

**PREDICCIÓN DEL NÚMERO DE BROMO DE NAFTAS COLOMBIANAS  
UTILIZANDO ESPECTROSCOPIA INFRARROJO EN LA REGIÓN CERCANA Y  
METODOS QUIMIOMÉTRICOS**

**CARLOS ALBERTO APARICIO RAMÍREZ**

**UNIVERSIDAD INDUSTRIAL DE SANTANER  
FACULTAD DE CIENCIAS  
ESCUELA DE QUÍMICA  
BUCARAMANGA**

**2013**

**PREDICCIÓN DEL NÚMERO DE BROMO DE NAFTAS COLOMBIANAS  
UTILIZANDO ESPECTROSCOPIA INFRARROJO EN LA REGIÓN CERCANA Y  
METODOS QUIMIOMÉTRICOS**

**CARLOS ALBERTO APARICIO RAMÍREZ**

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

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

**UNIVERSIDAD INDUSTRIAL DE SANTANER  
FACULTAD DE CIENCIAS  
ESCUELA DE QUÍMICA  
BUCARAMANGA**

**2013**

## DEDICATORIA

A mi familia, principalmente a mi mamá, por estar conmigo incondicionalmente, y apoyarme en todo momento.

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

## **AGRADECIMIENTOS**

Al Profesor Enrique Mejía Ospino por brindarme la oportunidad de trabajar en el Laboratorio de Espectroscopia Atómica y molecular (LEAM) y por su apoyo incondicional en la realización de este trabajo.

A mi familia por apoyarme durante mi carrera. A mi mamá principalmente.

A todos mis amigos y compañeros, quienes de una forma u otra me ayudaron a culminar este objetivo. Gracias por su compañía y ayuda en todo momento.

## TABLA DE CONTENIDO

	<b>pág</b>
INTRODUCCIÓN.....	13
<b>1. MARCO TEÓRICO .....</b>	<b>15</b>
1.1 EL PETRÓLEO .....	15
1.1.1 Generalidades del petróleo .....	15
1.1.2 Clasificación del petróleo.....	15
1.2 NAFTA.....	17
1.2.1. Generalidades.....	17
1.2.2. Naturaleza química .....	17
1.3. NÚMERO DE BROMO.....	18
1.4. ESPECTROSCOPIA INFRARROJA CERCANA (NIR) .....	20
1.4.1. Desarrollo histórico <sup>[12]</sup> .....	20
1.4.2. Teoría de absorción .....	21
1.4.3 Análisis espectral.....	24
1.5. QUIMIOMETRÍA .....	25
1.5.1. Calibración multivariable .....	27
1.5.2. Clasificación de los métodos de calibración multivariable .....	27
1.5.3. Construcción de modelos de calibración multivariable .....	28
1.5.3.1. Análisis por componentes principales (PCA) .....	31
1.5.3.2 Regresión por mínimos cuadrados parciales (PLS) .....	32
1.5.3.3. Regresión Polinomial Múltiple (RPM).....	34
1.6. ANTECEDENTES.....	35
<b>2. METODOLOGÍA.....</b>	<b>38</b>
2.1. SELECCIÓN DE MUESTRAS .....	38
2.2. PROCESAMIENTO DE DATOS.....	39
2.2.1. Pretratamiento de datos .....	39

2.2.2. Análisis de componentes principales (PCA) .....	39
2.2.3. Regresión por mínimos cuadrados parciales (PLS) .....	39
2.2.4. Regresión Polinomial Múltiple (RPM) .....	40
<b>3. RESULTADOS Y ANÁLISIS .....</b>	<b>41</b>
3.1. ANÁLISIS ESPECTRAL DE LAS NAFTAS .....	41
3.2. ANÁLISIS POR COMPONENTES PRINCIPALES (PCA) .....	43
3.2.1. Análisis por componentes principales para el grupo de muestras pesadas .....	48
3.2.2. Análisis por componentes principales para el grupo de muestras livianas .....	51
3.3. RESULTADOS DEL MÉTODO PLS .....	53
3.3.1. Calibración y validación del modelo PLS para naftas pesadas .....	53
3.3.2. Calibración y validación del modelo PLS para naftas livianas .....	57
3.4. RESULTADOS DEL MÉTODO RPM .....	59
<b>4. CONCLUSIONES .....</b>	<b>63</b>
<b>5. RECOMENDACIONES .....</b>	<b>65</b>
<b>6. REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>66</b>
Anexo A .....	71
Anexo B .....	72

## LISTA DE TABLAS

Tabla 1. <i>Clasificación del crudo según la gravedad API</i> <sup>[6]</sup> .....	16
Tabla 2. Isómeros en función del número de átomos de carbono .....	18
Tabla 3. Regiones de absorción en el infrarrojo .....	21
Tabla 4. Frecuencias de absorción de algunos grupos funcionales en la región NIR <sup>[17]</sup> .....	25
Tabla 5. Análisis del número de bromo de las 39 muestras .....	38
Tabla 6. Varianza de los componentes principales generados por las 39 muestras .....	43
Tabla 7. Separación según el número de bromo .....	45
Tabla 8. Varianza de los componentes principales generados por las muestras pesadas .....	49
Tabla 9. Varianza de los componentes principales generados por las muestras livianas .....	51
Tabla 10. Parámetros estadísticos del modelo seleccionado para la predicción de naftas pesadas .....	53
Tabla 11. Validación del modelo de predicción del número de naftas pesadas .....	56
Tabla 12. Parámetros estadísticos del modelo seleccionado para la predicción de naftas livianas .....	57
Tabla 13. Validación del modelo de predicción de naftas livianas .....	59
Tabla 14. Estructura de los polinomios empleados en el método de RPM .....	60
Tabla 15. Resultados del método de RPM .....	60
Tabla 16. Validación del modelo de predicción con RPM .....	62

## LISTA DE FIGURAS

Figura 1. Energía potencial según el modelo del oscilador (a) armónico, (b) anharmónico <sup>[14]</sup> .....	22
Figura 2. Bandas de absorción observadas en un espectro NIR de una nafta <sup>[15]</sup> .....	23
Figura 3. Representación geométrica de un PCA <sup>[22]</sup> .....	32
Figura 4. gráfica del modelo de regresión PLS <sup>[21]</sup> .....	33
Figura 5. de espectros NIR de naftas con hidrocarburos puros <sup>[33]</sup> .....	42
Figura 6. Gráfico de puntuaciones para las 39 muestras: descomposición de la información espectral en el PC2 vs PC1 .....	44
Figura 7. Gráfico de pesos de las 39 muestras: importancia de las diferentes variables para el PC1 y el PC2 .....	46
Figura 8. Gráfico de biplot: combinación de la gráfica de puntuaciones con la gráfica de pesos.....	47
Figura 9. Gráfico de puntuaciones con el límite de Hotelling $T^2$ para las 39 muestras .....	48
Figura 10. Gráfico de puntuaciones con el límite de Hotelling $T^2$ para las muestras pesadas .....	50
Figura 11. Gráfico de influencia para las muestras pesadas .....	50
Figura 12. Gráfico de puntuaciones con el límite de Hotelling $T^2$ para las muestras livianas .....	52
Figura 13. Gráfico de influencia para las muestras livianas .....	52
Figura 14. Varianza explicada en el modelo PLS para las naftas pesadas .....	54
Figura 15. Gráfica de los errores RMSEC y RMSECV calculados en función de los PCs usados para el modelo PLS para las naftas pesadas .....	55
Figura 16. Coeficientes de regresión para el primer componente principal del modelo PLS para las naftas pesadas .....	56
Figura 17. Curvas de calibración y validación del modelo de predicción para las naftas pesadas .....	57
Figura 18. Descripción gráfica del modelo con 9 componentes principales para la predicción de naftas livianas.....	58
Figura 19. Curvas de calibración y validación del modelo de predicción del número de bromo empleando el método de RPM.....	61

## RESUMEN

**TÍTULO: PREDICCIÓN DEL NÚMERO DE BROMO DE NAFTAS COLOMBIANAS UTILIZANDO ESPECTROSCOPIA INFRARROJO EN LA REGIÓN CERCANA Y METODOS QUIMIOMÉTRICOS\***

**AUTOR: CARLOS ALBERTO APARICIO RAMÍREZ\*\***

**PALABRAS CLAVES:** FT-NIR, NAFTAS, NÚMERO DE BROMO, PCA, PLS, RPM

**CONTENIDO:** En el presente trabajo se utilizaron 39 muestras de naftas proporcionadas por el Instituto Colombiano del Petróleo. Todas las muestras fueron caracterizadas por espectroscopia NIR y su señal espectral fue correlacionada, mediante el análisis por componentes principales (PCA), regresión por mínimos cuadrados parciales (PLS) y regresión polinomial múltiple (RPM), con la información del número de bromo suministrado por el ICP.

El análisis por PCA permitió clasificar las muestras en muestras pesadas y livianas 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 4000-4300, 4500-4700 y 5800-6000  $\text{cm}^{-1}$ . La regresión PLS y la RPM aplicadas sobre los espectros NIR, mostraron un desempeño satisfactorio para predecir el número de bromo. Se comprobó que la combinación de la espectroscopia FT-NIR junto con el análisis quimiométrico puede ser un método alternativo de análisis, ya que requiere menos de cinco minutos de análisis y un menor costo del mismo en comparación con los métodos tradicionales, los cuales requieren uso de solventes y mayor tiempo de análisis.

La validación mostró resultados satisfactorios para la predicción del número de bromo de las naftas. Para cada modelo se obtuvieron errores medios de validación cruzada (RMSECV) menores al 5%. En todos los casos el coeficiente de correlación entre los valores de referencia y predichos por los modelos fue superior a 0.95.

---

\* Trabajo de grado

\*\* Facultad de Ciencias: Escuela de Química. Director: Enrique Mejía Ospino.

## SUMMARY

**TITLE: PREDICTION BROMINE NUMBER OF THE COLOMBIANS NAPHTHA USING NEAR-  
INFRARED SPECTROSCOPY AND CHEMOMETRIC METHODS. \***

**AUTHOR: CARLOS ALBERTO APARICIO RAMÍREZ \*\***

**KEYWORDS: FT-NIR, NAPHTHA, BROMINE NUMBER, PCA, PLS, RPM**

**CONTENT:** In this work were used 39 samples of Colombian naphtha provided by the Colombian Petroleum Institute (ICP). All samples were characterized by spectroscopy near infrared region (NIR) and their spectra were correlated by similarity using principal component analysis (PCA), partial least squares regression (PLS) and regression polynomial multiple (RPM) with bromine number information supplied by the ICP.

PCA analysis allowed classifying the samples in light and heavy samples and determining the most important spectral variables which helped to form the principal components. These variables are located in the regions between 4000-4300, 4500-4700 and 5800-6000  $\text{cm}^{-1}$ . PLS regression and RPM applied on the NIR spectra, showed satisfactory performance to predict the number of bromine. Combination of the FT-NIR spectroscopy and chemometric analysis can be an alternative method of analysis, which requires less than five minutes of analysis and a lower cost compared to traditional methods which require the use solvent and longer analysis.

The validation showed satisfactory results for predicting the bromine number of the Colombian naphtha. For each model were obtained the Root-Mean-Square Error of Cross-Validation (RMSECV) less than 5%. In all cases the correlation coefficient between the values of reference and predicted by the models was higher than 0.95.

---

\* Work Degree

\* \*Science Faculty. Chemistry School. Directress: Enrique Mejía Ospino.

## INTRODUCCIÓN

En la industria petroquímica existe un gran interés en el desarrollo de metodologías analíticas que proporcionen información de calidad y que además lo hagan en el menor tiempo posible, debido a que en esta industria es habitual el monitoreo y control de los procesos utilizando la información que proporcionan los análisis de laboratorio.

Entre las técnicas espectroscópicas empleadas en la industria, el uso de la espectroscopia infrarroja en la región cercana (FT-NIR) es una alternativa prometedora para reemplazar los métodos tradicionales debido a que la muestra no requiere tratamiento, la toma de espectros tarda unos pocos minutos y presenta una repetibilidad aceptable en muestras líquidas<sup>[1]</sup>. Estos métodos tradicionales por lo general presentan, elevados tiempos de respuesta y no suelen suministrar información detallada sobre su composición química, lo cual hace que estos métodos sean poco prácticos para emplearlos como herramientas en el control de procesos, por esta razón se hace muy importante contar con métodos alternos de caracterización, para poder superar estas debilidades.

La aplicación de técnicas multivariadas a la espectroscopia FT-NIR ha permitido la interpretación cuantitativa de las señales espectrales, prediciendo y determinando parámetros y propiedades fisicoquímicas de gran interés en las fracciones del petróleo.

Este trabajo busca predecir el número de bromo en naftas, las cuales se caracterizan por poseer altos números de octano. El número de bromo de las naftas es muy importante, ya que cuando este número es alto, el porcentaje de olefinas es elevado, lo cual se traduce en números de octano por encima del 70% para la gasolina que se va a obtener a partir de estas naftas<sup>[2]</sup>.

La técnica FT-NIR junto con la aplicación de técnicas quimiométricas busca jugar un papel esencial para el análisis de productos derivados del petróleo, específicamente de la fracción de naftas, buscando desarrollar nuevas metodologías, que permitan operar en condiciones menos rigurosas que los procedimientos estándar con un menor costo de análisis.

El desarrollo del presente trabajo permitió clasificar y analizar las naftas colombianas a partir de su espectro FT-NIR, mediante el uso de técnicas quimiométricas de análisis, en este trabajo se escogen dos, que son de gran importancia: La regresión por mínimos cuadrados parciales (PLS-R) y la regresión polinomial múltiple (RPM), tanto el método lineal como el no-lineal mostraron buen desempeño para la predicción del número de Bromo.

# 1. MARCO TEÓRICO

## 1.1 EL PETRÓLEO

### 1.1.1 Generalidades del petróleo

El petróleo o también conocido como crudo es uno de los combustibles de mayor aplicación en el mundo, llegando a convertirse en la base económica de diversos sectores como la industria y el transporte, los cuales dependen de los productos que de él se derivan. El crudo es una mezcla compleja de cientos de hidrocarburos, los cuales pueden tener desde 1 hasta 60 átomos de carbono, junto con los compuestos derivados de éstos, presente en pequeñas cantidades, que pueden contener en su estructura, nitrógeno, oxígeno y metales como níquel, vanadio, hierro y cobre. La apariencia y composición de un crudo varía ampliamente, aunque se considera que un crudo promedio contiene aproximadamente de 84 a 87% de carbono, 11 a 14 % de hidrógeno, entre 1 a 3 % de azufre y menos del 1% de nitrógeno, oxígeno y metales. Estas diferencias en composición influyen de manera apreciable en los aspectos de diseño y localización de las plantas de refinación, en la determinación de los procesos de conversión, tratamiento requeridos y en la producción de los derivados de mayor demanda e importancia económica. <sup>[4,5]</sup>

### 1.1.2 Clasificación del petróleo

Los crudos tienen características físicas y químicas muy variables de un campo de producción a otro e incluso dentro de un mismo yacimiento. La clasificación más sencilla, pero no menos importante en cuanto a los resultados económicos, es la clasificación en crudos pesados y ligeros. Al estar formado principalmente por moléculas hidrocarbonadas, la densidad de un crudo será tanto menor cuanto mayor sea la relación atómica H/C. la densidad de los crudos puede oscilar entre 0,7 y 1, expresándose con mucha frecuencia en grados API (American Petroleum

Institute) cuyo valor varía entre 70 y 5; esta variabilidad de la densidad es consecuencia de composiciones en familias químicas muy diferentes.

Los crudos están constituidos por mezclas de un número muy elevado de componentes puros, aumentando la dificultad de la descripción de las distintas fracciones. En términos de la densidad API, los crudos ligeros o de baja gravedad específica presentan alto dicho valor. Los crudos con bajo contenido de carbono, alto contenido de hidrogeno y alta densidad API son generalmente ricos en hidrocarburos parafínicos y tienden a producir mayores cantidades de gasolina y productos ligeros; aquellos crudos con alto contenido de carbono, bajo contenido de hidrógeno y baja densidad API son ricos en hidrocarburos nafténicos y aromáticos. De acuerdo al contenido de compuestos con azufre, el crudo puede clasificarse como agrio, si presenta cantidades apreciables de estos compuestos o como dulce, si presenta cantidades muy pequeñas. <sup>[6]</sup>

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

**Tabla 1. Clasificación del crudo según la gravedad API <sup>[6]</sup>**

<b>ACEITE CRUDO</b>	<b>DENSIDAD</b>	<b>GRAVEDAD API</b>
<b>Extrapesado</b>	>1.0	10
<b>Pesado</b>	1.0-0.92	10.0 -22.3
<b>Mediano</b>	0.92-0.87	22.3 - 31.1
<b>Ligero</b>	0.87-0.83	31.1 – 39
<b>Superligero</b>	<0.83	> 39

## 1.2 NAFTA

### 1.2.1. Generalidades

La nafta es un producto líquido derivado del petróleo que, generalmente, constituye entre 15 a 30% en peso de éste. Considerada como una mezcla compleja de estructuras hidrocarbonadas con presencia de ciertos heteroátomos, su composición química depende de la naturaleza del crudo de origen, del rango de ebullición de la nafta y del proceso de obtención empleado. La nafta obtenida por destilación atmosférica del crudo en el rango de 30°C a 200°C se denomina nafta virgen y está conformada por hidrocarburos desde 5 hasta 12 átomos de carbono <sup>[7]</sup>. Éste producto usualmente se emplea en la generación de componentes de alto octanaje y compuestos aromáticos de alta pureza. La nafta también puede ser extraída de residuos o fracciones más pesadas del crudo por procesos de craqueo catalítico y viscorreducción que generan un producto con altos contenidos de olefinas.

### 1.2.2. Naturaleza química

La nafta está constituida por un gran número de compuestos (200 en promedio) que presentan en su estructura principalmente carbono, hidrógeno, nitrógeno y azufre. También es posible encontrar oxígeno y elementos metálicos dependiendo de la naturaleza de la nafta, aunque en baja proporción. Normalmente la composición química de la nafta se da en función de la distribución de hidrocarburos saturados lineales (parafinas), hidrocarburos saturados ramificados (isoparafinas), hidrocarburos saturados cíclicos (naftenos), hidrocarburos insaturados (olefinas) e hidrocarburos aromáticos<sup>[8]</sup>. Típicamente una nafta virgen contiene de 40 a 70% de parafinas, 20 a 50% de naftenos, 5 a 20% de aromáticos y menos del 2% de olefinas.

La complejidad en composición química es mayor ya que el número de posibles isómeros incrementa exponencialmente con el número de átomos de carbono <sup>[9]</sup> (Tabla 2).

**Tabla 2. Isómeros en función del número de átomos de carbono**

Átomos de Carbono	Número de Isómeros
1	1
2	1
3	1
4	2
5	3
6	5
7	9
8	18
9	35
10	75
15	4,347
20	366,319
25	36,797,588
30	4,111,846,763
40	62,491,178,805,831

### 1.3. NÚMERO DE BROMO

Con el número de bromo (NB) se determina el porcentaje de compuestos no saturados (mayormente enlaces dobles C-C) en derivados del petróleo. Para ello se separa el enlace doble mediante la adición de bromo.

El grado de insaturación de los compuestos orgánicos influye en distintas especificaciones de los compuestos, tales como la viscosidad, el punto de ebullición o las características de polimerización. Define la tendencia a formar gomas nocivas que pueden obstruir, por ejemplo, puertos y filtros de combustible. El bromo formará dobles enlaces bajo ciertas condiciones. Cuantos más dobles enlaces tenga una molécula, más bromo se consumirá en esta reacción. El número de bromo es la cantidad de bromo expresada en gramos absorbidos por 100 gramos de una muestra. El número indica el grado de insaturación.

El método se utiliza para los siguientes productos:

- Destilados con un punto de ebullición por debajo de 327 °C (620 °F) y un porcentaje de volumen del 90%, que son más livianos que 2-metilpropano (entre ellos, los combustibles con y sin aditivos de plomo, queroseno y gasóleos).
- Olefinas comerciales (mezclas de monoolefinas alifáticas) con un número de bromo de 95-165.
- Propenos (trímeros y tetrámeros), trímeros de buteno, mezclas de heptenos, octenos y nonenos<sup>[10]</sup>.

*Determinación del número de bromo de destilados del petróleo y olefinas alifáticas comerciales por titulación electrométrica.*

Este método se utiliza para la determinación del número de bromo de ciertos productos del petróleo y se realiza basado en la norma ASTM D 1159<sup>[11]</sup>. La magnitud del número de bromo es indicativa de la cantidad de constituyentes que reaccionan con el bromo, y no una identificación de los constituyentes. El número de bromo principalmente indica la concentración de dobles enlaces presentes en el producto.

Para la determinación del número de bromo se titula un peso conocido de la muestra, disuelto en un solvente especificado y manteniendo a una temperatura de 0 a 5°C con una solución patrón de bromuro-bromato. El punto final se determina con un equipo de titulación electrométrica de punto muerto, cuando la presencia de bromo libre produce un cambio repentino en la conductividad eléctrica del sistema.

## **1.4. ESPECTROSCOPIA INFRARROJA CERCANA (NIR)**

### **1.4.1. Desarrollo histórico <sup>[12]</sup>**

La existencia de radiación en la región del infrarrojo cercano (NIR) fue desarrollada hacia el año de 1800 por Sir Fredrick William Herschel, consagrado músico y astrónomo aficionado, cuando trataba de encontrar la contribución de cada una de las frecuencias que componen la luz blanca en el incremento de la temperatura de una sustancia expuesta a ella. Herschel observó experimentalmente que, al dispersar un haz de luz blanca por un prisma, el efecto de calentamiento relativo era mayor al desplazarse del color azul al rojo. Sorpresivamente encontró que la región más allá del color rojo, invisible para el ojo humano, presentaba un mayor efecto sobre el incremento de la temperatura de una sustancia que las demás regiones del haz de luz dispersado. Sin saberlo, Herschel había descubierto la radiación NIR a la cual se refirió, inicialmente, como la región de rayos caloríficos.

Aunque la existencia de la radiación NIR se demostró muchos años antes que la radiación en la región del infrarrojo medio (M-IR), fue esta última región la que recibió el mayor interés para su desarrollo teórico e instrumental por los primeros espectroscopistas gracias a la utilidad mostrada en la identificación de grupos funcionales en sustancias orgánicas. Las bandas anchas, débiles y superpuestas hacían creer a los espectroscopistas de la época que la región NIR no contenía ningún tipo de información valiosa para el análisis químico.

Sin embargo, hacia el año de 1938 se publican las primeras aplicaciones de la espectroscopia NIR mediante la identificación de una simple longitud de onda que obedecería la ley de Beer para el desarrollo de aplicaciones cuantitativas en diversas áreas. Tal vez uno de los primeros aportes, y más valiosos en esta área, fue el realizado en el año de 1965 por Karl Norris en el desarrollo de un método para determinar la humedad en productos agrícolas. Su trabajo describe, por primera vez, la generación de un modelo con dos longitudes de onda para

correlacionar, al mismo tiempo, los datos espectrales y la concentración del analito, dejando de lado la aplicación de la ley de Beer como un requisito para el análisis cuantitativo. A partir de este momento, el número de aplicaciones desarrolladas por espectroscopia NIR crecieron exponencialmente.

#### 1.4.2. Teoría de absorción

La región infrarroja cubre las radiaciones en el intervalo de 780 nm a 1.000  $\mu\text{m}$  (o en unidades de número de onda de  $12.800\text{ cm}^{-1}$  a  $10\text{ cm}^{-1}$ ), la cual se divide en tres áreas más pequeñas identificables debido a su interacción con la materia (tabla 3).

**Tabla 3. Regiones de absorción en el infrarrojo**

Región Infrarroja	Longitud de onda	Número de onda ( $\text{cm}^{-1}$ )	Transición característica
Cercana (NIR)	780 a 2500 nm	12800 a 4000	Sobretonos y Combinaciones
Medio (MIR)	2.5 a 50 $\mu\text{m}$	4000 a 200	Vibraciones fundamentales
Lejano (FIR)	50 a 1000 $\mu\text{m}$	200 a 10	Rotaciones

La espectroscopia NIR, es una clase de espectroscopia vibracional que emplea fotones de energía en el rango de  $2.65 \times 10^{-19}$  a  $7.96 \times 10^{-20}$  Julios (J) correspondientes al rango de frecuencias mostrado en la tabla 3. Esta energía es mayor que la necesaria para promover las moléculas a su primer estado vibracional excitado y menor que la requerida para llevar a cabo una excitación electrónica. Solo las frecuencias que puedan suplir la diferencia energética entre dos estados vibracionales en una molécula serán completamente absorbidas, mientras otras frecuencias serán parcialmente o no serán absorbidas<sup>[13]</sup>.

Matemáticamente la energía de los estados vibracionales en una molécula diatómica se puede describir por el modelo del oscilador armónico (figura 1a) mediante la ecuación 1.

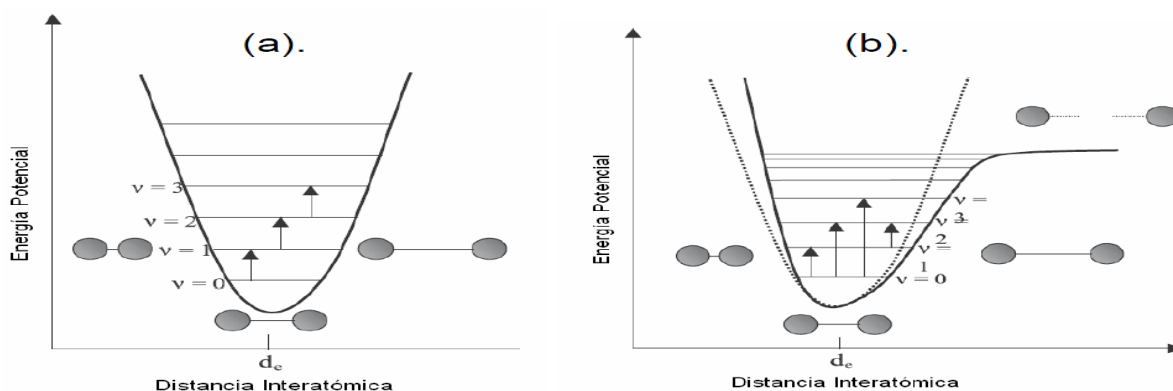
$$E_{vib} = \left[ v + \frac{1}{2} \right] \frac{h}{2\pi} \sqrt{\frac{K}{\mu}} \quad \text{Ecuación (1)}$$

Donde “ $\nu$ ” es el número cuántico vibracional, “ $h$ ” es la constante de Planck, “ $k$ ” es la constante de fuerza y  $\mu$  es la masa reducida del sistema. Este modelo permite únicamente las transiciones entre niveles energéticos adyacentes ( $\Delta\nu = \pm 1$ ) y asume que los niveles energéticos se encuentran igualmente espaciados siendo su diferencia en energía siempre la misma. Esta diferencia energética puede ser calculada por la ecuación 2.

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

Donde “ $\nu$ ” es la frecuencia de vibración fundamental del enlace que produce una banda de absorción en la región del infrarrojo medio.

**Figura 1. Energía potencial según el modelo del oscilador (a) armónico, (b) anarmónico [14]**



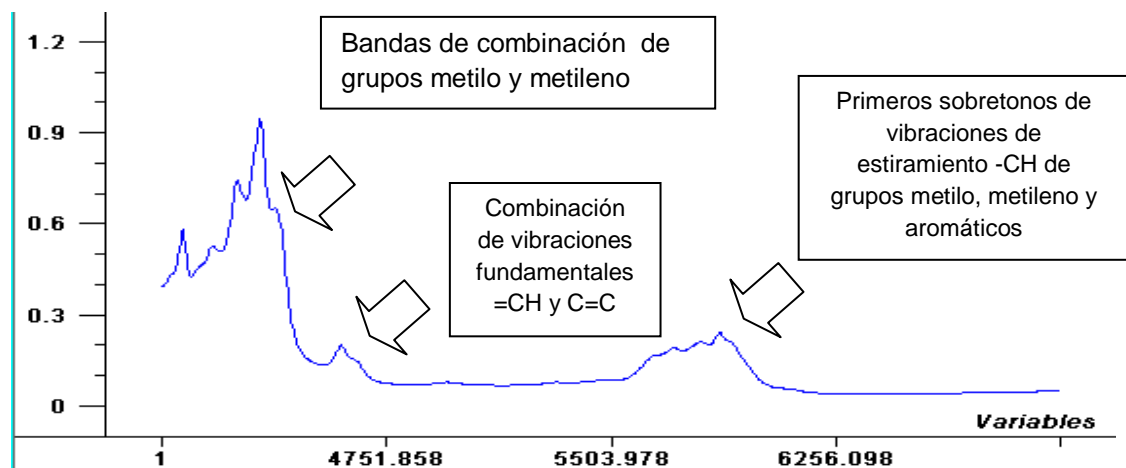
Aunque el modelo del oscilador armónico es una buena aproximación, no explica muchos aspectos sobre el comportamiento molecular. El modelo del oscilador anarmónico representa de manera más aproximada estos aspectos y asume que los niveles energéticos no se encuentran igualmente espaciados (figura 1b). De esta manera la diferencia energética disminuye a medida que aumenta  $\nu$ , y se puede calcular por medio de la ecuación 3.

$$\Delta E_{\nu ib} = h\nu[1 - (2\nu + \Delta\nu + 1)\gamma] \quad \text{Ecuación (3)}$$

Donde “ $y$ ” es el factor de anharmonicidad. Esta anharmonicidad permite las transiciones entre estados de energía vibracional no consecutivos ( $\Delta\nu = \pm 2, \pm 3, \dots$ ), generando las bandas de absorción conocidas como sobretonos las cuales son, aproximadamente, múltiplos de las frecuencias fundamentales de vibración aunque su intensidad es menor. La intensidad de las bandas para el primer sobretono puede ser, dependiendo del enlace, diez a cien veces menor que la frecuencia fundamental. Aunque teóricamente son posibles transiciones entre cualquier par de niveles energéticos, experimentalmente sólo se observan las bandas de absorción correspondientes a las frecuencias de vibración fundamental en el MIR ( $\nu$ ) y a los dos primeros sobretonos en el NIR ( $2\nu$ ,  $3\nu$ ).

Cuando se tienen moléculas poliatómicas, dos o más modos de vibración pueden interactuar para causar cambios simultáneos en energía y generar las llamadas bandas de combinación, cuyas frecuencias corresponden a la suma de múltiplos de cada frecuencia que interactúa ( $\nu_{comb} = n_1\nu_1 \pm n_2\nu_2 \pm \dots$ ). Estas últimas bandas y los sobretonos son las señales típicamente observadas en la región del infrarrojo cercano (figura 2).

**Figura 2. Bandas de absorción observadas en un espectro NIR de una nafta<sup>[15]</sup>**



Otros fenómenos tales como la resonancia Fermi (interacción entre una banda fundamental y un sobretono con diferencia en energía muy baja) y la resonancia Darling-Denninson (interacción entre dos sobretonos de orden superior y una banda de combinación) pueden generar bandas en la región cercana <sup>[16]</sup>

### **1.4.3 Análisis espectral**

La presencia de bandas de combinación y sobretonos en la región NIR limita su aplicación para la identificación cualitativa de grupos funcionales inorgánicos y orgánicos, en contraste con las señales en la región M-IR donde las bandas de absorción pueden ser fácilmente atribuidas a estructuras específicas. Pese a lo anterior, se han desarrollado tablas de asignación de bandas de absorción a grupos funcionales en la región NIR que pueden ser empleadas, como una primera aproximación, para el análisis cualitativo en muestras orgánicas (tabla 4).

La información presente en un espectro NIR puede ser empleada para estimar la concentración de un componente o para estimar una propiedad física cuando estas reflejan cambios significativos en las características espectrales generadas por la muestra. Para lograr esto, es necesario recurrir a diversos métodos multivariados de análisis que se encuentran agrupados en una nueva rama de la química analítica, la Quimiometría, encargada de usar técnicas matemáticas y estadísticas para extraer información relevante de datos analíticos, en este caso de la información espectral obtenida en la región del infrarrojo cercano.

**Tabla 4. Frecuencias de absorción de algunos grupos funcionales en la región NIR <sup>[17]</sup>**

Grupo Funcional	Banda cm <sup>-1</sup>	Comentario
Acetilenos	9800-9430 6580-6400	Sobretono =C estiramiento
Alcoholes	7140-7010	Sobretono OH estiramiento
Aldehídos Alifáticos	4640-4520 8000	combinación C=O y CH estiramiento
Aromáticos	4525 4445	
Formaldehido	4775-4630	
Alcanos	9000-8350	
-CH3-	5850-5660 4510-4280	Bandas muy débiles
-CH2-	9170-8475 5830-6640 4420-4070	
=CH-	8550-8130 7000-6800	
Ciclopropano	5650-5560 6160-6060 4500-4400	
Alquenos C=C	6850-6370 7580-7300	
C=CH2 CH=CH2	6140-5980 4760-4700	

## 1.5. QUIMIOMETRÍA

El análisis de muestras que contienen mezclas de compuestos químicos constituye uno de los retos más importantes de la química analítica. La dificultad aumenta cuando se pretende determinar simultáneamente estos compuestos a

bajas concentraciones o en presencia de otras sustancias que tienen estructuras y propiedades similares, las cuales pueden interferir en su determinación. El análisis de muestras complejas requiere utilizar un conjunto de muestras de referencia que tenga las mismas características que las muestras a analizar y de algún método instrumental que mida alguna propiedad característica de los analitos. En la utilización de un sistema de detección para las muestras a analizar nos podemos encontrar con problemas debido a interferencias, a efectos de matriz, ruido de fondo, etc., que precisan una especial atención.

El método tradicionalmente utilizado para resolver la falta de selectividad en muestras complejas, consiste en la separación previa de los compuestos químicos puros antes de analizarlos. Si la separación ha sido completa (ej. Mediante cromatografía) tan solo es necesaria una medida univariable para tener la información cuantitativa deseada. No obstante, ésta presenta un costo elevado, ya que se necesita disponer de una instrumentación analítica sofisticada e invertir bastante tiempo para optimizar las condiciones de separación. Como alternativa, se han desarrollado los métodos quimiométricos, los cuales permiten extraer la información requerida sobre los componentes de interés de las muestras a partir del tratamiento de datos, sin necesidad de su separación previa.

La quimiometría es una disciplina ampliamente utilizada en el área de la química, nacida aproximadamente al final de los años 70, y que, actualmente, está siendo ampliamente utilizada en diversos campos, especialmente en química analítica. El desarrollo rápido de la quimiometría en química analítica se debe fundamentalmente a la utilización masiva de los ordenadores acoplados a la moderna instrumentación. Se ha pasado de hacer medidas puntuales a una longitud de onda (univariable), a medidas espectrales (multivariables) con gran contenido de información donde los métodos tradicionales de análisis y procesado de datos resultan totalmente insuficientes.

Por tanto, uno de los campos de trabajo fundamental de la quimiometría es el desarrollo y aplicación de los nuevos métodos de calibración multivariable, que permiten extraer la máxima información química del análisis de muestras a partir de señales o respuestas instrumentales con muy poca selectividad. [18]

### 1.5.1. 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. [19]

$$y = g(x) \quad \text{Ecuación (4)}$$

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 un método de análisis químico. La función “**g(x)**” puede ser difícil de determinar por el número de variables en la matriz “**x**” y por la incertidumbre asociada a la determinación de los parámetros “**y**” y “**z**”. 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.

### 1.5.2. Clasificación de los métodos de calibración multivariable

Aunque no existe un criterio unificado de clasificación de los métodos de calibración multivariable, el propuesto por Martens & Naes [20] ha sido uno de los más aceptados. Su clasificación se basa en los siguientes aspectos fundamentales:

- Según la relación entre las variables dependientes e independientes; ésta relación puede ser descrita mediante un modelo lineal o no lineal.
- Según la forma de encontrar la relación entre las variables: pueden ser métodos directos, donde los parámetros de calibración se calculan

directamente a partir de la señal de cada uno de los analitos en forma individual, o métodos indirectos, donde tales parámetros se calculan a partir de las señales analíticas de las mezclas de los componentes

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

Debido a la gran cantidad de métodos multivariados existentes, la selección y aplicación de un método específico depende del objetivo propuesto, de las características de las muestras y de la complejidad del sistema en estudio.

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

La norma ASTM E-1655<sup>[21]</sup> sugiere las siguientes etapas para la construcción de modelos de calibración multivariable a partir de mediciones espectrales sobre el analito de interés:

- Selección de muestras para calibración: se debe contar con muestras altamente representativas que contenga la máxima variabilidad física y química esperada en las muestras para las cuales será aplicado el modelo.
- Caracterización de muestras de calibración: la caracterización se debe realizar por un método de referencia previamente establecido, el cual 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 en la muestra.
- Cálculo del modelo matemático: en esta etapa se realizan pretratamientos a la señal espectral y aplicación de técnicas de regresión sobre los datos espectrales. Los pretratamientos incluyen: el suavizado para reducir el ruido

en los datos, la normalización para lograr que los datos estén aproximadamente a la misma escala, la centralización para evitar que ciertos puntos tengan más peso que otros en el modelo y la derivación de diferente orden para extraer información detallada que no puede ser observada en el espectro normal. Dentro de las técnicas de correlación más comunes se tiene la regresión lineal múltiple (MLR), la regresión por componentes principales (PCR), y la regresión por mínimos cuadrados parciales (PLS).

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

$$RMSEC = \sqrt{\frac{\sum_{i=1}^c (y_i - \hat{y}_i)^2}{n_c}} \quad \text{ecuación (5)}$$

Donde  $y_i$  es el valor de referencia de la espuesta,  $\hat{y}_i$  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:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^c (y_i - \hat{y}_{CV,i})^2}{n_c}} \quad \text{ecuación (6)}$$

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

El fundamento matemático para el desarrollo de métodos de calibración multivariable aplicado a técnicas espectroscópicas de análisis instrumental tiene su origen en el álgebra matricial. Si “**S**” espectros de calibración son medidos a “**W**” discretas longitudes de onda (o frecuencias), es posible construir una matriz de datos espectrales “**X**” de dimensiones “**W x S**” que contiene un espectro en columna. De la misma manera es posible construir un vector “**Y**” de dimensiones “**S x 1**” que contiene los valores de referencia de las muestras de calibración. El objetivo de la calibración multivariable es calcular un vector “**p**” de dimensión “**W x 1**” que resuelva la ecuación 7.

$$y = X^t p + e \quad \text{Ecuación (7)}$$

Donde “ $X^t$ ” es la transpuesta de la matriz “ $X$ ” y “ $e$ ” es un vector de dimensión “ $S \times 1$ ”, llamado vector error. Este último vector se calcula como la diferencia entre los valores de referencia y los valores estimados por el modelo. Generalmente el vector “ $p$ ” se estima minimizando la suma de los cuadrados de los errores mediante la ecuación 8.

$$e^t e = |e^2| = (y - X^t p) \quad \text{Ecuación (8)}$$

#### **1.5.3.1. 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.

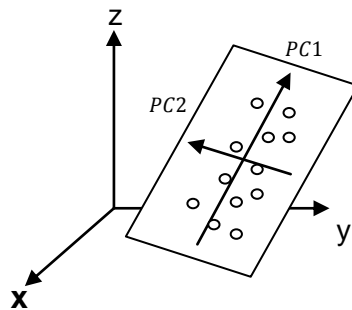
Para esto se parte de una matriz de datos  $X$ , donde cada fila es una muestra, y cada columna una variable, que puede ser la absorbancia o la longitud de onda. La dimensionalidad de esta matriz, si se han registrado  $m$  muestras y  $k$  variables, será  $m \times k$ . La situación de estas variables 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 parciales (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.

Geoméricamente, el PCA es un cambio en la dirección y dimensionalidad de los ejes (figura 3). 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 X; el segundo se escoge de tal forma que sea perpendicular al primero y que explique la máxima variabilidad de la matriz 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 gráfico de puntuaciones (scores) revela las coordenadas de las muestras en estos nuevos ejes <sup>[22]</sup>.

**Figura 3. Representación geométrica de un PCA <sup>[22]</sup>**

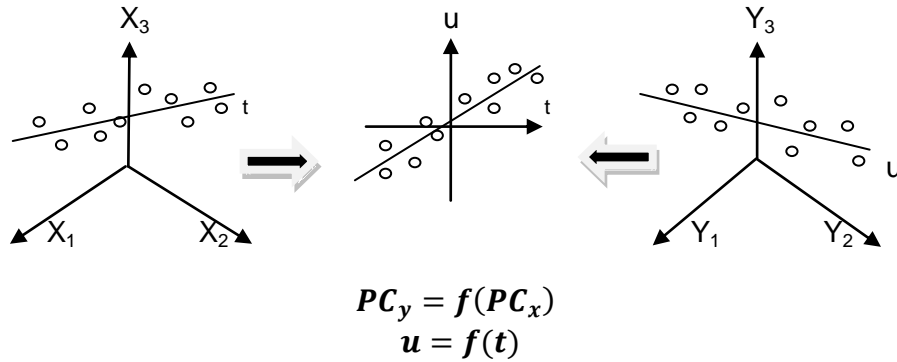


### 1.5.3.2 Regresión por mínimos cuadrados parciales (PLS)

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, número de bromo), para encontrar un

conjunto de variables auxiliares llamadas variables latente VL o factores en X que mejor predicen las variables latentes en Y. En la figura 4 se muestra la descripción del PLS.

**Figura 4. gráfica del modelo de regresión PLS<sup>[21]</sup>**



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 9 y 10.

$$X = TP^T + E = \sum t_a p_a^T + E \quad \text{Ecuación (9)}$$

$$Y = UQ^T + F = \sum u_a q_a^T + F \quad \text{Ecuación (10)}$$

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 x a),  $P^T$  (a x k) y  $Q^T$  (a x 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:

$$u_a = b_a t_a \quad \text{Ecuación (11)}$$

Donde  $b_a$  es el coeficiente de regresión para cada VL obtenido a través de la ecuación (12).

$$b_a = \frac{u_a^T t_a}{t_a^T t_a} \quad \text{Ecuación (12)}$$

A partir de la ecuación (11) se calcula el valor de Y utilizando la relación interna  $u_a$ .

$$Y = TBQ^T + F \quad \text{Ecuación (13)}$$

Donde B es la matriz de los regresores de  $b_a$ , de dimensiones (a x 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>[22]</sup>.

### 1.5.3.3. Regresión Polinomial Múltiple (RPM).

Esta es una forma especial de ajuste de mínimos cuadrados, donde la función de caracterización es aproximada por un polinomio.

Para realizar un mapeo entre los componentes principales (PC) y la propiedad de número de bromo, se definen relaciones como las siguientes:

$$NB = a_{11}PC1 + a_{12}PC2 + a_{13}PC3 + a_{14}(PC1)^2 + a_{15}(PC2)^2 + \dots \quad \text{Ecuación (14)}$$

Donde los valores  $a_{ij}$  (i,j = 1,2,3,...) son los coeficientes a determinar.

El planteamiento se puede realizar de forma matricial:

$$[NB] = [PC1 \ PC2 \ PC3 \dots (PC1)^2 \dots (PC1)(PC2)(PC3)] \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \quad \text{Ecuación (15)}$$

Si se tienen n valores de número de bromo y sus correspondientes componentes principales, se hace necesario determinar la matriz **A**:

$$T = SA \quad \text{Ecuación (16)}$$

Donde **T** es la matriz de los valores de número de bromo, **S** es la matriz de valores de los componentes principales

$$A = S^{-1}T \quad \text{Ecuación (17)}$$

Donde **S**<sup>-1</sup> es la matriz inversa de **S**. Si **S** no es una matriz cuadrada, es decir, si el sistema es sobre-determinado (más ecuaciones que variables), se debe determinar la pseudoinversa de **S**. Teniendo los valores de la matriz **A**, es posible calcular los valores de número de bromo, para cualquier grupo nuevo de componentes principales<sup>[23]</sup>.

## 1.6. ANTECEDENTES

Es importante tener en cuenta los cambios que se han introducido con el paso de los años a la espectroscopia infrarroja, En 1943, los discos de KBr empezaron a emplearse en sólidos. Después se empleo la reflexión total atenuada (ATR), especialmente útil para muestras gruesas, muestras fuertes de absorber, y estudios de superficie. En la década de 1960, se empleo la espectroscopia infrarroja con trasformada de Fourier (FT-IR). Todos estos avances fueron posibles gracias a la aplicación de nuevos elementos ópticos, el desarrollo y la miniaturización de los láseres, la introducción del algoritmo para transformada rápida de Fourier (FFT), el desarrollo de las microcomputadoras.<sup>[24]</sup>

En el LEAM (Laboratorio de Espectroscopia Atómica y Molecular) de la Universidad Industrial de Santander, se han venido realizando estudios para la predicción de propiedades (densidad, análisis SARA, entre otros) a los fondos de vacío.

Se han realizado estudios para caracterizar las fracciones del crudo por diversas técnicas entre las que se encuentra la espectroscopia infrarroja en la región cercana (NIR), la cromatografía de gases (técnica que se emplea comúnmente para determinar la composición química del nafta, esta es una fracción liviana obtenida por destilación atmosférica); para este análisis se obtuvieron resultados satisfactorios, debido a que la fracción liviana de los crudos presenta moléculas más pequeñas y tiene un menor número de componentes que las fracciones pesadas como los fondos de vacío los cuales contiene un mayor número de moléculas grandes y complejas, por ende la región cercana del infrarrojo es capaz de analizar estos componentes, moléculas pequeñas y cumple con los requisitos para el análisis de componentes livianos del crudo (gasolina, gasóleo, entre otros).<sup>[25]</sup>

En el año 2010 se generó un modelo de predicción de la densidad en asfáltenos colombianos utilizando espectroscopia fotoacústica en la región del infrarrojo medio y métodos quimiométricos, con el objetivo de reducir los costos y el tiempo de análisis y poder más adelante reemplazar las técnicas tradicionales. Para este estudio, se obtuvo un error estándar de predicción calculado para el conjunto de muestras de validación de 0.01, lo que garantiza la predicción de la densidad de los asfáltenos de crudos Colombianos<sup>[26]</sup>.

Además se realizó un trabajo de determinación de la composición elemental de los fondos de vacío de crudos colombianos, pero empleando la espectroscopia de plasma inducido por laser (LIBS), en este estudio se determinó de manera cuantitativa la relación %p(C/H), además la influencia de los diversos metales

presentes Fe, V, Ni, entre otros, obteniéndose resultados bastante confiables para la determinación cuantitativa de la relación %p (C/H) a partir de la relación de sus áreas. <sup>[27]</sup>.

El estudio de la determinación SARA de crudos colombianos utilizando espectroscopia infrarrojo en la región cercana (NIR) y el análisis multivariado, permitió la ampliación del análisis SARA (la separación de saturados (S), aromáticos (A) resinas (R) y asfáltenos (A)) hasta crudos súper-ligeros (>39°API), incluyendo para la cuantificación total, el análisis de las fracciones livianas que se pierden durante el desarrollo de la metodología convencional (el PIANO de la nafta y los aromáticos totales del keroseno)<sup>[28]</sup>. Además se realizó un método de predicción del análisis SARA de crudos colombianos también se realizó un método de predicción utilizando espectroscopia infrarrojo en la región media con reflectancia total atenuada (FTIR-ATR) y métodos quimiométricos, la cual ha mostrado ser una técnica alternativa a la cromatografía de exclusión molecular para la determinación del análisis SARA; con la ventaja de evitar el pretratamiento de la muestra e incrementar la reproducibilidad en los resultados. Para este estudio, se obtuvo un error de predicción calculado para el conjunto de muestras de validación entre 3.03 correspondiente a los saturados y 5.17 correspondiente a los asfáltenos <sup>[29]</sup>. Para este mismo estudio de predicción del análisis SARA se utilizó la espectroscopia de fluorescencia inducida por láser (LIF) y técnicas multivariadas, en este estudio, se obtuvo un error de predicción calculado para el conjunto de muestras de validación entre 1.29 correspondiente a las resinas y 2.42 correspondiente a los saturados.<sup>[30]</sup>

Otra técnica utilizada para la predicción del análisis SARA es la espectroscopia fotoacústica en la región del infrarrojo medio (FTIR/PAS) y métodos quimiométricos, la cual demostró ser muy útil para el estudio estructural de muestras sólidas y opacas, proporcionando espectros repetibles con una desviación estándar de 0.002.<sup>[31]</sup>

## 2. METODOLOGÍA

### 2.1. SELECCIÓN DE MUESTRAS

Para la realización del presente trabajo se utilizaron treinta y nueve (39) espectros NIR de naftas colombianas (tabla 5); tanto los espectros NIR como el número de bromo de cada nafta fueron suministrados por el Instituto Colombiano de Petróleo (ICP)

**Tabla 5. Análisis del número de bromo de las 39 muestras**

Nafta	NB	Nafta	NB
1	118,78	21	28,3
2	161,48	22	24,52
3	131,25	23	22,19
4	113,41	24	22,2
5	89,16	25	19,57
6	59,19	26	145,04
7	46,08	27	124,99
8	45,53	28	115,17
9	32,73	29	90,77
10	28,9	30	56,95
11	24,42	31	24,67
12	44,01	32	15,07
13	112,45	33	25,7
14	127,65	34	90,17
15	120,59	35	82,06
16	100,94	36	62,73
17	55,44	37	43,56
18	27,57	38	21,34
19	72,73	39	12,5
20	41,23		

## **2.2. PROCESAMIENTO DE DATOS**

### **2.2.1. Pretratamiento de datos**

Los datos espectroscópicos suministrados por el ICP se analizaron utilizando las herramientas del software The Unscrambler versión 9.7, suministrado por el Laboratorio de Espectroscopia Atómica y Molecular (LEAM) de la UIS. Para verificar la calidad de la información espectral obtenida se realizó inicialmente una inspección visual de los espectros NIR para las 39 muestras en estudio (figura 4) con el fin de descartar la presencia de muestras atípicas o la influencia de posibles errores instrumentales y humanos durante la toma de muestras y posterior adquisición de espectros. La desviación estándar observada en las absorbancias medidas para cada familia de espectros fue en todos los casos inferior a 0,012 para el rango de longitudes de onda empleado. Por la baja dispersión en las señales espectrales se decidió excluir cualquier clase de procesamiento previo al análisis matemático de las señales.

### **2.2.2. Análisis de componentes principales (PCA)**

Se realizó una descomposición por componentes principales (PC) al conjunto total de intensidades espectrales. Esto permitió discriminar las muestras de acuerdo al número de bromo (grafico de puntuaciones).

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

Mediante la regresión PLS se evaluó el potencial de la espectroscopia NIR para predecir el número de bromo a 39 naftas colombianas.

El desarrollo del modelo de predicción abarcó dos grandes etapas: calibración y validación interna.

- Calibración y validación interna:

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

#### **2.2.4. Regresión Polinomial Múltiple (RPM)**

Como se explicó previamente en el capítulo anterior, por medio de esta técnica se pretende encontrar la relación entre los componentes principales del espectro NIR y el número de bromo a través de un polinomio. La relación se puede apreciar en la ecuación 1.7.

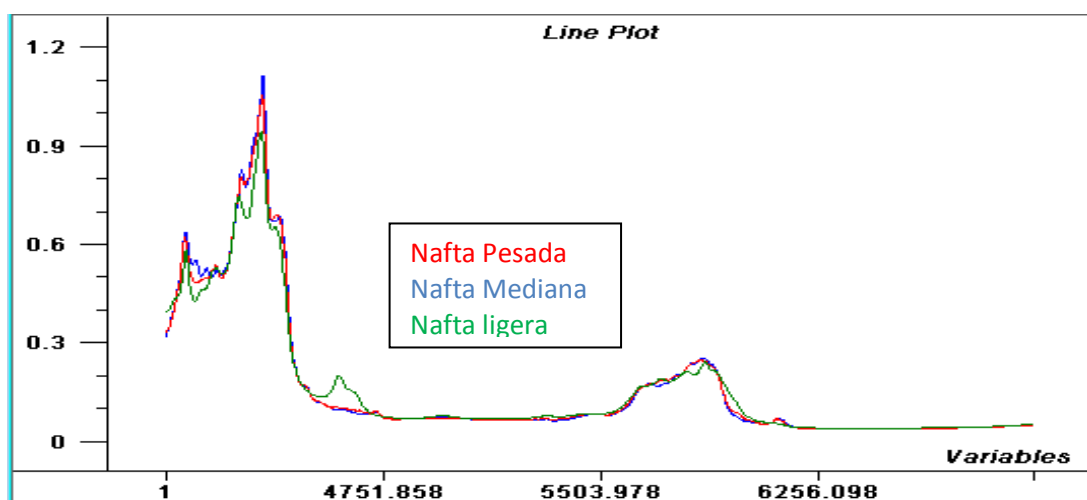
La tarea de encontrar los coeficientes  $a_{ij}$  (matriz de caracterización o de transformación) permite determinar el número de bromo para cualquier vector nuevo de valores de componentes principales (PC1, PC2, PC3), los cuales se determinaron a través del código en MATLAB expuesto en el anexo A. El procedimiento de cálculo de la matriz de coeficientes es posible a través del código en MATLAB expuesto en el anexo B.

Seis clases de polinomios son utilizados para el proceso de mapeo. Tres de ellos de primer grado, y los demás de segundo, tercer y cuarto grado respectivamente; contando con tres hasta 35 términos. Con las 39 muestras se procede a encontrar la matriz de coeficientes que proporcione la mejor transformación. Los resultados son comparados con los valores dados por el ICP a través del cálculo del RMSEC Y RMSECV y el coeficiente de regresión.

### 3. RESULTADOS Y ANÁLISIS

#### 3.1. ANÁLISIS ESPECTRAL DE LAS NAFTAS

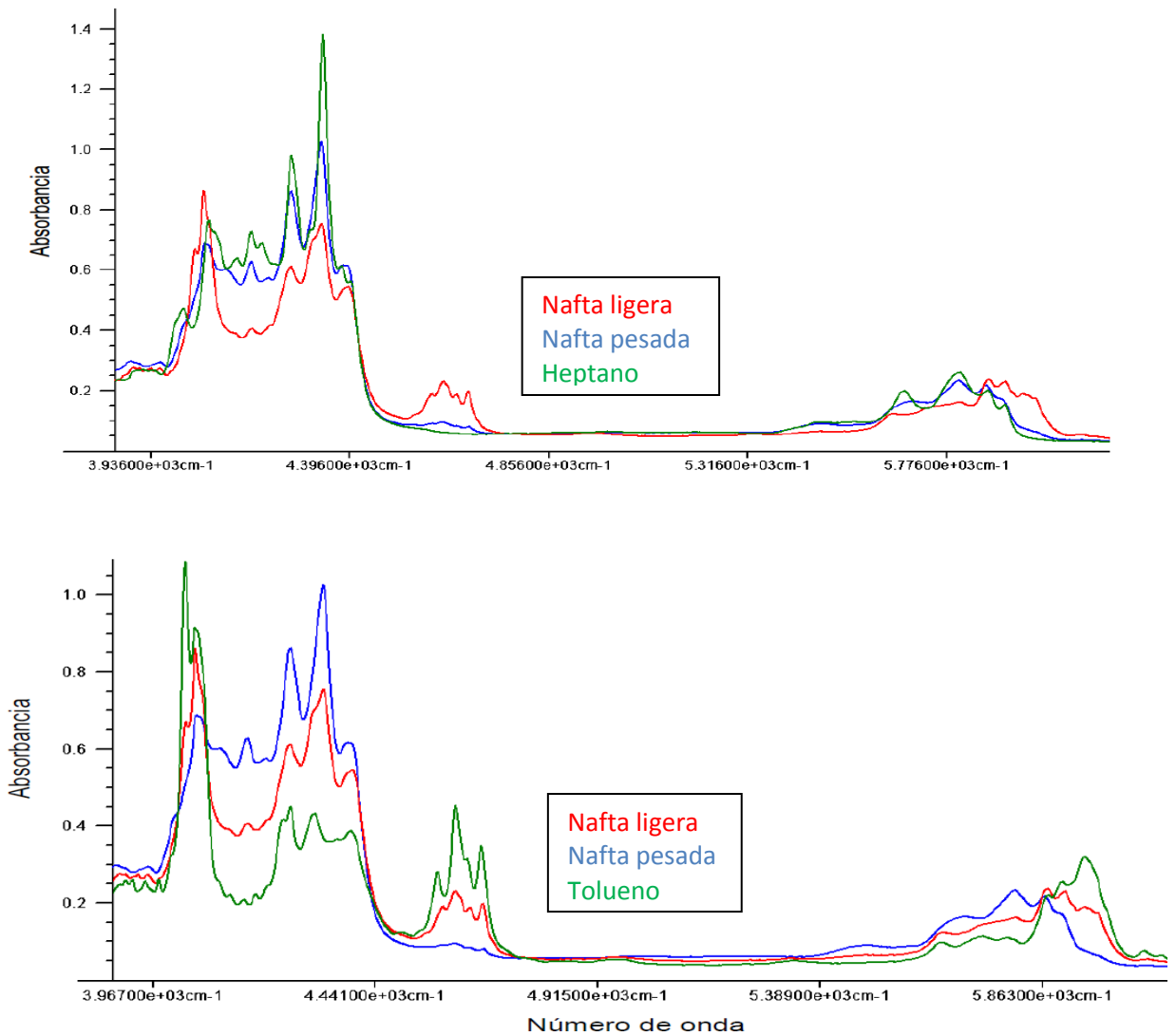
Figura 4. Descripción gráfica del modelo de regresión PLS



Las naftas pesadas y medianas son prácticamente indistinguibles a partir de una inspección visual de sus espectros NIR debido a su similitud en el número de bromo en concordancia con los resultados obtenidos por el ICP. Las naftas livianas presentan marcadas diferencias con las anteriores debido a su gran diferencia en el número de bromo. Estas diferencias se observan en la banda de absorción en la región de  $4500\text{-}4600\text{cm}^{-1}$ , atribuidas a una combinación de vibraciones fundamentales  $=\text{CH-}$  Y  $\text{C}=\text{C}$  en componentes aromáticos y olefinas. Otras diferencias apreciables se observan en las regiones de  $4000\text{-}4400\text{cm}^{-1}$ , donde surgen bandas de combinación de grupos metilo y metileno, y en la región de  $5400\text{-}6300\text{cm}^{-1}$ , donde aparecen los primeros sobretonos de vibraciones de estiramiento  $-\text{CH}$  de grupos metilo, metileno y aromáticos <sup>[32]</sup>.

Una comparación con los espectros NIR de hidrocarburos puros confirma las asignaciones anteriores y permite identificar claramente la contribución de grupos estructurales presentes en las muestras de nafta (figura 5)

Figura 5. de espectros NIR de naftas con hidrocarburos puros<sup>[33]</sup>



El espectro NIR del n-heptano es muy similar al de una nafta pesada en la mayor parte del rango espectral analizado, siendo clara la contribución de las vibraciones  $-CH$  en esta clase de compuestos. El tolueno presenta bandas de absorción características en las regiones de  $4500$  a  $4700\text{ cm}^{-1}$  y de  $5900$  a  $6200\text{ cm}^{-1}$  que concuerdan claramente con la alta presencia de componentes aromáticos y olefinas de una nafta ligera.

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

Una descomposición por componentes principales realizada sobre la información espectral permitió identificar las muestras analizadas de acuerdo a su naturaleza química, ya que a simple vista, es casi imposible diferenciarlas a partir de su espectro de absorción NIR.

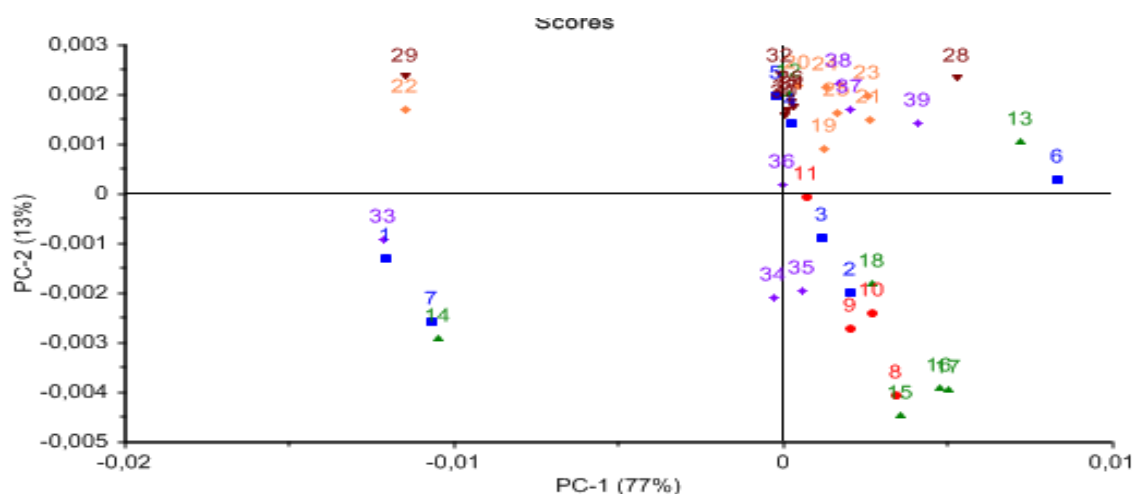
**Tabla 6. Varianza de los componentes principales generados por las 39 muestras**

Componente principal	Varianza Explicada (%)	Varianza Explicada Acumulada (%)
PC 1	77,1	77,1
PC 2	13,42	90,52
PC 3	4,06	94,58
PC 4	2,24	96,82
PC 5	1,59	98,41
PC 6	0,4	98,81
PC 7	0,32	99,13
PC 8	0,23	99,36
PC 9	0,18	99,54

Observando el porcentaje de varianza de los datos espectrales que explica los primeros 9 componentes principales se determinó que 7 es el número máximo de componentes a utilizar en los modelos de predicción ya que a partir del PC8 el aumento de la varianza explicada acumulada no aumenta en más de una unidad.

La representación de los dos primeros componentes principales (PC1 y PC2) se muestra en la figura 6.

**Figura 6. Gráfico de puntuaciones para las 39 muestras: descomposición de la información espectral en el PC2 vs PC1**



La figura de scores, es un gráfico de dispersión en dos dimensiones de los componentes principales PC1 y PC2 en los cuales se recoge el 77% y el 13% de varianza respectivamente, lo que indica que estos dos componentes contienen el 90% de la varianza total de las intensidades espectrales. Además éste permite establecer las relaciones entre las muestras según su cercanía en el plano PC1-PC2, ya que cuanto más cerca están las muestras al eje de coordenadas del gráfico, más similares son con respecto a los dos componentes en cuestión.

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

Tanto el PC1 como el PC2 no presentan un efecto claro sobre la distribución para las naftas livianas, muestras que tienen un número de bromo alto, entre 82.06 Y 161.48 (grupo de color marrón), ya que estas muestras no se agrupan en un solo cuadrante sino que aparecen dispersas. Sin embargo, si contienen información que permite diferenciar las naftas pesadas, muestras que tienen un número de bromo bajo, entre 12.5 y 72.73 (grupo de color anaranjado), ya que pertenecen a los cuadrantes I y IV, ubicados en la zona positiva para el PC1, mientras que en el

cuadrante I tienen un valor positivo para PC2, y negativo para el cuadrante IV. Las muestras en color verde fueron muestras con un valor de número de bromo inferior a 72.73 pero que se encontraban dispersas del grupo de color anaranjado.

De acuerdo con lo anterior se encontró que las muestras se agrupan de acuerdo a la tabla 7.

**Tabla 7. Separación según el número de bromo**

Nafta	NB	Nafta	NB
6	59,19	33	25,7
8	45,53	22	24,52
9	32,73	1	118,78
10	28,9	2	161,48
11	24,42	3	131,25
12	44,01	4	113,41
19	72,73	5	89,16
20	41,23	13	112,45
21	28,3	14	127,65
23	22,19	15	120,59
24	22,2	16	100,94
25	19,57	26	145,04
31	24,67	27	124,99
32	15,07	28	115,17
37	43,56	29	90,77
38	21,34	30	56,95
39	12,5	34	90,17
17	55,44	35	82,06
18	27,57	36	62,73
7	46,08		

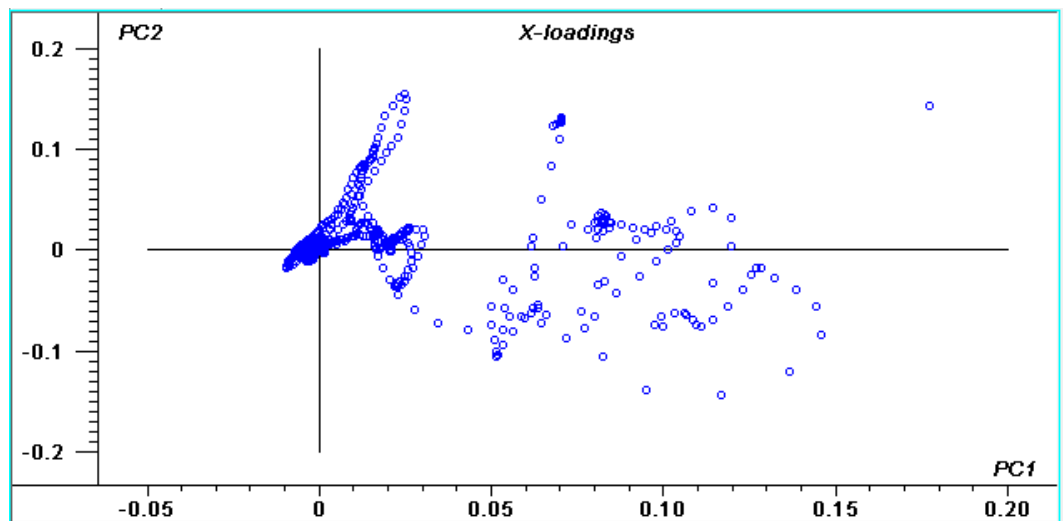
De la figura 6 se encontró que del grupo de naftas pesadas, con número de bromo entre 12.5 y 72.5, la muestra 22 se aleja considerablemente de las demás, analizando su valor de número de bromo se estableció que esto se debe a que dicha muestra presenta un valor similar a la muestra 11, por lo que se define como un posible outliers, y es posible que deba sacarse de la calibración del modelo de predicción. Para el grupo de naftas livianas, con número de bromo entre 82.06 y

161.48, las muestras 7, 22 y 33 se alejan considerablemente de las demás, analizando su valor de número de bromo, se estableció que esto se debe a que dichas muestras presentan un valor inferior a este rango, por lo que se definen como posibles outliers, y es posible que deban sacarse de la calibración del modelo de predicción.

En la figura 7 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.

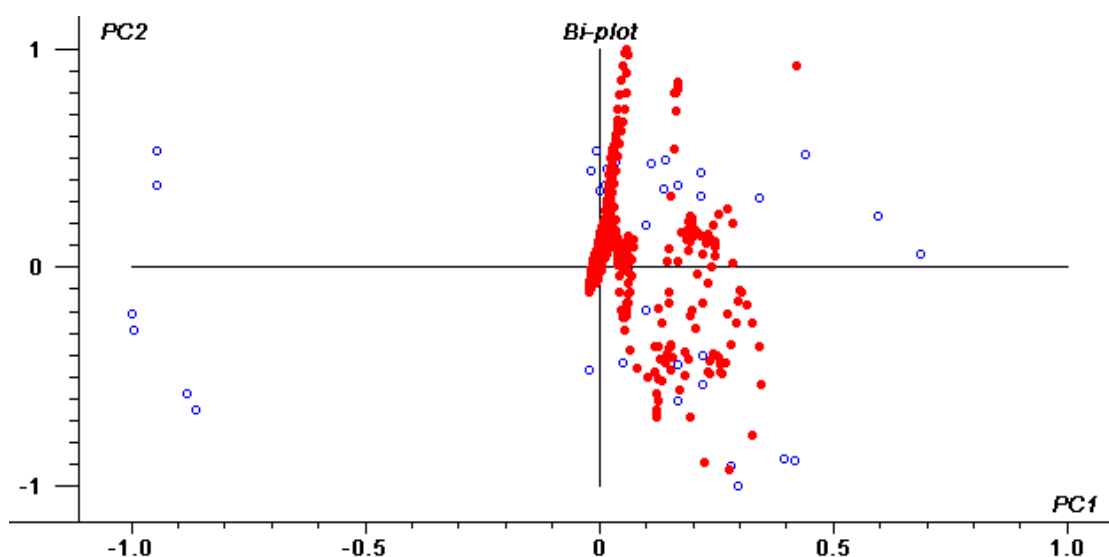
Si se utiliza junto con la gráfica de puntuaciones entonces; las variables a la derecha de la gráfica de pesos serán variables x con valores altos para las muestras a la derecha de la gráfica de puntuaciones.

**Figura 7. Gráfico de pesos de las 39 muestras: importancia de las diferentes variables para el PC1 y el PC2**



En la figura 8 muestra cuales son las variables responsables de las diferencias entre las muestras. Para esto se combinaron la gráfica de puntuaciones (figura 6) junto con la gráfica de pesos (figura 7) para los dos primeros componentes principales (PC2 vs PC1), ya que estos representan las mayores variaciones de la matriz de datos X.

**Figura 8. Gráfico de biplot: combinación de la gráfica de puntuaciones con la gráfica de pesos**

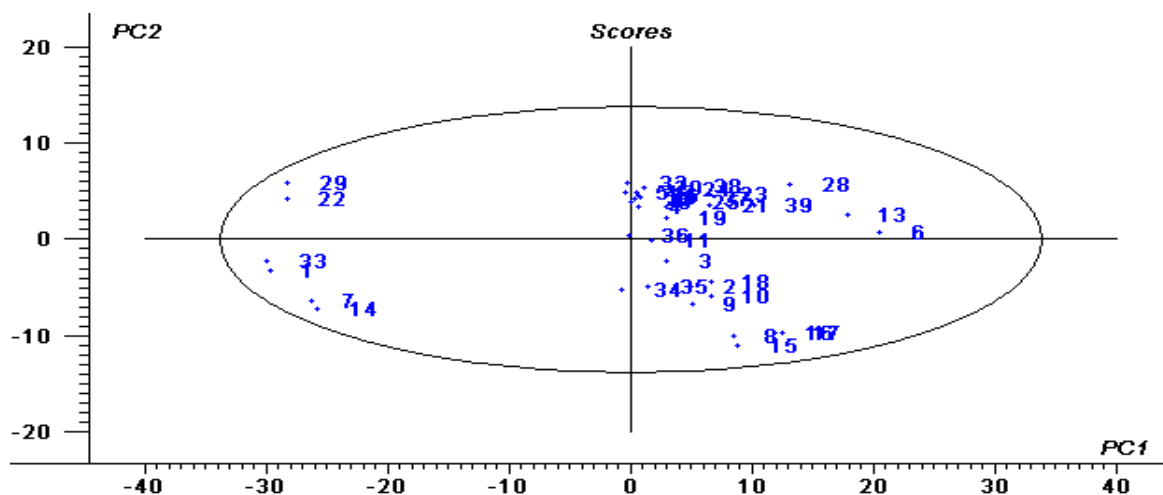


La ubicación de las variables  $x$  en la gráfica de pesos afecta la ubicación de las muestras en la gráfica de puntuaciones. Las variables que se encontraron a la derecha de la gráfica de pesos (figura 7), son explicadas de una forma positiva por el PC1, por esta razón tiene un valor mayor que las variables 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).

De la figura 8 se encontró que para las naftas pesadas, muestras con número de bromo entre 12.5 y 72.73, las regiones que más afectan su ubicación en el gráfico de puntuaciones, son las regiones de 4000 a 4300  $\text{cm}^{-1}$  donde surgen bandas de combinación de grupos metilo y metileno, y la región de 4500 a 4700  $\text{cm}^{-1}$ , atribuidas a una combinación de vibraciones fundamentales  $=\text{CH}-$  Y  $\text{C}=\text{C}$  en componentes aromáticos y olefinas. Además de estas regiones, para las naftas livianas, muestras con número de bromo entre 82.06 y 161.45, se encontró otra región que afecta su ubicación en el gráfico de puntuaciones, la región de 5800 a 6000  $\text{cm}^{-1}$  donde aparecen los primeros sobretonos de vibraciones de estiramiento  $-\text{CH}$  de grupos metilo, metileno y aromáticos

Por otro lado en la gráfica de puntuaciones aparece la elipse de confianza de Hotelling  $T^2$  (figura 9). Los datos que quedan fuera de la elipse indican datos que no quedarían bien explicados por las nuevas variables.

**Figura 9. Gráfico de puntuaciones con el límite de Hotelling  $T^2$  para las 39 muestras**



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

Debido a la diferente naturaleza de las muestras de naftas se estableció que lo más recomendable es dividir las 39 muestras en los dos grupos determinados en los resultados anteriores y generar el PCA para cada uno.

### 3.2.1. Análisis por componentes principales para el grupo de muestras pesadas

Se determinó que el grupo de muestras con número de bromo entre 12.5 y 72.72 está comprendido por 24 muestras que son: 6-12,17-25,30-33,36-39.

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

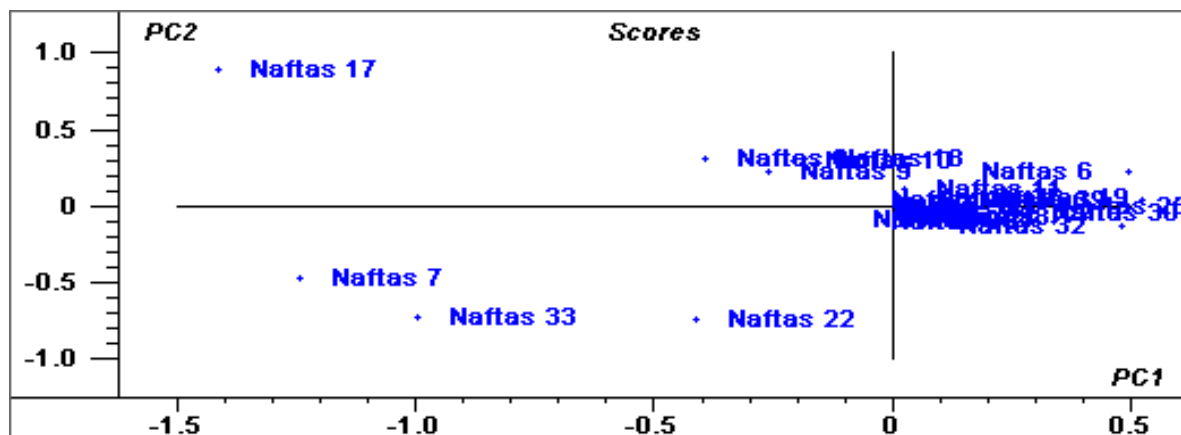
**Tabla 8. Varianza de los componentes principales generados por las muestras pesadas**

Componente principal	Varianza Explicada (%)	Varianza Explicada Acumulada (%)
PC 1	44,64	44,64
PC 2	24,92	69,56
PC 3	7,79	77,35
PC 4	4,07	81,42
PC 5	3,04	84,46
PC 6	6,42	90,88
PC 7	1,5	92,38
PC 8	0,47	92,85
PC9	0,6	93,45
PC10	1,12	94,57
PC 11	0,02	94,59

Observando el porcentaje de varianza de los datos espectrales que explica los primeros 11 componentes principales se determinó que 10 es el número máximo de componentes principales a utilizar en los modelos de predicción ya que a partir del PC10 el aumento de la varianza explicada acumulada es bajo.

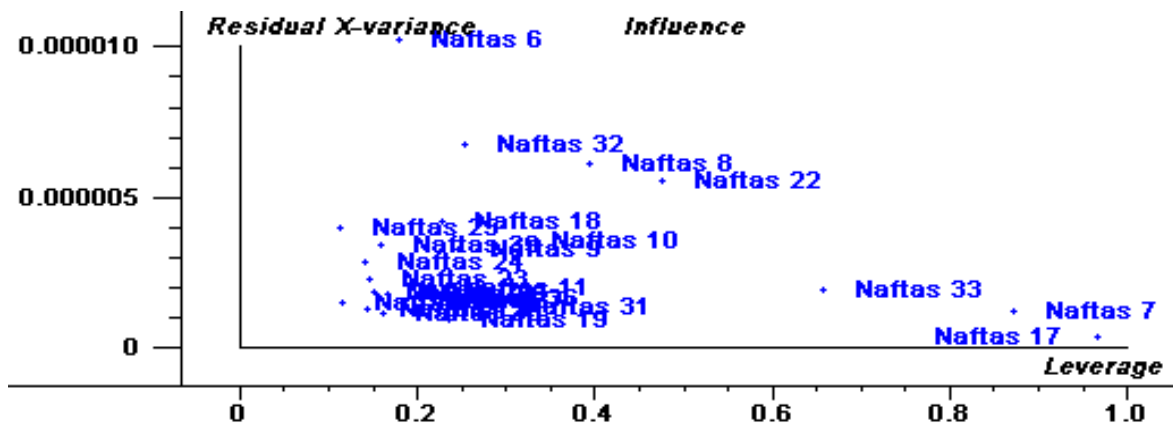
El gráfico de scores con la aplicación del estadístico de Hotelling  $T^2$  (Figura 10) permite visualizar que las muestras 7, 17, 22, y 33 se alejan considerablemente de las demás; analizando la propiedad del número de bromo disponible se estableció que esto se debe posiblemente por efecto de alguna propiedad fisicoquímica diferente al número de bromo, por lo que se definen como outliers, y es posible que se deban sacar del modelo de predicción.

**Figura 10. Gráfico de puntuaciones con el límite de Hotelling  $T^2$  para las muestras pesadas**



El gráfico de influence (figura 11) permite clasificar de forma más detallada los valores atípicos (outliers), influyentes y peligrosos (dangerous outliers) de las muestras analizadas.

**Figura 11. Gráfico de influencia para las muestras pesadas**



Como la muestra 6 presenta alta varianza residual (parte superior izquierda del gráfico) se define como un posible outlier, analizando la propiedad del número de bromo disponible se estableció que esto se debe posiblemente por efecto de alguna propiedad fisicoquímica diferente al número de bromo. Las muestras que presentan alto leverage, es decir las que se extienden a la derecha de la gráfica son influyentes, lo que significa que son las que mejor describirán el modelo de

predicción; para este caso se observó las muestras 7, 17 y 33; con lo cual se pueden llegar a descartar como posibles outliers. Las muestras que presentan alta varianza residual y alto leverage como las muestras 22, 8 y 32 se definen como dangerous outliers, lo que significa que puede que no sean descritas por el modelo que describe la mayoría de las muestras.

### 3.2.2. Análisis por componentes principales para el grupo de muestras livianas

Se determinó que el grupo de muestras con número de bromo entre 82.06 y 161.45 está comprendido por 15 muestras que son: 1-5,13-16,26-29,34-35.

El porcentaje de varianza explicado por cada componente principal se muestra en la tabla 9.

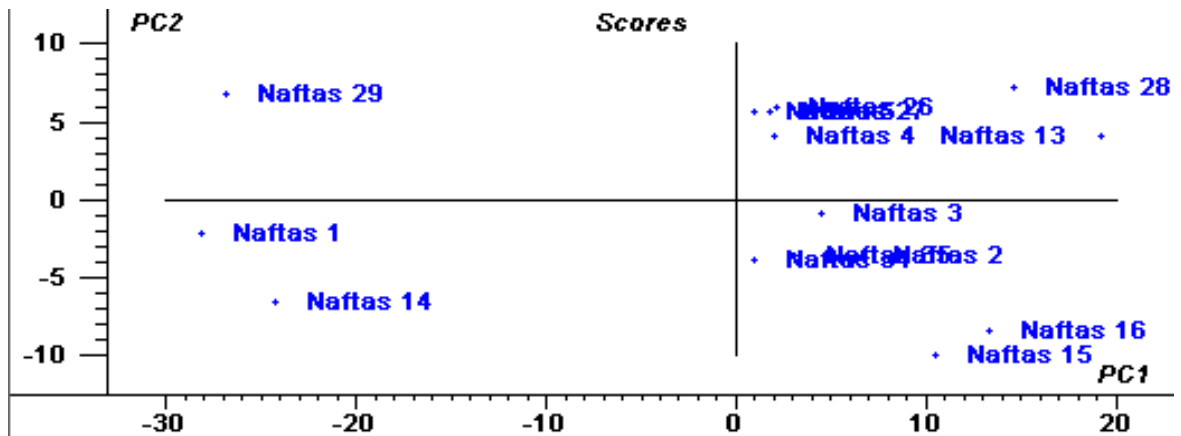
**Tabla 9. Varianza de los componentes principales generados por las muestras livianas**

Componente principal	Varianza Explicada (%)	Varianza Explicada Acumulada (%)
PC 1	75,97	75,97
PC 2	14,59	90,56
PC 3	0,84	91,40
PC 4	2,29	93,69
PC 5	2,64	96,33
PC 6	0,32	96,65
PC 7	0,59	97,24
PC 8	0,44	97,68
PC 9	0,23	97,93
PC 10	0,03	97,96

Observando el porcentaje de varianza de los datos espectrales que explica los primeros 11 componentes principales se determinó que 9 es el número máximo de componentes principales a utilizar en los modelos de predicción ya que a partir del PC9 el aumento de la varianza explicada acumulada es bajo.

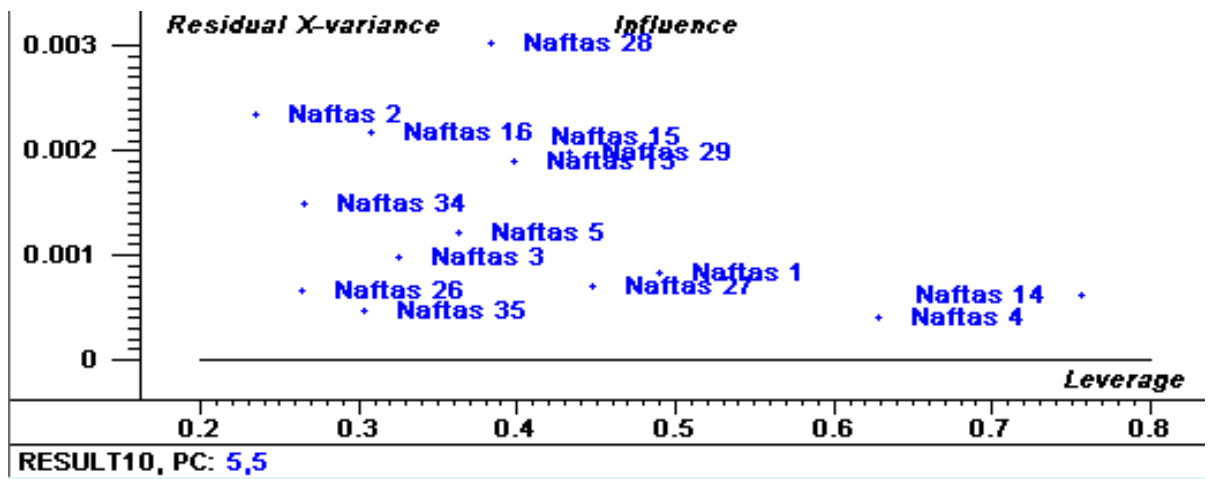
Como se aprecia en figura 12, ninguna muestra presenta valores extremo (ninguna es atípica) de la composición química con respecto a las demás.

**Figura 12. Gráfico de puntuaciones con el límite de Hotelling T<sup>2</sup> para las muestras livianas**



En la gráfica de leverage se observa que la muestra 28 presenta alta varianza residual (parte superior izquierda del gráfico) se define como probable outlier. Las muestras que se extienden a la derecha de la gráfica son influyentes, lo que significa que son las que mejor describirán el modelo de predicción; en este caso las muestras 14 y 4.

**Figura 13. Gráfico de influencia para las muestras livianas**



En la figura 13 no se observó muestras con alta varianza residual y alto leverage, es decir, no hay posibles dangerous outliers.

### 3.3. RESULTADOS DEL MÉTODO PLS

Mediante el algoritmo de regresión PLS, y empleando el programa de análisis multivariado The Unscrambler versión 9.7 (CAMO), se evaluó el potencial de la espectroscopia NIR para predecir el número de bromo de las naftas colombianas.

El desarrollo de los modelos de predicción abarcó dos grandes etapas: calibración y validación. A continuación se describe el procedimiento empleado, tomando como ejemplo el desarrollo del modelo para determinar el número de bromo de naftas, usando 24 muestras, las cuales presentan un rango de 12.5 a 72.73. Para el otro modelo se siguió la misma metodología, por tanto serán mostrados los resultados más relevantes en la sección 3.2.3

#### 3.3.1. Calibración y validación del modelo PLS para naftas pesadas

Teniendo en cuenta el número óptimo de componentes principales y las regiones espectrales más influyentes, fueron desarrollados modelos de predicción variando el número de muestras de calibración en función del análisis PCA descrito en la sección 3.2. La selección del modelo de predicción se realizó evaluando parámetros estadísticos como el RMSECV, RMSEC y la varianza explicada (tabla 10).

**Tabla 10. Parámetros estadísticos del modelo seleccionado para la predicción de naftas pesadas**

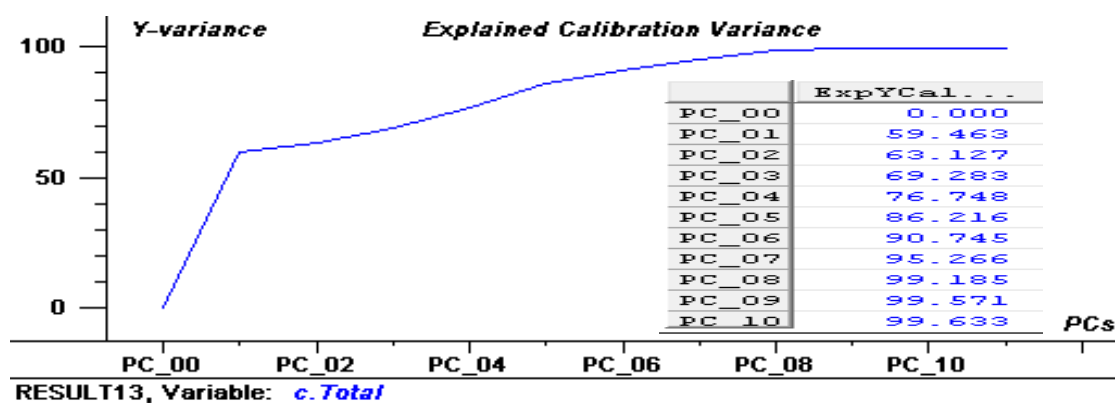
Rangos usados (cm <sup>-1</sup> )	Muestras excluidas	Componentes	Varianza Explicada	RMSEC (%)	RMSECV (%)
5974.15-5804.67	6,7,17,18	10	99,63 %	0,89	3,33
4670.87-4566.45	22,30,33,36				
4339.23-4049.54					

Éste modelo (tabla 10) fue seleccionado por presentar por presentar el menor % de error tanto de validación (RMSECV) como de calibración (RMSEC) y la mayor varianza explicada por los componentes (99.63%).

La muestra 6, 7, 17, 18, 22, 30, 33 y 36 fueron excluidas del modelo de predicción, por diferentes razones como ser potenciales outliers, entre otras descritas anteriormente en el análisis por componentes principales (PCA).

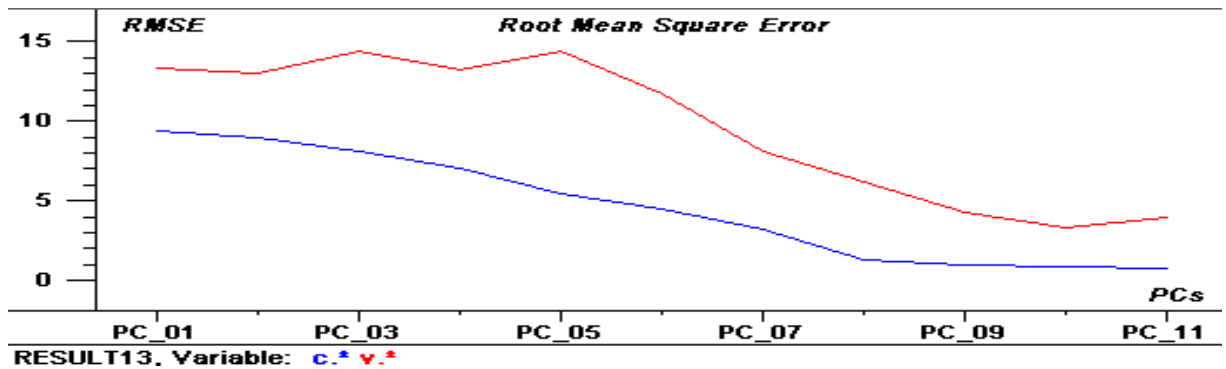
El modelo de predicción explica en las regiones espectrales mostradas en la tabla 10, más del 99% en la variabilidad de los datos, con un mínimo de 10 componentes principales (figura 14).

**Figura 14. Varianza explicada en el modelo PLS para las naftas pesadas**



Otra gráfica que permite visualizar el software, con la que se comprobó el número óptimo de componentes principales (PCs) para generar el modelo de predicción, es la de la raíz cuadrada del error de validación (RMSECV) y calibración (RMSEC) figura 15, ya que, el RMSEC es una medida de la desviación estándar de los residuales obtenidos por la diferencia entre los valores de referencia y predichos por el modelo para las muestras de calibración, y el RMSECV se basa en un algoritmo iterativo que selecciona muestras dentro del grupo de calibración para desarrollar el modelo de predicción y posteriormente evaluarlo sobre la muestra restante.

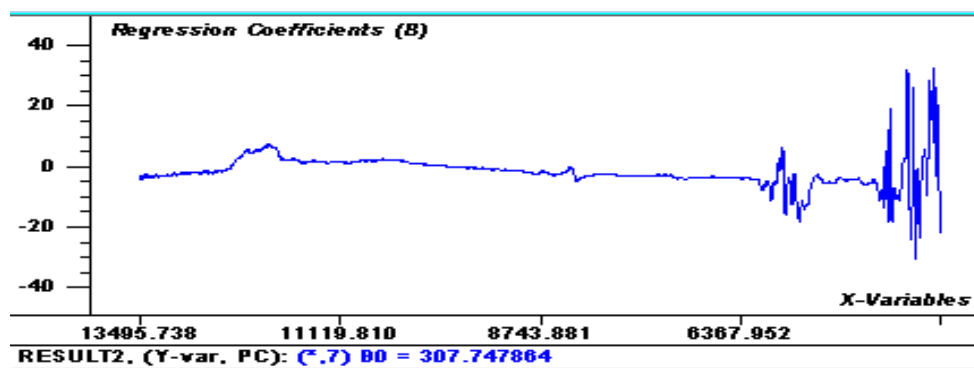
**Figura 15. Gráfica de los errores RMSEC y RMSECV calculados en función de los PCs usados para el modelo PLS para las naftas pesadas**



Como se observa en la figura 15, se logró comprobar que 10 es el número óptimo de PCs usados en el modelo de predicción del número de bromo, ya que para este caso, a pesar que el RMSEC disminuye a medida que aumenta el número de variables latentes, el RMSECV aumenta a partir del PC10, teniendo este último como valor numérico de 3.33.

Claramente se identificaron regiones espectrales como las de mayor influencia sobre las componentes principales y que concuerdan con las regiones de mayor absorbancia en los espectros NIR obtenidos para las muestras de nafta: la región de de 4000 a 4700  $\text{cm}^{-1}$  y la región de 5500 a 6000  $\text{cm}^{-1}$ . Además, los coeficientes de regresión (figura 16), permitieron identificar que frecuencias de absorbancia contribuyen tanto positiva como negativamente a la componente principal y, por ende, a la cuantificación del número de bromo entre 12.5 y 72.73. Se observó que las regiones de 4000 a 4100  $\text{cm}^{-1}$ , de 4500 a 4700  $\text{cm}^{-1}$  y de 5700 a 5800  $\text{cm}^{-1}$  presentan un efecto positivo sobre la componente principal (absorción debida a grupos CH aromático), la región de 4100 a 4400  $\text{cm}^{-1}$  presenta un efecto negativo sobre un efecto negativo (absorción debida a grupos CH<sub>2</sub> y CH<sub>3</sub> en alcanos).

**Figura 16. Coeficientes de regresión para el primer componente principal del modelo PLS para las naftas pesadas**



La validación del modelo PLS para naftas pesadas se realizó empleando el grupo de muestras utilizadas para generar el modelo de calibración.

La tabla 11 muestra que los valores predichos por el modelo propuesto y los valores de referencia determinados por el ICP están cercanos entre sí, para 1 de las 14 muestras evaluadas se obtuvo un porcentaje de error superior al 5%, siendo 5.94% el mayor porcentaje de error, que lo presenta la muestra 24.

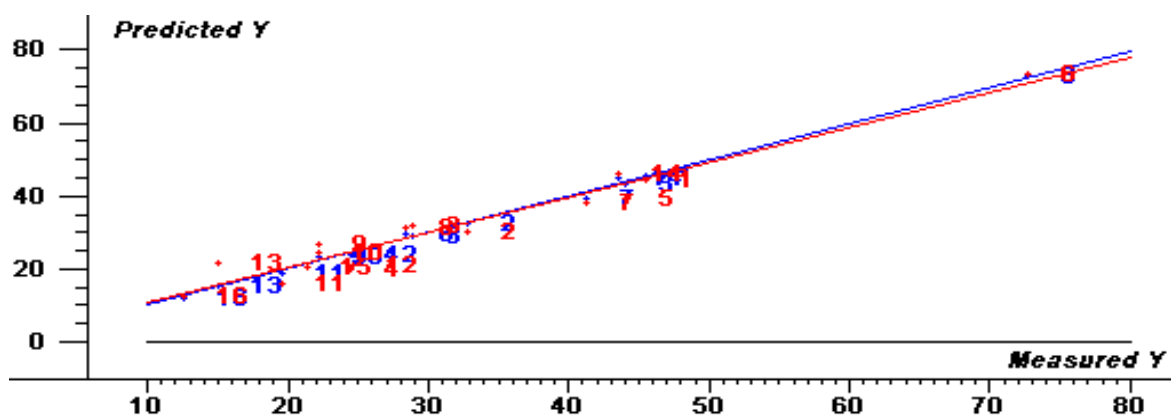
**Tabla 11. Validación del modelo de predicción del número de naftas pesadas**

Muestras	Referencia (NB)	Predicho (NB)	Residual	Error relativo (%)
Nafta 8	45,53	44,58	-0,95	0,04
Nafta 9	32,73	30,33	-2,4	0,64
Nafta 10	28,9	31,69	2,79	1,31
Nafta 11	24,42	19,96	-4,46	2,17
Nafta 12	44,01	39,12	-4,89	2,15
Nafta 19	72,73	73,15	0,42	0,14
Nafta 20	41,23	38,03	-3,2	4,46
Nafta 21	28,3	31,17	2,87	4,91
Nafta 23	22,19	26,91	4,72	4,64
Nafta 24	22,2	24,57	2,37	5,94
Nafta 25	19,57	15,86	-3,71	4,34
Nafta 31	24,67	21,12	-3,55	2,03
Nafta 32	15,07	21,37	6,3	1,66
Nafta 37	43,56	45,96	2,4	3,03
Nafta 38	21,34	20,51	-0,83	2,81
Nafta 39	12,5	12,61	0,11	2,24

El coeficiente de correlación entre los valores de referencia y de predicción fue superior al 0.99 para la curva de calibración y superior a 0.95 para la curva de validación (figura 17).

Los resultados hallados en esta etapa demostraron que el modelo para la predicción del número de bromo de naftas pesadas es satisfactorio para todas las muestras.

**Figura 17. Curvas de calibración y validación del modelo de predicción para las naftas pesadas**



### 3.3.2. Calibración y validación del modelo PLS para naftas livianas

El rango de calibración en el que se desarrollaron los análisis fue de 82.06 a 131.25 número de bromo.

La tabla 12 muestra el error estándar de calibración (RMSEC), de validación (RMSECV) y la varianza explicada para el modelo de predicción seleccionado con 9 PCs.

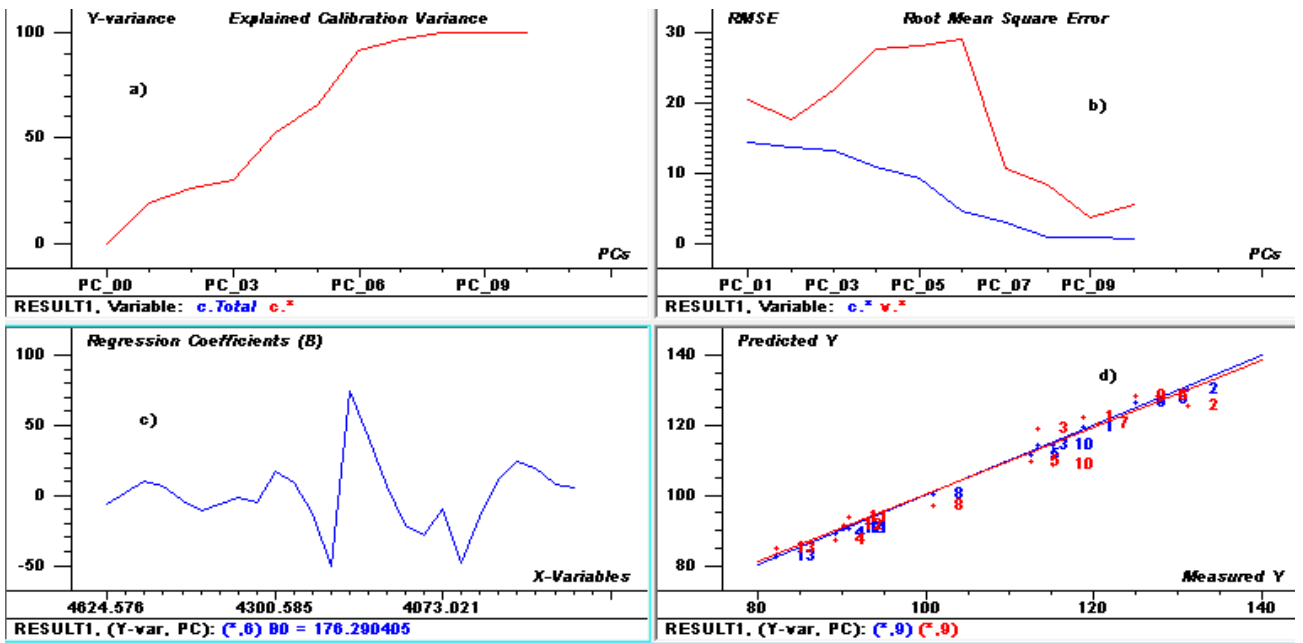
**Tabla 12. Parámetros estadísticos del modelo seleccionado para la predicción de naftas livianas**

Rangos usados (cm <sup>-1</sup> )	Muestras excluidas	componentes	Varianza Explicada	RMSEC (%)	RMSECV (%)
4000-4300	2, 26	9	99,78 %	0,74	3,68
4500-4700					

Las muestras 2 y 26 fueron excluidas del modelo de predicción por presentar los mayores número de bromo, 161.45 y 145.04 respectivamente, lo que indica que son muestras influenciables

El modelo seleccionado con 9 PCs explica, a partir de la señal NIR en las regiones mostradas en la tabla 12, más del 99% en la variabilidad de los datos de número de bromo (figura 18a). El error estándar en la etapa de calibración y validación cruzada disminuye apreciablemente hasta el noveno componente donde alcanza valores de 0.74 % y 3.68% respectivamente (figura 18b). De los coeficientes de regresión calculados para el PC6, que explica la mayor variabilidad en los datos (25%), se determinó que la región de 4100 a 4300  $\text{cm}^{-1}$  presenta un efecto positivo para la predicción del número de bromo (figura 18c). La validación cruzada del modelo mostró un desempeño favorable para la predicción de naftas livianas, ya que el coeficiente entre los valores de referencia y de predicción fue hallado superior al 0.99 para la curva de calibración y superior a 0.95 para la curva de validación (figura 18d).

**Figura 18. Descripción gráfica del modelo con 9 componentes principales para la predicción de naftas livianas**



Como se observa en la tabla 13 en la etapa de validación, para 1 de las 13 muestras evaluadas se obtuvo un porcentaje de error relativo superior al 1%, siendo 1,24% el mayor porcentaje de error, que lo presenta la muestra 27.

**Tabla 13. Validación del modelo de predicción de naftas livianas**

Muestras	Referencia (NB)	Predicho (NB)	Residual	Error relativo (%)
Nafta 1	118,78	119,62	0,84	0,71
Nafta 3	131,25	130,13	-1,12	0,85
Nafta 4	113,41	114,17	0,76	0,67
Nafta 5	89,16	88,99	-0,17	0,19
Nafta 13	112,45	111,47	-0,98	0,87
Nafta 14	127,65	127,49	-0,16	0,12
Nafta 15	120,59	120,58	-0,01	0,008
Nafta 16	100,94	100,37	-0,57	0,56
Nafta 27	124,99	126,55	1,56	1,24
Nafta 28	115,17	114,43	-0,74	0,64
Nafta 29	90,77	90,69	-0,08	0,088
Nafta 34	90,17	90,29	0,12	0,13
Nafta 35	82,06	82,60	0,54	0,65

### 3.4. RESULTADOS DEL MÉTODO RPM

La elaboración del modelo de predicción a través de esta técnica se lleva a cabo desarrollando un código que realice un mapeo entre los componentes principales. Se plantean polinomios de diferentes grados y se busca determinar el valor de los coeficientes de la ecuación 16.

Se utilizaron los tres (3) primeros componentes principales que son lo que explican la mayor varianza del espectro (95 %). De las treinta y nueve (39) muestras se eliminaron tres muestras para realizar el modelo, las muestras 24, 25 y 27, debido a que por prueba y error estas muestras afectan significativamente tanto el error de calibración como el de validación. Para estas muestras se calculó el RMSEC, el RMSECV y los coeficientes de correlación. Los polinomios utilizados junto con sus tamaños y grados se presentan en la siguiente tabla.

**Tabla 14. Estructura de los polinomios empleados en el método de RPM**

No. Del método	Polinomio	No de términos	Grado
1	[PC1 PC2 PC3]	3	1
2	[PC1 PC2 PC3 1]	4	1
3	[PC1 PC2 PC3 PC1PC2PC3 1]	5	1
4	[PC1 PC2 PC3 PC1PC2 PC1PC3 PC2PC3 PC1 <sup>2</sup> PC <sup>2</sup> PC3 <sup>2</sup> 1]	10	2
5	[PC1 PC2 PC3 PC1PC2 PC1PC3 PC2PC3 PC1 <sup>2</sup> PC <sup>2</sup> PC3 <sup>2</sup> PC1PC2PC3 PC1 <sup>2</sup> PC2 PC2 <sup>2</sup> PC3 PC3 <sup>2</sup> PC1 PC1 <sup>2</sup> PC3 PC2 <sup>2</sup> PC1 PC3 <sup>2</sup> PC2 PC1 <sup>3</sup> PC2 <sup>3</sup> PC3 <sup>3</sup> 1]	20	3
6	[PC1 PC2 PC3 PC1PC2 PC1PC3 PC2PC3 PC1 <sup>2</sup> PC <sup>2</sup> PC3 <sup>2</sup> PC1PC2PC3 PC1 <sup>2</sup> PC2 PC2 <sup>2</sup> PC3 PC3 <sup>2</sup> PC1 PC1 <sup>2</sup> PC3 PC2 <sup>2</sup> PC1 PC3 <sup>2</sup> PC2 PC1 <sup>3</sup> PC2 <sup>3</sup> PC3 <sup>3</sup> PC1 <sup>3</sup> PC2 PC1 <sup>3</sup> PC3 PC2 <sup>3</sup> PC1 PC2 <sup>3</sup> PC3 PC3 <sup>3</sup> PC1 PC3 <sup>3</sup> PC2 PC1 <sup>2</sup> PC2PC3 PC1PC2 <sup>2</sup> PC3 PC1PC2PC3 <sup>2</sup> PC1 <sup>2</sup> PC2 <sup>2</sup> PC1 <sup>2</sup> PC3 <sup>2</sup> PC2 <sup>2</sup> PC3 <sup>2</sup> PC1 <sup>4</sup> PC2 <sup>4</sup> PC3 <sup>4</sup> 1]	35	4

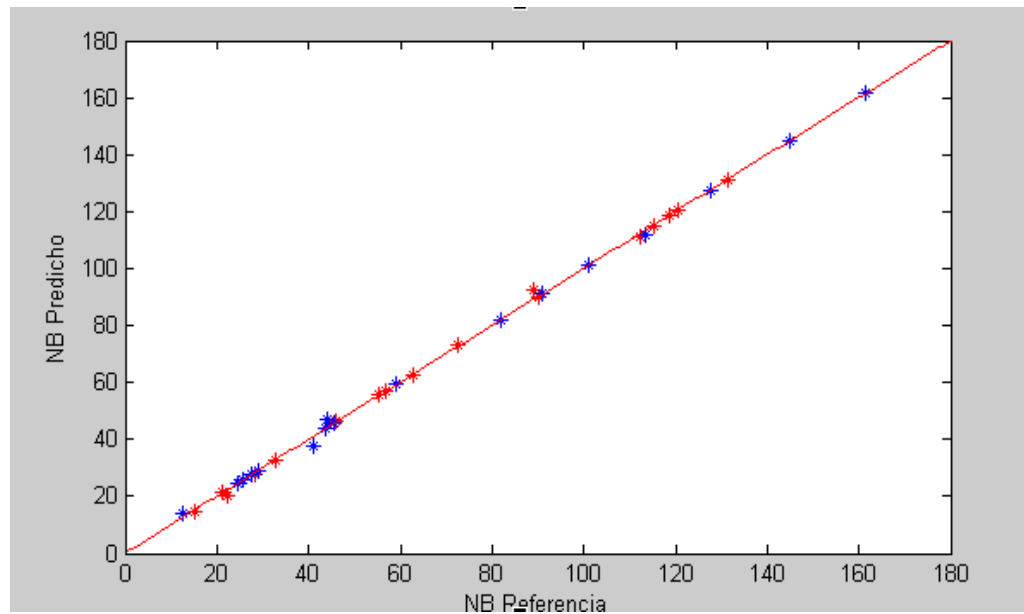
Los resultados de los cálculos del error de calibración y del error de validación se muestran a continuación.

**Tabla 15. Resultados del método de RPM**

No de términos del polinomio	RMSEC	RMSECV
3	77.44	78.56
4	40.75	42.24
5	40.16	43.04
10	39.06	41.17
20	28.9	33.82
35	1.07	1.22

El coeficiente de correlación entre los valores de referencia y de predicción fue superior al 0.99 tanto para la curva de calibración como para la curva de validación (figura 19).

**Figura 19. Curvas de calibración y validación del modelo de predicción del número de bromo empleando el método de RPM**



Como se observa en la tabla 16 en la etapa de validación, para 4 de las 37 muestras evaluadas se obtuvo un porcentaje de error relativo superior al 5%, siendo 11,6% el mayor porcentaje de error, que lo presenta la muestra 39, lo cual puede ser debido a que esta muestra presenta el menor valor del rango de calibración.

**Tabla 16. Validación del modelo de predicción con RPM**

<b>Muestras</b>	<b>Referencia (NB)</b>	<b>Predicho (NB)</b>	<b>Residual</b>	<b>Error relativo (%)</b>
Nafta 1	118,78	118,76	0,02	0,02
Nafta 2	161,48	161,48	0	0
Nafta 3	131,25	131,26	-0,01	0,01
Nafta 4	113,41	111,69	1,72	1,51
Nafta 5	89,16	92,24	-3,08	3,45
Nafta 6	59,19	59,48	-0,29	0,49
Nafta 7	46,08	46,11	-0,03	0,06
Nafta 8	45,53	45,47	0,06	0,13
Nafta 9	32,73	33,66	-0,93	2,84
Nafta 10	28,89	28,85	0,04	0,13
Nafta 11	24,42	24,76	-0,34	1,39
Nafta 12	44,01	47,11	-3,1	7,04
Nafta 13	112,45	111,18	1,27	1,13
Nafta 14	127,65	127,63	0,02	0,01
Nafta 15	120,59	120,64	-0,05	0,04
Nafta 16	100,94	100,94	0	0
Nafta 17	55,43	55,44	-0,01	0,02
Nafta 18	27,57	27,5	0,07	0,25
Nafta 19	72,73	73,14	-0,41	0,56
Nafta 20	41,23	37,78	3,45	8,37
Nafta 21	28,29	27,81	0,48	1,69
Nafta 22	24,52	24,46	0,06	0,24
Nafta 23	22,19	20,54	1,65	7,43
Nafta 26	145,04	144,92	0,12	0,08
Nafta 28	115,17	115,19	-0,02	0,02
Nafta 29	90,77	91,02	-0,25	0,28
Nafta 30	56,95	56,85	0,1	0,18
Nafta 31	24,67	24,91	-0,24	0,97
Nafta 32	15,07	14,7	0,37	2,46
Nafta 33	25,7	25,72	-0,02	0,08
Nafta 34	90,17	90,18	-0,01	0,01
Nafta 35	82,06	82,21	-0,15	0,18
Nafta 36	62,73	62,51	0,22	0,35
Nafta 37	43,56	43,69	-0,13	0,29
Nafta 38	21,34	21,5	-0,16	0,75
Nafta 39	12,5	13,95	-1,45	11,6

#### 4. CONCLUSIONES

La espectroscopia FT-NIR en combinación con técnicas de calibración multivariable como PLS y RPM se ha mostrado como una técnica alternativa, a los métodos tradicionales para la determinación del número de bromo en naftas; con la ventaja de evitar el pretratamiento de la muestra.

El análisis por componentes principales permitió identificar las agrupaciones naturales de los espectros de las muestras y el número de componentes principales optimo a incluir en el modelo de predicción; además se determino que las regiones del espectro de mayor importancia se encuentran en los rangos de  $4000-4300\text{ cm}^{-1}$ ,  $4500\text{ a }4700\text{ cm}^{-1}$  y  $5800-6000\text{ cm}^{-1}$ .

Tanto la técnica de regresión parcial por mínimos cuadrados (PLS) como la regresión polinomial múltiple (RPM) aplicadas sobre el espectro FT-NIR mostraron un desempeño satisfactorio para predecir el número de bromo en las naftas, demostrando que los datos espectroscópicos pueden proporcionar información detallada de esta propiedad.

Los modelos obtenidos por regresión PLS tuvieron que ser ajustados por eliminación de un número alto de outliers debido a que los datos de referencia del número de bromo no presentaban una relación lineal, mientras que para el modelo obtenido por regresión polinomial múltiple (RPM) sólo hubo la necesidad de eliminar tres muestras debido a que este método utiliza relaciones no-lineales.

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.

Los modelos desarrollados así como su aplicación, deben hacerse en muestras de la misma naturaleza, en el caso de la regresión PLS, para muestras pesadas y livianas.

Durante el proceso de elaboración del modelo de predicción con el método de RPM, mientras mayor es el número de términos mejor es el resultado obtenido.

## 5. RECOMENDACIONES

Los resultados del estudio anteriormente presentado muestran el gran potencial de la espectroscopia NIR y técnicas multivariantes para la determinación del número de bromo en naftas colombianas. Aunque los resultados de validación cruzada muestran un buen desempeño de los modelos desarrollados, se recomienda que éstos sean evaluados y alimentados con nuevas muestras para hacerlos más robustos y asegurar que siempre estarán actualizados.

## 6. REFERENCIAS BIBLIOGRÁFICAS

1. SASTRY, S; CHOPRA, A; SARPAL, S; JAIN, K; SRIVASTAVA, P and BHATNAGAR, K. Determination of physicochemical properties and carbón-type analysis of base oils using Mid-IR spectroscopy and partial least-squares regression analysis. *Energy & Fuels* 1998. Vol 12. 304-11 p.
2. <http://www.iae.org.ar/archivos/educ1205.pdf> [citado el 29 de septiembre del 2011]
3. PARISOTTO, G., FERRAO, M. Total Acid Number Determination in Residues of Crude Oil Distillation Using ATR-FTIR and Variable Selection by Chemometric. *Energy & Fuels* 2010, 24, 5474-5478 p.
4. SPEIGHT, J. G. The chemistry and technology of petroleum. Cuarta edición: Laramie: 2006. Chemical Industries Series. 916 p.
5. ALTGELT, K.H. and BODUSZYNSKI, M. M. Composition and analysis of heavy petroleum fractions. Editorial Marcel Dekker. Segunda Edición. New York. 1994. 495 p.
6. WAQUIER, J.P. El refino del petróleo. Petróleo crudo, productos petrolíferos, esquemas de fabricación. EN: Composición de los petróleos crudos y de los productos petrolíferos. España.: Editorial Díaz de Santos 2004. 1-15 p.
7. SPEIGHT, J. G. The chemistry and technology of petroleum. Cuarta Edición. Laramie: CRC, 2006. 916 p.
8. ANTOS, G. J. & AITANI, A. M. Catalytic naphtha reforming. Segunda edición. New York: CRC Press, 1994. 528 p.

9. SPEIGHT, J. G. Environmental analysis and technology for the refining industry. Laramie: John Wiley & Sons, 2005. 349 p.
10. [http://petro.metrohm.com/pdf\\_Download/Prosp\\_Petrochem\\_Analytik\\_ES\\_w eb.pdf](http://petro.metrohm.com/pdf_Download/Prosp_Petrochem_Analytik_ES_w eb.pdf) [citado el 21 de septiembre del 2011]
11. AMERICAN SOCIETY FOR TESTING AND MATERIALS. Standard Test Method for Bromine Numbers of Petroleum Distillates and Commercial Aliphatic Olefins by Electrometric Titration. Philadelphia: ASTM, 2005.
12. DAVIES, T. The history of near infrared spectroscopic analysis: past, present and future "From sleeping technique to the morning star of spectroscopy". Analisis Magazine. Vol 26, N°4 (1998). M17-M19 p.
13. CIURCZAK, E. W. Handbook of near-infrared analysis. New York. Taylor & Francis Group, 2001.
14. PASQUINI, C. Near Infrared spectroscopy fundamentals, practical aspects and analytical applications. EN: J. Braz. Chem Soc. 2003. Vol 14, N°2. 198-219 p.
15. SACORAGUE, L. A. Avaliação de diferentes regiões do espectro do infravermelho próximo na determinação de parâmetros de qualidade de combustíveis empregando ferramentas quimiométricas. Campinas, 2004. Tese de Doutorado: Instituto de Química, Universidade Estadual de Campinas.
16. BARTON, F. E. Theory and principles of near infrared spectroscopy Europe. Vol 14, N°1 (2002). 12-18 p.
17. PATNAIK, P. Dean's Analytical Chemistry Handbook. Segunda edición. New York: McGraw-Hill Professional, 2004. 1280 p.

18. GEMPERLINE P. J. Introduction to Chemometrics. Practical Guide to Chemometrics,. Second Edition. 2006.
19. GELADI, P. Some recent trends in the calibration literatura. Chemometrics and Intelligent Laboratory Systems. Vol. 60 (2002). 211-224 p.
20. MARTENS, H., NAES, T. Multivariate calibration. New York. Jhon Wiley & Sons, 1989. 438 p.
21. AMERICAN SOCIETY FOR TESTING AND MATERIALS. Standard practices for infrared multivariate quantitative analysis. Philadelphia: ASTM, 2005.
22. MILLER, J.N., MILLER, J.C. Estadística y Quimiometría para Química Analítica. Cuarta Edición. Prentice Hall. 2002. 595 p.
23. CHEUNG, V. "A comparative study of the characterization of colour cameras by means of neural networks and polynomial transforms". Color. Technol. 120. 2004. 19-25 p.
24. RYCZKOWSKI Janusz. IR spectroscopy in catalysis. Catalysis Today 68 (2001) 263–381p.
25. BUENO Alexis. Caracterización de la planta de Aromáticos de la GCB Mediante Espectroscopía NIR y Métodos Quimiométrico de Análisis. Bucaramanga 2008. Tesis de Maestría. Universidad industrial de Santander. Facultad de ciencias. Escuela de Química.
26. MONROY L. Predicción de la densidad en Asfáltenos Colombianos utilizando espectroscopia fotoacústica en la región del infrarrojo medio y métodos quimiométricos. Bucaramanga 2010. Tesis de grado (químico). Universidad Industrial de Santander. Facultad de ciencias. Escuela de Química.

27. SANCHEZ M, MOLINA D, CABANZO R, MEJIA OSPINO E. Determinación de la Composición Elemental de Fondos de Vacío de Crudos Colombianos por Espectroscopía de Plasma Inducido por Láser (LIBS), rev. col. fís., vol. 40, No. 2, (2008).
28. LAMUS C. Uso de análisis multivariado en La determinación SARA de crudos mediante la espectroscopia NIR. Bucaramanga 2010. Tesis de grado (Química). Universidad Industrial de Santander. Facultad de ciencias. Escuela de Química.
29. LACHE, A., MELENDEZ, V. Predicción del análisis SARA de crudos colombianos aplicando espectroscopía FTIR-ATR y métodos quimiométricos. Bucaramanga 2010. Tesis de grado (Química). Universidad industrial de Santander. Facultad de ciencias. Escuela de Química.
30. BALLESTEROS, Sandra. Predicción de las Fracciones SARA de Fondos de vacío de Crudos Colombianos, por Métodos Quimiométricos Utilizando Espectroscopia de Fluorescencia Inducida por Láser (LIF). Bucaramanga 2010. Tesis de grado (Química). Universidad industrial de Santander. Facultad de ciencias. Escuela de Química.
31. VARGAS, Gloria. Predicción del análisis SARA de fondos de vacío colombianos utilizando espectroscopia fotoacústica en la región del infrarrojo medio y métodos quimiométricos. Bucaramanga 2011. Tesis de grado (Química). Universidad industrial de Santander. Facultad de ciencias. Escuela de Química.
32. WORKMAN, J., WEYER, L. Practical Guide to Interpretive Near-Infrared Spectroscopy. CRC Press. 2008. 318 p.
33. BUENO, A. Caracterización de la planta de Aromáticos de la GCB Mediante Espectroscopia NIR y Métodos Quimiométrico de Análisis. Bucaramanga

2008. Tesis de Maestría. Universidad industrial de Santander. Facultad de ciencias. Escuela de Química.

## Anexo A

### Calculo de los componentes principales

```
clear all %borrar todos los datos
clc % (limpia la pantalla)
close all
%%% Llamado de los archivos en excel

%%cd C:\Users\Seven\Dropbox\Proyecto_Carbones
cd C:\Users\Gallo\Dropbox\Proyecto_Carbones
esp_cal=xlsread('Datos.xlsx','Hojal','A1:CPQ36');
ss=esp_cal;
[NF NC]=size(ss);
mes=mean(ss,1);ms=repmat(mes,NF,1);
[coef_cal score_cal]=princomp(ss,'econ');
%%% coef, vectores propios
%%% score, PCs
reconstruidos=coef_cal*score_cal'+ms';

% % for i=1:NF
% %     drawnow
% %     plot(reconstruidos(:,i),'b'); hold on
% %     pause
% %     plot(ss(i,:), 'k');pause
% % end

SRMS=mean(sqrt(sum((reconstruidos'-ss).*(reconstruidos'-ss))/length(ss)))
xlswrite('pcs.xlsx',score_cal,'PCs_cal');
xlswrite('pcs.xlsx',coef_cal,'VecP');
xlswrite('pcs.xlsx',mes,'MediaEsp');
```

## Anexo B

### Calculo de la matriz de coeficientes de regresión

```
clc
close all
clear all

pc=xlsread('pcs.xlsx','PCs_cal');
A=pc(:,1:3);   %%% Grupo de calibración
C=pc(2:2:36,1:3); %%% Grupo de validación
NB=xlsread('pcs.xlsx','NB_cal');
NBcal=NB(:,1);
NBval=NB(2:2:36,1);

% % A1=[A(:,1) A(:,2) A(:,3)]; % Polinomio de grado1, 3 términos

% % A2=[A(:,1) A(:,2) A(:,3) ones(length(A),1)]; % Polinomio de grado1,
4 términos

% % A3=[A(:,1) A(:,2) A(:,3) A(:,1).*A(:,2).*A(:,3) ones(length(A),1)];
%% Polinomio de grado 1, 5 términos

% % A4=[A(:,1) A(:,2) A(:,3) A(:,1).*A(:,2) A(:,1).*A(:,3) A(:,2).*A(:,3)
A(:,1).*A(:,1),...
% % A(:,2).*A(:,2) A(:,3).*A(:,3) ones(length(A),1)]; % Polinomio
de grado 2, 10 términos

% % A5=[A(:,1) A(:,2) A(:,3) A(:,1).*A(:,2) A(:,1).*A(:,3) A(:,2).*A(:,3)
A(:,1).*A(:,1),...
% % A(:,2).*A(:,2) A(:,3).*A(:,3) A(:,1).*A(:,2).*A(:,3)
A(:,1).*A(:,1).*A(:,2) A(:,2).*A(:,2).*A(:,3),...
% % A(:,3).*A(:,3).*A(:,1) A(:,1).*A(:,1).*A(:,3)
A(:,2).*A(:,2).*A(:,1) A(:,3).*A(:,3).*A(:,2) A(:,1).*A(:,1).*A(:,1),...
% % A(:,2).*A(:,2).*A(:,2) A(:,3).*A(:,3).*A(:,3) ones(length(A),1)];
%%Polinomio de grado 3, 20 términos

A6=[A(:,1) A(:,2) A(:,3) A(:,1).*A(:,2) A(:,1).*A(:,3) A(:,2).*A(:,3)
A(:,1).*A(:,1),...
A(:,2).*A(:,2) A(:,3).*A(:,3) A(:,1).*A(:,2).*A(:,3)
A(:,1).*A(:,1).*A(:,2), ...
A(:,2).*A(:,2).*A(:,3) A(:,3).*A(:,3).*A(:,1) A(:,1).*A(:,1).*A(:,3),
...
A(:,2).*A(:,2).*A(:,1) A(:,3).*A(:,3).*A(:,2) A(:,1).*A(:,1).*A(:,1),
...
A(:,2).*A(:,2).*A(:,2) A(:,3).*A(:,3).*A(:,3)
A(:,1).*A(:,1).*A(:,1).*A(:,2), ...
```

```

    A(:,1).*A(:,1).*A(:,1).*A(:,3) A(:,2).*A(:,2).*A(:,2).*A(:,1)
A(:,2).*A(:,2).*A(:,2).*A(:,3), ...
    A(:,3).*A(:,3).*A(:,3).*A(:,1) A(:,3).*A(:,3).*A(:,3).*A(:,2)
A(:,1).*A(:,1).*A(:,2).*A(:,3), ...
    A(:,1).*A(:,2).*A(:,2).*A(:,3) A(:,1).*A(:,2).*A(:,3).*A(:,3)
A(:,1).*A(:,1).*A(:,2).*A(:,2), ...
    A(:,1).*A(:,1).*A(:,3).*A(:,3) A(:,2).*A(:,2).*A(:,3).*A(:,3)
A(:,1).*A(:,1).*A(:,1).*A(:,1), ...
    A(:,2).*A(:,2).*A(:,2).*A(:,2) A(:,3).*A(:,3).*A(:,3).*A(:,3)
ones(length(A),1)]; %% Polinomio de grado 4, 35 términos

```

```

B=A6;
t=pinv(B)*NBcal;   %% coeficientes

```

```

%%Validación del modelo

```

```

C1=[C(:,1) C(:,2) C(:,3)]; %% Polinomio de grado1, 3 términos

```

```

% % C2=[C(:,1) C(:,2) C(:,3) ones(length(C),1)]; %% Polinomio de grado1,
4 términos

```

```

% % C3=[C(:,1) C(:,2) C(:,3) C(:,1).*C(:,2).*C(:,3) ones(length(C),1)];
%% Polinomio de grado 1, 5 términos

```

```

% % C4=[C(:,1) C(:,2) C(:,3) C(:,1).*C(:,2) C(:,1).*C(:,3) C(:,2).*C(:,3)
C(:,1).*C(:,1), ...
% % C(:,2).*C(:,2) C(:,3).*C(:,3) ones(length(C),1)]; %% Polinomio
de grado 2, 10 términos

```

```

% % C5=[C(:,1) C(:,2) C(:,3) C(:,1).*C(:,2) C(:,1).*C(:,3) C(:,2).*C(:,3)
C(:,1).*C(:,1), ...
% % C(:,2).*C(:,2) C(:,3).*C(:,3) C(:,1).*C(:,2).*C(:,3)
C(:,1).*C(:,1).*C(:,2) C(:,2).*C(:,2).*C(:,3), ...
% % C(:,3).*C(:,3).*C(:,1) C(:,1).*C(:,1).*C(:,3)
C(:,2).*C(:,2).*C(:,1) C(:,3).*C(:,3).*C(:,2) C(:,1).*C(:,1).*C(:,1), ...
% % C(:,2).*C(:,2).*C(:,2) C(:,3).*C(:,3).*C(:,3) ones(length(C),1)];
%%Polinomio de grado 3, 20 términos

```

```

C6=[C(:,1) C(:,2) C(:,3) C(:,1).*C(:,2) C(:,1).*C(:,3) C(:,2).*C(:,3)
C(:,1).*C(:,1), ...
    C(:,2).*C(:,2) C(:,3).*C(:,3) C(:,1).*C(:,2).*C(:,3)
C(:,1).*C(:,1).*C(:,2), ...
    C(:,2).*C(:,2).*C(:,3) C(:,3).*C(:,3).*C(:,1) C(:,1).*C(:,1).*C(:,3),
...
    C(:,2).*C(:,2).*C(:,1) C(:,3).*C(:,3).*C(:,2) C(:,1).*C(:,1).*C(:,1),
...
    C(:,2).*C(:,2).*C(:,2) C(:,3).*C(:,3).*C(:,3)
C(:,1).*C(:,1).*C(:,1).*C(:,2), ...

```

```

        C(:,1).*C(:,1).*C(:,1).*C(:,3) C(:,2).*C(:,2).*C(:,2).*C(:,1)
C(:,2).*C(:,2).*C(:,2).*C(:,3), ...
        C(:,3).*C(:,3).*C(:,3).*C(:,1) C(:,3).*C(:,3).*C(:,3).*C(:,2)
C(:,1).*C(:,1).*C(:,2).*C(:,3), ...
        C(:,1).*C(:,2).*C(:,2).*C(:,3) C(:,1).*C(:,2).*C(:,3).*C(:,3)
C(:,1).*C(:,1).*C(:,2).*C(:,2), ...
        C(:,1).*C(:,1).*C(:,3).*C(:,3) C(:,2).*C(:,2).*C(:,3).*C(:,3)
C(:,1).*C(:,1).*C(:,1).*C(:,1), ...
        C(:,2).*C(:,2).*C(:,2).*C(:,2) C(:,3).*C(:,3).*C(:,3).*C(:,3)
ones(length(C),1)]; %% Polinomio de grado 4, 35 términos

```

```

TV=C6*t;
TC=A6*t;
figure(1)
plot(NBcal,TC,'*r',NBval,TV,'*b')
% % figure(2)
% % plot(NBval,TV,'*b')
% % plot(TC,NBcal,'*r')
% % plot(TV,NBval,'*r')

% % p=polyfit(NBcal,TC,1)
% % pv=polyfit(NBval,TV,1)
% % R=corrcoef(NBcal,TC)
% % RV=corrcoef(NBval,TV)

```

```

Nc=length(TC);
Nv=length(TV);

```

```

PRESS=sum((TC-NBcal).^2); %Predictive residual sums of squares
SECPRESS=sqrt(PRESS/Nc); %Standar error of calibration

```

```

PRESSV=sum((TV-NBval).^2); %Predictive residual sums of squares
SEPPRESS=sqrt(PRESSV/Nv); %Standar error of prediction

```

```

metricas=[SECPRESS SEPPRESS]

```