 3700-3300                                   | Vibraciones de tensión con hidrógeno:<br>$\nu_{OH}$ , $\nu_{NH}$ , $\nu_{CH}$ |
| 3090-3030                                   | vibraciones de tensión $(\text{C-H})_{ar}$                                    |
| 2978-2960                                   | R-CH <sub>3</sub> estiramiento asimétrico                                     |
| 2925-2918                                   | R <sub>2</sub> -CH <sub>2</sub> estiramiento asimétrico                       |

|           |  |
|-----------|--|
| 2860      | CH alifático (estiramiento)  |
| 2854      | RCH <sub>2</sub> - estiramiento simétrico  |
| 1654      | -C=O conjugado   |
| 1650-1570 | (C-H) <sub>ar</sub> sistema aromático polinuclear  |
| 1500-1450 | (C-C) <sub>ar</sub> estiramiento   |
| 1460-1450 | Flexión C-H alifático  |
| 1380      | Deformación simétrica -CH <sub>2</sub> - (flexión)   |
| 1300-1000 | Deformación C-O fenólica (estiramiento)  |
| 925-670   | Enlaces aromáticos; (C-H) <sub>ar</sub> (flexión fuera del plano)                              |
| 873       | Anillo bencénico sustituido con H "aislado"  |
| 860-750   | Anillos aromáticos simples y condensados   |
| 820       | Sistema angular de anillos condensados   |
| 758       | Anillos bencénico <i>o</i> -sustituido, <i>m</i> -sustituido<br>Sistemas de anillos aromáticos |
| 720       | Banda de balanceo de los CH <sub>2</sub>   |

En la figura 7 se observan bandas de adsorción características del agua a 3222 cm<sup>-1</sup> (vibraciones de tensión O-H), provenientes del agua absorbida por el KBr usado en la preparación de la pastilla, afectando la región espectral donde aparecen las bandas para las señales alifáticas y aromáticas. Cabe mencionar que la total eliminación del agua absorbida por el KBr es muy difícil de lograr aun secándolo en la mufla o tomando los espectros en atmosfera inerte.

De esta manera la calidad del espectro IR depende de varios factores: como el tamaño de las partículas de la muestra, el índice de refracción de la muestra, la cantidad de KBr utilizado (si se desea cuantificar) y la humedad de este. La toma

de los espectros IR que realizamos para fondos de vacío presentaron algunos inconvenientes, debido a que las muestras son oscuras y viscosas. La preparación de las pastillas resultó difícil, ya que éstas permanecían opacas aun después de agregar una representativa cantidad de KBr.

Por otra parte, la reproducibilidad de los espectros FT-IR también dependen de la utilización de las mismas cantidades de muestra y de KBr, ya que es común que los espectros de hidrocarburos presenten corrimiento de línea base debido a la dispersión de la radiación <sup>[15]</sup>.

Una técnica que resuelve estos problemas es la espectroscopia Fotoacústica PAS (del inglés PhotoAcoustic Spectroscopy) en el MIR, porque sirve para el análisis de muestras ópticamente opacas, mostrando una repetibilidad aceptable en la obtención de los espectros IR <sup>[15,29]</sup>.

Por tal motivo el presente trabajo se realizó combinando la técnica FT-IR con la PAS, ya que de este modo se aprovecha la elevada relación señal-ruido, la velocidad de operación y la gran capacidad del procesamiento de los datos que presenta la técnica FT-IR, mientras que con la PAS se elimina la preparación de la muestra y obtienen espectros reproducibles.

### **1.3.1. Efecto Fotoacústico (PA)**

El termino fotoacústica (PA), se refiere a la generación de ondas acústicas en una muestra debido a la absorción de radiación electromagnética de forma pulsada o modulada. La señal fotoacústica se origina en la fracción no radiativa (producción de calor) de la energía liberada por el decaimiento desde un estado excitado al estado fundamental; es una técnica complementaria con las técnicas convencionales en las cuales se utilizan los procesos radiativos (absorción-emisión, fluorescencia etc.). De esta manera, proporciona espectros similares a los de absorción óptica en cualquier tipo de material líquido, sólido, semisólido, gaseoso, cristalino, pulverizado o amorfo. Además, puesto que sólo la luz

absorbida se convierte en sonido, la luz dispersada no genera mayores inconvenientes.

El efecto PA fue descubierto por Alexander Graham Bell (1847-1922) en 1890 mientras intentaba transmitir la voz a grandes distancias utilizando la luz solar como portadora de la información, experimentando descubrió que una señal acústica puede producirse iluminando periódicamente con radiación modulada, una muestra colocada en una celda cerrada <sup>[16]</sup>.

Aunque el efecto PA en esos años ganó el interés de algunos investigadores, como Röntgen, Tyndall y (Lord) Rayleigh, permaneció como una curiosidad científica por casi medio siglo hasta que:

Se dio el desarrollo del micrófono, que sustituiría al tubo de escucha en el montaje experimental de Bell.

Surgieron fuentes luminosas intensas como los láseres.

Se desarrollaron sensibles sistemas de detección (amplificadores sensitivos a fase o sincrónicos) y

Aparecieron los equipos para el procesamiento de datos.

De este modo comenzaron a gestarse las primeras aplicaciones prácticas. No obstante fue de gran importancia el modelo desarrollado años después por Rosencwaig y Gersho <sup>[19]</sup> con el cual demostraron que la principal fuente para la generación de la señal PA es 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, ésta 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.

Para dar explicación al fenómeno se basaron en la opacidad, la difusividad térmica y el grosor total de la muestra. Para muestras ópticamente opacas, la magnitud de

la señal fotoacústica,  $q$ , y la fase  $\phi$ , se pueden expresar como se muestra en las ecuaciones 1 a 3.

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

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

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

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

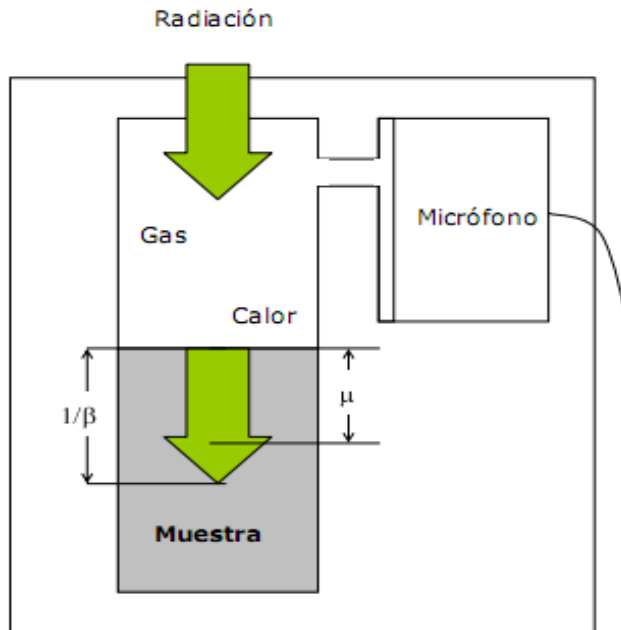
En los últimos años el efecto PA ha despertado un enorme interés en todos los ámbitos de la industria, porque abarca muchas otras disciplinas, no exclusivamente acústicas o técnicas de detección térmicas, sino también se ha aplicado con gran satisfacción a varios problemas de la física, química, biología, medicina e ingeniería <sup>[17]</sup>. Además varios investigadores han mejorado el diseño de alguna de las partes de la celda PA. Por ejemplo se ha recurrido al uso de múltiples micrófonos para mejorar el límite de detección de las señales, lo que ha mostrado una excelente mejora en la relación señal-ruido (S/N) <sup>[18]</sup>.

En la industria petrolera la aplicación de la PAS nace por la necesidad de caracterizar y analizar las fracciones pesadas de los crudos, y así asegurar que estos crudos sean bienes rentables para las industrias que los extraen y refinan,

ya que las técnicas con las que se analizan los crudos livianos no se han podido usar con éxito para estos fines.

Como se mencionó anteriormente la fotoacústica es una técnica espectroscópica no destructiva, la muestra que se va a analizar se coloca en una celda cerrada que contiene aire u otro gas. La señal PA se conduce desde  $\mu$  hasta la superficie donde calienta una delgada capa del gas que se encuentra más próxima a la muestra (Figura. 8). La cantidad de calor transportada de este modo es dependiente tanto de las características termofísicas de la muestra como de las del gas. La capa calentada del gas actúa como pistón acústico que se expande 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 <sup>[15]</sup>.

**Figura 8. Generación de la señal fotoacústica en un sólido.**



Por otra parte la diferencia fundamental entre la espectroscopía fotoacústica y la espectroscopía de absorción convencional, es la forma como se detecta la absorción. En la espectroscopía convencional esta medición es indirecta, en realidad se detecta la radiación transmitida y posteriormente se realiza la conversión a unidades de absorbancia. Las muestras con bajos coeficiente de absorción,  $\beta$  (figura 7), difícilmente pueden ser analizadas por esta espectroscopía 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 altos valores de  $\beta$ , porque la magnitud de la señal transmitida es muy pequeña comparada con la señal de referencia y fácilmente se pierde en el ruido de fondo. El rango de detección lineal de éste tipo de experimentos, está limitado a las muestras con un rango medio de  $\beta$ .

En contraste la absorción de la radiación electromagnética en espectroscopía 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  $\beta$ , debido a que el cambio inducido de presión es directamente proporcional a la energía absorbida.

Por último 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. La señal que se obtiene es procesada por el equipo IR y obtenido el interferograma, el cual es transformado para producir un espectro de absorción <sup>[15]</sup>.

#### **1.4. QUIMIOMETRÍA**

La quimiometría es la aplicación de métodos estadísticos y matemáticos a los problemas químicos con el objetivo de extraer la máxima y más útil información de los mismos. El desarrollo de la instrumentación química ha llevado a la necesidad

de utilizar métodos avanzados de diseño de experimentos, calibración y análisis de los datos resultantes. Por esta razón ha venido despertando el interés como herramienta de análisis, clasificación y calibración multivariable, debido a la imposibilidad de describir y modelar sistemas químicos mediante la estadística univariable (calibración convencional).

Recientemente, la calibración multivariable se ha aplicado para el análisis de muchos sistemas químicos, incluyendo aguas residuales, alimentos, formulaciones de medicamentos, aceites y combustibles, sin la necesidad de la separación de 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. En la calibración convencional la información química de un conjunto de soluciones depende de sólo una variable (es decir, un número de onda) que se relaciona con la concentración del soluto en la solución. Para el caso específico de la espectroscopía MIR, raramente se puede emplear el análisis univariable para fines cuantitativos; si se requiere un análisis cuantitativo, se debe emplear más de una o todas las variables espectrales con el fin de obtener información suficiente para el desarrollo del procedimiento analítico (análisis multivariable) <sup>[11,13]</sup>.

Esto indica que los procedimientos de calibración convencional tendrían una aplicación limitada para las determinaciones cuantitativas que se desean desarrollar en este trabajo. Por lo tanto la determinación simultánea de varias variables requeriría la aplicación de métodos ASTM o una técnica de calibración multivariable para la resolución del sistema complejo. La segunda opción fue elegida en este estudio debido a la sencillez, rapidez y bajo costo que presenta.

El fin último de los métodos de calibración multivariable es encontrar relaciones útiles entre los datos espectroscópicos obtenidos mediante las mediciones en FT-

IR, y el análisis SARA usando la norma ASTM que es costosa y con mediciones prolongadas. Las relaciones encontradas mediante la calibración multivariable se utilizan para predecir el análisis SARA de las muestras de fondos de vacío rápidamente y con una buena exactitud, evitando la aplicación tediosa de los métodos ASTM.

En la mayoría de los análisis químicos los métodos multivariados aplicados son, la regresión de componentes principales (PCR), regresión por mínimos cuadrados parciales (PLS-R) y el análisis por componentes principales (PCA) [11, 13,14,20,24 34].

#### **1.4.1. Calibración multivariable**

La calibración multivariable se puede definir como la actividad de encontrar relaciones entre una o más variables de respuesta “y” y una matriz de variables predictoras “x”, de manera que se cumpla la ecuación 4.

$$Y = g(X) \quad \text{(Ecuación 4)}$$

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

#### **1.4.2. Clasificación de los métodos de calibración multivariable**

Según Martens & Naes, [22] - los métodos de calibración multivariable más relevantes se clasifican basándose en los siguientes aspectos fundamentales:

- Según la relación entre las variables dependiente e independiente:

*Calibración lineal:* descrita mediante un modelo lineal.

$$Y = b_0 + \sum_{k=1}^k b_k X_k \quad (\text{Ecuación 5})$$

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:

$$Y = \alpha X^\beta \quad (\text{Ecuación 6})$$

Donde  $\alpha$  y  $\beta$  son los parámetros a determinar.

- Según la forma de encontrar la relación entre las variables: pueden ser

*Métodos directos*, donde los parámetros de calibración se calculan directamente a partir de la señal de cada uno de los analitos en forma individual, o

*Métodos indirectos*, donde tales parámetros se calculan a partir de las señales analíticas de mezclas de los componentes.

- Según la variable que se defina como dependiente o independiente:

*Calibración clásica:* si la calibración sigue el criterio directamente relacionado con la ley de Beer-Lambert. Donde la señal analítica actúa como variable dependiente y la propiedad a predecir como variable independiente.

*Calibración inversa:* la propiedad a predecir actúa como variable dependiente y la señal analítica como variable independiente.

Teniendo en cuenta esta última clasificación, cabe destacar como calibración clásica la regresión lineal múltiple clásica (CLS, *Classical Least Squares*) y los métodos de regresión sesgados, basados en la reducción de variables. Estos

últimos, teniendo en cuenta la ordenación de los datos espectroscópicos utilizados en este trabajo se pueden dividir en dos grandes bloques:

*Métodos “unfolded”*: entre los que se destaca la regresión en componentes principales (PCR), la regresión por mínimos cuadrados parciales (PLS) y la regresión continua (CR)- *Métodos “N-Way”*: entre los que se destaca el *parallel factor analysis* (PARAFAC) y el PLS multidimensional (N-PLS).

### **1.4.3. Construcción de modelos de calibración multivariable**

El objetivo de los métodos de calibración multivariable es establecer modelos de calibración que sean capaces de predecir el valor de la propiedad estudiada para nuevas muestras, a las que se ha registrado la señal analítica.

Según La Norma ASTM E-1655 <sup>[23]</sup>, la construcción de modelos de calibración multivariable a partir de espectroscopía IR debe regirse por las siguientes etapas:

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

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

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

-Pretratamiento: el suavizado para reducir el ruido en los datos, la normalización para lograr que los datos estén aproximadamente a la misma escala, la centralización para evitar que ciertos puntos tengan más peso que otros en el modelo y la derivación de diferente orden para extraer información detallada que no puede ser observada en el espectro normal.

-Cálculo del modelo matemático: después de encontrar el pretratamiento óptimo de la señal, se da la aplicación de técnicas de regresión sobre los datos espectrales.

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

-Validación del modelo de calibración: Después de calcular la ecuación para una determinada calibración es esencial determinar la capacidad del modelo para predecir valores desconocidos. Un parámetro estadístico utilizado es el error medio de predicción, basado en el cuadrado de las diferencias entre los valores de referencia y los predichos por el modelo de calibración. Otro es el error medio de calibración definido como:

$$RMSEC = \sqrt{\frac{\sum_{i=1}^c (y_i - \hat{y}_i)^2}{n_c}} \quad (\text{Ecuación 7})$$

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.

Un modelo de calibración se puede validar mediante validación cruzada, si no se cuenta con un número suficiente de muestras o mediante un conjunto externo de muestras de validación.

-Validación cruzada: La validación cruzada, consiste en la utilización de los datos del conjunto de calibración para validar el modelo. 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 validación cruzada total o *leave-one-out*. Cuando se utiliza validación cruzada para validar un modelo se puede definir la raíz cuadrada de la media del error al cuadrado (RMSECV) como:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^c (y_i - \hat{y}_{CV,i})^2}{n_c}} \quad (\text{Ecuación 8})$$

-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.

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

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

Otro parámetro muy utilizado en el caso de la espectroscopia IR para evaluar la calidad de un proceso de regresión es el parámetro SESGO o BIAS que se define como:

$$SESGO = \frac{\sum_{i=1}^{n_p} (\hat{y}_i - y_i)}{n_p} \quad (\text{Ecuación 10})$$

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

-Comparación de la exactitud de dos métodos: La exactitud del método de referencia y el método espectroscópico se compara a través del test de Student de la t para las diferencias entre el valor de referencia y el estimado para (n-1) grados de libertad y nivel de significación  $\alpha$  para un grupo n de muestras. Con este test se determinará si el valor de la media para las diferencias de las muestras pareadas es estadísticamente igual a cero, lo cual indicará si los dos métodos miden igual. Si el valor de la t calculada es menor que el valor de la t crítica entonces se cumple la hipótesis nula (la media de las diferencias es igual a cero) y se podrá afirmar que ambos métodos poseen la misma exactitud.

#### **1.4.4. Técnicas de Pretratamiento de Datos**

##### **1.4.4.1. Suavizado espectral**

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

#### **1.4.4.2. Normalización**

Ésta se usa para lograr que los datos estén aproximadamente a la misma escala, y puede ser mediante el uso de:

-Normalización por rangos: con ésta 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.

-Normalización media: Este es el caso más clásico de la normalización. Consiste en dividir cada fila de una matriz de datos por su media, neutralizando así la influencia de los factores ocultos.

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

-Propiedad de un máximo de muestras normalizadas: Si todos los valores son positivos: el valor máximo se convierte en 1. Si todos los valores son negativos: el valor mínimo se convierte en -1. Si el signo de los cambios de valores en la curva: o bien el valor máximo se convierte en 1 o el mínimo valor se convierte en -1.

#### **1.4.4.3. Corrección de línea base**

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

#### **1.4.4.4. Centralización**

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

$$X'_{i,m} = X_{i,m} - X_m \quad (\text{Ecuación 11})$$

Siendo  $X'_{i,m}$  el dato centrado,  $X_{i,m}$  el dato de la fila  $i$  y la columna  $m$  antes del centrado,  $X_m$  la media de la columna  $m$  ( $X_m = \sum x_{i,m} / I$ ). La propiedad fundamental de los datos centrados es que el valor medio de cada una de las variables es igual a cero.

#### 1.4.4.5. Derivación

La derivada tiene como función extraer información detallada que no puede ser observada en el espectro normal y ésta puede ser de diferente orden <sup>[25]</sup>:

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

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

#### **1.4.5. Métodos basados en reducción de variables**

Estos métodos se basan en que la información contenida en las variables de la señal puede estar contenida en un número menor de variables sin que haya pérdida de información relevante. El proceso de calibración se realiza, no sobre los datos originales, sino sobre estas nuevas variables, simplificando el modelo y la interpretación de los resultados.

Este tipo de métodos de calibración son de espectro completo no presentan problemas de colinealidad ni las consecuencias derivadas de ella, por estas razones, la tendencia actual es la utilización de métodos de calibración basados en una reducción de variables previamente al cálculo del modelo. Generalmente, al igual que en el análisis de componentes principales (PCA), los procedimientos de reducción de variables no son realizados sobre los datos originales sino que se centran o autoescalan previamente, uno de estos métodos es PLS-R (regresión parcial por mínimos cuadrados) <sup>[7,21]</sup>.

#### **1.4.6. Análisis Por Componentes Principales (PCA)**

El análisis por componentes principales es el primer paso cuando se emplean métodos multivariantes de análisis de datos, y se aplica con diferentes finalidades entre ellas cabe destacar la clasificación y la reducción de la dimensionalidad de los datos.

Para esto se parte de una matriz de datos  $\mathbf{X}$ , donde cada fila es una muestra, y cada columna una variable, que puede ser la absorbancia o la longitud de onda. La dimensionalidad de esta matriz, si se han registrado  $m$  muestras y  $k$  variables, será  $m \times k$ . La situación de estas variables podría representarse geoméricamente identificando cada variable con un vector y la no correlación lineal de dos variables como la ortogonalidad (ángulo de  $90^\circ$ ) entre los vectores correspondientes. Por tanto estarán más correlacionadas cuanto más difiera su posición relativa del ángulo recto. Mediante el *análisis de componentes principales* (PCA) se lleva a cabo una diagonalización de la matriz de covarianzas, que transforma las  $m$

variables originales en un número menor de nuevas variables incorrelacionadas que contienen la información principal del sistema, eliminando la información redundante o reduciendo la contribución por parte del ruido.

La ASTM define, en su norma E131 <sup>[30]</sup>, 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), cuya combinación lineal se aproxima a los datos originales de acuerdo al grado de exactitud deseado. 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 <sup>[26]</sup>.

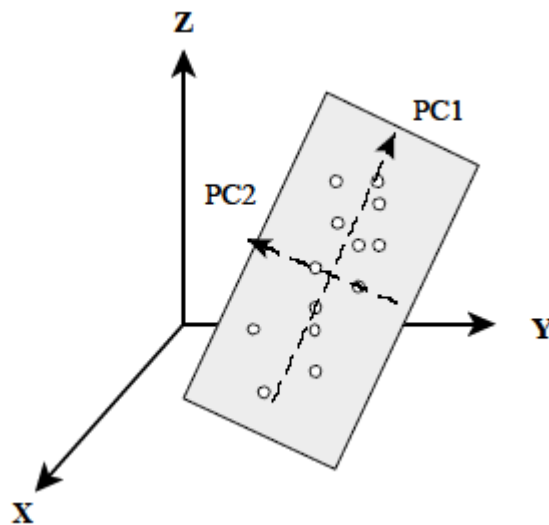
Geoméricamente, el PCA es un cambio en la dirección y dimensionalidad de los ejes (figura 9). Busca las direcciones ortogonales que explican la máxima variabilidad de las muestras y las utiliza como nuevos ejes de coordenadas (PC) para representarlas. Los componentes principales se determinan con base en el criterio de varianza máxima. El primer componente principal es la dirección que explica la máxima variabilidad de la matriz  $\mathbf{X}$ ; el segundo se escoge de tal forma que sea perpendicular al primero y que explique la máxima variabilidad de la matriz  $\mathbf{X}$  no contenida en el primer componente, y así sucesivamente. De este modo unas pocas variables explicarán un porcentaje suficientemente alto como para prescindir de las variables originales y trabajar a partir de ese momento con las variables nuevas; de esta manera se podría reducir considerablemente la dimensión del análisis posterior.

Para poder definir los componentes principales se utiliza la gráfica de *pesos* (*loadings*) y la gráfica de *puntuaciones* (*scores*). La gráfica de *pesos* contiene los coeficientes de las transformaciones lineales de las variables originales que posteriormente darán lugar a ellos mismos, y geoméricamente son los cosenos de los ángulos que forman con los antiguos ejes, mientras que el grafico de

*puntuaciones (scores)* revela las coordenadas de las muestras en estos nuevos ejes.

Cuanto más cerca están las muestras en la gráfica de las *puntuaciones*, más similares son con respecto a los dos componentes en cuestión. Por el contrario, las muestras que se encuentran muy distanciadas (no pertenecen al mismo grupo) son diferentes unas de otras. La gráfica de puntuaciones se puede utilizar para interpretar las diferencias y similitudes entre las muestras.

**Figura 9. Interpretación geométrica de un PCA.**



A partir de la matriz  $\mathbf{X}$  construida de la información espectral obtenida para  $m$  muestras medidas a  $k$  longitudes de onda, se realiza una descomposición por componentes principales que proporciona una aproximación a la matriz  $\mathbf{X}$  como un producto de dos matrices,  $\mathbf{T}$  (matriz de *puntuaciones*) y  $\mathbf{P}$  (matriz de *cargas*), más una matriz  $\mathbf{E}$  de residuales de  $\mathbf{X}$ . (ecuación 12):

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

La matriz  $\mathbf{T}$  contiene información pertinente a las relaciones entre muestras y está constituida por  $m$  filas, que corresponden al número de muestras u objetos, y  $A$

columnas, que corresponden al número de componentes principales más dominantes. La matriz **P** explica la relación existente entre variables originales y está constituida por  $A$  filas y  $k$  columnas, y **E** representa el ruido o la información no relevante.

El producto  $TP^T$  se puede representar como la suma de  $A$  términos (ecuación 12), que corresponde a cada una de las columnas y filas de las matrices **T** y **P** respectivamente cada uno de dichos términos se denomina factor o componente principal [24].

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

Como se dijo anteriormente 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, y los últimos solo describen variaciones debidas al ruido, permitiendo una reducción importante del número de variables. Para tener una reproducción mas perfecta de la matriz **X** se necesitan  $k$  vectores de *cargas* y *puntuaciones*, sin embargo es posible representar la matriz original con un numero  $a < k$  de vectores.

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

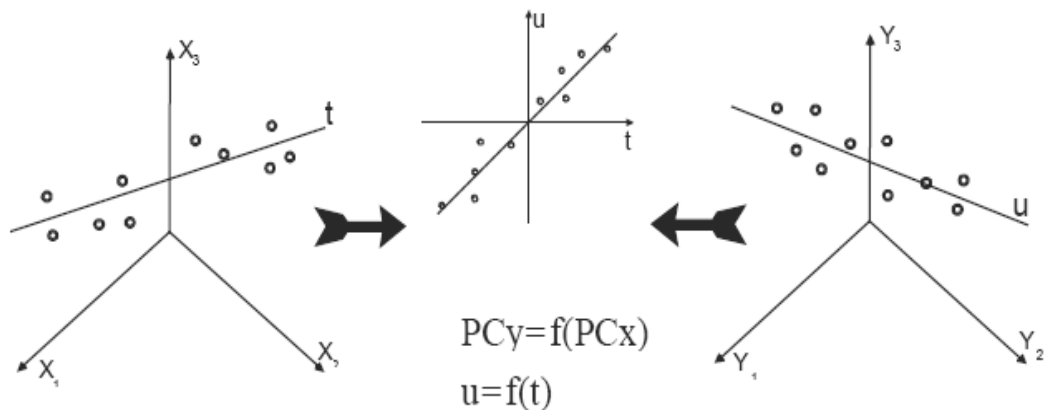
Existen diferentes algoritmos de cálculo para obtener las matrices **T** y **P**. uno de ellos es el NIPALS (*Nonlinear Iterative Partial Least Squares*) [Wold, 1966], y es el algoritmo que utiliza el software *The Unscrambler* [CAMO, 1996] utilizado en este trabajo.

#### **1.4.7. Regresión por Mínimos Cuadrados Parciales (PLS)**

Este método pretende que los primeros componentes contengan la mayor información para la predicción de la matriz **Y**. Para ello, durante la etapa de

calibración, el PLS utiliza tanto la información contenida en la matriz de datos (matriz  $X$ . Datos espectroscópicos) como la información contenida en la matriz de la propiedad a determinar (matriz  $Y$ , análisis SARA), para encontrar un conjunto de variables auxiliares llamadas variables latentes  $VL$  o factores en  $X$  que mejor predicen las variables latentes en  $Y$  <sup>[21,11]</sup>. En la figura 10 se muestra la descripción del PLS.

**Figura 10. 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 14 y 15.

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

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

Donde:

$T$  y  $U$  son las matrices de *puntuaciones* (*scores*) de  $X$  y  $Y$  respectivamente;

$P$  y  $Q$  son las matrices de *pesos* (*loadings*) de  $X$  y  $Y$  respectivamente;

**E** y **F** son los residuos.

Si tenemos  $m$  muestras,  $a$  factores,  $k$  variables y  $p$  analitos, la dimensionalidad de las matrices es la siguiente: **T** y **U** ( $m \times a$ ), **P**<sup>T</sup> ( $a \times k$ ) y **Q**<sup>T</sup> ( $a \times p$ ). En este caso los *loadings* no coinciden exactamente con la dirección de máxima variabilidad de las muestras como en el caso de PCA, ya que están corregidos para obtener la máxima capacidad predictiva para la matriz **Y**.

La descomposición de ambas matrices no es independiente, sino que se realiza de forma simultánea, estableciéndose una relación interna entre los scores de los bloques **X** y **Y** de forma que para cada **VL**  $a$ , se cumpla que:

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

Donde  $b_a$  es el coeficiente de regresión para cada **VL** obtenido a través de la ecuación 17.

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

A partir de la ecuación 16 se calcula el valor de **Y** utilizando la relación interna  $u_a$

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

Donde **B** es la matriz de los regresores de  $b_a$ , de dimensiones ( $a \times a$ ), y **F** la matriz de los residuales de **Y**. En el caso de calcular una sola propiedad de la matriz **Y** el algoritmo recibe el nombre de PLS 1 y si se determinan simultáneamente varias propiedades recibe el nombre de PLS 2 <sup>[21]</sup>.

## 2. PARTE EXPERIMENTAL

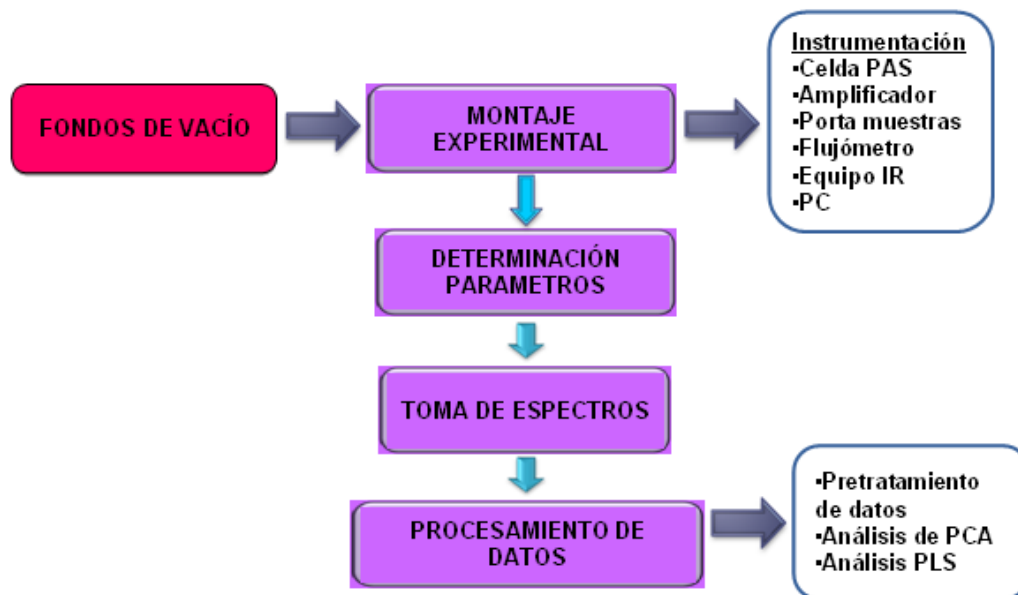
En el Laboratorio de Espectroscopía Atómica y Molecular (LEAM) de la Universidad Industrial de Santander, se han venido desarrollando en los últimos años investigaciones originadas por el convenio 003 de 2007 ICP-UIS, generando aportes científicos y tecnológicos en la caracterización del petróleo colombiano.

Actualmente, el laboratorio desarrolla investigaciones del crudo y sus fracciones como el análisis de asfaltenos por medio de la técnica de PAS, crudo por FTIR-ATR y fondos de vacío por LIF, entre otras.

### 2.1. ESQUEMA EXPERIMENTAL

En la figura 11 se muestra un esquema metodológico de los pasos que se realizaron para el desarrollo del proyecto.

Figura 11. Esquema metodológico



## 2.2. SELECCIÓN DE MUESTRAS

Para la realización del presente trabajo se utilizaron treinta y seis (36) muestras de fondos de vacío de diferentes crudos colombianos, veintiocho de ellas fueron usadas como grupo de calibración (tabla 3), y las ocho restantes se usaron como grupo de validación externa (tabla 4); tanto las muestras como el análisis SARA (realizado por medio de la norma ASTM D-2007) fueron suministradas por el Instituto Colombiano de Petróleo (ICP)

**Tabla 3. Análisis SARA del grupo de calibración**

| Fondo de vacío | Saturados | Aromáticos | Resinas | Asfaltenos |
|----------------|-----------|------------|---------|------------|
| 1              | 11,1      | 47,7       | 30,7    | 10,5       |
| 2              | 16,4      | 42,1       | 32,8    | 8,7        |
| 3              | 14,6      | 48,8       | 27,5    | 9,1        |
| 4              | 13,1      | 38         | 29,8    | 19,1       |
| 5              | 19,1      | 40,3       | 29,9    | 10,7       |
| 6              | 19,3      | 47,5       | 28,4    | 4,8        |
| 7              | 12,4      | 45,3       | 31,8    | 10,5       |
| 8              | 9,9       | 59,6       | 22,1    | 8,4        |
| 9              | 4,5       | 56,3       | 26,2    | 13         |
| 10             | 4,1       | 37,1       | 27,5    | 31,3       |
| 11             | 6         | 42         | 32,8    | 19,2       |
| 12             | 10,4      | 59,2       | 16,8    | 13,6       |
| 13             | 17,8      | 49,9       | 23,4    | 8,9        |
| 14             | 16,9      | 48,7       | 27      | 7,4        |
| 15             | 35,8      | 44,7       | 18,2    | 1,3        |
| 16             | 15        | 45,1       | 26,4    | 13,5       |
| 17             | 4,2       | 34,1       | 32,9    | 28,8       |
| 18             | 29,8      | 45,7       | 20,3    | 4,2        |
| 19             | 19,2      | 44,9       | 25,2    | 10,7       |
| 20             | 16        | 45,4       | 25,4    | 13,2       |
| 21             | 15,2      | 42,3       | 36,6    | 5,9        |
| 22             | 11        | 44,4       | 31,1    | 13,5       |
| 23             | 17,6      | 44,3       | 28,9    | 9,2        |
| 24             | 2,8       | 35,9       | 31,6    | 29,7       |
| 25             | 13        | 39,5       | 18,2    | 29,3       |
| 26             | 24,3      | 49,8       | 20,1    | 5,9        |

|    |      |      |      |     |
|----|------|------|------|-----|
| 27 | 15,8 | 49,5 | 25   | 9,7 |
| 28 | 21,1 | 40,9 | 29,2 | 8,8 |

**Tabla 4. Análisis SARA del grupo de validación**

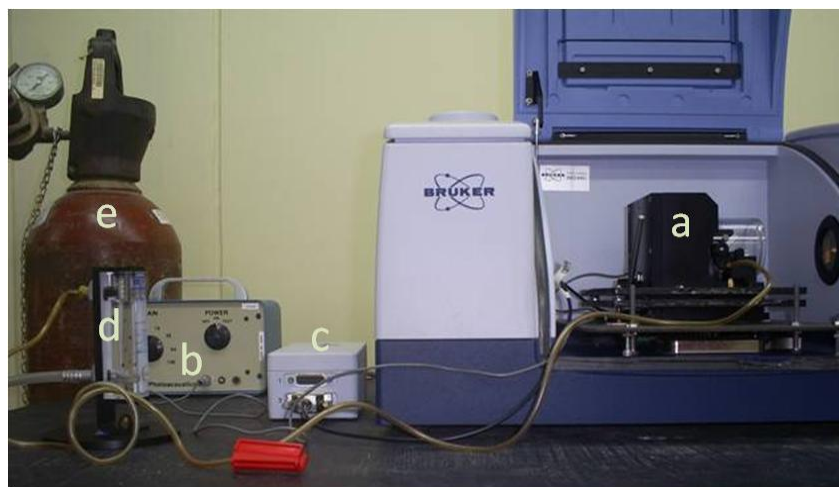
| Fondo de vacío | Saturados | Aromáticos | Resinas | Asfaltenos |
|----------------|-----------|------------|---------|------------|
| 1              | 11,8      | 41,7       | 29,8    | 16,7       |
| 2              | 7,3       | 40,7       | 26,9    | 25,1       |
| 3              | 5,4       | 54,5       | 28,8    | 11,3       |
| 4              | 14,4      | 43,6       | 24,8    | 17,2       |
| 5              | 19,9      | 49,7       | 21      | 9,4        |
| 6              | 13,8      | 49,7       | 25,6    | 11         |
| 7              | 16,1      | 45,8       | 26,5    | 11,6       |
| 8              | 14,8      | 51,4       | 24,8    | 9,1        |

## 2.3. MONTAJE EXPERIMENTAL

### 2.3.1. Instrumentación

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

**Figura 12. Sistema PAS acoplado al equipo IR Tensor 27 marca Bruker**



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

1. *Celda PAS* (figura 12a). Está formada principalmente por un compartimiento para la muestra, 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 13 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 13. Celda PAS modelo MTEC 200**



2. *Amplificador* (figura 12b) 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. *Interfaz* (figura 12c) usada para llevar la señal fotoacústica hasta el procesador de la señal del equipo FT-IR.

4. *Flujómetro* (Figura 12d) se encarga de controlar el flujo de gas He 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 H<sub>2</sub>O y CO<sub>2</sub>). El flujo de gas debe ser controlado para minimizar el riesgo de daño del micrófono.

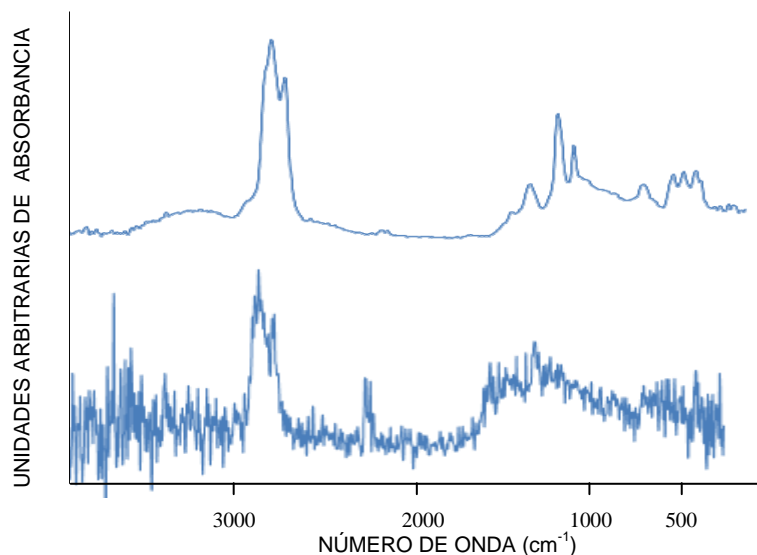
## 2.4. DETERMINACIÓN DE PARAMETROS

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. En general, el ruido degrada la exactitud y la precisión, y establece un límite inferior en la cantidad de analito que se puede detectar. 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.

$$\frac{S}{N} = \frac{\text{media}}{\text{desviación estándar}} = \frac{\mu}{\sigma} \quad (\text{Ecuación 18})$$

La figura 14 muestra el aspecto de una muestra de fondo de vacío con diferente relación S/N.

**Figura 14. Efecto de la relación señal ruido en un espectro de una muestra de fondo de vacío**



Las variables que se tuvieron en cuenta para obtener los espectros IR con la mejor relación señal- ruido fueron:

- Número de barridos.
- Velocidad de espejo.
- Ganancia de la señal.
- Resolución

Las mejores condiciones para la toma de espectros fueron: una velocidad de espejo de 0.0583 cm/s, una resolución de 8  $\text{cm}^{-1}$ , 256 barridos y una ganancia de 1. Debido a que bajo estas condiciones se obtuvo el máximo valor de la relación señal/ruido, se fijaron estos valores y se realizó la toma de los espectros de los 37 fondos de vacío.

#### **2.4.1. Verificación del desempeño del Espectrofotómetro**

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

Cuando se hace un experimento por segunda vez, éste se repite en algunos aspectos pero no en todos; se puede considerar que las condiciones externas e internas del experimento no son exactamente las mismas, y que por tanto el experimento es diferente para cada evento. Por tal motivo se calcularon las desviaciones estándar para las absorbancias medidas a 1616,57 y 1462,12  $\text{cm}^{-1}$  en 7 espectros adquiridos sobre una misma muestra. Las desviaciones encontradas para cada uno de los espectros fueron inferiores a 0.002 (Figura 15), asegurando de esta manera la repetibilidad del espectrofotómetro.

**Figura 15. Prueba de repetibilidad del espectrofotómetro FTIR**



## 2.5. ADQUISICIÓN Y ANÁLISIS DE ESPECTROS PAS

El espectro PAS-FTIR de cada muestra fue el resultado de las siguientes condiciones: intervalo espectral de 4000 a 400 cm<sup>-1</sup>, 256 barridos, velocidad de espejo de 0.0583 cm/s y resolución de 8 cm<sup>-1</sup>. El negro de humo se usó como material de referencia (background) y helio como gas de purga de la celda PAS.

### 2.5.1. Espectros MIR de fondos de vacío

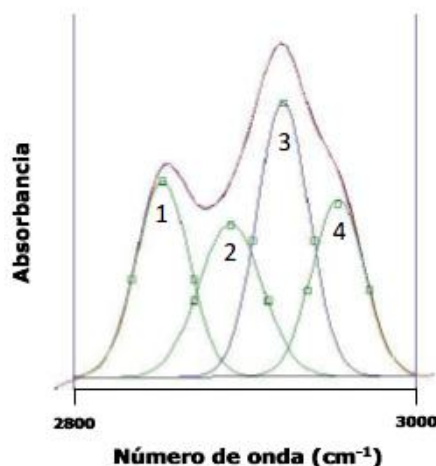
En general, los espectros MIR de los fondos de vacío presentan las mismas bandas apreciadas en los crudos y sus fracciones livianas.

#### Vibraciones alifáticas:

Las flexiones asociadas a la cadena carbonada ( $\delta_{C-C}$ ) son vibraciones de baja energía, fuertemente acopladas y fuera del rango del IR medio. Las vibraciones de tensión ( $\nu_{C-C}$ ) son en general débiles (baja polaridad de estos enlaces) y fuertemente acopladas, observándose en la zona de la huella de la molécula (1600-1300 cm<sup>-1</sup>). Las vibraciones de estiramiento ( $\nu_{C-H}$ ) se presentan en la zona de 3000-2800 cm<sup>-1</sup>, su posición es muy estable y corresponden

esencialmente a los movimientos localmente acoplados de los hidrógenos unidos a un mismo carbono.

**Figura 16. Deconvolución de la región entre 3000-2800  $\text{cm}^{-1}$**



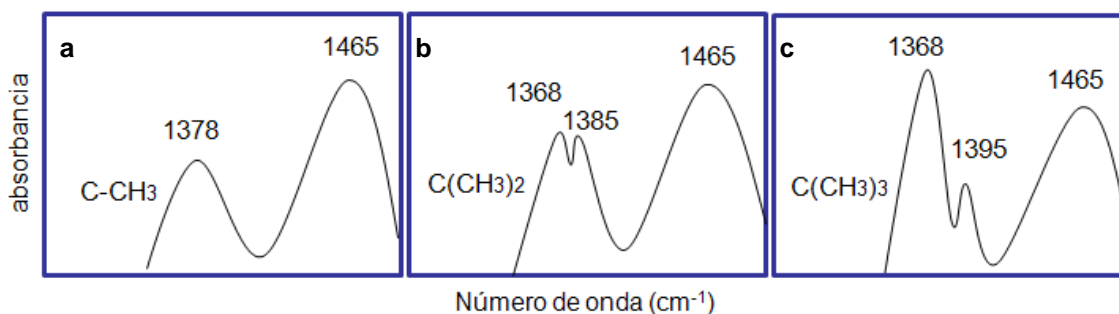
Fuente: ORREGO, J. Estudio de la estructura de cinco carbonos por espectroscopia fotoacústica en el infrarrojo medio<sup>[15]</sup>

Los picos 1 y 2 corresponden a las tensiones asimétricas de los grupos  $\text{CH}_2$  y  $\text{CH}_3$  respectivamente, mientras que los picos 3 y 4 a las tensiones simétricas de estos grupos.

- *Vibraciones asociadas al grupo metilo:*

Las vibraciones  $\nu_{\text{C-H}}$  del grupo metilo corresponden a movimientos acoplados. Se observan 2 bandas: la vibración de estiramiento en el plano de los tres enlaces  $\text{CH}_3$  a  $2870 \text{ cm}^{-1}$  (pico 2 de la figura 16), y a la vibración asimétrica fuera del plano a  $2960 \text{ cm}^{-1}$  (pico 4 de la figura 16), esta última más intensa. Las vibraciones de flexión más relevantes de los grupos metilo son las correspondientes a la flexión de deformación en el plano  $\delta_{\text{CH}_3}$  a  $1378 \text{ cm}^{-1}$ ; esta banda permite la detección de ramificaciones; y la flexión de deformación asimétrica fuera del plano  $\delta_{\text{CH}_3}$  a  $1463 \text{ cm}^{-1}$ . 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 17.

Figura 17. Bandas de vibración para el grupo CH<sub>3</sub>



- a. no hay ramificación si la banda es sencilla. b. La presencia de un grupo iso-propilo se detecta por la aparición de un doblete simétrico (1385 y 1368 cm<sup>-1</sup>). c. la presencia de un grupo ter-butilo se detecta por un doblete asimétrico (1395 y 1368 cm<sup>-1</sup>).

- *Vibraciones asociadas al grupo metileno:*

Al igual que en el grupo metilo, las vibraciones de estiramiento  $\nu_{C-H}$  de los grupos metileno tienen carácter acoplado. Se observan dos bandas que corresponden a la vibración de estiramiento en el plano  $\nu_{CH_2}$  a 2815 cm<sup>-1</sup> (figura 16, pico 1), y a la vibración fuera del plano  $\nu_{CH_2}$  a 2926 cm<sup>-1</sup> (pico 3). La vibración de flexión en el plano más interesante es la de tijera  $\delta_{CH_2}$  a 1463 cm<sup>-1</sup>, que usualmente aparece superpuesta con la banda fuera del plano  $\delta_{CH_3}$  de los grupos metilos. La banda de balanceo  $\tau_{CH_2}$  de los grupos CH<sub>2</sub> 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<sup>-1</sup>.

- *Vibraciones asociadas a los grupos aromáticos:*

Los hidrocarburos aromáticos presentan numerosas bandas de absorción en el MIR. Las vibraciones de tensión  $\nu_{C-H}$  causadas por la hibridación  $sp^2$  de los carbonos aromáticos aparecen a  $3100-3000\text{ cm}^{-1}$ , generalmente son varias bandas cercanas, a veces no resueltas:  $3096-3048$ ,  $3039-3025$  y  $3020-3000\text{ cm}^{-1}$ . A  $1604$ ,  $1585$ ,  $1500$  y  $1450\text{ cm}^{-1}$  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 no observarse.

Las vibraciones de flexión  $\gamma_{C-H}$  dan lugar a bandas intensas en la zona de bajas frecuencias y dependen del número de hidrógenos vecinos en el anillo. En la Tabla 5 se muestra la posición de las bandas  $\gamma_{C-H}$  de acuerdo al número de hidrógenos vecinos.

**Tabla 5. Vibraciones de flexión del C-H aromático**

| H vecinos | Posición de las bandas ( $\text{cm}^{-1}$ ) |
|-----------|---|
| 1         | 900-860                                     |
| 2         | 860-800                                     |
| 3         | 810-750                                     |
| 4         | 770-735                                     |
| 5         | 770 - 730                                   |

## 2.6. PROCESAMIENTO DE DATOS

### 2.6.1. Pretratamiento de datos

Los datos espectroscópicos obtenidos se analizaron utilizando las herramientas del software The Unscrambler® versión X <sup>[32]</sup> donde inicialmente se realizó un procesamiento de los datos para disminuir la fuente de variabilidad causadas por efectos del ruido, movimiento de la línea base, diferencias de escala entre otros.

Debido a que el software presenta diferentes opciones para el pretratamiento de los datos espectrales, se realizó como primera medida un análisis exploratorio para determinar el valor de la desviación estándar relativa (RSD) a tres números de onda característicos en los espectros, como se muestra en la tabla 6.

**Tabla 6. Cálculo de la RSD para tres números de onda característicos aplicando diferentes pretratamientos**

| Número de Onda (cm <sup>-1</sup> ) | 744,6      | 1454,4      | 2924,2     |
|------------------------------------|------------|-------------|------------|
| Datos Originales                   | 28,7       | 12          | 6,8        |
| Normalización Máxima               | 24,2       | 17,2        | 24,1       |
| Normalización Rango                | 32,2       | 14          | 4,4        |
| <b>Normalización Media</b>         | <b>5,6</b> | <b>10,5</b> | <b>1,2</b> |
| Primera Derivada                   | 26,4       | 10          | 10,8       |
| Segunda Derivada                   | 24,9       | 22          | 13,3       |

De esta manera, el único pretratamiento realizado fue la normalización media, ya que generó los menores valores de RSD para las bandas evaluadas.

### 2.6.2. Análisis de componentes principales (PCA)

Se realizó una descomposición por componentes principales (PC) al conjunto total de intensidades espectrales (grupo de calibración) corregidos por el pretratamiento escogido. Esto permitió discriminar las muestras de acuerdo al contenido de asfaltenos (grafico de *puntuaciones*).

### **2.6.3. Regresión por mínimos cuadrados parciales (PLS)**

Mediante la regresión PLS se evaluó el potencial de la espectroscopia PAS-FTIR para predecir el análisis SARA a 36 fondos de vacío colombianos.

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

- *calibración y validación interna:*

La selección de cada modelo se realizó con base en las regiones espectrales más influyentes, evaluando parámetros estadísticos como el RMSEP, RMSEC y la varianza explicada para encontrar el número óptimo de variables latentes.

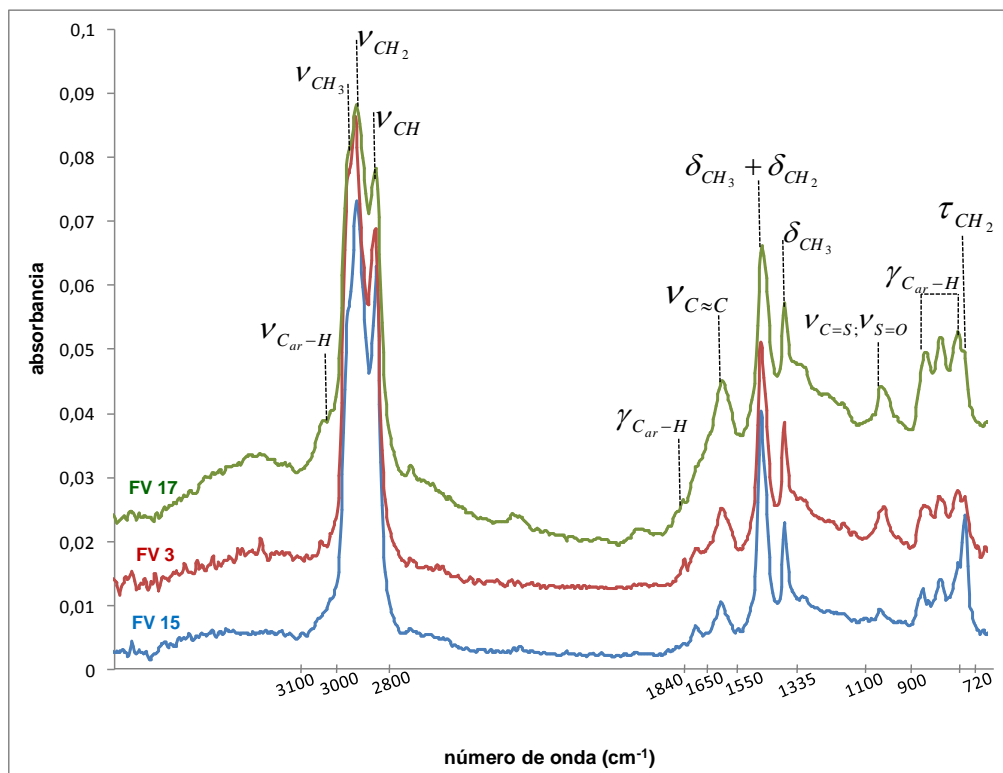
- *Validación externa:*

La validación de los modelos se realizó empleando un grupo de 8 fondos de vacío que no fueron empleados en la calibración.

### 3. RESULTADOS Y ANÁLISIS

#### 3.1. ANÁLISIS ESPECTRAL DE FONDOS DE VACÍO

Figura 18. Espectros de Fondos de vacío



Es importante notar antes de iniciar con el análisis cualitativo de los espectros IR, que los fondos de vacío pueden diferenciarse principalmente según su contenido de saturados y asfaltenos. Ya que, si en los saturados es habitual encontrar cadenas alifáticas de mas de cinco grupos metileno, en los asfaltenos este tipo de estructuras es poco común. Además, a diferencia de los saturados, en los asfaltenos los anillos aromáticos forman estructuras policondensadas de entre tres y cinco anillos <sup>[33]</sup>. Así, los fondos de vacío pueden fácilmente distinguirse por el contenido de estas dos fracciones. En la figura 18 aparecen los espectros de los fondos de vacío 17, 3 y 15, que corresponden a muestras con un contenido de asfaltenos de 28.8, 9.1, 1.3% en peso y un contenido de saturados de 4.2, 14.6, 35.8% en peso respectivamente. Una simple inspección visual de los espectros permite establecer diferencias significativas que pueden asociarse con la

composición química. De 3100 a 3000  $\text{cm}^{-1}$  se puede notar la presencia de una banda atribuida a la vibración de tensión entre el  $\text{C}_{\text{ar}}\text{-H}$  intensificada para la muestra 17. Es decir, que dicha muestra concentra una elevada cantidad de carbonos con hibridación  $\text{sp}^2$  con respecto a las muestras 15 y 3. En la región de 1840-1650  $\text{cm}^{-1}$  se observan diferencias en cuanto a la resolución y área de las bandas correspondientes a las vibraciones de flexión de los enlaces  $\mathcal{V}_{\text{C}_{\text{ar}}-\text{H}}$ . A 1650-1550  $\text{cm}^{-1}$  se observa una banda debida a la tensión del enlace  $\mathcal{V}_{\text{C}_{\text{ar}}\approx\text{C}_{\text{ar}}}$ ; Esto demuestra que el aumento en la aromaticidad aumenta el contenido de los asfaltenos y disminuye el contenido de saturados. Otras bandas características de este tipo de muestras se observan bien resueltas en la región de 1535-1335  $\text{cm}^{-1}$ . A 1462  $\text{cm}^{-1}$  se encuentra una banda intensa producto de la combinación entre la vibración de flexión fuera del plano y la vibración de flexión en el plano, para los grupos metilo y metileno ( $\delta_{\text{CH}_3} + \delta_{\text{CH}_2}$ ). A 1375  $\text{cm}^{-1}$  se aprecia la banda correspondiente a la vibración de flexión en el plano debido a la presencia del grupo metilo ( $\delta_{\text{CH}_3}$ ). En la región comprendida entre 1100-980  $\text{cm}^{-1}$  aparece una banda común a las tres muestras con una intensidad relativa menor en el espectro de la muestra 15. Esto podría ser atribuido a una baja concentración de grupos funcionales con enlaces  $\text{C}=\text{S}$  y  $\text{S}=\text{O}$ , lo cual está directamente relacionado con su bajo contenido de asfaltenos.

Aparecen una serie de vibraciones de flexión fuera del plano  $\mathcal{V}_{\text{C}_{\text{ar}}-\text{H}}$  para los enlaces aromáticos que son sensibles a la naturaleza del sustituyente, en la región de baja energía entre 900-730  $\text{cm}^{-1}$  y que difícilmente diferencian las muestras. Por último, se observa que la muestra 15 tiene una banda sobresaliente a 721  $\text{cm}^{-1}$  que corresponde al balanceo ( $\tau_{\text{CH}_2}$ ) cuando están presentes cadenas carbonadas con 4 o más grupos metilenos contiguos.

Las intensidades espectrales en las regiones 3000-2800, 1535-1335 y 729-720  $\text{cm}^{-1}$ , reflejan que la muestra 15 contiene la mayor concentración de saturados. Las intensidades relativas de las bandas presentes en las regiones 3100-3000, 1840-1550, y 900-730 demuestran una vez más que la muestra 17 contiene la mayor concentración de asfaltenos. De esta manera se puede concluir que a partir de la espectroscopia PAS-FTIR es posible hacer análisis composicional de fondos de vacío.

### 3.2. ANÁLISIS POR COMPONENTES PRINCIPALES (PCA)

Cada espectro corresponde a 881 variables  $x$ , de las cuales 717 se eliminaron porque pertenecen al ruido espectral. Dentro de esas variables se incluyen las que corresponden a la región de 1650-1550 y 1100-980  $\text{cm}^{-1}$ , ya que para algunas muestras la intensidad relativa de esas bandas era igual al ruido. Finalmente el PCA y el PLS se realizaron a partir de 164 variables  $x$ .

En la tabla 7 se registra la varianza de los datos que explica cada PC.

**Tabla 7. Varianza explicada por los primeros 7 componentes principales de las muestras de calibración**

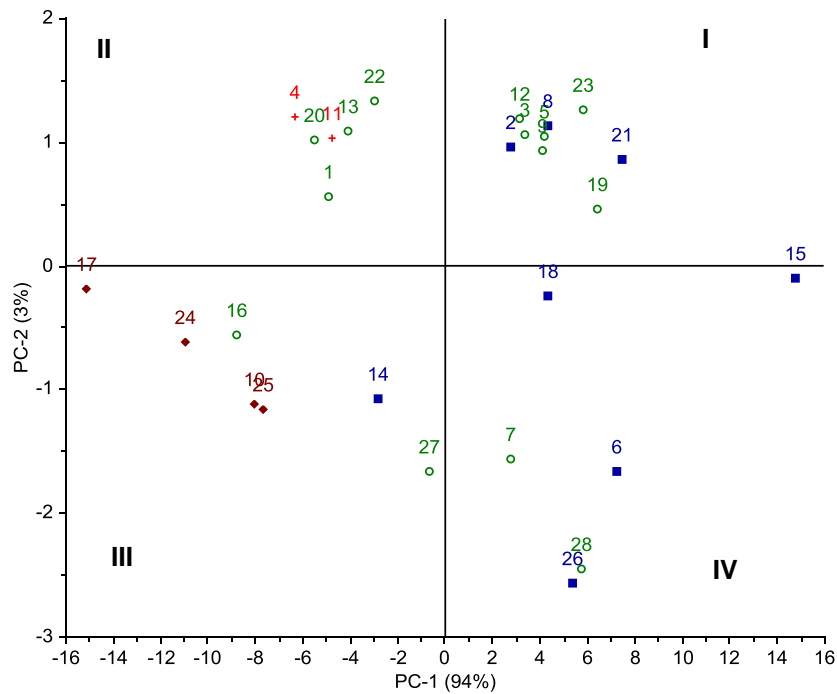
| Componente principal | Varianza Explicada (%) | Varianza Explicada Acumulada (%) |
|----------------------|------------------------|----------------------------------|
| PC 1                 | 94.01                  | 94.01                            |
| PC 2                 | 3.04                   | 97.05                            |
| PC 3                 | 1.44                   | 98.49                            |
| PC 4                 | 0.76                   | 99.25                            |
| PC 5                 | 0.19                   | 99.45                            |
| PC 6                 | 0.1                    | 99.55                            |
| PC 7                 | 0                      | 99.55                            |

De la tabla 7 se observa que a partir del 4 PC la varianza acumulada no aumenta en más de una unidad, siendo tres el número óptimo de componentes principales.

El gráfico de *puntuaciones* (figura 19) muestra cómo las muestras se organizaron entre los componentes principales 1 y 2.

En el primer cuadrante aparecen las muestras que son explicadas de una forma positiva tanto por el PC1 como por el PC2. En el segundo aparecen las muestras que son explicadas de una forma negativa por el PC1 pero positiva por el PC2. En el tercero aparecen las muestras que son explicadas de una forma negativa para los dos PC. Por último, aparecen las muestras que son explicadas de una forma positiva para el PC1 pero negativa para el PC2 (cuarto cuadrante).

**Figura 19. Gráfico de *puntuaciones*: descomposición de la información espectral en el PC2 vs PC1**



Como los componentes principales dan información acerca de las semejanzas o diferencias químicas que siguen las muestras, se realizó una agrupación de acuerdo al contenido de asfaltenos.

Tanto el PC1 como el PC2 no presentan un efecto claro sobre la distribución para las muestras que tienen entre 1.3-13.6 % en peso (grupo verde y azul), ya que estas muestras no se agrupan en un solo cuadrante sino que aparecen dispersas. Sin embargo, sí contienen información que permite diferenciar las muestras agrupadas entre 19.1-31.3 % (grupo marrón y anaranjado), ya que pertenecen a un solo cuadrante (II o III), los dos grupos están ubicadas en la zona negativa para el PC1, mientras que el PC2 tiene un valor positivo para los FV 4 y 11 (grupo anaranjado), y negativo para los FV 17, 24, 10, y 25 (grupo marrón).

De acuerdo con lo anterior se encontró que las muestras se agrupan de acuerdo a la tabla 8.

**Tabla 8. Separación de las muestras según el contenido Asfaltenos**

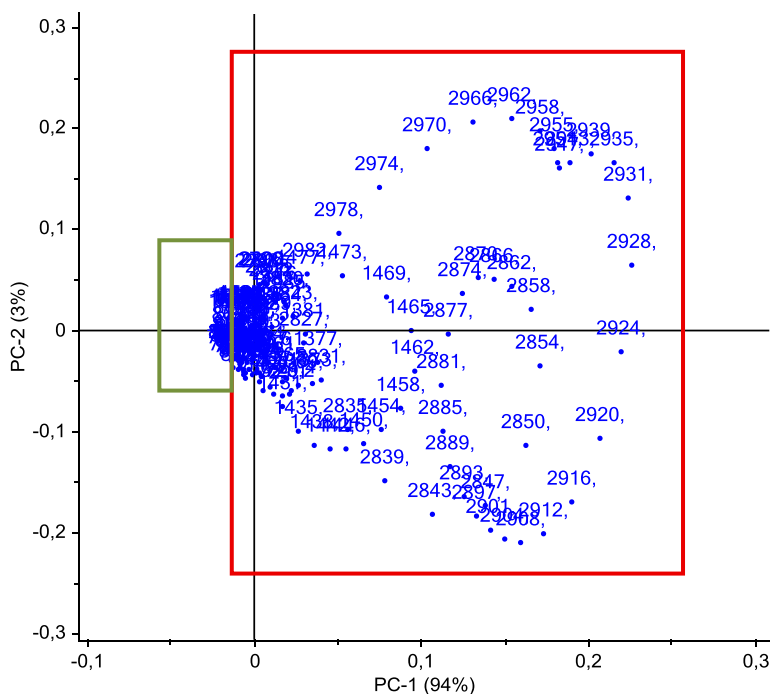
| <b>FV</b> | <b>Asfaltenos</b> |
|-----------|-------------------|
| 10        | 31,3              |
| 24        | 29,7              |
| 25        | 29,3              |
| 17        | 28,8              |
| 11        | 19,2              |
| 4         | 19,1              |
| 12        | 13,6              |
| 16        | 13,5              |
| 22        | 13,5              |
| 20        | 13,2              |
| 9         | 13                |
| 5         | 10,7              |
| 19        | 10,7              |
| 1         | 10,5              |
| 7         | 10,5              |
| 27        | 9,7               |
| 23        | 9,2               |
| 3         | 9,1               |
| 13        | 8,9               |
| 28        | 8,8               |
| 2         | 8,7               |
| 8         | 8,4               |
| 14        | 7,4               |
| 21        | 5,9               |

|    |     |
|----|-----|
| 26 | 5,9 |
| 6  | 4,8 |
| 18 | 4,2 |
| 15 | 1,3 |

En la figura 20 se muestra el gráfico de pesos. Es un gráfico de dispersión en dos dimensiones para los dos primeros componentes, que ayuda a detectar las variables-x importantes.

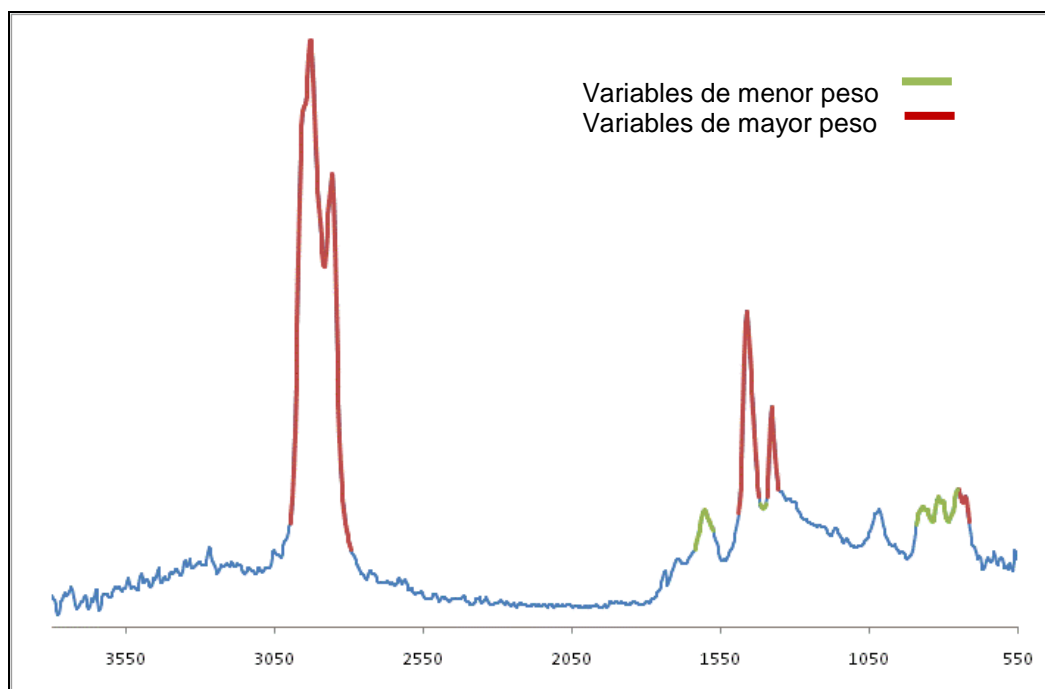
Si se utiliza junto con la gráfica de puntuaciones entonces; las variables a la derecha de la gráfica de pesos serán variables x con valores altos para las muestras a la derecha de la gráfica de puntuaciones (fondos de color azul en la figura 19).

**Figura 20. Gráfico de pesos: Importancia de las diferentes variables para el PC1 y el PC2**



En la figura 21 se muestra cuáles son las variables responsables de las diferencias entre las muestras. Para esto se combinaron la gráfica de *puntuaciones* (figura 19) junto con la gráfica de *pesos* (figura 20) para los dos primeros componentes principales (PC2 vs PC1), ya que estos representan las mayores variaciones de la matriz de datos **X**.

**Figura 21. Contribución de las variables en la formación PC1 y el PC2**



La ubicación de las variables  $x$  en la gráfica de *pesos* afecta la ubicación de las muestras en la gráfica de *puntuaciones*.

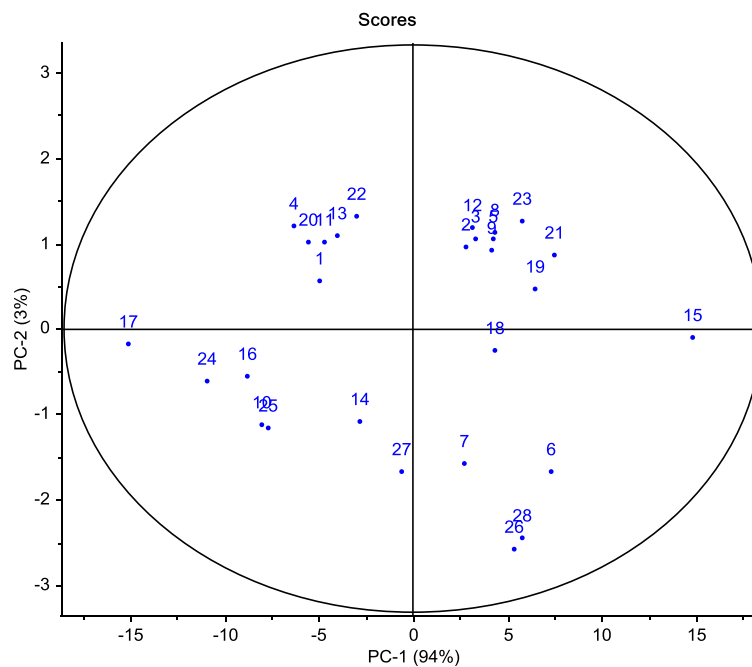
De la figura 21 se puede ver que las variables  $x$  de color verde son debidas a contribuciones puramente aromáticas de las vibraciones entre los enlaces =C-H y C=C; y las variables  $x$  de color rojo son debidas a contribuciones alifáticas (sección 3.1). Esto quiere decir que las muestras difieren entre sí por las diferencias estructurales de los grupos alifáticos presentes en ellas en donde existen mayores posibilidades en el ordenamiento molecular respecto a las estructuras aromáticas.

Las variables que se encontraron a la derecha de la gráfica de pesos (figura 20), son explicadas de una forma positiva por el PC1, por esta razón tienen un valor mayor que las variables que se encuentran a la izquierda. De esta manera las variables  $x$  de mayor peso tienen un mayor valor en las muestras y esto hace que se ubiquen en la zona positiva de la gráfica de puntuaciones (cuadrante I y IV).

Las muestras que están a la derecha de la gráfica de puntuaciones presentan altos valores en el contenido de saturados (FV 15 y 18: 35.8 y 29.8%p de saturados, respectivamente), mientras que las variables  $x$  de color verde que se encontraron a la izquierda de la gráfica de pesos tienen menor valor para las muestras, por lo tanto se ubican a la izquierda de la gráfica de puntuaciones (muestras que son explicadas de manera negativa por los dos primeros PC).

Por otro lado en la gráfica de puntuaciones aparece la elipse de confianza de Hotelling  $T^2$  (figura 22). Los datos que quedan fuera de la elipse indican datos que no quedarían bien explicados por las nuevas variables.

**Figura 22. Gráfico de puntuaciones con el límite de Hotelling  $T^2$ .**



Como se aprecia en esta figura, ninguna muestra presenta valores extremos (ninguna es atípica) de la composición química con respecto a las demás.

### 3.3. DESARROLLO DE LOS MODELOS PLS

#### 3.3.1. Calibración y validación del modelo PLS para la fracción de saturados

El modelo se calculó eliminando las variables  $x$  de menor importancia (con coeficientes de regresión cercanos a cero) hasta que el coeficiente de correlación ( $R^2$ ) y la raíz cuadrada media del error de calibración y de predicción (RMSEC y RMSEP, respectivamente) permanecieron estables. Las muestras alejadas de la curva de calibración y validación también fueron eliminadas. Para comprobar que las muestras excluidas eran atípicas, se realizó de nuevo el modelo pero sacando inicialmente las muestras elegidas y seguidamente se eliminaron las variables  $x$ .

En la tabla 9 se muestra el resumen de los parámetros que se obtuvieron para el modelo.

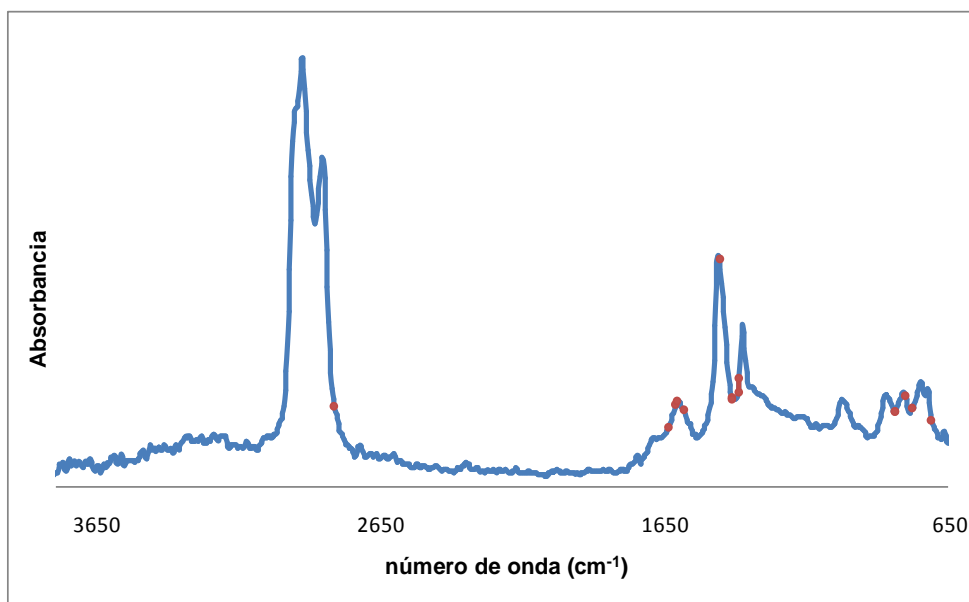
**Tabla 9. Parámetros del modelo seleccionado para la predicción del % de Saturados**

| Variables<br>$X$ | Muestras<br>excluidas | Variables<br>Latentes (VL) | Varianza<br>Explicada | RMSEC<br>(%) | RMSEP<br>(%) | $R^2$ |
|------------------|-----------------------|----------------------------|-----------------------|--------------|--------------|-------|
| 15               | 8,9                   | 5                          | 98.3 %                | 0.83         | 1.36         | 0.99  |

Este modelo requirió el uso de 15 variables  $x$  agrupadas en 5 VL que explican el 98.3% de la variación de los datos. Al excluir los fondos 8 y 9 se obtuvieron errores de predicción y de calibración inferiores a 1.4%, mientras que el coeficiente de correlación fue de 0.99 para la calibración y 0.97 para la de validación. El modelo de predicción tiene un rango de calibración de 2.8 a 35.8 % en peso de saturados.

En la figura 23 se puede observar las 15 variables  $x$  que explican la máxima variabilidad de los datos iniciales, mostrando a cuál región espectral pertenece cada una de ellas.

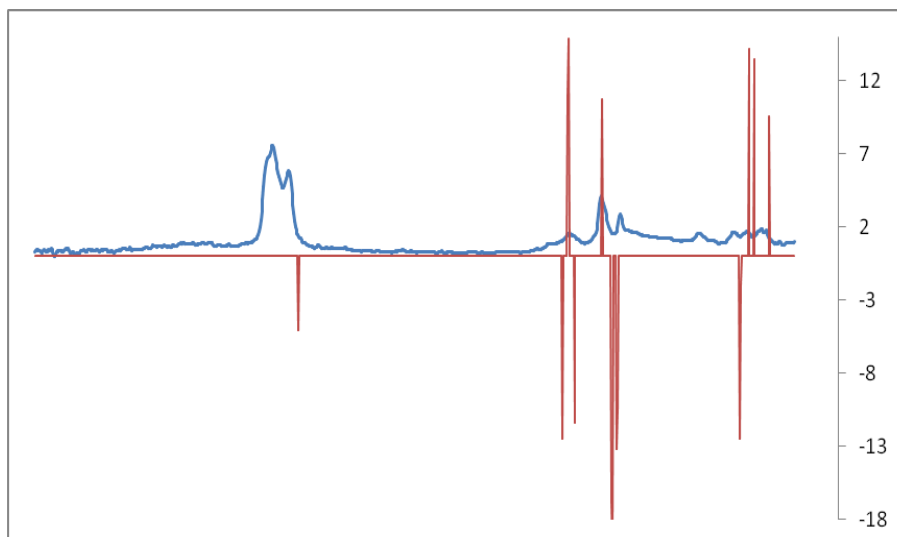
**Figura 23. Ubicación de las variables  $x$  en la región MIR para el modelo de saturados**



Se puede ver cómo con una sola variable se explica la región donde ocurren las vibraciones de estiramiento de los grupos alifáticos ( $\nu_{CH_2}$  a  $2816\text{ cm}^{-1}$ ). Las 14 variables restantes están distribuidas en la región de  $1635$  a  $700\text{ cm}^{-1}$ , donde absorben tanto los grupos alifáticos (6 variables) como los aromáticos (8 variables).

En la figura 24 se muestran los coeficientes de regresión. Estos coeficientes presentan un valor positivo para las intensidades de las señales ubicadas en  $709$ ,  $779$ ,  $802$ ,  $1458$ ,  $1608$ , y  $1612\text{ cm}^{-1}$ . Es decir que a medida que las intensidades ubicadas en estas posiciones aumenten, el contenido de saturados aumenta; mientras que las intensidades que se encuentran a  $837$ ,  $841$ ,  $1388$ ,  $1392$ ,  $1411$ ,  $1415$ ,  $1581$ ,  $1635$  y  $2816\text{ cm}^{-1}$  (ver tabla 10) tienen un coeficiente de regresión negativo, es decir que cuando estas intensidades aumenten proporcionarían una disminución en la fracción de saturados.

**Figura 24. Ubicación de los Coeficientes de regresión para el modelo de saturados**

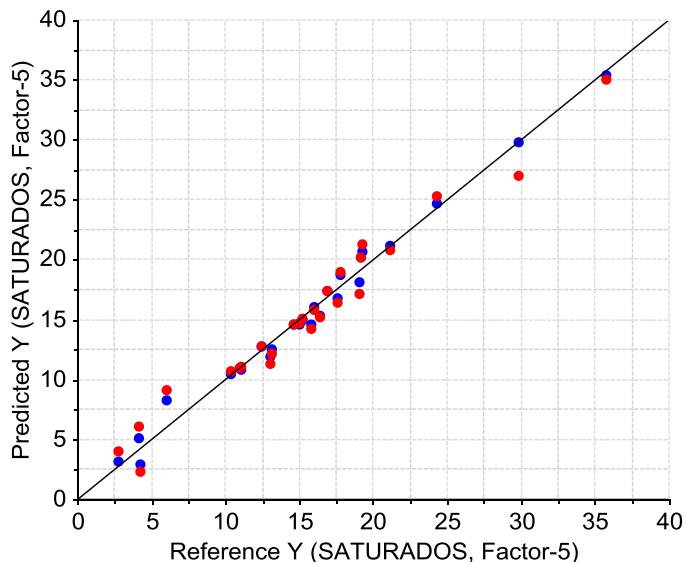


**Tabla 10. Valor de los coeficientes de regresión para 5 VL**

| Variables x<br>(Cm <sup>-1</sup> ) | Coefficientes<br>Regresión |
|------------------------------------|----------------------------|
| 2816,22                            | -5,07                      |
| 1635,72                            | -12,47                     |
| 1612,57                            | 11,06                      |
| 1608,72                            | 14,85                      |
| 1581,71                            | -11,38                     |
| 1458,26                            | 10,68                      |
| 1415,82                            | -16,89                     |
| 1411,97                            | -19,96                     |
| 1392,68                            | -13,22                     |
| 1388,82                            | -10,42                     |
| 841,01                             | -12,47                     |
| 837,15                             | -1,97                      |
| 802,43                             | 14,17                      |
| 779,28                             | 13,42                      |
| 709,84                             | 9,52                       |

En la figura 25 se muestra la curva de predicción obtenida al finalizar la creación del modelo. Los puntos de color rojo representa el valor dado para las muestras de validación cruzada y los puntos azules representan los de las muestras de calibración.

**Figura 25. Curva de calibración del modelo seleccionado para Saturados**



**3.3.1.1. Predicción de la composición de saturados para el conjunto de muestras de validación externa.**

En la tabla 11 se muestra el error relativo entre los valores predichos y los valores de referencia.

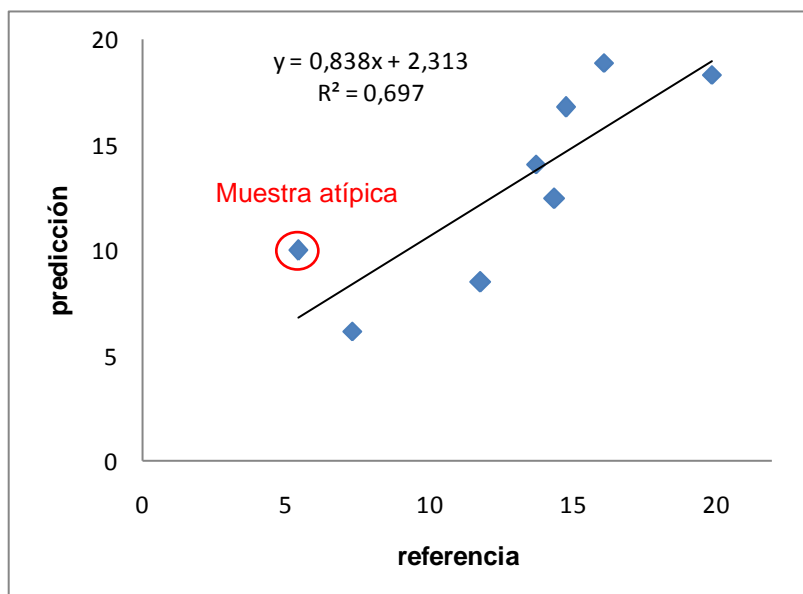
**Tabla 11. Validación del modelo de predicción de %p de Saturados**

| Muestra N° | Referencia (%p) | Predicho (%p) | Residual | RMSEP |
|------------|-----------------|---------------|----------|-------|
| FV 1       | 11,80           | 8,46          | 3,34     | 2,71  |
| FV 2       | 7,30            | 6,17          | 1,13     |       |
| FV 3       | 5,40            | 9,97          | 4,57     |       |
| FV 4       | 14,40           | 12,43         | 1,97     |       |
| FV 5       | 19,90           | 18,36         | 1,54     |       |

|             |       |       |      |
|-------------|-------|-------|------|
| <b>FV 6</b> | 13,75 | 14,11 | 0,36 |
| <b>FV 7</b> | 16,10 | 18,89 | 2,79 |
| <b>FV 8</b> | 14,77 | 16,79 | 2,02 |

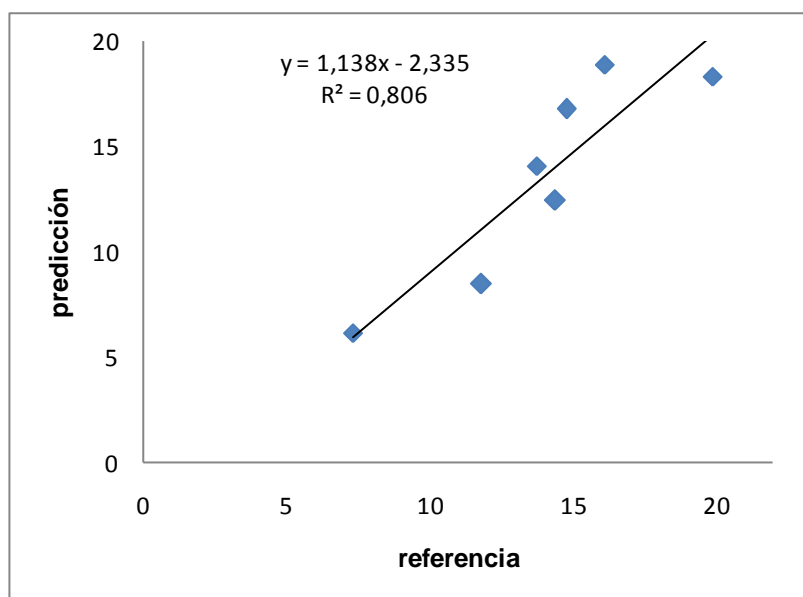
En la figura 26 se muestra la gráfica del valor predicho contra el valor de referencia, con un coeficiente de correlación ( $R^2$ ) de 0.67. Sin embargo, de la tabla 11 se observa que el fondo de vacío 3 presenta el residual más alto, indicando que ésta muestra puede ser atípica.

**Figura 26. Gráfico del valor predicho contra el valor de referencia para el modelo de saturados**



En la figura 27 se muestra cómo el  $R^2$  pasó de 0.67 a 0.81, indicando la favorabilidad de la calibración excluyendo esa muestra.

**Figura 27. Gráfico del valor predicho contra el valor de referencia, eliminando una muestra**



### 3.3.2. Calibración y validación del modelo PLS para la predicción de Aromáticos

En la tabla 12 se muestra el resumen de los parámetros que se obtuvieron para el modelo final.

**Tabla 12. Parámetros del modelo seleccionado para la predicción del %p de aromáticos**

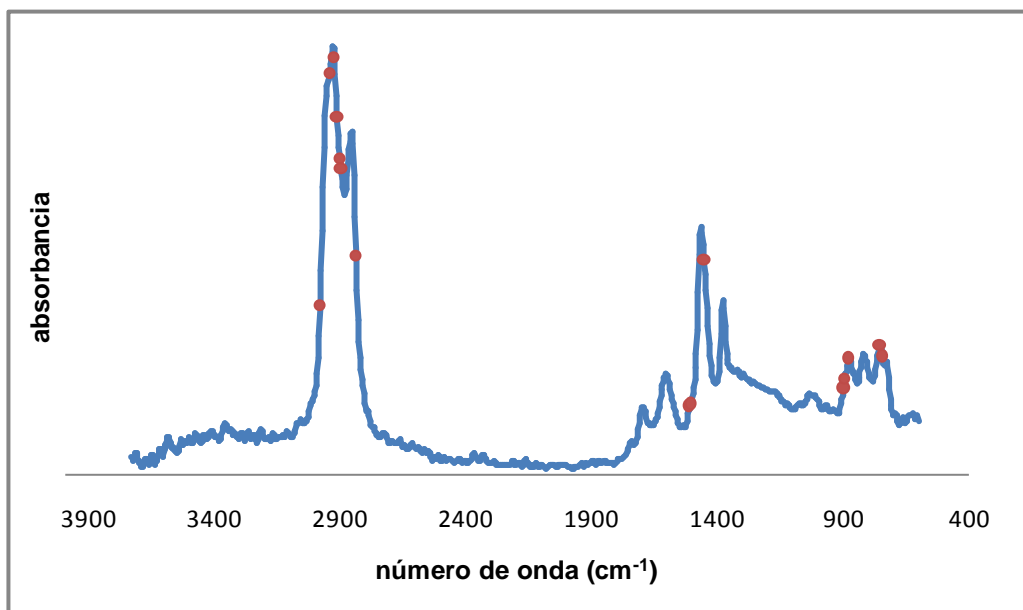
| Variables X | Muestras excluidas | Variables Latentes (VL) | Varianza Explicada | RMSEC (%) | RMSEP (%) | R <sup>2</sup> |
|-------------|--------------------|-------------------------|--------------------|-----------|-----------|----------------|
| 15          | 1,2                | 7                       | 96.6%              | 1.16      | 2.20      | 0.97           |

Este modelo requirió el uso de 15 variables x agrupadas en 7 VL que explican el 96.6% de la variación de los datos. Al excluir las muestras 1 y 2 se obtuvieron errores de predicción y de calibración menores a 2.2%, mientras que el coeficiente

de correlación fue de 0.97 para la calibración y 0.89 para la validación. El modelo de predicción tiene un rango de calibración de 35.9 a 59.6 %.

En la figura 28 se puede observar cómo se ubican las 15 variables  $x$ , mostrando a cuál región espectral pertenece cada una de ellas.

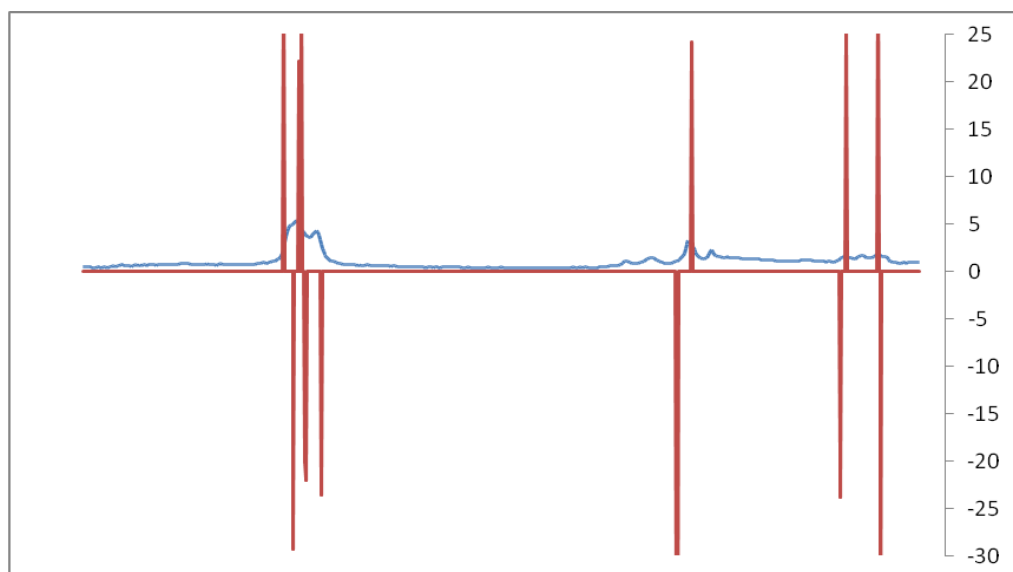
**Figura 28. Ubicación de las variables  $x$  en la región MIR para el modelo de aromáticos**



De la figura 28 se puede ver cómo siete variables explican la región donde se aprecian las vibraciones de tensión de los grupos alifáticos. Las 8 variables restantes están distribuidas entre la región de 1635 a 700  $\text{cm}^{-1}$ , donde absorben tanto los grupos alifáticos ( $\delta_{CH_3} + \delta_{CH_2}$ ) como los grupos aromáticos (5 variables).

En la figura 29 se muestran los coeficientes de regresión de las cinco variables latentes.

**Figura 29. Ubicación de los Coeficientes de regresión para 5 VL en el modelo de aromáticos**



Se determinó que las variables  $x$  pertenecientes a 752, 871, 1450, 2912, 2924, y 2982  $\text{cm}^{-1}$  presentan un efecto positivo para el modelo de aromáticos, es decir, que a medida que aumenten las intensidades a estos número de onda, aumenta el porcentaje de aromáticos en las muestras; mientras que las intensidades que se encuentran a 740, 891, 895, 1504, 1508, 2839, 2897, 2901 y 2943  $\text{cm}^{-1}$  influyen de manera negativa (ver tabla 13). Es decir, cuando estas variables aumentan disminuye el contenido de aromáticos.

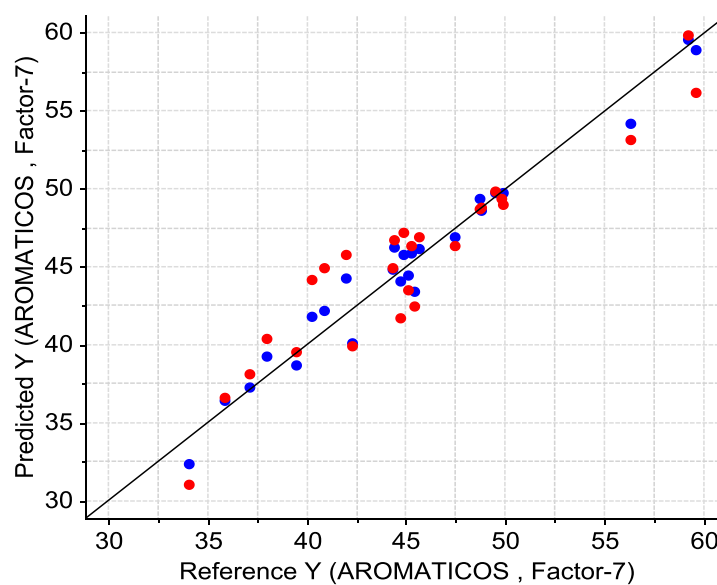
**Tabla 13. Valor de los coeficientes de regresión para 7 VL en el modelo de aromáticos**

| Variabes $x$<br>$\text{Cm}^{-1}$ | Coefficientes<br>Regresión |
|----------------------------------|----------------------------|
| 2982,10                          | 31,25                      |
| 2943,52                          | -29,18                     |
| 2924,24                          | 22,12                      |
| 2912,66                          | 31,79                      |
| 2901,09                          | -20,02                     |
| 2897,23                          | -21,94                     |
| 2839,36                          | -23,56                     |
| 1508,41                          | -36,88                     |

|         |        |
|---------|--------|
| 1504,55 | -45,19 |
| 1450,54 | 24,17  |
| 895,02  | -20,50 |
| 891,16  | -23,79 |
| 871,87  | 26,08  |
| 752,28  | 32,43  |
| 740,70  | -43,43 |

En la figura 30 se muestra la curva de predicción del modelo de aromáticos.

**Figura 30. Curva de calibración del modelo seleccionado para aromáticos**



### 3.3.2.1. Predicción de la composición de aromáticos para el conjunto de muestras de validación externa.

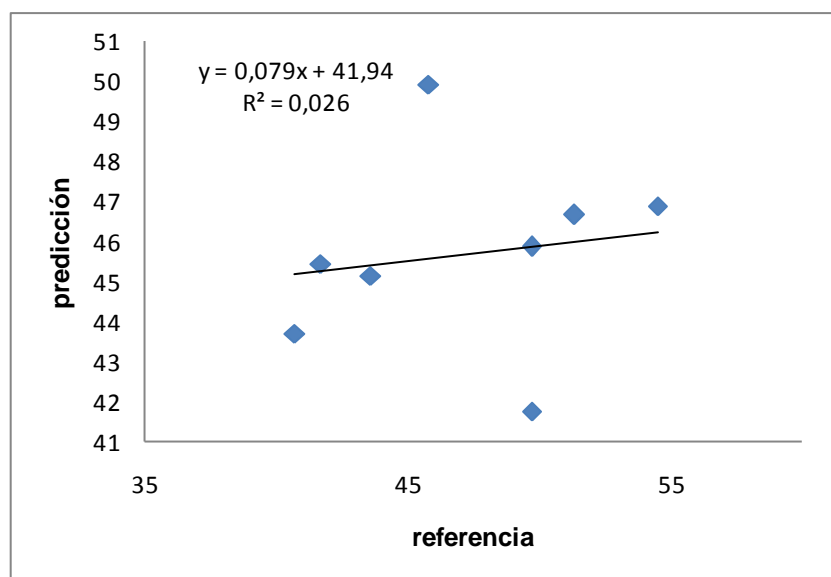
En la tabla 14 se muestra el error para los valores predichos.

**Tabla 14. Validación del modelo de predicción de %p de Aromáticos**

| Muestra<br>Nº | Referencia<br>(%p/p) | Predicho<br>(%p/p) | Residual | RMSEP |
|---------------|----------------------|--------------------|----------|-------|
| <b>FV 1</b>   | 41,70                | 45,45              | 3,75     | 5,36  |
| <b>FV 2</b>   | 40,70                | 43,69              | 2,99     |       |
| <b>FV 3</b>   | 54,50                | 46,87              | 7,63     |       |
| <b>FV 4</b>   | 43,60                | 45,14              | 1,54     |       |
| <b>FV 5</b>   | 49,70                | 45,86              | 3,84     |       |
| <b>FV 6</b>   | 49,72                | 41,73              | 7,99     |       |
| <b>FV 7</b>   | 45,80                | 49,93              | 4,13     |       |
| <b>FV 8</b>   | 51,35                | 46,67              | 4,68     |       |

En la figura 31 se muestra la gráfica del valor predicho versus el valor de referencia, con un coeficiente de correlación ( $R^2$ ) de 0.03

**Figura 31. Gráfico del valor predicho vs valor de referencia para aromáticos**



La fracción de aromáticos tiene más probabilidades de obtener modelos con baja correlación (figura 31). Esto podría explicarse si se considera el procedimiento de

referencia empleado para obtener la fracción de aromáticos (norma ASTM 4124), ya que esta fracción no eluye completamente de la columna cromatográfica sino que, para muestras pesadas como fondos de vacío, parte de los compuestos aromáticos podrían eluir con la fracción de saturados y resinas.

### 3.3.3. Calibración y validación del modelo PLS para la predicción de resinas

En la tabla 15 se muestra el resumen de los parámetros que se obtuvieron para el modelo final.

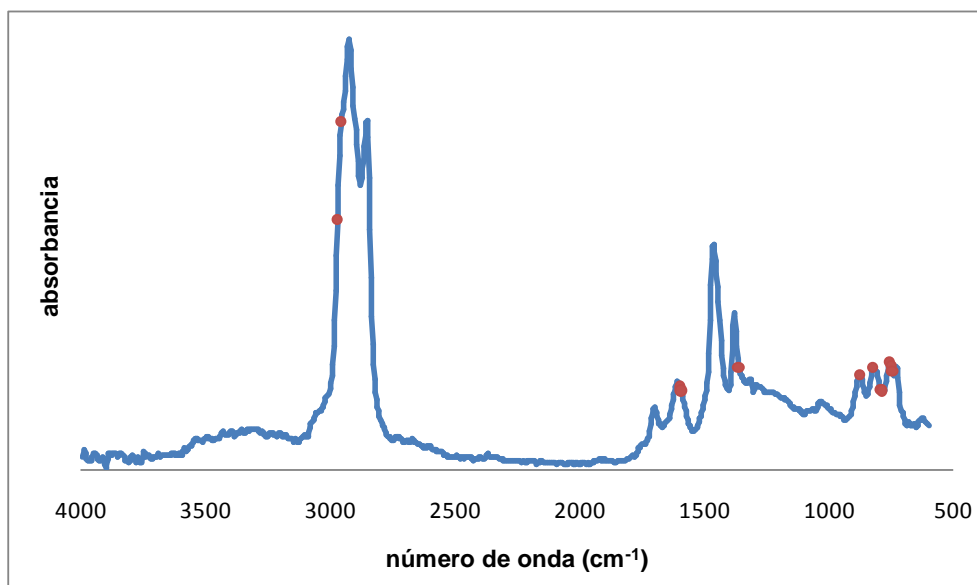
**Tabla 15. Parámetros del modelo seleccionado para la predicción del %p de Resinas**

| Variables<br>x | Muestras<br>excluidas | Variables<br>Latentes (VL) | Varianza<br>Explicada | RMSEC<br>(%) | RMSEP<br>(%) | R <sup>2</sup> |
|----------------|-----------------------|----------------------------|-----------------------|--------------|--------------|----------------|
| 12             | 8,28                  | 7                          | 95.3%                 | 1.00         | 1.72         | 0.96           |

Este modelo requirió el uso de 12 variables x agrupadas en 7 VL que explican el 95.3% de la variación de los datos. Al excluir las muestra 8 y 28 se obtuvieron errores de predicción y de calibración inferiores a 1.7%. El coeficiente de correlación fue de 0.96 para la calibración y 0.90 para la validación. El modelo de predicción tiene un rango de calibración de 16.8 a 36.6 % en peso de resinas.

En la figura 32 se puede observar la ubicación de las 12 variables x, mostrando a cuál región espectral pertenece cada una de ellas.

Figura 32. Ubicación de las variables x en la región MIR para resinas



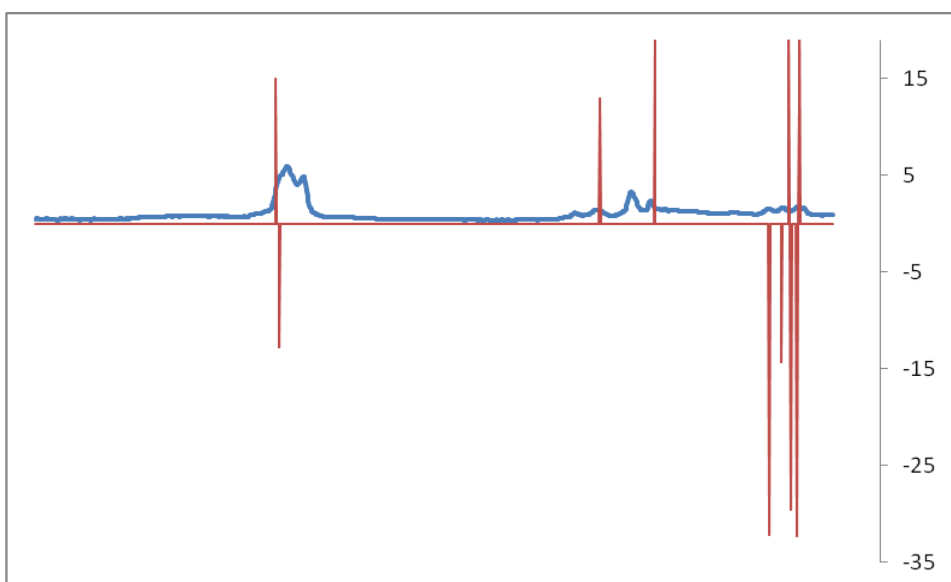
Se puede ver cómo dos variables explican la región donde ocurren las vibraciones de estiramiento de los grupos alifáticos. Una de ellas corresponde a la vibración de estiramiento asimétrico fuera del plano de los grupo metilo ( $\nu_{CH_3}$ ) a  $2970\text{ cm}^{-1}$ , y la otra, a la vibración de estiramiento fuera del plano de los metileno ( $\nu_{CH_2}$ ) a  $2955\text{ cm}^{-1}$ . Dos variables x explican la región donde sucede el acoplamiento de las vibraciones C=C a  $1589$  y  $1593\text{ cm}^{-1}$ . Y una sola explica la región donde se presentan las vibraciones de flexión en el plano por parte de los grupos metilo ( $\delta_{CH_3}$ ) a  $1357\text{ cm}^{-1}$ . Las siete variables x restantes explican las regiones que son debidas a vibraciones de flexión fuera del plano por parte de grupos aromáticos.

**Tabla 16. Valores de los coeficientes de correlación**

| Variables x<br>(cm <sup>-1</sup> ) | Coefficiente<br>correlación |
|------------------------------------|-----------------------------|
| 2970,529                           | 14,97046                    |
| 2955,097                           | -12,71043                   |
| 1593,284                           | 11,78263                    |
| 1589,426                           | 12,89058                    |
| 1357,956                           | 19,98299                    |
| 871,8695                           | -32,10303                   |
| 817,8599                           | -14,26534                   |
| 786,9973                           | 23,68122                    |
| 779,2816                           | -29,5904                    |
| 752,2768                           | -32,31778                   |
| 744,5611                           | 12,98822                    |
| 740,7033                           | 24,08528                    |

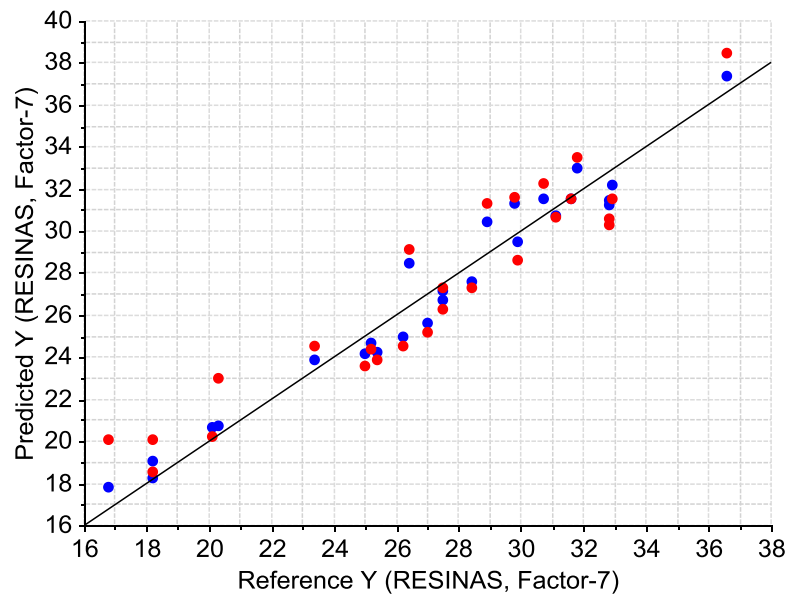
En la figura 33 se muestra la gráfica de los coeficientes de regresión de las 7 variables latentes.

**Figura 33. Ubicación de los Coeficientes de regresión para el modelo de resinas**



De esta manera se puede ver cómo las intensidades de las señales ubicadas en 740, 744 786, 1357, 1589, 1593 y 2970  $\text{cm}^{-1}$  presentan un efecto positivo para el modelo de resinas, es decir, que a medida que las intensidades ubicadas en estas posiciones aumentan, el contenido de resinas aumenta. Las intensidades que se encuentran a 752, 779, 817, 871 y 2955  $\text{cm}^{-1}$  (ver tabla 16) influyen de manera negativa, es decir, cuando estas variables aumentan, disminuye la fracción de resinas. En la figura 34 se muestra la curva predicción obtenida para el modelo de resinas.

**Figura 34. Curva de calibración del modelo seleccionado para resinas**



### **3.3.3.1. Predicción de la composición de resinas para el conjunto de muestras de validación externa.**

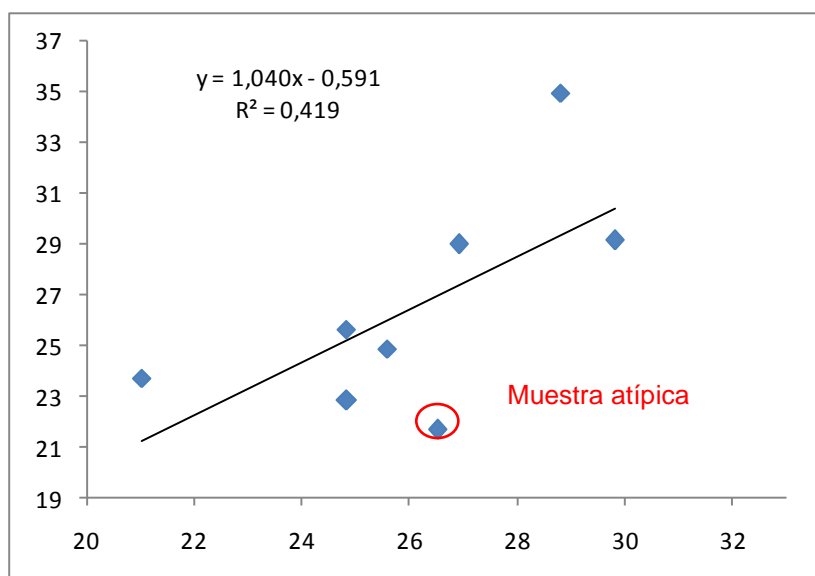
En la tabla 17 se muestran los errores relativos para el modelo de resinas.

**Tabla 17. Validación del modelo de predicción de %p de resinas**

| Muestra<br>Nº | Referencia<br>(%p/p) | Predicho<br>(%p/p) | Residual | Error relativo<br>(%) |
|---------------|----------------------|--------------------|----------|-----------------------|
| <b>FV 1</b>   | 29,8                 | 29,2               | 0,6      | 2,01                  |
| <b>FV 2</b>   | 26,9                 | 29,0               | 2,1      | 7,81                  |
| <b>FV 3</b>   | 28,8                 | 35,0               | 6,2      | 21,5                  |
| <b>FV 4</b>   | 24,8                 | 25,6               | 0,8      | 3,23                  |
| <b>FV 5</b>   | 21,0                 | 23,7               | 2,7      | 12,9                  |
| <b>FV 6</b>   | 25,6                 | 24,9               | 0,68     | 2,66                  |
| <b>FV 7</b>   | 26,5                 | 21,7               | 4,8      | 18,1                  |
| <b>FV 8</b>   | 24,8                 | 22,8               | 2,03     | 8,18                  |

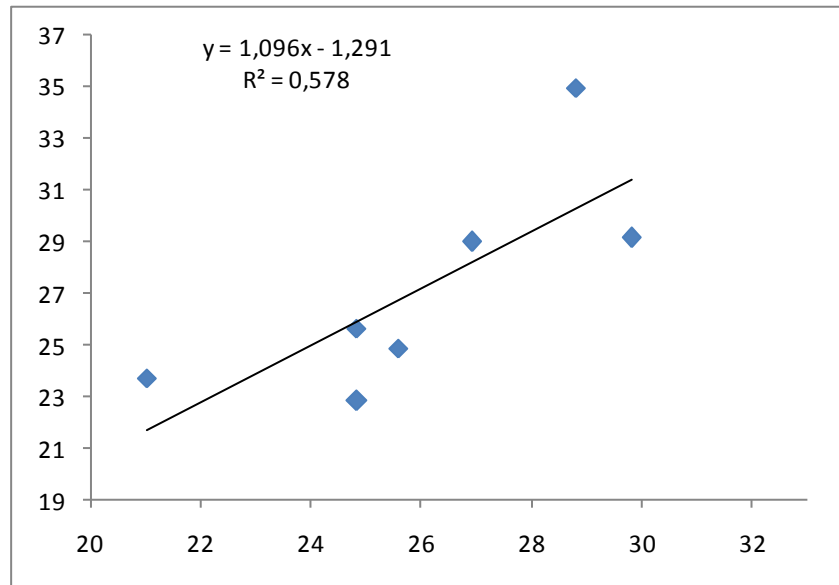
En la figura 35 se muestra la gráfica del valor predicho versus el valor de referencia, con un coeficiente de correlación ( $R^2$ ) de 0.42. Sin embargo se observa que el fondo 7 puede ser una muestra atípica, ya que presenta un error relativamente alto (ver tabla 17).

**Figura 35. Gráfico del valor predicho vs valor de referencia para el modelos de resinas**



En la figura 36 se muestra como el  $R^2$  mejoró, pasando de 0.4 a 0.6 indicando la favorabilidad de la calibración excluyendo esa muestra.

**Figura 36. Gráfico del valor predicho vs valor de referencia.**



### 3.3.4. Calibración y validación del modelo PLS para la predicción de asfaltenos

En la tabla 18 se muestra el resumen de los parámetros que se obtuvieron para el modelo.

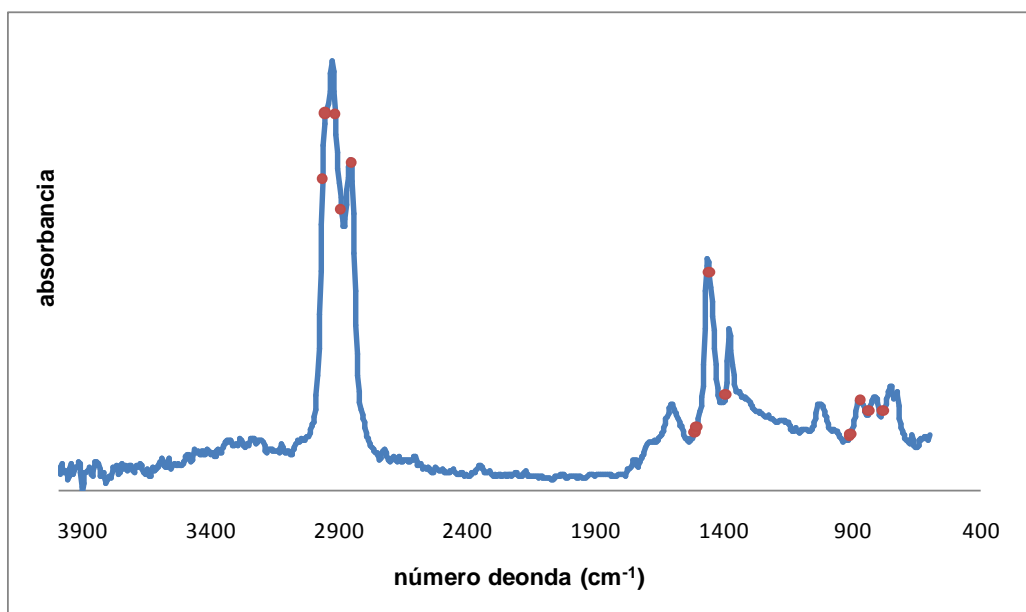
**Tabla 18. Parámetros del modelo seleccionado para la predicción del %p de Asfaltenos**

| Variables x | Muestras excluidas | Variables Latentes (VL) | Varianza Explicada | RMSEC (%) | RMSEP (%) | $R^2$ |
|-------------|--------------------|-------------------------|--------------------|-----------|-----------|-------|
| 16          | 3,23               | 6                       | 99.2%              | 0.67      | 1.26      | 0.99  |

Este modelo requirió el uso de 16 variables  $x$  agrupadas en 6 VL que explican el 99.2% de la variación de los datos. Al excluir las muestras 3 y 23 se obtuvieron errores de predicción y de calibración menores a 1.3%, mientras que el coeficiente de correlación fue de 0.99 para la calibración y 0.98 para la validación. El modelo de predicción tiene un rango de calibración de 1.3 a 31.3 % en peso de asfaltenos.

En la figura 37 se pueden observar la ubicación de las 16 variables  $x$ , mostrando a cuál región espectral pertenece cada una de ellas.

**Figura 37. Ubicación de las variables  $x$  en la región MIR para el modelo de asfaltenos**



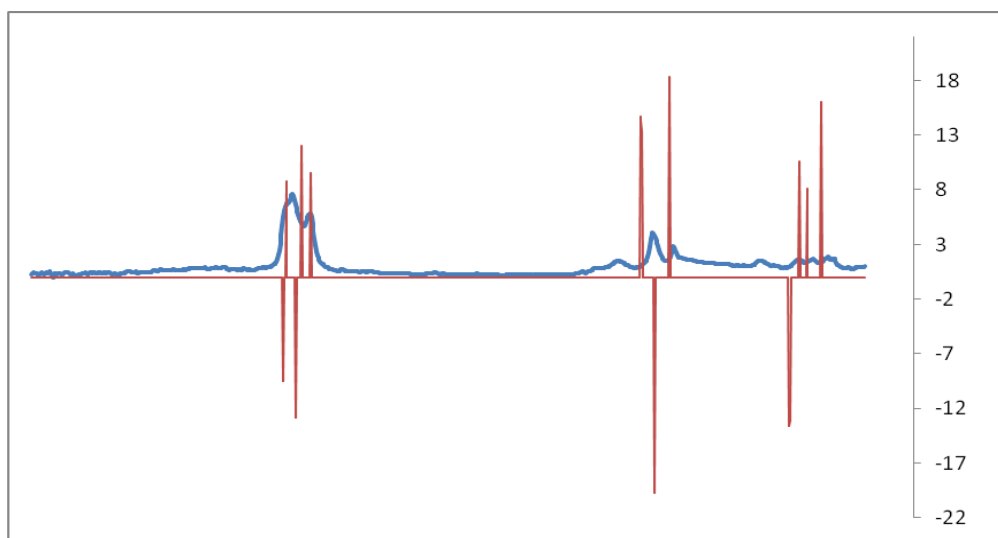
Se puede ver cómo cinco variables explican la región de 2966 a 2854  $\text{cm}^{-1}$  donde ocurren las vibraciones de tensión de los grupos alifáticos. Otras 5 variables explican la región de 1512-1504 a 1454 y 1392  $\text{cm}^{-1}$  que también pertenece a la absorción de los grupos alifáticos. Las 6 variables restantes explican las regiones donde suceden vibraciones por parte de grupos aromáticos (ver tabla 19).

**Tabla 19. Valores de los coeficientes de correlación para cada variable x**

| Variables x<br>(cm <sup>-1</sup> ) | Coefficiente<br>correlación |
|------------------------------------|-----------------------------|
| 2966,671                           | -9,457158                   |
| 2955,097                           | 8,808015                    |
| 2916,519                           | -12,79062                   |
| 2893,372                           | 12,04402                    |
| 2854,794                           | 9,546617                    |
| 1512,269                           | 14,65663                    |
| 1508,411                           | 14,31035                    |
| 1504,554                           | 13,33075                    |
| 1454,402                           | -19,72297                   |
| 1392,677                           | 18,32295                    |
| 906,59                             | -13,61891                   |
| 902,7321                           | -13,12384                   |
| 864,1538                           | 10,58343                    |
| 848,7225                           | 13,75118                    |
| 833,2912                           | 8,116161                    |
| 775,4238                           | 16,07816                    |

En la figura 38 se muestra la gráfica de los coeficientes de regresión obtenidas para 6 Variables Latentes.

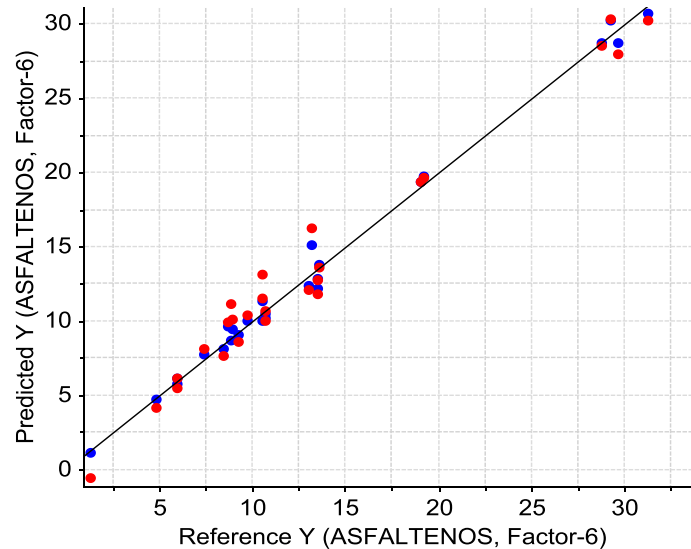
**Figura 38. Ubicación de los Coeficientes de regresión para el modelo de asfaltenos**



De esta manera, se puede ver cómo las intensidades a 2955, 2893, 2854, 1512-1504, 1392, 864, 848, 833 y 775  $\text{cm}^{-1}$  presentan un efecto positivo para asfaltenos, es decir, que a medida que estas variables aumentan, el contenido de asfaltenos aumenta; mientras que las intensidades que se encuentran a 2966, 2916, 1454, 906, 902  $\text{cm}^{-1}$  influyen de manera negativa, es decir cuando estas variables aumentan, disminuye la fracción de asfaltenos (ver tabla 19).

En la figura 39 se muestra la curva predicción obtenida para el modelo.

**Figura 39. Curva de calibración del modelo seleccionado para asfaltenos**



#### **3.3.4.1. Predicción de la composición de asfaltenos para el conjunto de muestras de validación externa.**

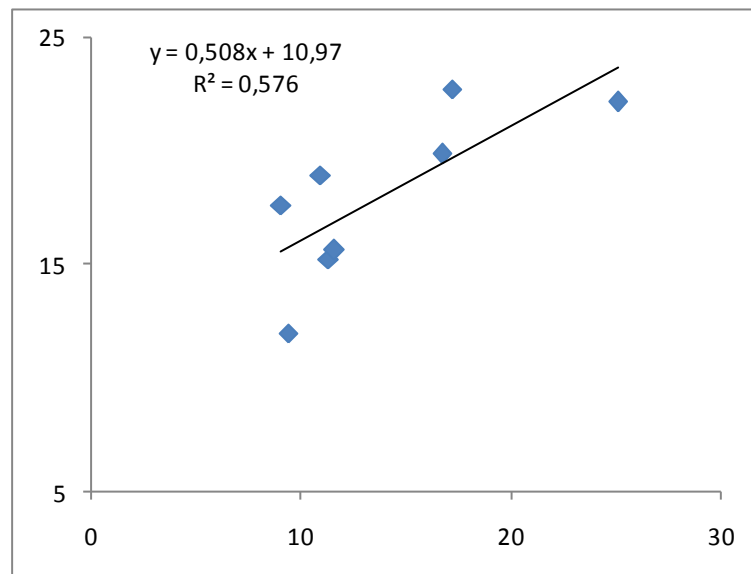
En la tabla 20 se muestra el porcentaje de error para los valores de predichos.

**Tabla 20. Validación del modelo de predicción de %p de Asfaltenos**

| Muestra<br>Nº | Referencia<br>(%p/p) | Predicho<br>(%p/p) | Residual | Error relativo<br>(%) |
|---------------|----------------------|--------------------|----------|-----------------------|
| <b>FV 1</b>   | 16,70                | 19,87              | 3,17     | 19,01                 |
| <b>FV 2</b>   | 25,10                | 22,18              | 2,92     | 11,62                 |
| <b>FV 3</b>   | 11,30                | 15,24              | 3,94     | 34,88                 |
| <b>FV 4</b>   | 17,20                | 22,73              | 5,53     | 32,16                 |
| <b>FV 5</b>   | 9,40                 | 12,02              | 2,62     | 27,82                 |
| <b>FV 6</b>   | 10,95                | 18,96              | 8,01     | 73,19                 |
| <b>FV 7</b>   | 11,60                | 15,64              | 4,04     | 34,87                 |
| <b>FV 8</b>   | 9,05                 | 17,66              | 8,61     | 95,17                 |

En la figura 40 se muestra la gráfica del valor predicho versus el valor de referencia, con un coeficiente de correlación ( $R^2$ ) de 0.58.

**Figura 40. Gráfico del valor predicho vs valor de referencia para el modelo de asfaltenos**



### 3.4. PRUEBAS T PARA MUESTRAS DE VALIDACIÓN.

Cinco fondos de la matriz de validación fueron usados para encontrar parámetros estadísticos como las pruebas t y el intervalo de confianza. Para el fondo de vacío 2 se tomaron diez repeticiones de su espectro, para evaluar la repetibilidad de los modelos (tabla 21). Las cuatro muestras restantes se tomaron por triplicado para evaluar la dependencia de la repetibilidad en función de la composición de las muestras (tabla 22).

**Tabla 21. Parámetros estadísticos usados para definir la calidad de las predicciones**

| saturados                  |                  | aromáticos       |                  | resinas          |                  | asfaltenos       |                  |
|----------------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| valor referencia           | valor predicho   | valor referencia | valor predicho   | valor referencia | valor predicho   | valor referencia | valor predicho   |
|                            | 5,91             |                  | 45,0             |                  | 25,9             |                  | 23,5             |
|                            | 5,73             |                  | 44,1             |                  | 27,8             |                  | 23,2             |
|                            | 6,16             |                  | 43,7             |                  | 29,0             |                  | 22,2             |
|                            | 5,29             |                  | 45,3             |                  | 28,4             |                  | 22,1             |
| 7,30                       | 8,02             | 40,7             | 38,9             | 26,9             | 29,0             | 25,1             | 20,1             |
|                            | 7,04             |                  | 44,5             |                  | 22,3             |                  | 25,6             |
|                            | 5,47             |                  | 40,7             |                  | 28,9             |                  | 22,1             |
|                            | 6,69             |                  | 43,5             |                  | 24,6             |                  | 24,1             |
|                            | 4,31             |                  | 39,0             |                  | 28,0             |                  | 23,8             |
|                            | 6,24             |                  | 44,8             |                  | 29,5             |                  | 19,8             |
| <b>Promedio</b>            | <b>6,09</b>      |                  | <b>42,9</b>      |                  | <b>27,3</b>      |                  | <b>22,6</b>      |
| <b>Error relativo (%)</b>  | <b>16,6</b>      |                  | <b>5,48</b>      |                  | <b>1,63</b>      |                  | <b>9,80</b>      |
| <b>Desviación</b>          | <b>1,02</b>      |                  | <b>2,47</b>      |                  | <b>2,34</b>      |                  | <b>1,77</b>      |
| <b>valor t</b>             | <b>3,58</b>      |                  | <b>2,71</b>      |                  | <b>0,56</b>      |                  | <b>4,17</b>      |
| <b>Intervalo confianza</b> | <b>6,09±1,21</b> |                  | <b>42,9±2,23</b> |                  | <b>27,3±0,44</b> |                  | <b>22,6±2,46</b> |

Según la tabla 21, la probabilidad de que ocurra un determinado valor para saturados está comprendido en un rango de 4,88 - 7,30. Para aromáticos en un rango de 40,7- 45,1. Para resinas en un rango de 26,9 - 27,7 y para asfaltenos de

20,1- 25,1. Estadísticamente los valores caen en el intervalo de confianza. Además para una confianza del 97,5% la t crítica es de 4,3, y según la tabla anterior ningún valor de predicción es mayor. Indicando que el valor de referencia y la media tienen la misma exactitud.

En la tabla 22 se muestran los valores de la prueba t para los fondos 1, 3, 4, y 5.

**Tabla 22. Parámetros estadísticos usados para evaluar la calidad de las predicciones a cuatro fondos de vacío**

|                     | saturados      |      | aromáticos |      | Resinas   |      | asfaltenos |      |
|---------------------|----------------|------|------------|------|-----------|------|------------|------|
| fondo<br>1          | 11,8           | 8,89 | 41,7       | 40,1 | 29,8      | 28,1 | 16,7       | 21,3 |
|                     |                | 10,2 |            | 37,0 |           | 37,0 |            | 16,8 |
|                     |                | 8,44 |            | 45,4 |           | 29,1 |            | 20,0 |
|                     | promedio       | 9,18 |            | 40,8 |           | 31,4 |            | 19,4 |
|                     | Error Relativo | 22,2 |            | 2,12 |           | 5,41 |            | 15,9 |
|                     | Desviación     | 0,92 |            | 4,22 |           | 4,91 |            | 2,29 |
|                     | Valor t        | 4,04 |            | 0,30 |           | 0,46 |            | 1,64 |
| Intervalo confianza | 9,18±2,62      |      | 40,8±0,88  |      | 31,4±1,61 |      | 19,4±2,66  |      |
| fondo<br>3          | 5,4            | 7,31 | 54,5       | 41,5 | 28,8      | 34,8 | 11,3       | 17,7 |
|                     |                | 6,64 |            | 45,8 |           | 51,3 |            | 10,1 |
|                     |                | 9,82 |            | 46,3 |           | 34,9 |            | 15,8 |
|                     | promedio       | 7,92 |            | 44,5 |           | 40,3 |            | 14,5 |
|                     | Error Relativo | 46,7 |            | 18,3 |           | 40,0 |            | 28,5 |
|                     | Desviación     | 1,68 |            | 2,61 |           | 9,48 |            | 3,92 |
|                     | Valor t        | 2,13 |            | 5,42 |           | 1,72 |            | 1,16 |
| Intervalo confianza | 7,92±2,52      |      | 44,5±9,99  |      | 40,3±11,5 |      | 14,5±3,22  |      |
| fondo<br>4          | 14,4           | 12,4 | 43,6       | 36,4 | 24,8      | 26,3 | 17,2       | 15,6 |
|                     |                | 12,0 |            | 33,4 |           | 25,1 |            | 21,2 |

|         |                     |           |      |           |      |           |     |           |
|---------|---------------------|-----------|------|-----------|------|-----------|-----|-----------|
|         |                     | 12,5      |      | 45,3      |      | 25,7      |     | 22,6      |
|         | Promedio            | 12,3      |      | 38,4      |      | 25,7      |     | 19,8      |
|         | Error Relativo      | 14,7      |      | 12,0      |      | 3,61      |     | 15,05     |
|         | Desviación          | 0,27      |      | 6,19      |      | 0,58      |     | 3,69      |
|         | Valor t             | 11,1      |      | 1,19      |      | 2,19      |     | 0,99      |
|         | Intervalo confianza | 12,3±2,12 |      | 38,4±5,23 |      | 25,7±0,89 |     | 19,8±2,59 |
| fondo 5 | 19,9                | 18,3      | 49,7 | 45,5      | 21,0 | 23,7      | 9,4 | 12,3      |
|         |                     | 15,4      |      | 45,1      |      | 23,5      |     | 15,3      |
|         |                     | 17,3      |      | 42,3      |      | 27,0      |     | 7,01      |
|         | promedio            | 17,0      |      | 44,3      |      | 24,7      |     | 11,5      |
|         | Error Relativo      | 14,6      |      | 10,9      |      | 17,7      |     | 22,8      |
|         | Desviación          | 1,49      |      | 1,77      |      | 1,95      |     | 4,20      |
|         | Valor t             | 2,76      |      | 4,31      |      | 2,70      |     | 0,72      |
|         | Intervalo confianza | 17,0±2,90 |      | 44,3±5,40 |      | 24,7±3,72 |     | 11,5±2,14 |

De la tabla 22 se puede ver que la repetibilidad se ve afectada con sólo aromáticos para el fondo 3, saturados para el fondo 4 y aromáticos para el fondo 5, ya que presentan valores t por encima del crítico.

En general se observan mejores resultados cuantas más repeticiones espectrales se tengan.

#### 4. CONCLUSIONES

El desarrollo el presente trabajo permitió clasificar y analizar fondos de vacío colombianos a partir de su espectro PAS-FTIR, mediante el análisis quimiométrico.

El análisis por componentes principales (PCA) permitió reducir la gran cantidad de variables iniciales con el uso de solo 3PC. Con los 3PC se pudo explicar aproximadamente el 100% de la variabilidad de los datos, clasificar las muestras de acuerdo al contenido en la fracción de asfaltenos y determinar las variables espectrales de mayor importancia que ayudaron a la formación de los PC. Estas variables se ubican en las regiones comprendidas entre 3000-2800, 1500-1350 y 950-700  $\text{cm}^{-1}$ .

Las variables  $x$  correspondientes a las regiones donde suceden las vibraciones para los grupos alifáticos mostraron un mayor aporte en la formación de los 3PCs a diferencia de las variables que describen las regiones donde suceden las vibraciones por parte de grupos aromáticos.

La regresión por mínimos cuadrados parciales (PLS-R) aplicada sobre los espectros mostró un desempeño satisfactorio para predecir la composición en peso del análisis SARA. El método propuesto de caracterización por PAS-FTIR proporciona información de la composición química en menos de cinco minutos, con el uso de una mínima cantidad de muestra y sin preparación de la misma. De esta manera la técnica usada es un método alternativo de análisis que permite hacer un estudio de muestras sólidas, complejas y opacas en un menor tiempo que las técnicas tradicionales.

La disminución de variables  $x$  empleada en la regresión PLS mostró una forma segura de trabajar sólo con las variables más importantes, es decir aquellas que más aportan en la formación de las variables latentes.

La espectroscopía Fotoacústica en la región del infrarrojo medio demostró ser muy útil para el estudio estructural de muestras solidas y opacas, proporcionando espectros repetibles con una desviación estándar de 0,002.

Para hacer uso de los modelos predictivos se debe tener en cuenta que sólo pueden ser aplicados a muestras de fondos de vacío, si se aplican a muestras de naturaleza diferente puede llevar a la obtención de resultados erróneos. Además la obtención de los espectros debe hacerse en el espectrofotómetro empleando las mismas condiciones y parámetros instrumentales que aquí se describen.

## 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 (gravidad API, relación C/H, entre otras).

Los resultados del estudio anteriormente presentados muestran el gran potencial de la espectroscopía PAS para la caracterización de fondos de vacío colombianos. Sin embargo es posible obtener mejores resultados en el proceso de calibración y validación de los modelos de predicción si se tienen triplicados de cada muestra.

## 6. REFERENCIAS BIBLIOGRÁFICAS

1. ALTGELT, K.H., BODUSZYNSKI, M. M. Composition and analysis of heavy petroleum fractions. Segunda Edición. New York: Marcel Dekker, 1994. 495p.
2. MORAWSKI, I., MOSIO-MOSIEWSKI, J. Effects of Parameters in Ni–Mo Catalysed Hydrocracking of Vacuum Residue on Composition and Quality of Obtained Products. En: Fuel Processing Technology 87 (2006) 659–669
3. CABRERA, E., FRANCO, F., MONDRAGÓN, F., y FERNÁNDEZ, J., Conversión de fondos de vacío de petróleo a semicoque. Química de recursos Energéticos y Medio Ambiente. Universidad de Antioquia. Medellín, Colombia. 2008.
4. VERSTRAETE, J.J., SCHNONGS., DULOT, H., HUDEBINE, D. Molecular Reconstruction of Heavy Petroleum Residue Fractions. En: Chemical Engineering Science (2009), doi:10.1016/j.ces.2009.08.033
5. MARTINEZ, A., MARTIN, N. Betunes Asfálticos: Materiales muy Utilizados y Poco Conocidos Químicamente. En: Anales de la real sociedad española de química. España, octubre-diciembre (2003). P 43-52.
6. 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: Díaz de Santos, 2004. P 1-15.
7. LACHE, A., MELENDEZ, V. Predicción del análisis SARA de crudos colombianos aplicando espectroscopía FTIR-ATR y métodos quimiométricos. Tesis de grado (Química). Bucaramanga. Universidad industrial de Santander. Facultad de ciencias. Escuela de Química, 2010
8. HUSSEIN, Alboudwarej., FELIX, Joao., TAYLOR Shawn. La Importancia del Petróleo Pesado. En: Oilfield Review; Edmonton, Alberta, Canada, otoño de 2006.

9. KHARRAT, A.,† ZACHARIA, J.,† CHERIAN, J., AND ANYATONWU, A. Issues with Comparing SARA Methodologies. En: Energy & Fuels (2007), 21, 3618–3621
10. AMERICAN SOCIETY FOR TESTING AND MATERIALS. Standard Test Method for Separation of Asphalt into four fractions. ASTM D4124-09.
11. MACHO-APARICIO, Santiago. Metodologías analíticas basadas en espectroscopía de infrarrojo y calibración multivariante. Aplicación a la industria petroquímica. Tarragona. Tesis doctoral (doctor en Química). Univertat Rovira i Virgili. Área de Química analítica. Departament de Química analítica i Química orgànica, 2002.
12. (En línea)  
[http://www.fq.uh.cu/dpto/qf/docencia/pregrado/estruc\\_2/ir/descargas/1\\_ir\\_tecnica.pdf](http://www.fq.uh.cu/dpto/qf/docencia/pregrado/estruc_2/ir/descargas/1_ir_tecnica.pdf) [citado el 6 de septiembre del 2010]
13. AHAMD AL-GHOUTI, M., SALIM AL-DEGS, Y., and AMER, M. Application of chemometrics and FTIR for determination of viscosity index and base number of motor oils. En: Talanta 81 (2010) 1096–110.
14. ABBAS, O., REBUFA, C., DUPUY, N., PERMANYER, A. Assessing petroleum oils biodegradation by chemometric analysis of spectroscopic data. En: Talanta 75 (2008) 857–871.
15. ORREGO, J.A. Estudio de la estructura de cinco carbones colombianos por Espectroscopía Fotoacústica en el infrarrojo medio. Bucaramanga. Tesis de maestría (maestro en Química). Universidad Industrial de Santander. Facultad de ciencias. escuela de Química, 2008.
16. 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.

17. GARCÍA, Roberto. Espectroscopía Fotoacústica: Una Teoría Antigua con nuevas Aplicaciones. Chiapas. Tesis de Maestría. Instituto Tecnológico de Tuxtla Gutiérrez. Departamento de Metal Mecánica.
18. 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. En: Anales AFA-62 vol. 18. (2006): ISSN 1850-1158.
19. ROSENCWAIG, A., GERSHO, A. Theory of the Photoacoustic effect with solids *J. Appl. Phy.* 1976, 47(1), 64.
20. PEREIRA, C., SKROBO, T., CASTRO, R., FORTES, P., and, PASA, M. Determination of Gasoline Adulteration by Principal Components Analysis-Linear Discriminant Analysis Applied to FTIR Spectra. En: Energy & Fuels (2006), 20, 1097-1102.
21. MILLER, J.N., MILLER, J.C. Estadística y Quimiometría para química analítica. Cuarta edición. España: Prentice Hall, 2002. P.595
22. MARTENS, H., MARTENS, M. Multivariate analysis of quality an introduction. Segunda edición: John Wiley & Sons, 2001. P. 445.
23. AMERICAN SOCIETY FOR TESTING AND MATERIALS. Standard practices for infrared multivariate quantitative analysis. Philadelphia: ASTM, 2005. 29h. (ASTM E 1655).
24. MONROY, L. Predicción de la densidad en asfaltenos colombianos utilizando Espectroscopía Fotoacústica en la región del infrarrojo medio y métodos quimiométricos. Bucaramanga. Tesis de grado (Química). Universidad industrial de Santander. Facultad de ciencias. Escuela de Química, 2010
25. STEINER, J; TERMONIA, Y; DELTOUR, J; Anal. Chemometrics. Vol 44. P. 1906, 1972.

26. BUENO Alexis. Caracterización de la planta de Aromáticos de la GCB Mediante Espectroscopía NIR y Métodos Quimiométricos de Análisis. Bucaramanga. Tesis de Maestría. Universidad industrial de Santander. Facultad de ciencias. Escuela de Química, 2008.
27. COELHO R., HOVELL I., de MELLO M., MIDDEA A., and de SOUZA A., Characterisation of aliphatic chains in Vacuum Residues (VRs) of asphaltenes and Resins Using Molecular Modelling and FTIR Techniques. En: Fuel Processing Technology 87 (2006) 325-333.
28. Comparing Venezuelan and Canadian Heavy Oil and Tar Sands. (2001-061: Calgary, Canada). Petroleum Society's Canadian International Petroleum Conference 2001, Calgary, Alberta, Canada.
29. FOSTER N., AMONETT J., AUTREY T., HO J., Detection of trace levels of water in oil by photoacoustic spectroscopy. En: sensors and Actuators B 77 (2001) 620-624
30. AMERICAN SOCIETY FOR TESTING AND MATERIALS. Standard terminology relating to molecular spectroscopy. Philadelphia: ASTM, 2005. 10 h. (ASTM E 131).
31. CURTIS C., DECOSTER E., GUZMAN-GARCIA A., HUGGINS C., Yacimientos de petróleo pesado. En: Oilfield Review 2002-2003. P 32-55
32. THE UNSCRAMBLER X METHODS. Software para diseño de experimentos y análisis multivariado. (En línea). <http://www.camo.com> [citado el 22 de septiembre de 2010].
33. MULLINS, Oliver C. The Modified Yen Model. En: Energy and Fuels, 2010, 24, 2179-2207.
34. BALLESTEROS, S. Predicción de las Fracciones SARA de Fondos de vacío de Crudos Colombianos, por Métodos Quimiométricos Utilizando

Espectroscopía de Fluorescencia Inducida por Láser (LIF). Bucaramanga. Tesis de grado (Química). Universidad industrial de Santander. Facultad de ciencias. Escuela de Química, 2010.

35. SPEIGHT, J.G. The chemistiy and thechnology of petroleum. Cuarta edición. USA: Taylor and Francis Group, 2007. ISBN 10: 0-8493-9067-2