

**Desarrollo de un modelo quimiométrico con aplicación de tecnología NIR para la
predicción de la calidad de crudos en el área de mezclado de la refinería de
Barrancabermeja**

Jholvis Sosa Gutiérrez

Trabajo de grado para la Maestría de Aplicación en Ingeniería Química

Director

Víctor Gabriel Baldovino Medrano

Ph.D. en Ingeniería biológica, agronómica y ambiental

Universidad Industrial de Santander

Facultad de Ingenierías Físicoquímicas

Escuela de Ingeniería Química

Maestría en Ingeniería Química

Bucaramanga

2018

Contenido

	Pág.
Introducción	13
1. Marco teórico conceptual y estado del arte.....	15
1.1 Tratamiento del crudo en la refinería de Barrancabermeja.....	15
1.2 Principios de la espectroscopía infrarroja	21
1.3 Definición de quimiometría	30
1.3.1 Análisis en componentes principales (PCA).....	32
1.3.2 Regresión por mínimos cuadrados parciales (PLS)	34
1.3.2.1 Etapas del desarrollo de un modelo de calibración.....	40
1.4 Estado del arte.....	43
2. Metodología	45
2.1 Logística de muestreo	46
2.1.2 Ejecución de relaciones de mezcla	46
2.1.3 Toma de muestras	46
2.2 Caracterización de muestras	46
2.3 Análisis estadístico de los espectros medidos.....	47
2.4 Análisis por componentes principales (PCA).....	47
2.5 Construcción del modelo de predicción con regresión por mínimos cuadrados parciales	48
2.5.1 Calibración.....	48

2.5.2 Validación	48
3. Análisis de Resultados	49
3.1 Caracterización de muestras y adquisición de espectros	49
3.1.1 Medidas espectroscópicas	51
3.2 Análisis estadístico de los espectros medidos.....	53
3.2.1 Región para la predicción de °API	54
3.2.2 Región para la predicción del %p/p de azufre	54
3.3 Análisis por componentes principales (PCA) en la región de predicción de la gravedad API	55
3.4 Modelo de predicción (R-PLS) para la gravedad API.....	59
3.4.1 Análisis de muestra excluidas	60
3.5 análisis por componentes principales (PCA) en la región de predicción del %P/P de azufre.....	64
3.6 Modelo de predicción (R-PLS) del %p/p de azufre.....	66
3.6.1 Análisis de muestra excluidas	67
4. Conclusiones	71
5. Recomendaciones	72
Referencias Bibliográficas	73
Apéndice	81

Lista de Tablas

	Pág.
Tabla 1. Relación de crudos analizados con valores de gravedad API y %p/p de Azufre	49
Tabla 2. Varianza asociada a cada uno de los componentes principales estudiados	56
Tabla 3. Parámetros estadísticos del modelo de predicción de la gravedad API en muestras de mezclas de crudos	62
Tabla 4. Validación cruzada del modelo de predicción de la gravedad API	63
Tabla 5. Varianza de los componentes principales PCA	65
Tabla 6. Parámetros estadísticos del modelo de predicción del %p/p de azufre en muestras de mezclas de crudos	69
Tabla 7. Validación cruzada del modelo de predicción del %p/p de azufre	70

Lista de Figuras

	Pág.
Figura 1. Diagrama general simplificado de la Refinería de Barrancabermeja.	16
Figura 2. Área de mezclado de crudo CB8 de la Refinería de Barrancabermeja	17
Figura 3. Componentes de la radiación electromagnética	21
Figura 4. Espectro Electromagnético	23
Figura 5. Componentes principales de un espectrofotómetro NIR	26
Figura 6. Ejemplo de un espectro de absorbancia.....	28
Figura 7. Ejemplo de un espectro de Transmitancia.....	28
Figura 8. Bandas de absorción en la región NIR	29
Figura 9. Representación gráfica de la matriz de datos espectrales.....	33
Figura 10. Descripción gráfica del modelo R-PLS	37
Figura 11. Gráfico de la varianza residual frente al leverage	42
Figura 12. Esquema metodológico experimental.....	45
Figura 13. Región espectral NIR sin tratamiento para la predicción de la gravedad API	52
Figura 14. Región espectral NIR sin tratamiento para la predicción del %p/p de azufre.....	53
Figura 15. Región espectral NIR con pretratamiento para predecir la gravedad API	54
Figura 16. Región espectral NIR con tratamiento para la predicción del %p/p de azufre.....	55
Figura 17. Gráfica de puntuación para PC1 y PC2 de las mezclas en la región de 4000 a 4500 cm ⁻¹	57

Figura 18. Gráfica de Influencia en PCA (región de 4000 a 4500 cm^{-1})	58
Figura 19. Varianza explicada en el modelo PLS para predicción de gravedad API	59
Figura 20. Error calculado en función del número de VL para el modelo de la gravedad API....	60
Figura 21. Gráfico del estadístico de Hotelling T^2 para la gravedad API.....	61
Figura 22. Curva de calibración y validación del modelo de predicción de la gravedad API.....	62
Figura 23. Gráfica de puntuación para PC1 y PC2 de las mezclas en región 3400 a 3500 cm^{-1} ..	66
Figura 24. Varianza explicada en el modelo PLS para predicción del %p/p de azufre	67
Figura 25. Gráfico del estadístico de Hotelling T^2 para el %p/p de azufre	68
Figura 26. Curva de calibración y validación del modelo de predicción del %p/p de azufre	69

Lista de Apéndices

	Pág.
Apéndice A. Estadística Unidades de Mezcla	81
Apéndice B. Identificación según Anexo 8 del procedimiento CID-CID-P-001 (ver Pág. 25)....	84
Apéndice C. Región espectral total del conjunto de mezclas de crudo con selección de rangos espectrales	85
Apéndice D. Gráfico de influencia en PLS en el modelo de predicción de la gravedad API.....	86
Apéndice E. Gráfico de influencia en PLS en el modelo de predicción del %p/p de azufre.....	87

Acrónimos y Abreviaturas

ASTM:	American Society for Testing and Materials (Sociedad Americana de Pruebas y Materiales)
GRB:	Gerencia Refinería de Barrancabermeja
NIR:	Near Infrared (Infrarrojo Cercano)
NIRS:	Near Infrared Spectroscopy (Espectroscopia de Infrarrojo Cercano)
PC:	Principal Component (Componentes principales)
PCA:	Principal Component Analysis (Análisis de Componentes principales)
PLS:	Partial Least Squares (Mínimos Cuadrados Parciales)
R²:	Coefficiente de determinación
RMSEC:	Root Mean Square Error of Calibration (Raíz Cuadrada Media del Error de Calibración)
RMSEP:	Root Mean Square Error of Prediction (Raíz Cuadrada Media del Error de Predicción)
RMSEV:	Root Mean Square Error of Validation (Raíz Cuadrada Media del Error de Validación)
SEC:	Standard Error of Calibration (Error Estándar de Calibración)
SECV:	Standard Error of Cross Validation (Error Estándar de Validación)
SEP:	Standard Error of Prediction (Error Estándar de Predicción)

Resumen

Título: Desarrollo de un modelo quimiométrico con aplicación de tecnología NIR para la predicción de la calidad de crudos en el área de mezclado de la refinería de Barrancabermeja *

Autor: Jholvis Sosa Gutiérrez **

Palabras claves: Crudo LCT, espectroscopia de infrarrojo NIR, quimiometría, PCA, PLS

Descripción:

En el presente trabajo se desarrolló un modelo de predicción para mezclas de crudos con gravedad API entre 21,10 a 25,50 y con porcentaje en peso de azufre entre 0,96 y 1,27 del área de mezclado de la refinería de Barrancabermeja, mediante espectroscopia en la región del infrarrojo cercano NIR.

La caracterización de las cuarenta (40) muestras de mezclas de crudo según métodos estandarizados fue suministrada por el sistema de información del laboratorio de la refinería; la información espectral permitió clasificar el grupo de muestras mediante análisis por componentes principales (PCA) y se desarrollaron dos modelos predictivos por regresión por mínimos cuadrados parciales (PLS). Con el uso del software THE UNSCRAMBLER se aplicaron diferentes pretratamientos para el rango espectral de 3400 a 3500 cm^{-1} y de 4000 a 4600 cm^{-1} .

La descomposición de los datos de infrarrojo por componentes principales permitió identificar las muestras analizadas de acuerdo con su composición porcentual similar en cada mezcla. Se conformaron tres grupos de muestras de acuerdo a las tendencias observadas en: livianas, medias y tipo LCT.

El método de regresión PLS aplicado a los espectros NIR mostró un desempeño satisfactorio en cada modelo de predicción, el valor RMSECV para la gravedad API es de 0.258 y para el %p/p de azufre de 0.026, valores aceptables en comparación a la reproducibilidad del método de ensayo en el laboratorio de la refinería.

Estos modelos de predicción pueden ser una alternativa a las técnicas convencionales empleadas comúnmente e incrementar la reproducibilidad de los métodos con disminución de costos y en un menor tiempo de análisis.

* Trabajo de Grado

** Facultad de Ingeniería Físicoquímicas. Escuela de Ingeniería Química. Maestría en Ingeniería Química Director: Víctor Gabriel Baldovino Medrano.

Abstract

Title: Development of a chemometric model with application of nir technology for the prediction of raw quality in the mixing area of the Barrancabermeja refinery*

Author: Jholvis Sosa Gutiérrez**

Keywords: Crude LCT, infrared spectroscopy of NIR, chemometry, PCA, PLS

Description:

In the present work, a prediction model was developed for crude mixtures with API gravity between 21.10 to 25.50 and with weight percentage of sulfur between 0.96 and 1.27 of the mixing area of the Barrancabermeja refinery, by means of spectroscopy in the near infrared NIR region.

The characterization of the forty (40) samples of crude mixtures according to standardized methods was provided by the information system of the refinery laboratory; the spectral information allowed to classify the group of samples by principal component analysis (PCA) and two predictive models were developed for partial least squares regression (PLS). With the use of software THE UNSCRAMBLER different pre-treatments were applied for the spectral range from 3400 to 3500 cm⁻¹ and from 4000 to 4600 cm⁻¹.

The decomposition of the infrared data by main components allowed to identify the samples analyzed according to their similar percentage composition in each mixture. Three groups of samples were formed according to the trends observed in: light, medium and LCT type.

The PLS regression method applied to the NIR spectra showed a satisfactory performance in each prediction model, the RMSECV value for the API gravity is 0.258 and for the % p / p of sulfur of 0.026, acceptable values in comparison to the reproducibility of the test method in the refinery laboratory.

These prediction models can be an alternative to the conventional techniques commonly used and increase the reproducibility of the methods with lower costs and in a shorter analysis time.

* Project of grade

** Faculty of Physicochemical Engineering. School of Chemical Engineering. Magister of Chemical Engineering
Directress: Víctor Gabriel Baldovino Medrano

Introducción

La Refinería de Ecopetrol en Barrancabermeja es el principal centro de recibo de crudos en Colombia, con una infraestructura que se desarrolló e instaló a finales de los años 70 e inicio de los años 80. La Refinería a través de su historia ha sufrido muchos cambios en su estructura orgánica y de infraestructura (Ecopetrol S.A., 2014). Desde 1921, cuando la Tropical Oil Company inició la actividad petrolera en Colombia (Ecopetrol S.A, 2014) instalando unos alambiques para destilar el crudo con una capacidad de procesar 1.500 barriles, que fueron traídos de Talara, Perú. (Vásquez C, 1994), aumentando progresivamente a medida que entraban en operación nuevos procesos, llegando, hasta el 2007, a una capacidad de destilar 250.000 barriles por día (Ecopetrol S.A., 2009).

Casa de Bombas 8 (CB8) es la sección encargada de recibir los diferentes tipos de crudos que llegan a la Refinería de Barrancabermeja. CB8 se encarga de segregar, almacenar, mezclar y entregar por medio de bombas y un área de mezclado los diferentes crudos mezclados a Casa de Bombas 2 (CB2) (Ecopetrol S.A., 2009). CB2 es una unidad intermedia de tanques de almacenamiento de crudos que hace enlace entre Casa Bombas 8 y las unidades de refinación lo cual permite dar un mejor manejo a los crudos de carga según disponibilidad de inventario de los mismos (Ecopetrol S.A., 2009).

Para los crudos, actualmente se tiene una rutina de muestreo con una frecuencia muy baja para análisis de contenido total de azufre y gravedad API. La logística del muestreo está estructurada para desarrollarse en un tiempo muy largo que abarca desde la toma de la muestra hasta la

obtención del resultado; para este caso, gravedad API y contenido de azufre, generando pérdidas difíciles de cuantificar respecto al aprovechamiento de la conversión del crudo en productos valiosos. Esta misma logística no permite aumentar la frecuencia de la toma de muestra porque al realizarlo genera sobre costos asociados a todo el proceso de análisis de los crudos mezclados en la CB8. Dadas estas circunstancias, no es posible hacer un mejor seguimiento y control de estos parámetros en los crudos mezclados en CB8. El hecho de no tener datos de manera oportuna baja la eficiencia de los procesos de producción de la Refinería, impide un control adecuado del contenido de azufre total de los crudos mezclados y, por tanto, aumenta la probabilidad de impactar este parámetro desde el proceso de refinación hasta los productos derivados del crudo que se entregan a los clientes de la Refinería de Barrancabermeja. A su vez, la falta de control de los contenidos de azufre posibilita la generación de mezclas ácidas corrosivas que impactan la confiabilidad de los equipos de proceso y disminuyen la disponibilidad operacional. Además, un bajo control de la gravedad API de las mezclas formuladas conlleva posibles pérdidas energéticas por baja fluencia de los crudos en las tuberías que los conducen.

La industria en general, y particularmente la Refinería de Barrancabermeja, necesita mejorar la eficiencia de los procesos de producción, para asegurar la obtención de productos con una elevada calidad y a la vez reducir consumos, lo que redundará en una mejora de la productividad. Entre los costos más significativos se encuentran los asociados al proceso de análisis de muestras que va desde el pago al conductor del laboratorio, quien recoge las muestras, los analistas, materiales, equipos de transporte y de laboratorio. La rapidez de los métodos analíticos mejora la eficiencia del proceso de producción al aumentar la frecuencia de la obtención de datos de API y azufre, necesarios para evitar inestabilidad de las variables del proceso en la Unidades de destilación

atmosférica y a su vez minimizar la posibilidad de presentarse paradas de plantas no programadas por problema de corrosión.

La espectroscopía de infrarrojo cercano (NIR por sus siglas en inglés) se ha empleado con éxito en la caracterización de distintas formas de materia orgánica y sus derivados, considerándose como una herramienta adecuada en el área de geoquímica orgánica, con la ventaja operativa de la rapidez y bajo costo de análisis (Kister, Pieri, Alvarez, Diez, & Pis, 1996).

El desarrollo de este trabajo permitió clasificar por análisis de componentes principales (PCA) y analizar mediante la regresión por mínimos cuadrados parciales (PLS-R) la correlación de la gravedad API y del contenido total de azufre en la mezcla de crudos de la CB8 del área de mezclado de la Refinería de Barrancabermeja con la espectroscopia de infrarrojo cercano.

1. Marco teórico conceptual y estado del arte

1.1 Tratamiento del crudo en la refinería de Barrancabermeja

La Figura 1 es un diagrama general simplificado de la Refinería de Barrancabermeja. La representación del tanque marcado como “CRUDO” hace referencia a las áreas del Departamento de Materias Primas y Productos denominadas *Casa de bombas 8* (CB8) y *Casa de bombas 2* (CB2). CB8 entrega los crudos mezclados a CB2. Esta última es una unidad intermedia que hace enlace con las unidades de refinación y permite dar un mejor manejo a los crudos de carga según disponibilidad de inventario de los mismos (Ecopetrol S.A., 2009).

En la ruta del diagrama aparece la representación gráfica de una torre marcada como “ATMOSFÉRICO” y hace referencia al Departamento de Refinación de Crudos, específicamente a las Unidades de Destilación Atmosférica. (Ecopetrol S.A., 2009). Estas son unidades de destilación compuestas por una sección de destilación atmosférica y otra de vacío. Para este trabajo, se hace referencia sólo a las unidades denominadas U-200, U-250, U-2000 y U-2100. En ellas se hace la transformación del crudo en productos más valiosos como son Nafta, Jet, ACPM, Gasóleo y fondos de vacío.

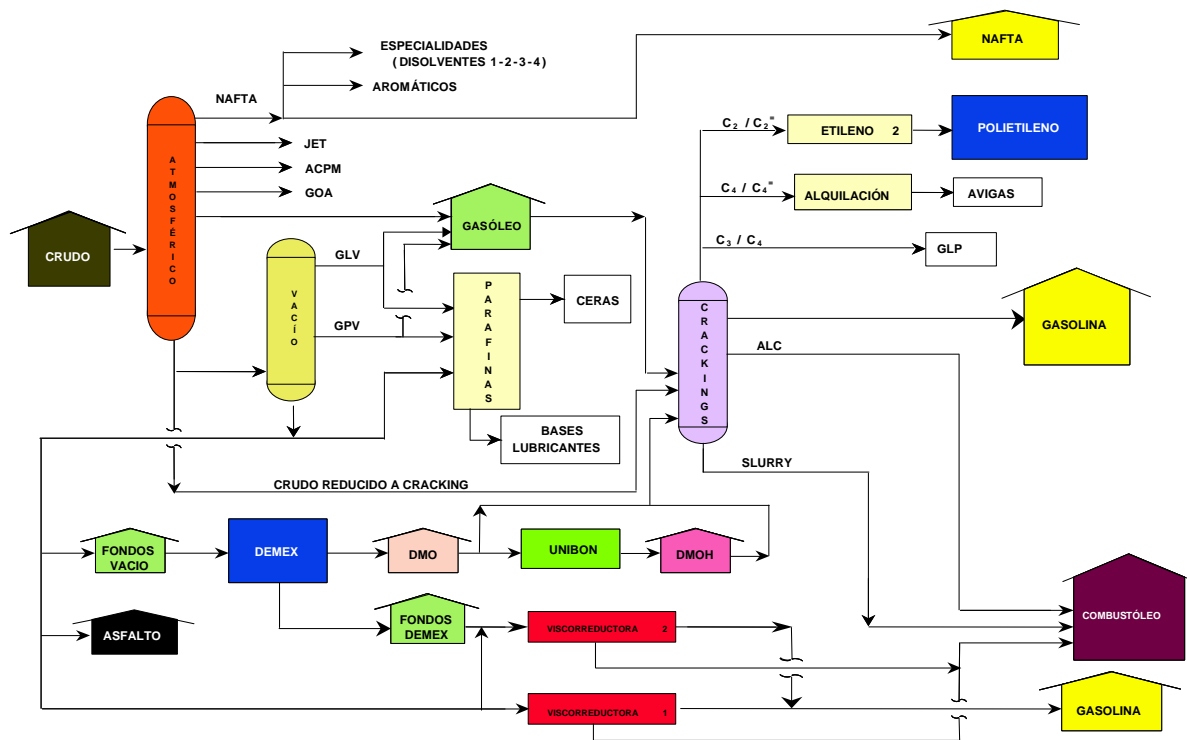


Figura 1. Diagrama general simplificado de la Refinería de Barrancabermeja.

Nota. Adaptado de: Empresa Colombiana de Petróleos, 2009

Cada cabezal de mezcla se conforma de una sección de tubería que recibe los 5 tipos de premezclas y a su salida pasa a través de un mezclador estático. Este último produce cambios

súbitos en la dirección de desplazamiento ocasionando turbulencia y produciendo una mezcla homogénea. A la salida del Mezclador Estático, la corriente se dirige por el ramal principal hacia el límite de CB8, con dirección a CB2 (Ecopetrol S.A., 2009).

La Figura 2 muestra el área de mezclado de CB8; donde se hace énfasis en los puntos de muestreo ubicados en los cabezales de mezcla de crudo. De acuerdo al Manual de Descripción del Proceso de la Unidad de Materias Primas y Productos (Ecopetrol S.A., 2009), el proceso de mezclado de crudos los crudos recibidos se agrupan en los tanques de almacenamiento según sus propiedades en premezclas: LCT (crudo de bajo punto de enfriamiento), Premezcla liviana “PML”, Premezcla pesada “PMP”, Caño Limón “CL” y Premezcla ácida “PMA”, hasta la salida de los 4 tipos de mezcla: parafínica, nafténica, pesada y liviana, hacia CB2 para almacenamiento y posterior entrega a las unidades de destilación atmosférica. En ella también se identifican los puntos de muestreo de los crudos mezclados; mezclas realizadas en la CB8 y tomadas en línea para su posterior envío al laboratorio y respectivo análisis gravedad API y %p/p de azufre.

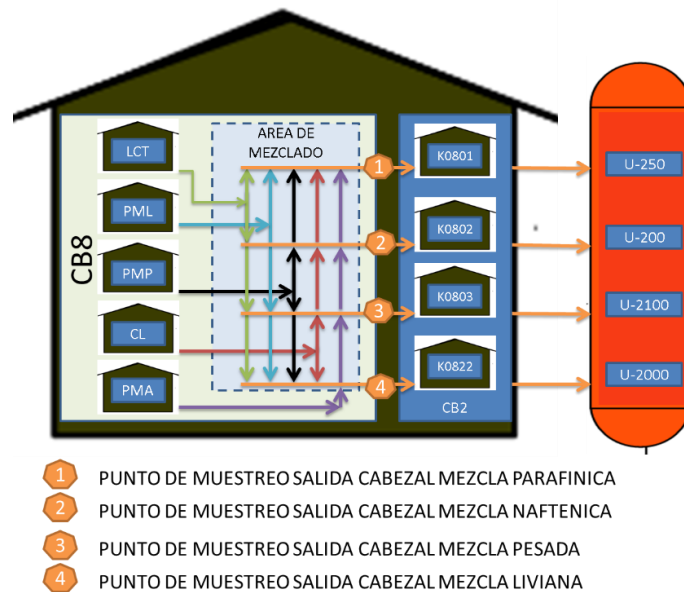


Figura 2. Área de mezclado de crudo CB8 de la Refinería de Barrancabermeja

- Propiedades de los Crudos relevantes a este trabajo

Los crudos se pueden clasificar de acuerdo con la denominada gravedad API. La gravedad API se define como una función de densidad relativa (gravedad específica -SG-) 60/60°F, representada por:

$$^{\circ}API = \frac{141,5}{\text{Densidad relativa (60}^{\circ}F/60^{\circ}F)} - 131,5 \text{ (Ecuación 1)}$$

Aquí, La densidad relativa se define como la densidad de la sustancia a 60°F (15,56°C) dividida por la densidad del agua a esta misma temperatura:

$$SG = \frac{\text{Densidad de la sustancia a } 60^{\circ}F}{\text{Densidad del agua a } 60^{\circ}F} \quad \text{(Ecuación 2)}$$

Ecopetrol S.A. realizó su propia clasificación de crudos de acuerdo a la gravedad API así (Ecopetrol S.A., 2009):

- Crudo Pesado: $API < 24^{\circ}$
- Crudo Intermedio: $24^{\circ} \leq API \leq 34^{\circ}$
- Crudo Liviano: $API > 34^{\circ}$

Además, según su contenido de azufre, los crudos también se pueden clasificar como dulces; aquellos que presentan porcentaje de azufre menor al 0.7%, y agrios; con porcentajes mayores al 0.7%. El contenido de azufre se expresa en porcentaje en peso (p/p). En general, los crudos livianos tienen bajos contenidos de azufre y los pesados tienen altos contenidos de azufre (Ecopetrol, Plantilla para elaborar documentos técnicos - 81131_ECP-VIN-P-MBC-MT-008-MMH, 2013)

- Laboratorio de la Refinería de Barrancabermeja

El laboratorio es una sección de la Refinería que se encarga de realizar ensayos sobre productos con la finalidad de caracterizarlos, analizarlos e indagar sobre los resultados obtenidos (Castro Zaldúa, 2010). Los resultados que se obtienen de análisis en laboratorios deben ser confiables. Para esto, es necesaria la utilización de métodos estandarizados y normalizados (OTT, 1992).

- Métodos de ensayo

El método de ensayo o método de prueba es un procedimiento definitivo que produce un resultado de prueba (ASTM, Form and style for ASTM standards, 2015). Una prueba puede ser considerada como una operación técnica que consiste en la determinación de una o más características de un determinado producto, proceso o servicio de acuerdo con un procedimiento especificado (ASTM, Standard Guide for Proficiency Testing by Interlaboratory Comparisons, 2015). Muchas veces una prueba hace parte de un experimento. El resultado de la prueba puede ser cualitativa, es decir, si presenta o no algún atributo; cuantitativa, o sea, que determina y mide el atributo.

- Métodos Estandarizados ASTM

ASTM es la sigla de American Society for Testing Materials (Sociedad Americana para Pruebas y Materiales). Los métodos ASTM cubren todos los materiales que se utilizan en la industria y el comercio. El Comité de la ASTM que se ocupa de lo concerniente a la industria del petróleo, se identifica con la letra D, y así los métodos ASTM que se refieren al petróleo se designan de la siguiente manera: Primero la letra D, luego, el número consecutivo correspondiente al método y, por último, las dos últimas cifras del año en que el método fue revisado y aprobado por última vez. Así, por ejemplo, el método que se utiliza para determinar la gravedad API mediante el método del hidrómetro se identifica con las siglas D-1298, lo cual significa que fue estudiado y aprobado por el Comité D, que se asignó el número 12 y fue revisado por última vez el año 1998 (HILARRAZA BRITO, 2011).

- Determinación de la gravedad API

Para calcular la gravedad API, el laboratorio de la Refinería de Barrancabermeja utiliza el método de prueba estándar para la densidad y la densidad relativa de crudos mediante el uso de un

Analizador de Densidad Digital (ASTM International, 2013) usando el estándar ASTM D5002 - 13. Este método cubre la determinación de la densidad o densidad relativa de aceites crudos que pueden ser manejados de una manera normal como líquidos a temperaturas de ensayo entre 15°C y 35°C. El método se aplica a los aceites crudos con presiones de vapor altas siempre que se tomen las precauciones adecuadas para evitar la pérdida de vapor durante la transferencia de la muestra al analizador densidad. Las unidades de medición de la densidad son gramos por mililitro y kilogramos por metro cúbico (ASTM, D5002 -13, 2013).

- Determinación de la concentración de azufre

La determinación de azufre en el laboratorio de la Refinería de Barrancabermeja se efectúa mediante el Método de prueba estándar para azufre en petróleo y sus derivados por energía dispersiva de rayos X espectrometría de fluorescencia según la norma ASTM D4294-10 (ASTM International, 2010). Este método proporciona un medio para determinar si el contenido de azufre de petróleo o de un producto de petróleo cumple con las especificaciones o con los límites normativos vigentes. En general, el límite máximo de contenido de azufre para los crudos cargados en las Unidades de destilación atmosférica de la refinería de Barrancabermeja es 1.40 g/100 g (Ecopetrol, 2006). El método usa la técnica de quimioluminiscencia, donde la radiación característica se derivada a partir de una fuente de Rayos X y es comparada con la de muestras patrón, cuyo contenido de azufre (mg/Kg o g/100g) es conocido (Gustato, 2003). La repetibilidad es del 95% (ASTM International, 2010). Como se planteó en la Introducción de este documento, este trabajo plantea como alternativa de análisis de las propiedades arriba destacadas el uso de espectroscopía infrarroja. Por ello, a continuación, se discutirán diferentes aspectos ligados a esta técnica de análisis instrumental.

1.2 Principios de la espectroscopía infrarroja

La figura 3 muestra los componentes de la radiación electromagnética de acuerdo a la interpretación de la teoría clásica de Maxwell. De acuerdo con esta teoría, la radiación se considera como dos campos eléctricos y magnéticos perpendiculares entre sí que oscilan en planos individuales en ángulos rectos. Estos campos están en fase y se propagan como una onda sinusoidal. Las magnitudes de los vectores eléctricos y magnéticos están representadas por E y B, respectivamente. Los métodos espectroscópicos hacen uso de los principios antes expuestos. En general, Monfort Fraga proveen la siguiente definición: “Los métodos espectroscópicos son un amplio grupo de métodos analíticos que se basan en las espectroscopias atómica y molecular. La espectroscopia es el término general para la ciencia que trata de las distintas interacciones de la radiación con la materia” (Monfort Fraga, 2014).

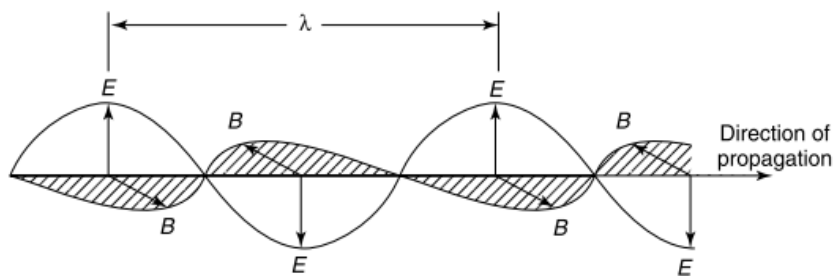


Figura 3. Componentes de la radiación electromagnética

Nota. Adaptado de: (Stuart, 2004)

- Absorción de la radiación

Cuando la radiación atraviesa una capa de sólido, un líquido o un gas, ciertas frecuencias pueden ser eliminadas selectivamente por absorción. La absorción se define como un proceso en el que la

energía electromagnética se transfiere a los átomos, iones o moléculas que componen la muestra. La absorción provoca que estas partículas pasen de su estado normal a temperatura ambiente, o estado fundamental, a uno o más estados excitados de energía superior. Desde el punto de vista del análisis instrumental, se realiza experimentalmente una representación gráfica de la absorbancia en función de la longitud de onda o frecuencia y se aprovecha la ley de Lambert-Beer para extraer información de la muestra a analizar. La ley; expuesta en la Ecuación 3, propone que para una radiación monocromática, la absorbancia (A) es directamente proporcional al camino óptico (b) a través del medio y la concentración (c) de la especie absorbente. Donde, ε es la constante de absortividad molar (Skoog, et al., 2001).

$$A = \varepsilon bc = -\log T = -\log(I_i/I_f) \quad (\text{Ecuación 3})$$

Para que una muestra absorba en la región infrarroja, se debe cumplir lo siguiente:

- La radiación debe tener la energía precisa para satisfacer los requerimientos del paso al estado excitado (cuantización).
- Debe producirse una interacción efectiva entre la radiación y la materia; i.e., debe haber un cambio neto del momento dipolar como resultado de la vibración o rotación de la molécula para que se verifique la absorción (Monfort Fraga, 2014).
- Espectroscopía NIR

La figura 4 presenta las diferentes regiones espectrales de radiación electromagnética.

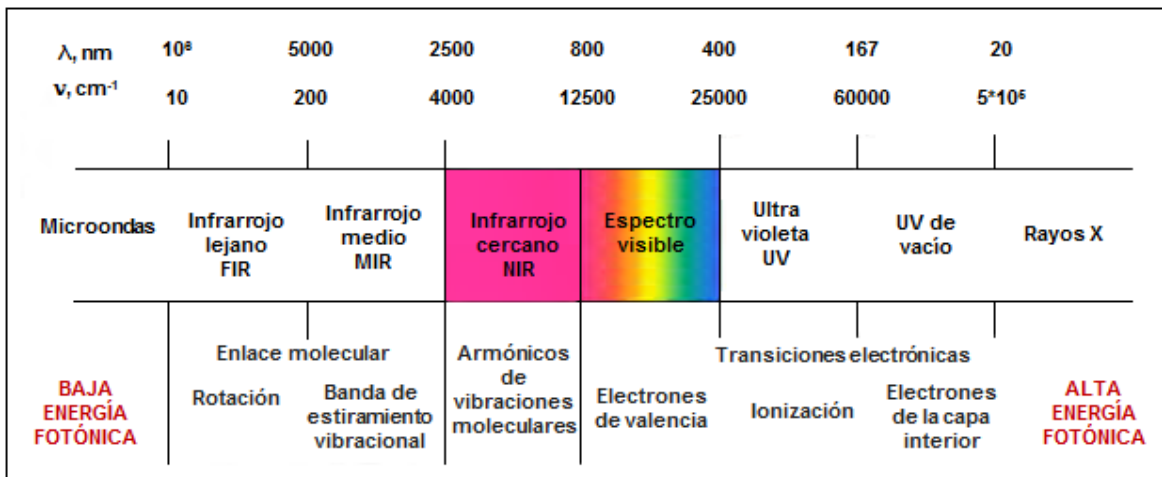


Figura 4. Espectro Electromagnético

Nota. Adaptado de: (Castillo Martínez, 2007)

La región espectral del infrarrojo cercano se extiende desde el extremo superior de longitudes de onda del visible, de alrededor desde 4000 cm^{-1} hasta 13000 cm^{-1} . Este rango tiene pequeñas variaciones, ya que algunos definen la región hasta 12800 cm^{-1} (ASTM, 2015) (Zamora Zamora & Blanco Romia, 2012), mientras que para otros va hasta $14,000 \text{ cm}^{-1}$ (Falla, et al., 2006).

El origen de las bandas de absorción en NIR son sobretonos, es decir, transiciones energéticas en las cuales la variación del número cuántico vibracional, n , es mayor de ± 1 , y combinaciones lineales de las bandas vibracionales de tensión fundamentales en moléculas poliatómicas y debidas a cambios simultáneos en la energía de uno o más modos de vibración. Esto es diferente de lo que sucede en el infrarrojo medio y en el infrarrojo lejano donde la absorción se debe a vibraciones moleculares fundamentales y a rotaciones moleculares, respectivamente (Peguero Gutiérrez, 2010).

La espectroscopia NIR emplea fotones de energía en el intervalo de $2.65 \cdot 10^{-19}$ a $7.96 \cdot 10^{-20}$ J correspondientes al rango de frecuencias mencionado anteriormente. 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 por la molécula, mientras que otras frecuencias serán absorbidas parcialmente o no lo serán (LAMUS BLANCO, 2010).

Desde el punto de vista matemático, la energía de los estados vibracionales, E_{vib} en una molécula diatómica se puede describir por el modelo del oscilador armónico mediante la siguiente ecuación:

$$E_{vib} = \left(v + \frac{1}{2} \right) \frac{h}{2\pi} \sqrt{\frac{k}{\mu}} \quad (\text{Ecuación 4})$$

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

$$\Delta v = E_{v2} - E_{v1} = \Delta v h \nu \quad (\text{Ecuación 5})$$

Donde, Δv es la frecuencia de vibración fundamental del enlace que produce una banda de absorción en el infrarrojo medio.

Las vibraciones en moléculas poliatómicas implican el movimiento complejo de sus átomos constituyentes. En la práctica, tales vibraciones moleculares tienden a ser anarmónicas, es decir, las vibraciones sobre la posición de equilibrio son asimétricas. Esta anarmonicidad permite las transiciones entre estados de energía vibracional no consecutivos, ($\Delta v = \pm 2, \pm 3, \dots$), lo que genera bandas de absorción conocidas como sobretonos las cuales son, aproximadamente, múltiplos de las frecuencias fundamentales de vibración aunque su intensidad es mucho menor. Si bien teóricamente son posibles transiciones entre cualquier par de niveles energéticos,

experimentalmente solo se observan las bandas de absorción correspondientes a las frecuencias de vibración fundamental en el MIR (ν) y a los dos primeros sobretonos en el NIR (2ν , 3ν).

- Instrumentación para NIR

La absorción en NIR es baja. Esto hace que el nivel de exigencia de los espectrómetros NIR deba ser mayor que en otros instrumentos en términos de nivel de ruido permisible y estabilidad instrumental, sobre todo si se quiere aplicar al análisis cuantitativo. El esquema básico de un espectrofotómetro NIR está compuesto de: fuente de radiación, un selector de longitudes de onda, un compartimento de la muestra y un detector. Las diferentes opciones para cada componente se muestran en la Figura 5 (Castillo Martínez, 2007). La salida del instrumento se conoce como espectro. La mayoría de los instrumentos comerciales presentan un espectro con el número de onda decreciente desde izquierda a derecha. En general, hay menos bandas infrarrojas en el cm^{-1} región de 4000 hasta 1800 con muchas bandas entre 1800 y 400 cm^{-1} . A veces, la escala se cambia de manera que se contrae la región entre 4000 y 1800 cm^{-1} y la región entre 1800 y 400 cm^{-1} se expande para enfatizar las características de interés. En el espectro, la escala de ordenadas se puede presentar en % de transmitancia con 100% en la parte superior del espectro (Stuart, 2004). Además, es común tener una opción de absorbancia o transmitancia como medida de la intensidad de las señales grabadas.

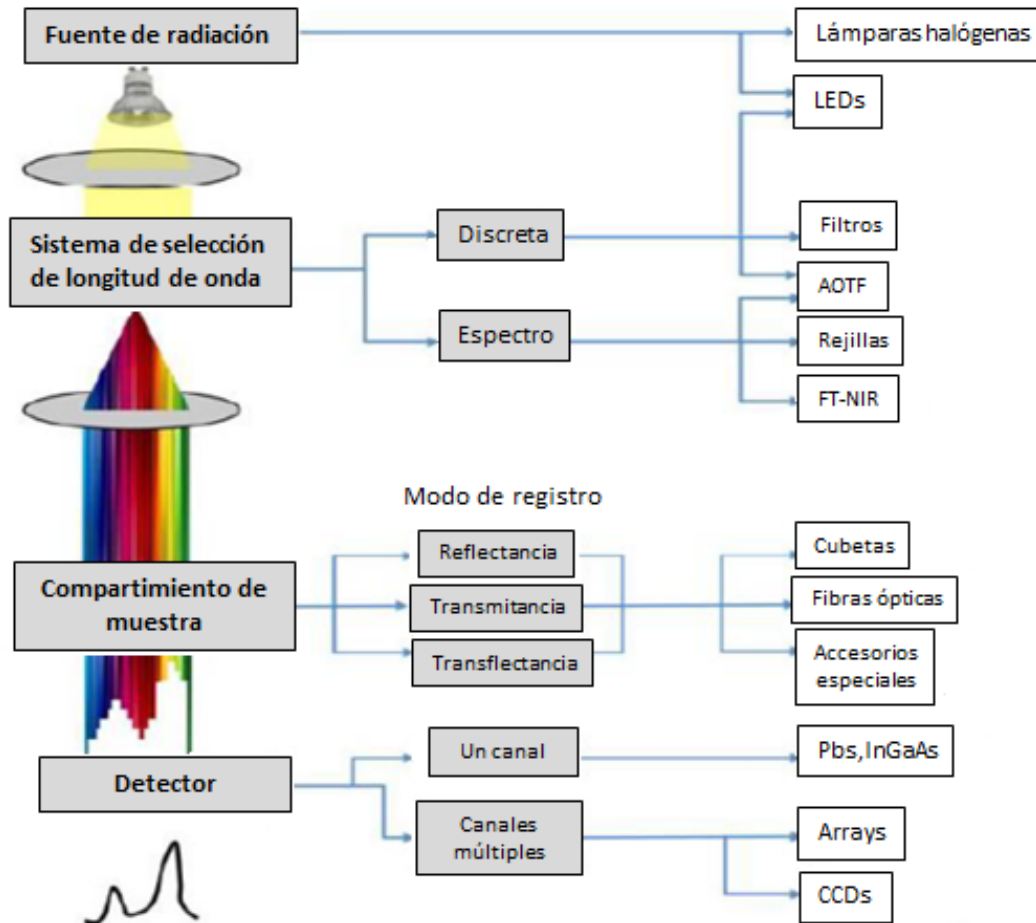


Figura 5. Componentes principales de un espectrofotómetro NIR.

Nota: Adaptado de: (Monfort Fraga, 2014)

De acuerdo a la anterior figura, la fuente de radiación que más se utiliza es la lámpara halógena de filamento de tungsteno con ventana de cuarzo, capaz de proporcionar un espectro continuo en la región de 320 nm a 2500 nm, otras fuentes de radiación que se pueden utilizarse son los llamados **LED's** (*Light Emitting Diodes*), que dependiendo de su composición pueden llegar a emitir hasta 1600 nm.

En función del dispositivo utilizado para la selección de longitudes de onda, los instrumentos NIR pueden ser clasificados en sistemas dispersos (Espectro) o no dispersos (Discreta). Los equipos no dispersos son los más ampliamente utilizados.

Existen instrumentos de filtros convencionales, filtros optoacústicos (**AOTF**, Acousto-Optic Tunable Filter) e instrumentos de transformada de Fourier (**FT**). (Peguero Gutiérrez, 2010)

Los detectores más habituales utilizados en Espectroscopia NIR están contruidos con semiconductores como **InGaAs**, **PbS**, InAs, InSb, Si...El detector más utilizado es el de PbS, que posee sensibilidad adecuada entre 900 y 2600 nm. Para medidas de transmisión en sólidos se utilizan detectores de Arseniuro de Indio y Galio (InGaAs), operativo de 600 a 1900 nm (Ciurczack, 2008)

Los detectores de canales múltiples pueden ser dispositivos acoplados por carga o **CCD**, acrónimo en inglés de Charge-Coupled Device, son detectores de elección para mediciones de alta precisión, en un amplio rango espectral que abarca desde el infrarrojo cercano (NIR) hasta la banda de rayos X.

Las figuras 6 y 7 muestran espectros infrarrojos de ácido láctico e ilustran la diferencia en apariencia entre los espectros presentados con absorbancia y con transmitancia. La transmitancia se utiliza tradicionalmente para la interpretación espectral, mientras que la absorbancia se utiliza para el trabajo cuantitativo.

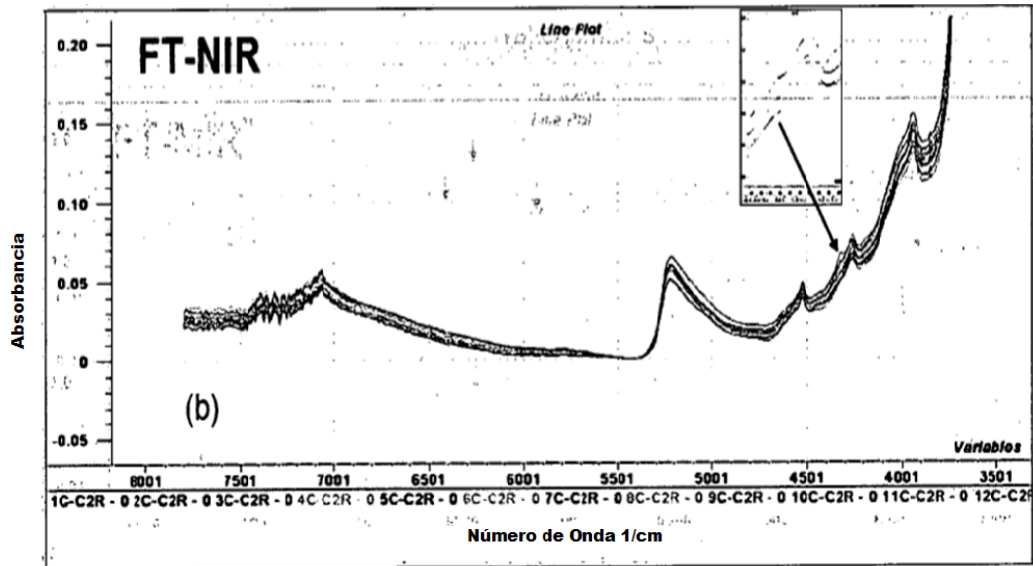


Figura 6. Ejemplo de un espectro de absorbancia

Nota: Adaptado de: Roeges, Noel PG (1994). A guide to the complete interpretation of infrared spectra of organic structures. Wiley

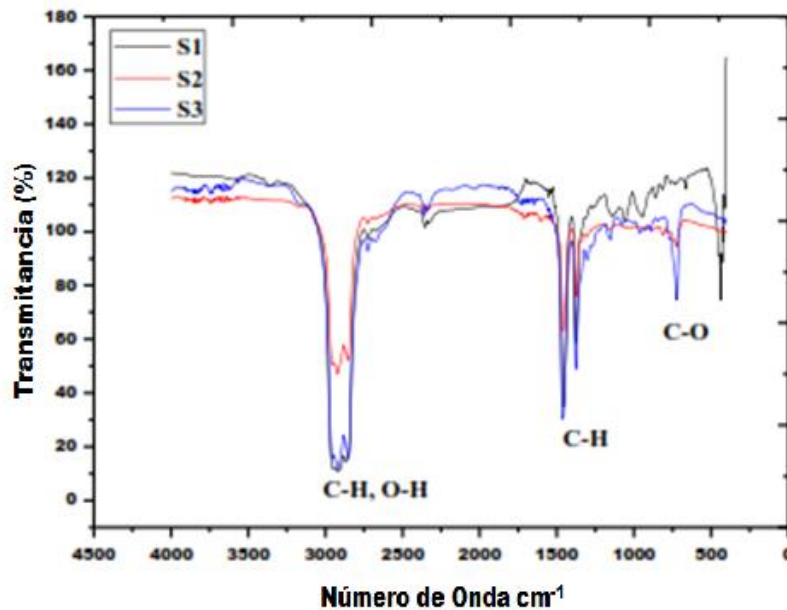


Figura 7. Ejemplo de un espectro de Transmitancia

Nota: Adaptado de: Spectral Response of Transformer Oil Using FTIR (s.f.) Recuperado de http://cybertesis.uni.edu.pe/bitstream/uni/1320/1/araujo_pp.pdf.

Para este trabajo es muy importante revisar las bandas de absorción de los grupos funcionales orgánicos que pueden registrarse en NIR puesto que son ellos los que se esperan encontrar en las muestras de crudo analizadas. En la figura 8 se muestran las regiones del NIR donde absorben los diferentes enlaces, así como si pertenecen al primer, segundo o tercer sobretono o a combinaciones de frecuencia de vibración de diferentes enlaces. Las bandas más frecuentes en NIR son debidas a sobretonos y combinaciones de enlaces en los que participa el hidrógeno y que contienen átomos ligeros como C-H, N-H, O-H, S-H, debido a su mayor armonicidad.

Las bandas de los grupos C=O, C-C, C-F, o C-Cl, en general son muy débiles o no aparecen en esta región. Las vibraciones fundamentales de estos grupos tienen bajas frecuencias en la región IR y por tanto, los primeros sobretonos también aparecen en esa región. La banda correspondiente al grupo -CHO que aparece entre los 4500 y 4600 cm^{-1} corresponde a la combinación de las vibraciones de los enlaces C-H y C=O del grupo aldehído.

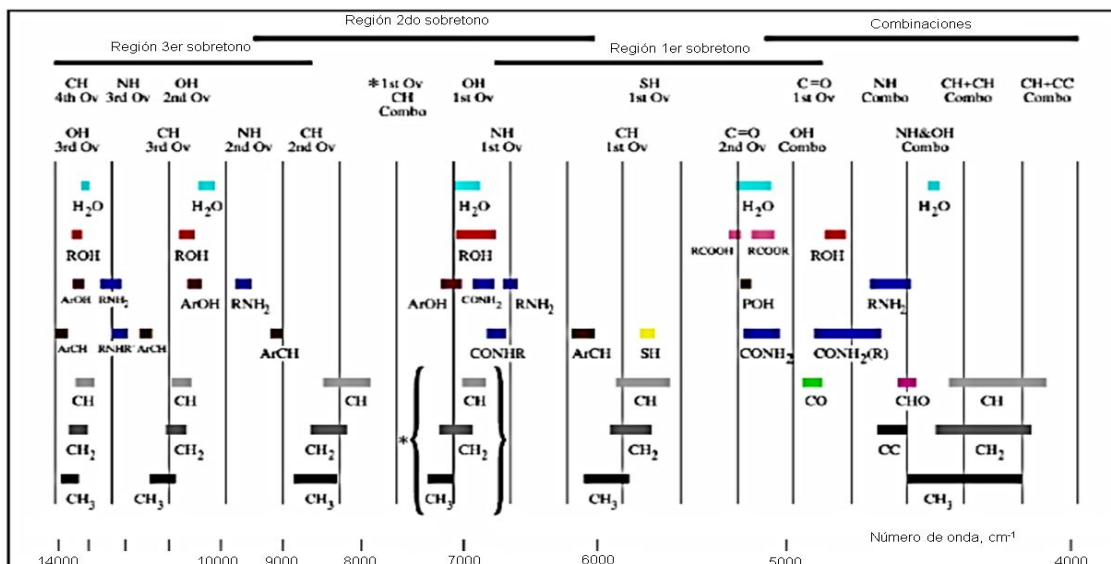


Figura 8. Bandas de absorción en la región NIR

Nota: Adaptado de: (Casal Martínez, 2013)

Los métodos que permiten el análisis químico de muestras complejas por NIR hacen parte de la denominada quimiometría. A continuación, se discuten aspectos respecto a este asunto.

1.3 Definición de quimiometría

La quimiometría corresponde al uso de herramientas matemáticas y estadísticas para la conversión de datos de análisis instrumental en información útil (Massart, et al., 1997) (Beebe, et al., 1998). Las técnicas instrumentales, y especialmente las espectroscópicas, generan una gran cantidad de datos relativos a las muestras analizadas en breves intervalos de tiempo. La quimiometría pretende convertir esos datos en información mediante el uso de (Horlick, 1972):

- **Filtrado de señales:** son pretratamientos aplicados a los datos experimentales.
- **Métodos de reconocimiento de pautas:** son técnicas multivariadas mediante las que se pueden establecer agrupaciones de muestras en función de su similitud y métodos de clasificación de nuevas muestras.
- **Métodos de calibración:** son técnicas multivariadas mediante las cuales se busca una relación cuantitativa entre la señal analítica y alguna propiedad de la muestra)

Entendiendo que la quimiometría trata con conjuntos de datos complejos que engloban los efectos de múltiples variables de entrada y respuesta, a continuación, se revisan algunos aspectos relacionados con el manejo estadístico de tales conjuntos.

- Análisis multivariable

El análisis multivariable es la investigación de muchas variables, simultáneamente, con el fin de entender las relaciones que existen entre ellas. Se trata sobre todo de un enfoque gráfico con

énfasis particular en el estudio de la variabilidad y sus fuentes. La variabilidad puede ser definida como deseada y no-deseada:

- La variabilidad deseada es el cambio premeditado de parámetros físicos o químicos que afectan al sistema de estudio con el fin de evaluar el impacto que dichos parámetros ejercen sobre él.
- La variabilidad no deseada resulta de la incapacidad para controlar algunos aspectos del sistema y es aleatoria.

Uno de los propósitos de estudiar la variabilidad es el desarrollo de modelos para predecir eventos futuros con una cierta confianza. Validar un modelo implica demostrar que se adecua al uso para el que fue pensado. Un modelo multivariable es capaz de mostrar la influencia que tiene en el sistema los dos tipos de variabilidad (deseada y no deseada) de modo que éste puede ser mejor comprendido (CAMO Software , 2012). Como beneficios de los análisis multivariable, se pueden nombrar: (i) la identificación de variables que contribuyen a la mayoría de la variabilidad de los datos; (ii) el aislamiento de variables que están relacionadas, en otras palabras, que co-varían las unas con las otras; (iii) manejo gráfico que permite una mejor interpretación. (CAMO Software , 2012).

A continuación, se presenta brevemente la base teórica de las principales herramientas quimiométricas utilizadas para la extracción de información de los datos espectroscópicos, ya sea de forma cualitativa o cuantitativa. Descripciones más detalladas se pueden encontrar en libros de texto donde se recopilan las técnicas quimiométricas más utilizadas (Jackson, 1980) & (Wold, et al., 1987).

1.3.1 Análisis en componentes principales (PCA) El PCA es una herramienta que permite condensar la variabilidad presente en una matriz de datos en unos pocos parámetros representativos, denominados componentes principales (PCs) o variables latentes. El PCA es uno de los métodos más empleados para reducción de variables, además de constituir la base a partir de la cual se han desarrollado multitud de técnicas quimiométricas de análisis cualitativo y cuantitativo (Jackson, 1980) (Wold, et al., 1987).

Una matriz de datos \mathbf{X} ($M \times N$) que está conformada por M espectros o muestras registradas a N longitudes de onda puede ser representada en un espacio de dimensión N donde cada muestra estaría representada por un solo punto. En este contexto, las características comunes entre las muestras o sus diferencias serán observables en función de su agrupación. Debido a que no es posible representar dimensiones superiores a 3, el PCA tiene como objetivo hallar las direcciones de máxima variabilidad en que están agrupados los M puntos en el espacio de dimensión N , utilizando unos nuevos ejes llamados componentes principales para representarlos. El primer componente principal se da con la combinación lineal de las N variables que explica la máxima variabilidad de las muestras; el segundo se escoge de forma que sea ortogonal al primero y que explique la máxima variabilidad una vez restada la del primero y así sucesivamente. Para definir matemáticamente el cambio de base, se definen los *loadings* (\mathbf{P}), que son los cosenos de los ángulos que forman los nuevos ejes con los originales. Los *scores* (\mathbf{T}) son las coordenadas de las muestras en los nuevos ejes (Castillo Martínez, 2007). Matemáticamente, la matriz de datos espectrales \mathbf{X} se descompone en el producto de las matrices de *scores* y *loadings* más una matriz \mathbf{E} de residuales según la siguiente ecuación:

La figura 9 es una representación gráfica de esta operación:

$$\mathbf{X} \mathbf{TP}^t = + \mathbf{E} \quad (\text{Ecuación 6})$$

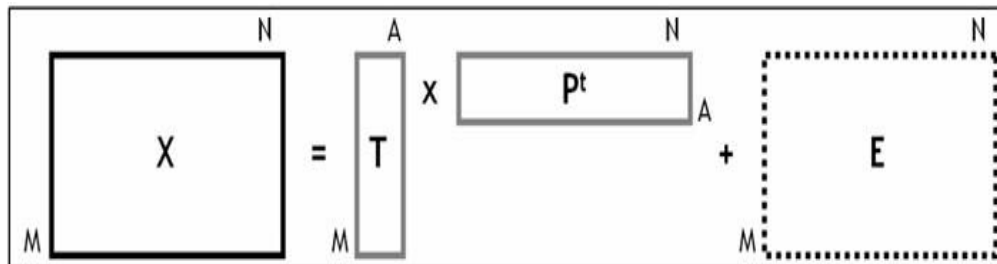


Figura 9. Representación gráfica de la matriz de datos espectrales

Nota: Adaptado de: (Castillo Martínez, 2007)

Los A componentes principales contienen diferente información debido a la condición de ortogonalidad. Los primeros describen las fuentes de variabilidad más relevantes, mientras que los últimos describen variaciones debidas al ruido, por lo que es posible reproducir la matriz de datos original con un pequeño número de componentes principales $A \ll N$. La matriz de datos E contiene la información no recopilada por los A componentes principales.

Uno de los aspectos fundamentales del PCA es la selección del número de componentes principales que contienen la información relevante de la matriz de datos. La forma más habitual de realizarlo es mediante un gráfico de la varianza explicada en función del número de PCs, escogiendo el menor número de PCs para el cual no se encuentra un incremento significativo de la varianza.

Después de haber establecido el número de PCs, el estudio de los gráficos de *scores* y *loadings* aporta información sobre variaciones en los datos en un menor número de dimensiones. El gráfico de dispersión de PC's (generalmente PC1 vs. PC2 cuando ambos explican un porcentaje elevado de la varianza explicada) revela la relación existente entre las muestras, mostrando la existencia de *clústers*, atípicos, etc.

Los *loadings* son útiles para determinar cuáles son las variables originales más importantes. Definido el *loading* como el coseno del ángulo entre un componente principal dado y la variable

n, cuando su valor se aproxima a 1, se indica que esa variable contribuye significativamente a la varianza explicada por ese componente. En cambio, cuando se aproxima a 0, ambos serán ortogonales indicando que la variable contribuye poco a la varianza explicada por el componente principal analizado. En particular, el estudio visual de los *loadings* es útil para identificar las regiones espectrales, y por tanto, los compuestos químicos con mayor importancia en los datos.

Además del estudio de datos, el análisis en componentes principales puede ser utilizado para filtrar el ruido en matrices de datos espectroscópicos, reconstruyendo la matriz con los A componentes principales que recogen las principales fuentes de variación y excluyendo los PCs poco significativos, que típicamente tienen una baja relación señal/ruido (Castillo Martínez, 2007).

1.3.2 Regresión por mínimos cuadrados parciales (PLS) La regresión por mínimos cuadrados parciales es una técnica de calibración multivariable inversa, lo que hace posible calibrar únicamente el componente deseado sin necesidad de conocer el resto de fuentes de variación. El PLS se basa en que la información contenida en las variables medidas se puede concentrar en un número menor de variables sin pérdida de información relevante, de forma similar al PCA, haciendo máxima la matriz de varianza-covarianza. La regresión se realiza sobre las nuevas variables, eliminando así los problemas de colinealidad que puedan tener los datos originales.

Durante la etapa de calibración, el PLS utiliza tanto la matriz de datos espectroscópicos (X) como la información contenida en la matriz de la propiedad a determinar (Y). Ambas matrices, una vez centradas como en el caso del PCA, se descomponen en una suma de factores de forma simultánea:

$$X = TP^t + E = \sum_{a=1}^A t_a p_a^t + E \text{ (Ecuación 7)}$$

$$Y = UQ^t + F = \sum_{a=1}^A u_a q_a^t + F \text{ (Ecuación 8)}$$

Donde, T es la matriz de *scores*, P la de *loadings*, E la de residuales correspondientes a la matriz de datos espectroscópicos (X); U es la matriz de *scores*, Q la de *loadings* y F la de residuales para la matriz de la propiedad a determinar (Y). Si tenemos M muestras, A componentes, N variables y K analitos, la dimensionalidad de las matrices es: T y U ($M \times A$), P^t ($A \times N$) y Q^t ($A \times K$).

La descomposición de ambas matrices se realiza de forma simultánea, estableciendo una relación interna entre los *scores* de los bloques X e Y , de forma que para cada componente se cumpla:

$$\hat{u} = b_a t_a \text{ (Ecuación 9)}$$

Donde, b_a es el coeficiente de regresión entre *scores* que los bloques X e Y para cada uno de los componentes. De esta manera, los nuevos ejes no coinciden con la dirección de máxima variabilidad espectral de las muestras como en el PCA, ya que están corregidos para obtener la máxima capacidad predictiva para la matriz Y .

Una vez establecido el modelo de calibración, se puede realizar la predicción de la propiedad modelada para nuevas muestras. A partir de los espectros originales, centrados o autoescalados, utilizando los valores calculados para las muestras de calibración, se calculan los *scores* de las muestras a predecir (t^*) y con los siguientes parámetros se evalúa el modelo de predicción construido.

- Selección del número de factores

La selección del número de factores que configura el modelo es un punto clave en la utilización de técnicas de reducción de variables, ya que el modelo ha de incorporar la información relevante,

pero ha de excluir el ruido. La forma más común de seleccionar el número de factores adecuado es utilizar la varianza explicada.

Los parámetros estadísticos más comunes para la evaluación del desempeño de los modelos de regresión, corresponden a:

- Raíz cuadrada del error cuadrado medio de calibración (*RMSEC, Root Mean Square Error of Calibration*): describe el grado de concordancia entre las propiedades estimadas por el modelo y los valores aceptados como verdaderos para las muestras de calibración. El RMSEC se define como:

$$RMSEC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-m-1}} \quad (\text{Ecuación 10})$$

- Raíz cuadrada del error cuadrado medio de calibración cruzada (*RMSECV, Root Mean Square Error of Cross-Validation*): Determina el desempeño cuando se utilizan los datos del conjunto de calibración para validar el modelo (validación interna). La siguiente ecuación define al RMSECV:

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

- R^2 , o coeficiente de correlación: Proporciona el porcentaje de varianza presente en los valores de los componentes, que después son reproducidos en la predicción. También es denominado varianza explicada. Cuanto más alto sea el coeficiente, mejor será la correlación entre los datos de la propiedad y los datos espectrales.
- Software analítico The Unscrambler

Este software se encarga de desarrollar modelos de regresión multivariados de una gran cantidad de muestras desarrollando un modelo de predicción y posteriormente validándolos.

(CAMO Software , 2012) La Figura 10 representa de forma esquemática el procedimiento que sigue el software en la construcción de un modelo de regresión R-PLS.

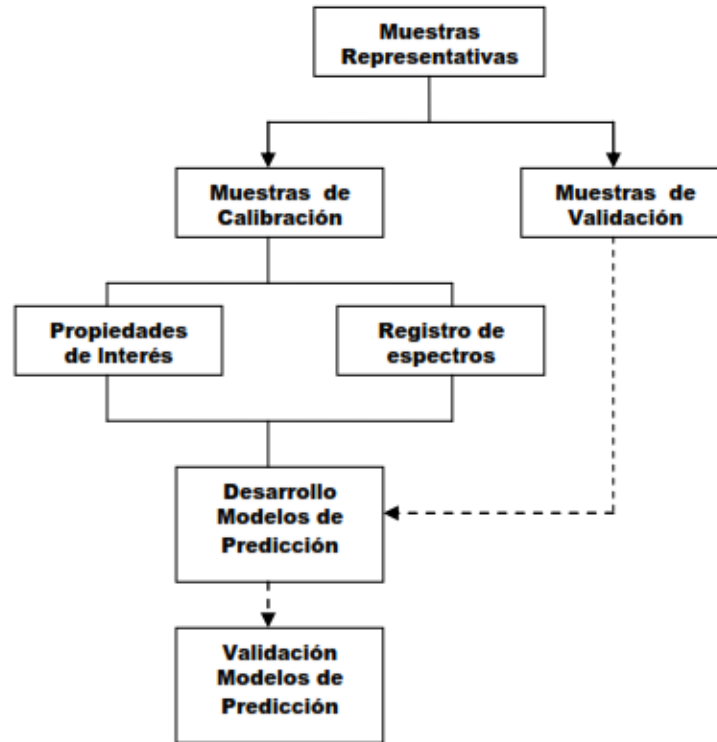


Figura 10. Descripción gráfica del modelo R-PLS

Nota: Adaptado de: (Ecopetrol S.A., 2009)

- Método cuantitativo: Uso de Análisis de Regresión y modelos predictivos

La calibración es definida clásicamente en la Química Analítica como el conjunto de operaciones que establecen, en las condiciones específicas, la relación entre los valores indicados por un instrumento de medida o por un sistema de medida y los valores conocidos correspondientes a un valor medido (Coello & Maspoch., 2007).

La expresión matemática que relaciona las variables de respuesta o señales analíticas con las concentraciones es la ecuación de calibración o modelo. El número de variables utilizadas es un

aspecto crítico de todas las alternativas matemáticas que relacionan la respuesta del instrumento con el parámetro que se desea determinar.

El objetivo de la calibración multivariable es la determinación simultánea de varios analitos o la de un único analito en presencia de una matriz compleja que participa en la señal analítica.

El análisis de regresión es el proceso de desarrollar un modelo a partir de los datos disponibles para predecir una respuesta deseada (o respuestas). El análisis de regresión requiere dos tablas de datos, una de variables independientes y otra de variables dependientes (CAMO Software, 2012).

Variables independientes son el conjunto de medidas disponibles que se usan para que un modelo prediga la respuesta de interés. Variables dependientes son las respuestas que se intentan extraer del modelo de predicción construido a partir de las variables independientes.

A continuación, se discute la aplicación del análisis multivariable al análisis de espectros infrarrojo.

- Prácticas estándar para análisis cuantitativo multivariado de espectros de infrarrojo

La norma ASTM E1655-05 (2012) establece prácticas para análisis cuantitativo multivariado e incluye una guía para la calibración multivariada de espectrómetros infrarrojos utilizados en la determinación de las características físicas o químicas de los materiales. Estas prácticas son aplicables a los análisis realizados en la región espectral del infrarrojo cercano y se usan para establecer la validez de los resultados obtenidos por un espectrómetro IR en el momento que se desarrolla la calibración. A continuación, se describen, de manera breve, los pasos necesarios para el análisis cuantitativo multivariado de espectros infrarrojo.

- Pretratamiento de datos

Una vez se adquieren los datos de infrarrojo, es conveniente procesar los espectros obtenidos con algunos pretratamientos matemáticos que permitan aumentar la relación señal/ruido y por ende

acrecentar la señal correspondiente al analito de interés. La elección de un tratamiento matemático u otro depende de las características de las muestras a clasificar, se escoge aquel tratamiento matemático que proporcione mejor discriminación entre las muestras, en otras palabras, la selección del pretratamiento más apropiado es experimental; i.e. a base de ensayo y error. Los tipos de pretratamientos posibles son (Peguero Gutiérrez, 2010):

- **Promedio de espectros.** Cada medida realiza un número alto de registros hasta que finalmente se obtiene un espectro promedio de todos los adquiridos.

- **Centrado.** Consiste en restar a cada valor de cada variable el valor promedio de la misma.

$$x' = x_{i,j} - \bar{x}_j \quad (\text{Ecuación 12})$$

Donde, x' es el dato centrado, $x_{i,j}$ es el dato del espectro i la variable j antes del centrado y \bar{x}_j es el promedio de la columna j .

- **Autoescalado.** Una vez centrados los datos se dividen por la desviación estándar de la variable correspondiente. Todas las variables tienen media = 0 y desviación estándar = 1

$$x' = \frac{x_{i,j} - \bar{x}_j}{s_j} \quad (\text{Ecuación 13})$$

Donde, x' es el autoescalado, $x_{i,j}$ es el dato de la fila i (espectro i) y la columna j (variable j) antes del autoescalado, \bar{x}_j es el promedio de la columna j y es la desviación estándar de la columna.

- **Normalización vectorial (SNV, Standard Normal Variate).** Se utiliza para corregir el efecto de dispersión producido, por ejemplo, por las diferencias espectrales en el tamaño de partícula de muestras sólidas ya que minimiza variaciones de desplazamiento de la línea base. Se basa en restar cada variable obtenida, absorbancia, del promedio y dividir el valor entre la desviación estándar.

$$x_{i,j}^{SNV} = \frac{x_{i,j} - \bar{x}_j}{s_j} \quad (\text{Ecuación 14})$$

Donde, x' es el autoescalado, $x_{i,j}$ es el dato de la fila i (espectro i) y la columna j (variable j) antes del autoescalado, \bar{x}_j es el promedio de la fila i y es la desviación estándar de la fila.

- **Derivadas.** El uso de las derivadas permite incrementar las diferencias entre las bandas anchas y solapadas del espectro. La primera derivada minimiza diferencias de desplazamiento de la línea base y la segunda minimiza diferencias de pendiente. La elección de más o menos puntos de suavizado tendrá que evaluarse según cada caso, ya que ventanas grandes hacen que pequeñas diferencias en los espectros queden diluidas, aunque también hacen que el ruido sea menor.

1.3.2.1 Etapas del desarrollo de un modelo de calibración Un modelo de calibración multivariable relaciona n variables experimentales (por ejemplo, un espectro NIR) frente a propiedades conocidas de las muestras (por ejemplo, la concentración de un analito), siendo el objetivo la predicción de dicha propiedad en nuevas muestras. Para la construcción de un modelo de calibración es conveniente seguir las siguientes etapas:

- **Selección de un conjunto de muestras**

Se debe disponer de un número de muestras suficiente para la creación y validación del modelo. Las muestras utilizadas para el desarrollo de un modelo de calibración deben ser representativas de la variabilidad que pueda darse durante el proceso y también entre procesos. Así, las muestras, además de cubrir todo el intervalo de la propiedad a determinar han de incorporar la variabilidad debida a otras variables del proceso, como puede ser el origen de la materia prima o la temperatura de reacción, entre otros.

- **Detección de outliers**

Debido a la capacidad que tienen los métodos multivariables de detectar la observación inconsistente con el resto de datos, la identificación de los outliers en esta etapa de establecimiento

del modelo, se puede realizar con la información de la influencia de las muestras en el conjunto de calibración (leverage) y de los residuales, tanto la propiedad de interés como en la respuesta instrumental.

Leverage. Es una medida de la influencia de una muestra en relación al modelo. Muestras con alto valor de leverage están muy alejadas del centro del modelo, por lo que tendrán una influencia muy alta respecto al mismo.

Residuales en la propiedad. En la etapa de establecimiento del modelo se dispone del valor de la propiedad de interés determinado por el método de referencia. Los residuales en esta propiedad comparan el valor predicho por el modelo \hat{c} con el valor considerado verdadero, c .

$$f = c - \hat{c} \quad (\text{Ecuación 15})$$

En varias ocasiones la detección de outliers se realiza combinando las herramientas descritas, un buen ejemplo es la Figura 11 en la que se representa el residual (propiedad) frente al leverage de las muestras.

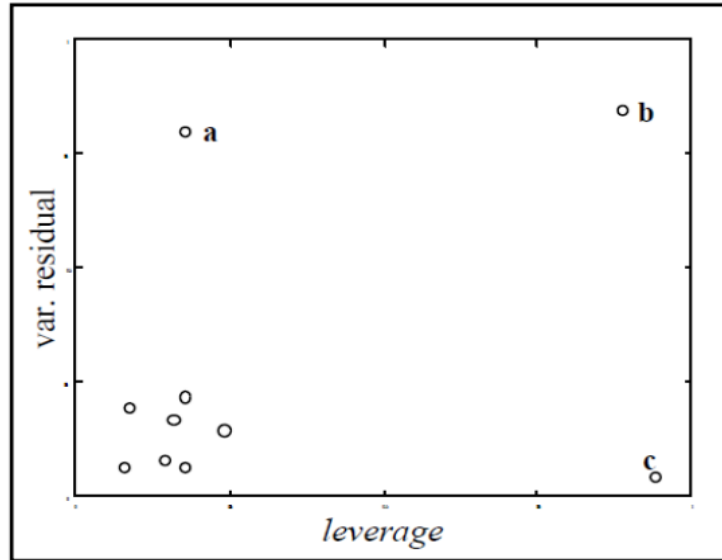


Figura 11. Gráfico de la varianza residual frente al leverage

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

Nota. Adaptado de: (MILLER & MILLER, 2002)

Otro de los valores estadísticos que pueden ser utilizados para la detección de datos *outlier* corresponde al **T² de Hotelling**. Este se obtiene a partir de la suma de los cuadrados de los scores y permite detectar si la variación incluida en las nuevas variables (PC o latentes) es más grande que la que le correspondería si solo influyeran variaciones aleatorias, su interpretación gráfica es como la de cualquier gráfico univariante, donde las muestras fuera de control aparecen más allá de la línea de control por poseer un valor de T² superior al límite.

- **Determinación del parámetro a modelar mediante el método de referencia**

Mediante metodologías analíticas adecuadas se determinan las concentraciones o propiedades a modelar en las muestras. Los métodos empleados deben proporcionar resultados precisos y exactos, ya que de éstos dependerá la bondad del modelo de calibración obtenido.

1.4 Estado del arte

En general, la espectroscopia de adsorción en el NIR, se ha empleado con éxito en la caracterización de distintas formas de materia orgánica y sus derivados, considerándose como una herramienta adecuada en el área de geoquímica orgánica, con la ventaja operativa de la rapidez y bajo costo de análisis (Kister, et al., 1996). A continuación, se discuten algunos de los trabajos encontrados en la literatura que evidencian la aplicación de NIR para la caracterización de crudos.

Varias muestras de crudos venezolanos con gravedades API entre 8 y 42, se caracterizaron a partir del análisis por Espectroscopia de Infrarrojo con Transformada de Fourier (FTIR). Los resultados indicaron que a partir de esta técnica se pueden obtener índices de composición que caractericen a los crudos con coeficientes de variación menores al 10%, lo que permite proponer su uso para estimar la gravedad API de crudos (Quintero, et al., 2014)

Por su parte, Aske et al. estudiaron la agregación de asfáltenos usando espectroscopia NIR sujetos a un PCA con el fin de detectar la presión de inicio de la agregación de asfáltenos (Aske, et al., 2002).

Hannisdal y colaboradores fraccionaron cuantitativamente en saturados, aromáticos, resinas y asfáltenos (SARA) veinte crudos pesados por precipitación de asfáltenos en n-hexano y cromatografía líquida de alta resolución (HPLC). Los crudos se caracterizaron por espectroscopia vibracional en las regiones NIR e IR medio. Se llevó a cabo un PCA de los conjuntos de datos. Se construyeron modelos de regresión PLS para cada componente SARA a partir de los datos de IR medio y NIR para predecir sus cantidades. Estos modelos se ajustaron satisfactoriamente con los datos experimentales de análisis NIR y mostraron una buena capacidad predictiva para la

composición del petróleo crudo. Los modelos de regresión de datos de IR no se modelaron correctamente para aromáticos y asfáltenos, pero sí modelaron bien los componentes saturados y resinas. Para la determinación SARA, la espectroscopia NIR parece ser una alternativa favorable frente al método de fraccionamiento que requiere más tiempo (Hannisdal, et al., 2005).

Falla y colaboradores (Falla, et al., 2006) propusieron una metodología para la estimación de las propiedades de Destilación simulada (SimDis) del petróleo crudo basados en espectroscopia NIR. Este análisis fue rápido en relación con la determinación True Boiling Point (TBP). En la metodología propuesta, se utilizó una sonda de transmisión NIR. Se emplearon cuarenta muestras de petróleo de grados API entre 31,1 y 36,4 con el fin de construir las correlaciones. Los datos espectrales NIR fueron correlacionados con las curvas SimDis (porcentaje en peso vaporizado a una temperatura dada) y el contenido de sal del petróleo mediante la técnica de redes neuronales. La red resultante reprodujo con bastante precisión las curvas SimDis.

En 2011, Lamus y colaboradores demostraron que la espectroscopia NIR puede predecir composiciones SARA, incluyendo el análisis de nafta y keroseno del crudo, con puntos de ebullición menores de 260°C. Se analizaron veinte tipos de crudos colombianos para el desarrollo de los modelos de calibración PLS. Esta metodología fue 500 veces más rápida que el método estándar. Los resultados mostraron una alta repetibilidad y una buena correlación con los datos del método por Cromatografía de absorción en Silica-Gel (Lamus, et al., 2011).

2. Metodología

A continuación, se presenta un esquema que resume la metodología desarrollada para cumplir los objetivos del proyecto (Figura 12).

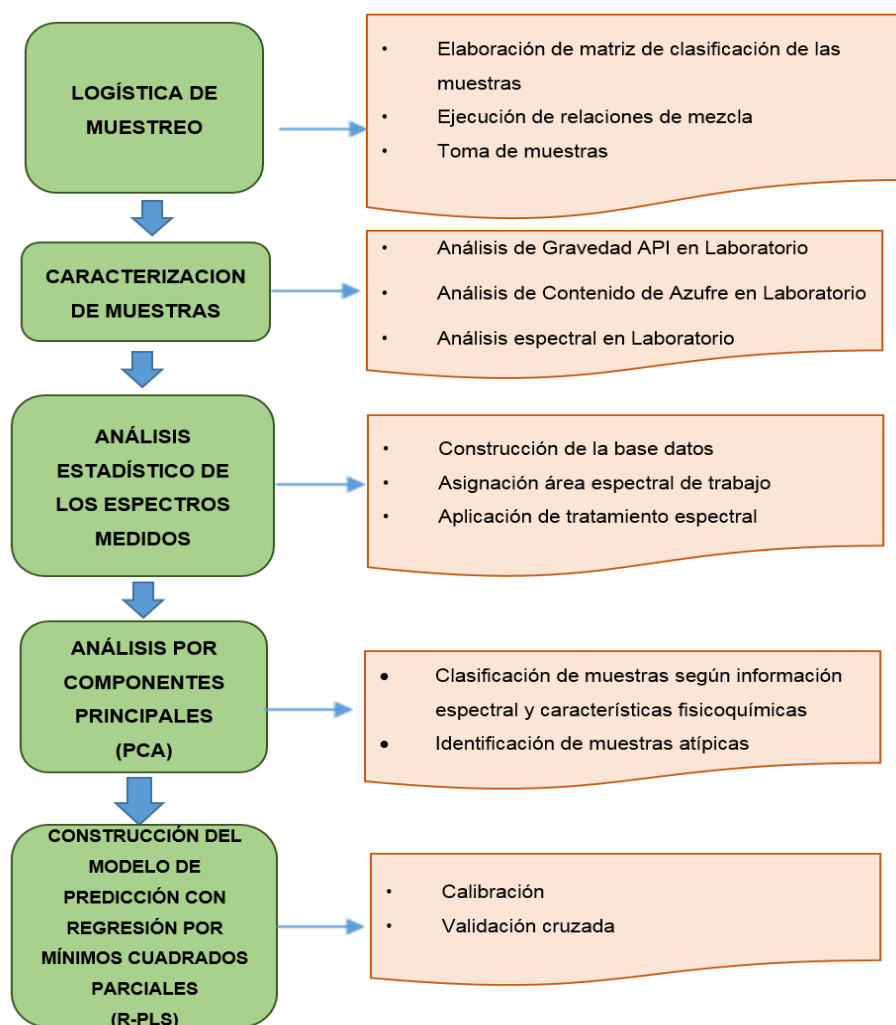


Figura 12. Esquema metodológico experimental

2.1 Logística de muestreo

2.1.1 Elaboración de matriz de clasificación de las muestras Se conformó una matriz con mezclas de crudo de las unidades de destilación cargadas en la refinería en un periodo de un año para la respectiva toma de muestras.

2.1.2 Ejecución de relaciones de mezcla De la matriz de mezclas se seleccionaron las más representativas (40 muestras) correspondientes a las recolectadas durante aproximadamente 90 días para cada unidad de destilación; en el Anexo A se presentan algunas gráficas de la estadística de las relaciones de mezcla para tres unidades de destilación.

2.1.3 Toma de muestras Se tomó una muestra por cada relación de mezcla representativa, la toma se realizó en recipientes nuevos y limpios de capacidad máxima de medio galón (1514 mL), los cuales se llenaron hasta el 80% de su capacidad en volumen y fueron etiquetados según la normatividad vigente en la GRB (ver Anexo B).

2.2 Caracterización de muestras

Se analizaron la gravedad API y el contenido de azufre (%p/p) para el conjunto total de 40 muestras de crudo utilizadas como componentes en la preparación de las mezclas de crudo que salen de CB8. Las propiedades mencionadas fueron determinadas por los métodos analíticos establecidos en los laboratorios de la GRB descritos en las normas ASTM D5002-13 y ASTM D4294-10 respectivamente (ASTM International, 2013) (ASTM International, 2010). Seguidamente las

muestras se analizaron por espectroscopia infrarroja cercana. Los espectros NIR se obtuvieron mediante la utilización de un equipo para análisis de muestras de crudo de línea **ABB** modelo **MB3000**, que presenta rango espectral de 485 cm^{-1} a 12000 cm^{-1} y una resolución de $0,7\text{ cm}^{-1}$. Este equipo se encuentra disponible en el laboratorio de la Refinería de Barrancabermeja (ABB, 2016). Para la medición de los NIR se siguió la norma ASTM E1655 -05 (2012)

2.3 Análisis estadístico de los espectros medidos

Pre-tratamiento de datos espectrales: En el programa *The Unscrambler X* versión 10.3 licenciado para su uso en Ecopetrol S.A. y la Universidad Industrial de Santander, se construyó una base de datos con los espectros IR de las 40 muestras analizadas. Se graficaron los espectros infrarrojo en conjunto y luego de una exploración visual se definió el área espectral para la predicción de cada propiedad. A continuación, se hizo una corrección de la línea base, normalización y derivada con el fin de detectar influencias de posibles errores instrumentales y humanos durante el manejo y adquisición espectral, además de mejorar la resolución de los mismos. En la sección 3.2 se especifica el desarrollo y resultado de cada pre-tratamiento aplicado en la respectiva zona espectral de trabajo.

2.4 Análisis por componentes principales (PCA)

Se realizó una descomposición por componentes principales al conjunto total de espectros con pre-tratamiento de tal forma que se determinaron los componentes (direcciones) que explican la máxima variabilidad de las muestras para utilizarlas posteriormente como nuevos ejes de

coordenadas (componentes principales), y así reducir la dimensionalidad de los datos obtenidos. Este análisis permitió discriminar las muestras según sus características fisicoquímicas por similitudes y/o diferencias entre las mismas, de igual forma con la gráfica de influencia en PCA se lograron identificar posibles muestras atípicas para el modelo de predicción de la gravedad API.

2.5 Construcción del modelo de predicción con regresión por mínimos cuadrados parciales

2.5.1 Calibración Empleando la regresión PLS se evaluó el potencial de la espectroscopia NIR para predecir la gravedad API y el contenido de azufre (%p/p) presente en las muestras de crudo. La calibración de cada modelo de predicción se realizó teniendo en cuenta el número óptimo de componentes principales y las regiones espectrales con longitudes de onda más representativas. La selección del modelo de predicción se realizó evaluando los parámetros estadísticos RMSEP, RMSECV y R^2 ; las gráficas de influencia y *leverage* resultantes del análisis fueron consideradas en la decisión para eliminar posibles muestras atípicas del conjunto de calibración.

2.5.2 Validación La validación de los modelos de predicción se desarrolló por validación cruzada. Este proceso permitió obtener predicciones independientes sin renunciar al uso de toda la información (total muestras). El principio del método consiste en dejar una parte de la información fuera del conjunto de calibración, se realiza el modelo con las muestras restantes y se predicen las muestras descartadas y así sucesivamente hasta calibrar y validar con cada una de las muestras.

3. Análisis de Resultados

3.1 Caracterización de muestras y adquisición de espectros

Un conjunto de 40 muestras de mezclas de crudos de diferentes campos petrolíferos de Colombia (ver Tabla 1) y de alta disponibilidad en la Gerencia Refinería de Barrancabermeja (GRB) de Ecopetrol fueron utilizadas para la finalidad del estudio. Las muestras se etiquetaron usando códigos de mezclas. Estos códigos permiten identificar los porcentajes de cada componente como: Liviana (L), Pesada (P), Caño Limón (C), Acida (A) y crudo de bajo punto de enfriamiento (LCT).

Tabla 1.

Relación de crudos analizados con valores de gravedad API y %p/p de Azufre

Crudo	Código Mezcla	Código Afinidad	°API	Contenido de azufre [% p/p]
MEZL4922	L60P40	L-P	21,8	1,23
MEZL4923	L55P45	L-P	21,5	1,23
MEZL4924	L75P25	L-P	22,5	1,19
MEZL4925	L100	L	23,5	1,16
MEZL4926	L50P20C30	L-C	24,4	1,01
MEZL4927	L55P15C30	L-C	24,1	1,02
MEZL4928	L85P15	L-P	26,6	1,16
MEZL4929	L75P25	L-P	22,5	1,19
MEZL4930	L85C15	L-C	23,4	1,13
MEZL4931	L75C25	L-C	24,0	1,09
MEZL4932	L55C45	L-C	24,9	1,00

Crudo	Código Mezcla	Código Afinidad	°API	Contenido de azufre [% p/p]
MEZL4933	L55P45	L-P	21,6	1,25
MEZL4934	L55P40A5	L-P-A	21,5	1,21
MEZL4935	L55P35A10	L-P-A	21,4	1,21
MEZL4936	L60P25A15	L-P-A	21,6	1,19
MEZL4937	L65P30A5	L-P-A	21,8	1,17
MEZL4938	L65P25A10	L-P-A	21,8	1,19
MEZL4939	L75C10A15	L-C-A	23,0	1,10
MEZL4940	L85A15	L-A	22,3	1,14
MEZL4941	L90A10	L-A	22,4	1,14
MEZL4942	L65P25A10	L-P-A	21,8	1,14
MEZL4943	L40P45A15	L-P-A	21,1	1,27
MEZL4944	L45P40C15	L-P-C	22,8	1,11
MEZL4945	L45P40C5A10	L-P-C-A	21,9	1,21
MEZL4946	L45P45A10	L-P-A	21,2	1,26
MEZL4947	L50P40A10	L-P-A	21,4	1,25
MEZL4948	L80P20	L-P	26,5	1,14
MEZL4949	L85A15	L-A	22,3	1,15
MEZL4950	L45P55	L-P	21,4	0,50
MEZL4951	P30LCT70	P-LCT	23,9	1,06
MEZL4952	P30LCT70	P-LCT	23,9	1,06
MEZL4953	L35LCT65	P-LCT	23,7	1,03
MEZL4954	P50LCT50	P-LCT	24,1	1,12
MEZL4955	LCT100	LCT	25,9	1,10
MEZL4956	P40LCT60	P-LCT	23,4	1,07
MEZL4957	A10LCT90	A-LCT	25,5	1,08
MEZL4958	A15LCT85	A-LCT	25,0	0,96
MEZL4959	A20LCT80	A-LCT	24,7	0,99
MEZL4960	A25LCT75	A-LCT	24,4	1,01
MEZL4961	P25LCT75	P-LCT	23,7	1,02

Crudo	Código Mezcla	Código Afinidad	°API	Contenido de azufre [% p/p]
		Valor Mínimo	21,10	0,50
		Valor Máximo	26,60	1,27
		Promedio	23,13	1,11
		Desviación Estándar	1,54	0,13

Las muestras presentan una distribución en un rango no muy amplio como lo señala la desviación estándar para cada propiedad con valores de 1,54 para la gravedad API y 0,13 para el %p/p de azufre.

3.1.1 Medidas espectroscópicas Los espectros NIR fueron tomados en el intervalo de 102 a 12000 cm^{-1} , a pesar de que las características de este tipo de espectros son bandas anchas y altamente superpuestas, se pueden identificar algunas vibraciones características. En el rango de 4500 a 4600 cm^{-1} se encuentran las bandas correspondientes a vibraciones de enlaces C=C (4600 - 4450 cm^{-1}) y C-H (3100 - 3000 cm^{-1}), ver figura 8. Las bandas de combinaciones de enlaces C-H+C-H (alifáticos o aromáticos) se ubican en el intervalo de 4000 a 4600 cm^{-1} .

La figura 13 presenta superpuestos los espectros NIR de todas las muestras estudiadas. Se observan dos bandas fuertes de absorción 4256.64 cm^{-1} y 4329.93 cm^{-1} correspondientes al primer sobretono C-H y primera banda de combinación, respectivamente.

Para correlacionar la información espectral con la gravedad API se seleccionó la región entre 4000 y 4602 cm^{-1} teniendo en cuenta los resultados aportados en el trabajo realizado por (ALVES, 2012), el uso de este rango espectral presenta bandas asignadas a combinaciones de modos vibracionales de los enlaces CH-metilo y grupos metilenos y a enlaces CH-anillos aromáticos.

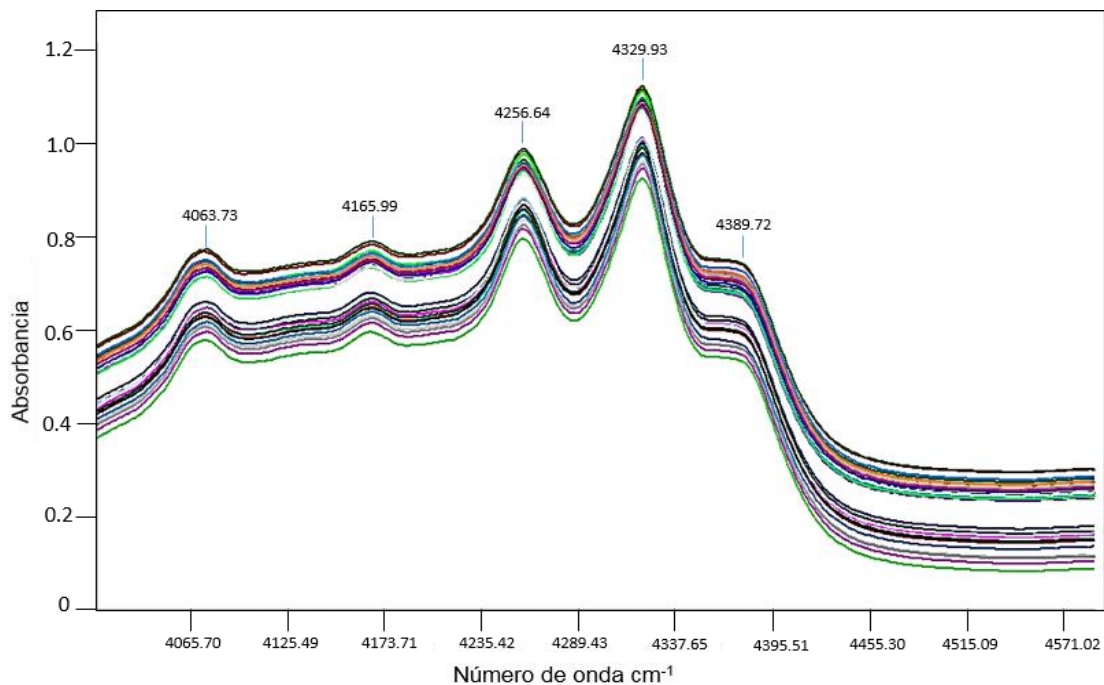


Figura 13. Región espectral NIR sin tratamiento para la predicción de la gravedad API

Por otra parte, para la predicción del porcentaje de azufre, se seleccionó la banda ubicada entre 3400 y 3500 cm^{-1} perteneciente a la región del espectro entre 3140 y 4000 cm^{-1} . La Figura 14 muestra los espectros infrarrojo de las muestras estudiadas para la región mencionada. La selección de las regiones mencionadas en el conjunto total de espectros se muestra en el Anexo C.

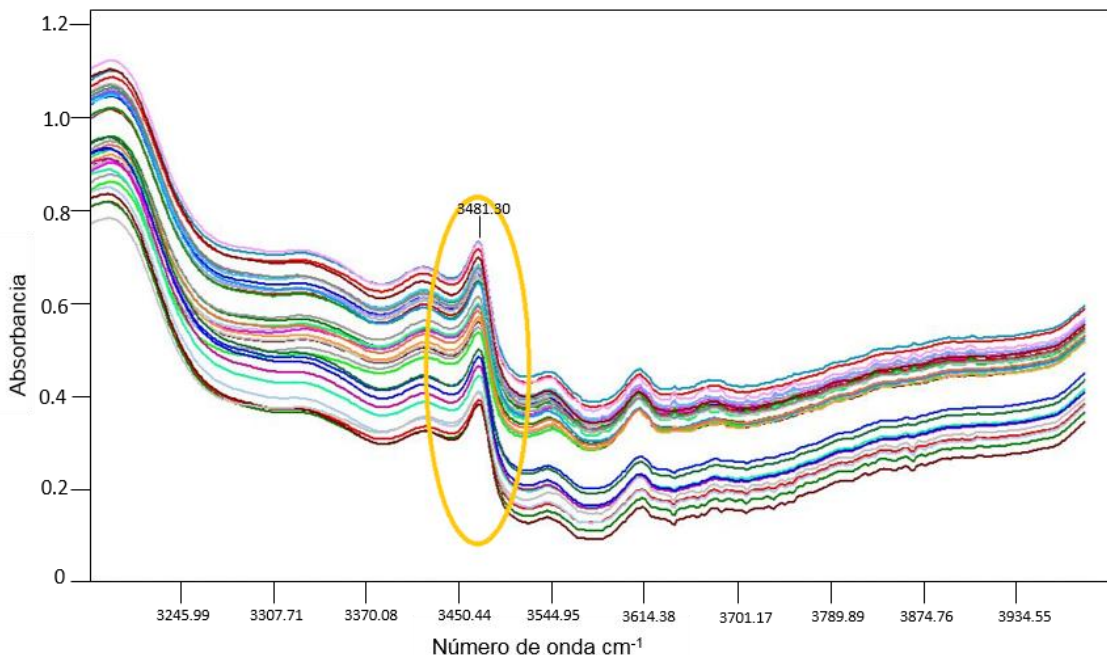


Figura 14. Región espectral NIR sin tratamiento para la predicción del %p/p de azufre.

A continuación se describe el pretratamiento aplicado a cada región del espectro seleccionada, para la predicción de la gravedad API y del contenido en porcentaje de azufre.

3.2 Análisis estadístico de los espectros medidos

Con la finalidad de corroborar la calidad de los espectros a analizar, inicialmente se realizó una exploración visual del conjunto de espectros NIR correspondiente a las 40 muestras de crudos con el fin de detectar agrupaciones, posibles errores durante la toma espectral y/o muestras con notables diferencias.

Nota: el software quimiométrico asociado al instrumento NIR utilizado realiza el centrado y autoescalado de los datos por defecto.

3.2.1 Región para la predicción de °API En la región espectral para la predicción de la gravedad API, primero se hizo una corrección de la línea base usando una horizontal del tipo *Baseline Offset and Linear Baseline Correction*, de acuerdo a las recomendaciones del manual del programa *The Unscrambler* (CAMO Software , 2012). El siguiente paso fue efectuar una normalización tipo unidad vector para luego aplicar una derivada de primer orden usando el algoritmo de Savitzky-Golay tomando como curva de ajuste un polinomio de segundo orden aplicado sobre cada 6 puntos. Los espectros pretratados de acuerdo con este procedimiento se presentan en la Figura 15.

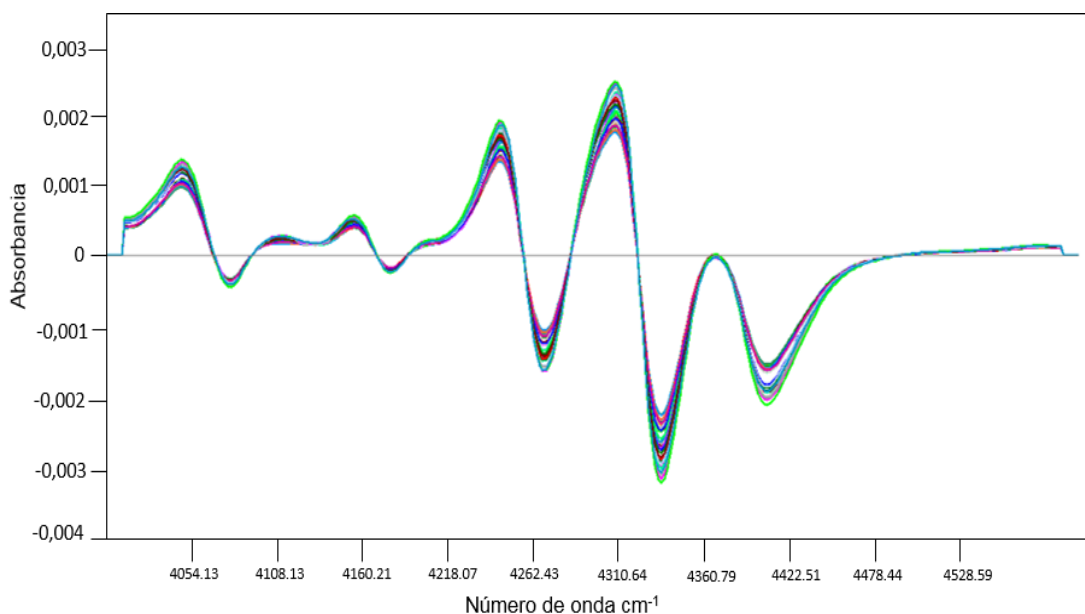


Figura 15. Región espectral NIR con pretratamiento para predecir la gravedad API

3.2.2 Región para la predicción del %p/p de azufre La Figura 16 presenta los espectros transformados; por el procedimiento mencionado anteriormente para el pretratamiento de los datos de la región espectral seleccionada para analizar los °API, para la región infrarroja de interés para el análisis del %p/p de azufre.

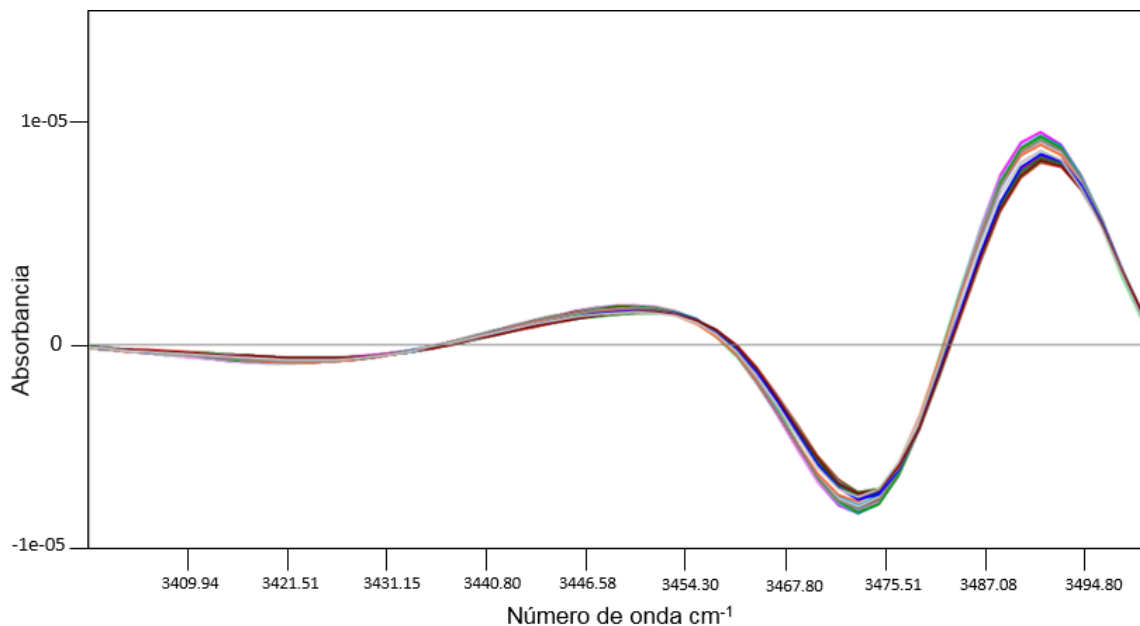


Figura 16. Región espectral NIR con tratamiento para la predicción del %p/p de azufre

En seguida, se presenta el análisis por componentes principales realizado sobre la región del espectro con pretratamiento.

3.3 Análisis por componentes principales (PCA) en la región de predicción de la gravedad API

La Tabla 2 relaciona los valores de las primeras cuatro componentes principales, de las cuales es posible explicar cerca del 99,14% de la variación en los datos espectrales con solamente las dos primeras. A partir del tercer componente el aumento en la variancia es mínimo. Cada una de estas componentes principales contiene cierta información que hace posible la clasificación de las muestras a partir de su espectro NIR (Figura 17).

Tabla 2.

Varianza asociada a cada uno de los componentes principales estudiados

<i>Componente Principal</i>	<i>Varianza Explicada (%)</i>	<i>Varianza Explicada Acumulada (%)</i>
PC 1	94,55	94,55
PC 2	4,59	99,14
PC 3	0,56	99,70
PC 4	0,03	99,73

La Figura 17 presenta un gráfico de puntuaciones de las componentes principales PC1 y PC2 que, como se comentó antes, explicaron respectivamente el 95% y el 4% de la varianza del conjunto de datos.

La gráfica identifica el grupo de las cuarenta muestras por colores y símbolos de acuerdo con los códigos definidos en la Tabla 1. Se encontró que las muestras se dividen en tres grupos de acuerdo con su porcentaje en composiciones similares, esto es: mezclas con mayor porcentaje de crudo liviano (círculo azul), mezclas de composición equitativa de crudos livianos y pesados “Medias” (círculo morado) y mezclas de crudos con mayor porcentaje de crudo tipo LCT (círculo verde).

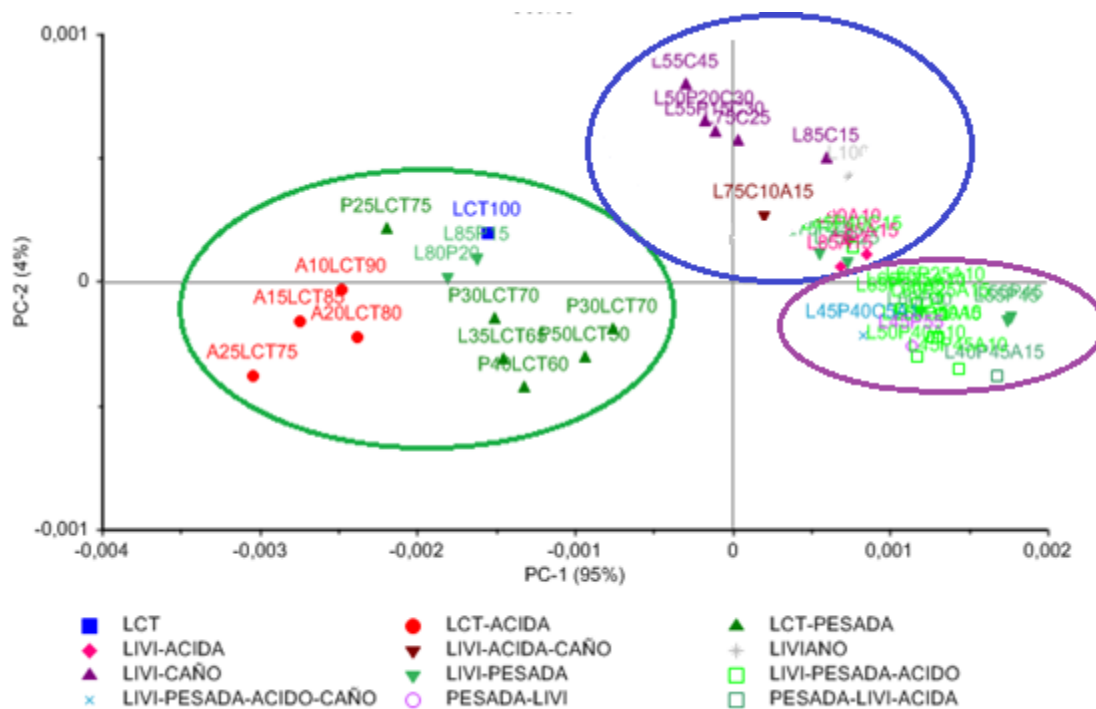


Figura 17. Gráfica de puntuación para PC1 y PC2 de las mezclas en la región de 4000 a 4500 cm^{-1}

1

A continuación, se presenta el análisis de la gráfica de influencia como herramienta para la identificación de posibles muestras atípicas. Este paso es importante, ya que al conocer estas muestras se puede validar su exclusión en la construcción del modelo de predicción.

La Figura 18 es la gráfica de influencia correspondiente a los datos analizados. En general, se encontraron dos muestras de crudo clasificadas como posibles *dangerous outliers*: **L85P15** y **L80P20**, por su alta varianza residual y alto *leverage*; por lo tanto las mezclas de crudo L85P15 y L80P20 deber ser consideradas como posibles muestras a excluir en la construcción del modelo de predicción PLS. Por otro lado, se observó un alto valor de *leverage* para la muestra **L55C45** lo cual explica su diferencia en la clasificación PCA en la gráfica de puntajes (Figura 17).

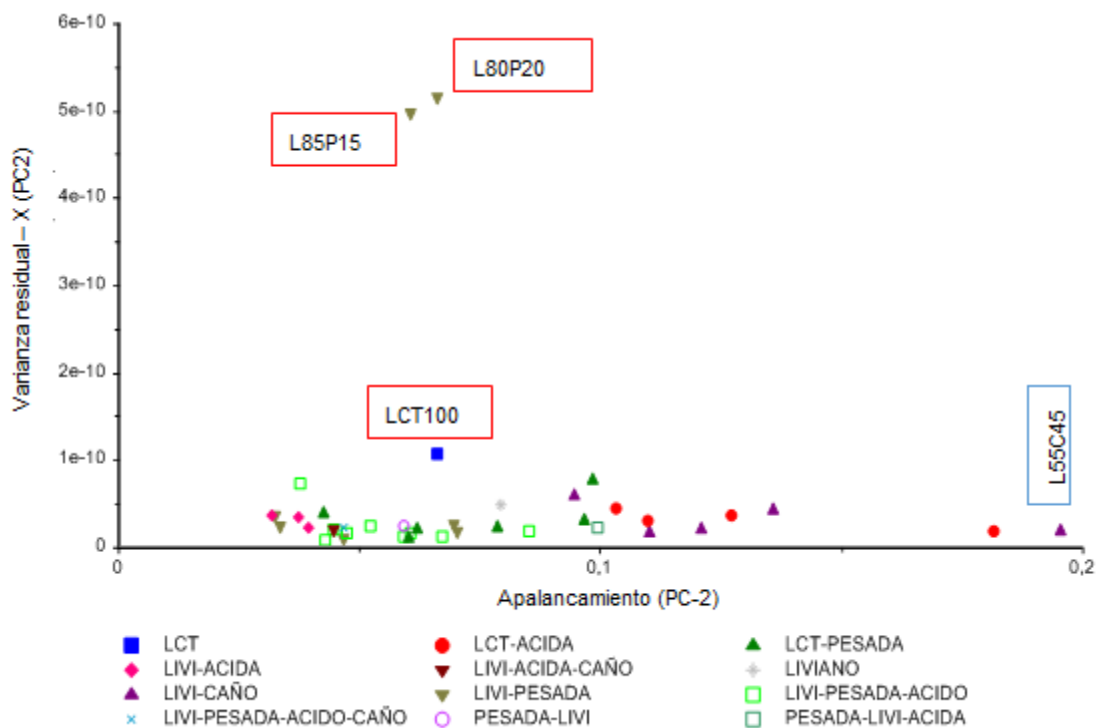


Figura 18. Gráfica de Influencia en PCA (región de 4000 a 4500 cm^{-1})

El análisis por componentes principales permitió clasificar el conjunto de muestras en tres grupos, debido a la similitud en la composición porcentual de cada mezcla de crudo. Además, se identificaron 2 muestras atípicas: **L85P15** y **L80P20**, consideradas para ser eliminadas durante la construcción del modelo de predicción de la gravedad API, teniendo en cuenta que este criterio se confirmará con otras herramientas estadísticas en cada etapa del proceso de predicción PLS.

En la siguiente sección, se describen las etapas de calibración y validación de la construcción del modelo de predicción para la gravedad API, el análisis muestras excluidas, de las gráficas de varianza, error de calibración y validación y por último un resumen de los parámetros estadísticos correspondiente al modelo.

3.4 Modelo de predicción (R-PLS) para la gravedad API

Mediante el algoritmo de regresión PLS se evaluó el potencial de la espectroscopia NIR para predecir la gravedad API de un conjunto inicial de 40 muestras de mezclas de crudos colombianos. Sin embargo, de la identificación de muestras discrepantes durante la construcción del modelo se eliminaron tres muestras; más adelante se detalla esta información.

El modelo de calibración se desarrolló en la región entre 4000 y 4600 cm^{-1} de los espectros NIR al conjunto total de muestras con rango de valores de gravedad API dentro del intervalo entre 21,10 y 26,60 $^{\circ}\text{API}$. La Figura 19 muestra la varianza explicada por el modelo propuesto sobre la variable de respuesta, que en este caso es la gravedad API. Las tres primeras variables latentes (VL) o Factores, explican el 96,20 % de la varianza observada en la variable de respuesta para la calibración (línea azul) y el 97,67% para la validación cruzada (línea roja). A partir de este componente el aumento en la varianza explicada es mínimo.

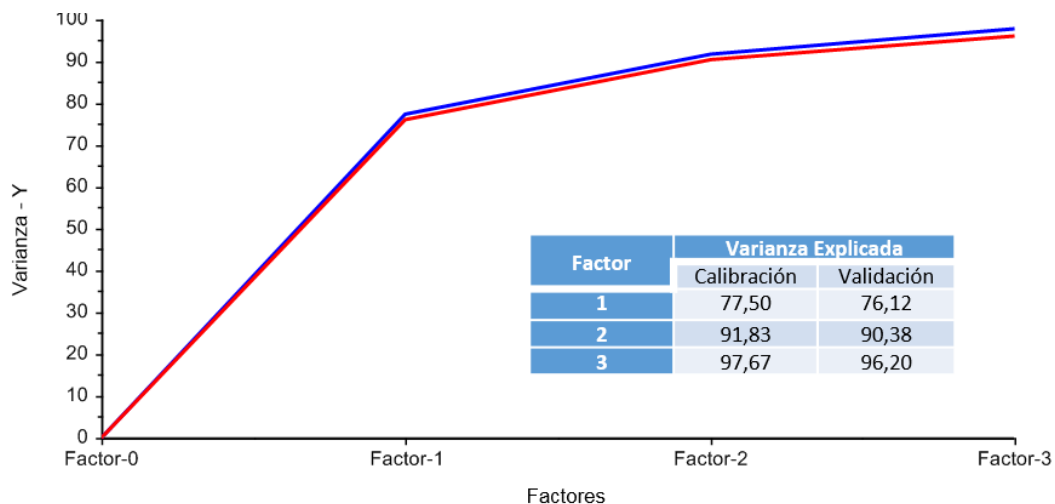


Figura 19. Varianza explicada en el modelo PLS para predicción de gravedad API

Para una mejor selección del número apropiado de variables latentes fue necesario evaluar los parámetros estadísticos de la variación de los errores de calibración (RMSEC) y validación (RMSECV). La Figura 20 indica que el error de estimación a partir del tercer factor, varía en muy poca proporción en valores de 0.19 y 0.26 para el RMSEC y el RMSECV, respectivamente. Este análisis sugiere que tres es un número recomendable de variables latentes para el modelo de estimación.

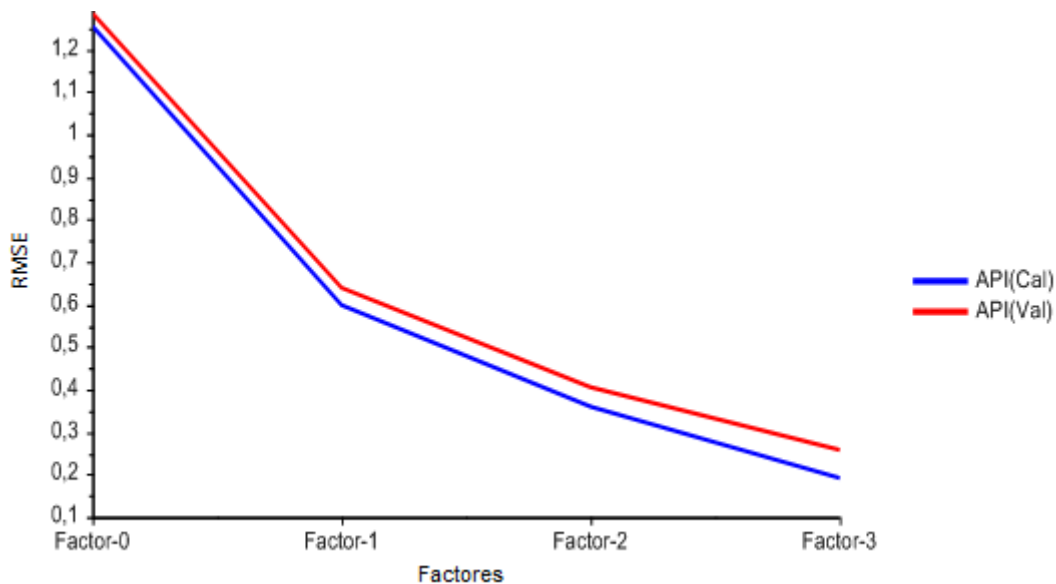


Figura 20. Error calculado en función del número de VL para el modelo de la gravedad API

3.4.1 Análisis de muestra excluidas Durante la construcción del modelo de predicción fue necesario eliminar tres muestras. Las mezclas de crudo con códigos: L85P15 y L80P20, fueron identificadas como muestras atípicas en la sección que describe el análisis por componentes principales y la tercera muestra (LCT100) se identificó por el análisis de la herramienta estadística de Hotelling T^2 que presenta a esta muestra en el factor 3 con una falta de ajuste para el modelo de predicción (Figura 21). En el Anexo D se presenta el gráfico de influencia en 3D del proceso

de construcción del modelo, que también permite identificar la muestra LCT100 como potencial *outlier*.

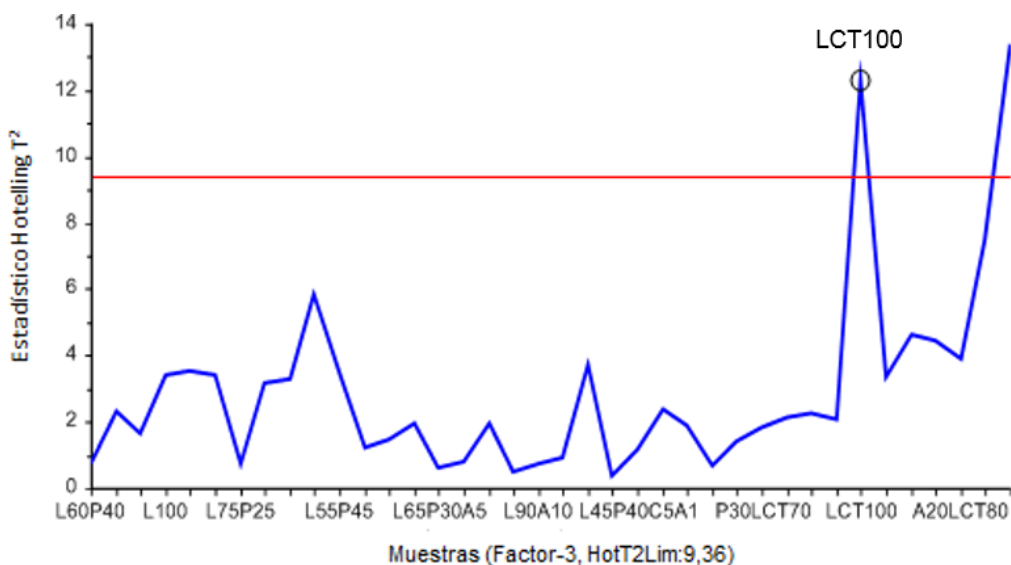


Figura 21. Gráfico del estadístico de Hotelling T^2 para la gravedad API

Se encontró una relación significativa entre las muestras eliminadas al presentar los tres valores más altos de gravedad API dentro del conjunto de las 40 muestras de mezclas de crudo (ver Tabla 1 – color amarillo).

El ajuste del modelo PLS fue mejorado al omitir progresivamente cada una de las muestras atípicas, reflejándose en la reducción del error tanto en la calibración como en la validación. El coeficiente de correlación (R^2) entre los valores de referencia y de predicción hallado, fue superior a 0.977 para la curva de calibración (color azul) y de 0.962 para la de validación (color rojo). El error estándar de predicción encontrado para el conjunto de muestras de validación fue de 0.258. Lo anterior valida la regresión PLS para la predicción de la gravedad API en el presente trabajo (Figura 22).

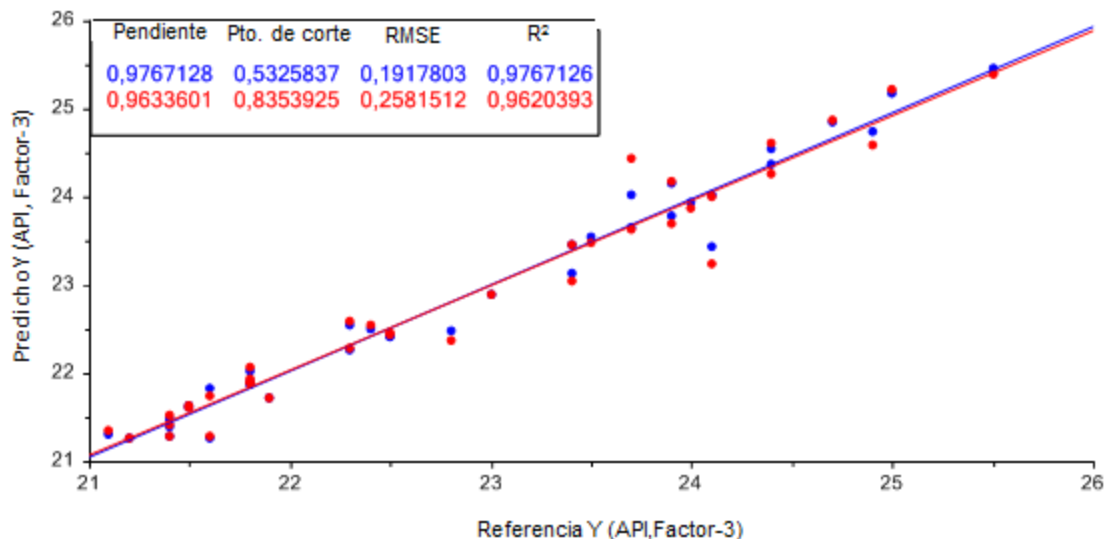


Figura 22. Curva de calibración y validación del modelo de predicción de la gravedad API

La Tabla 3 muestra los parámetros estadísticos de desempeño correspondientes al modelo de regresión obtenido para las 37 muestras de crudos con gravedad API entre 21,10 y 25,50.

Tabla 3.

Parámetros estadísticos del modelo de predicción de la gravedad API en muestras de mezclas de crudos

Región espectral (cm ⁻¹)	Variables X	Muestras excluidas	VL	RMSEC	RMSECV	R ²
4000- 4600	313	3(L85P15,L80P20,LCT100)	3	0,192	0,258	0,96

La validación del modelo de predicción de la gravedad API en mezclas de crudos se desarrolló utilizando el mismo grupo de muestras del conjunto para obtener el modelo de calibración (validación cruzada). El modelo propuesto mostró un desempeño adecuado para la predicción de

la gravedad API en los crudos de interés, al presentar valores muy similares entre los valores predichos y los de referencia con desviación estándar de 0.74. En la Tabla 4 se muestra el error relativo definido como: $\%ER = |((Referencia - Predicho)/Referencia) \times 100|$

Tabla 4.

Validación cruzada del modelo de predicción de la gravedad API

No.	Muestra	Referencia (°API)	Predicho (°API)	Residual	Error Relativo (%)
1	L60P40	21,80	21,89	0,09	0,41
2	L55P45	21,50	21,62	0,12	0,56
3	L75P25	22,50	22,45	-0,05	0,22
4	L100	23,50	23,49	-0,01	0,04
5	L50P20C30	24,40	24,27	-0,13	0,53
6	L55P15C30	24,10	24,00	-0,10	0,41
7	L75P25	22,50	22,43	-0,07	0,31
8	L85C15	23,40	23,45	0,05	0,21
9	L75C25	24,00	23,88	-0,12	0,50
10	L55C45	24,90	24,59	-0,31	1,24
11	L55P45	21,60	21,75	0,15	0,69
12	L55P40A5	21,50	21,62	0,12	0,56
13	L55P35A10	21,40	21,41	0,01	0,05
14	L60P25A15	21,60	21,29	-0,31	1,44
15	L65P30A5	21,80	21,93	0,13	0,60
16	L65P25A10	21,80	22,06	0,26	1,19
17	L75C10A15	23,00	22,90	-0,10	0,43
18	L85A15	22,30	22,58	0,28	1,26
19	L90A10	22,40	22,54	0,14	0,63
20	L65P25A10	21,80	21,89	0,09	0,41
21	L40P45A15	21,10	21,34	0,24	1,14
22	L45P40C15	22,80	22,37	-0,43	1,89

No.	Muestra	Referencia (°API)	Predicho (°API)	Residual	Error Relativo (%)
23	L45P40C5A10	21,90	21,72	-0,18	0,82
24	L45P45A10	21,20	21,27	0,07	0,33
25	L50P40A10	21,40	21,28	-0,12	0,56
26	L85A15	22,30	22,29	-0,01	0,04
27	L45P55	21,40	21,52	0,12	0,56
28	P30LCT70	23,90	24,18	0,28	1,17
29	P30LCT70	23,90	23,69	-0,21	0,88
30	L35LCT65	23,70	23,64	-0,06	0,25
31	P50LCT50	24,10	23,24	-0,86	3,57
32	P40LCT60	23,40	23,05	-0,35	1,50
33	A10LCT90	25,50	25,39	-0,11	0,43
34	A15LCT85	25,00	25,22	0,22	0,88
35	A20LCT80	24,70	24,86	0,16	0,65
36	A25LCT75	24,40	24,62	0,22	0,90
37	P25LCT75	23,70	24,42	0,72	3,04
<i>Desviación Estándar</i>					0,74

Los resultados del análisis por componentes principales en la región de predicción del %p/p de azufre se presentan a continuación, y el desarrollo del modelo PLS se muestra en la sección siguiente.

3.5 análisis por componentes principales (PCA) en la región de predicción del %P/P de azufre

De la base inicial de 40 muestras, se descartaron las tres identificadas en el análisis de muestras excluidas del anterior modelo (LCT100, L85P15, L80P20) y las otras dos fueron eliminadas por

presentar réplicas en la toma de sus espectros infrarrojo (L75P25, P30LCT70) .Se analizó el grupo de 35 muestras de mezclas de crudos por componentes principales en la región de 3400 a 3500 cm^{-1} . Los porcentajes de varianzas explicadas para las tres primeras componentes se muestran en la Tabla 5. Con el primer componente principal se explica más del 99.92% de la varianza. A partir del segundo componente, el aumento de la varianza es menor a 0.07%.

Tabla 5.

Varianza de los componentes principales PCA

<i>Componente Principal</i>	<i>Varianza Explicada (%)</i>	<i>Varianza Explicada Acumulada (%)</i>
PC 1	99,92	99,92
PC 2	0,07	99,99
PC 3	0,00	99,99

El gráfico de puntuaciones de las 35 muestras de mezclas de crudos se presenta en la Figura 23. Se conformaron tres grupos de acuerdo a las tendencias observadas: Livianas (azul), Medias (verde) y LCT (rojo).

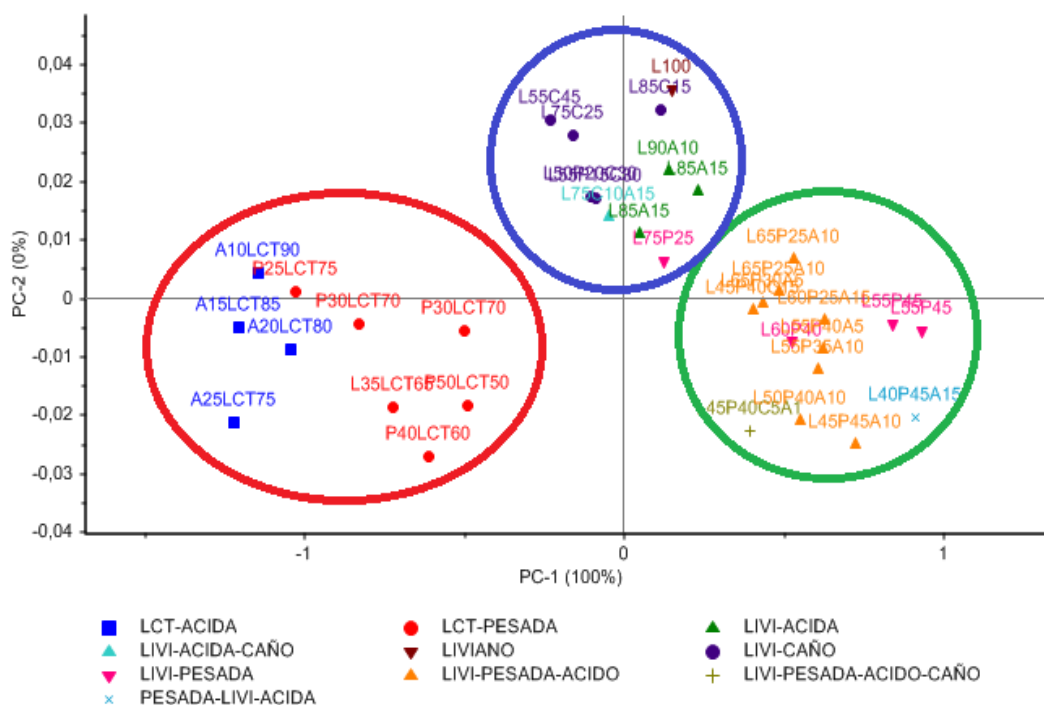


Figura 23. Gráfica de puntuación para PC1 y PC2 de las mezclas en región 3400 a 3500 cm^{-1}

La descomposición de los datos de infrarrojo por componentes principales permitió identificar claramente las muestras analizadas de acuerdo con su composición porcentual de mezcla, descrita por la región espectral elegida como objeto de estudio.

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

3.6 Modelo de predicción (R-PLS) del %p/p de azufre

La etapa de calibración se desarrolló en el rango espectral de 3400 a 3500 cm^{-1} para las muestras con rango de valores de %p/p de azufre entre 0,50 y 1,27. Los resultados de la regresión indican que con cinco variables latentes (VL) o factores, es posible explicar el 91,80% de la variabilidad de los datos de referencia (Figura 24).

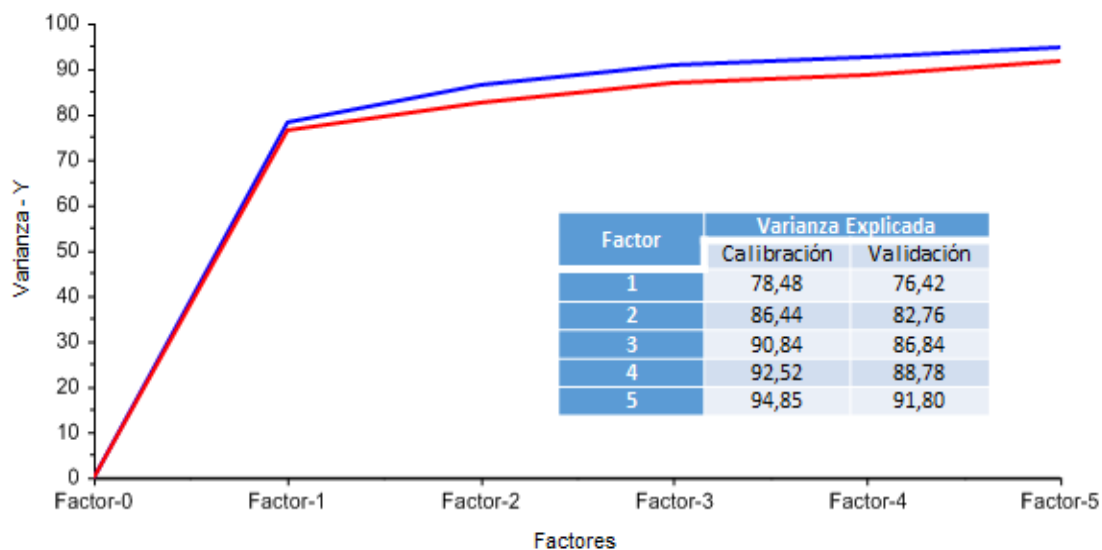


Figura 24. Varianza explicada en el modelo PLS para predicción del %p/p de azufre

3.6.1 Análisis de muestra excluidas Del conjunto inicial de 35 muestras se eliminaron tres: L45P55, A10LCT90 y L85C15, la mezcla de crudo L45P55 se descartó al comienzo del proceso al ser identificada como muestra discrepante con el menor valor de la propiedad a predecir (0.4981 %p/p de azufre). Las muestras con código A10LCT90 y L85C15, fueron eliminadas durante del desarrollo del modelo de predicción por tener alta varianza residual en Y; la Figura 25 presenta el gráfico del estadístico de Hotelling T^2 que permitió seleccionar cada una de las muestras para su respectiva exclusión. En el Anexo E con la gráfica de influencia en 3D se puede verificar la identificación de estas muestras discrepantes para el modelo de predicción del azufre.

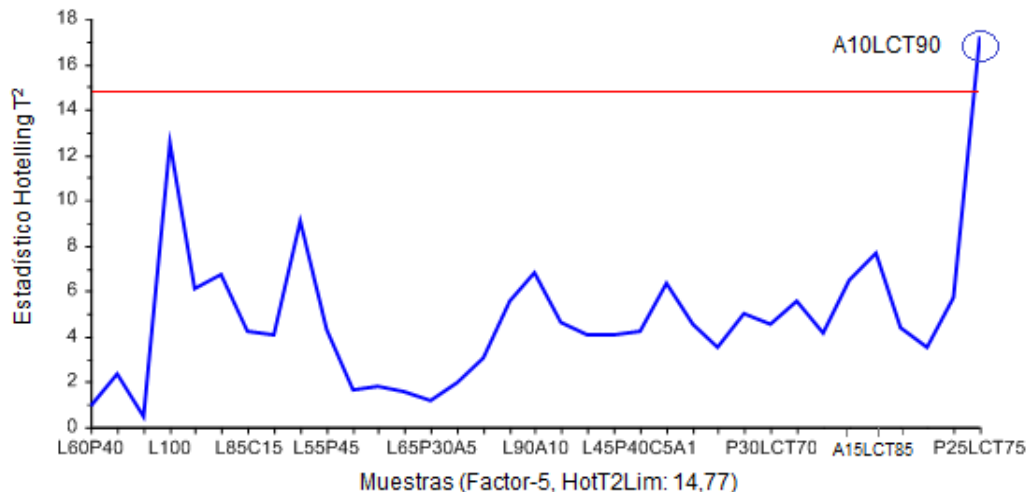


Figura 25. Gráfico del estadístico de Hotelling T^2 para el %p/p de azufre

Para las muestras A10LCT90 y L85C15, se puede inferir una posible diferencia en su caracterización fisicoquímica del conjunto total de muestras, pues, aunque sus valores de %p/p de azufre se sitúan en el rango medio del grupo, la información espectral y relación con la propiedad en predicción no aportan al ajuste del modelo.

Es necesario e importante excluir estas muestras discrepantes en el modelo para minimizar la degradación de la capacidad predictiva del mismo.

De esta forma, el modelo fue ajustado al omitir de forma secuencial cada una de las muestras atípicas, reflejándose en la reducción del error tanto en la calibración como en la validación. La Figura 26 muestra las curvas superpuestas de calibración y validación, el coeficiente de correlación (R^2) entre los valores de referencia y de predicción hallado fue de 0,95 para la curva de calibración (color azul) y de 0,92 para la de validación (color rojo). El error estándar de predicción encontrado para el conjunto de muestras de validación fue de 0,026. Lo anterior valida la regresión PLS para la predicción del %p/p de azufre en las muestras de mezclas de crudos en la región espectral de 3400 a 3500 cm^{-1} de espectros NIR.

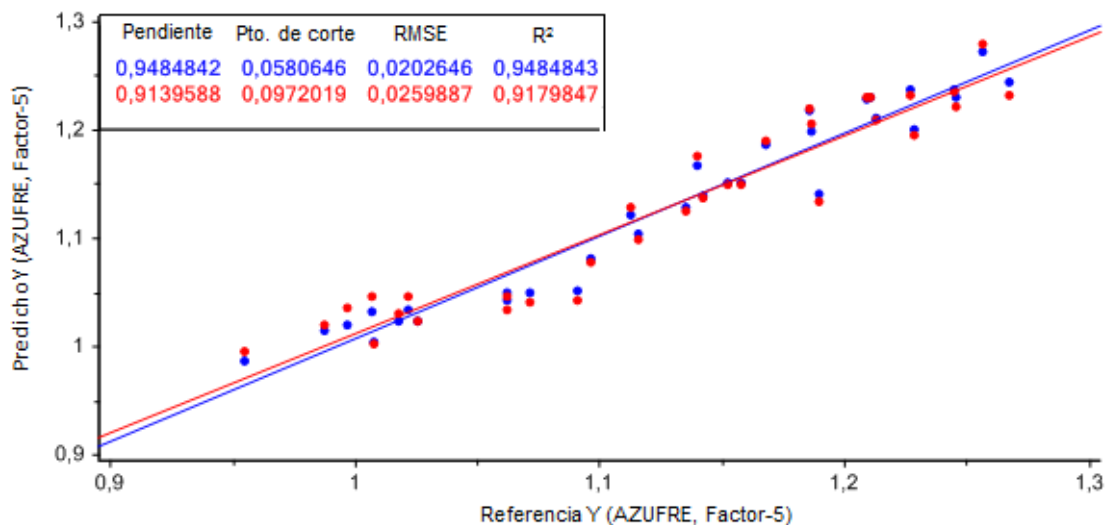


Figura 26. Curva de calibración y validación del modelo de predicción del %p/p de azufre

La Tabla 6 muestra el resumen de los parámetros estadísticos de desempeño correspondientes al modelo de regresión obtenido para las 32 muestras de crudos con %p/p de azufre entre 0,96 y 1,27.

Tabla 6.

Parámetros estadísticos del modelo de predicción del %p/p de azufre en muestras de mezclas de crudos

Región espectral (cm ⁻¹)	Variables X	Muestras excluidas	VL	RMSEC	RMSECV	R ²
3400- 3500	53	3(L45P55,A10LCT90, L85C15)	5	0,020	0,026	0,92

La validación del modelo de predicción del %p/p de azufre en mezclas de crudos se desarrolló utilizando el mismo grupo de muestras del conjunto para obtener el modelo de calibración

(validación cruzada). El desempeño del modelo construido se puede considerar satisfactorio al presentar un valor de error de validación de 0.026, significativamente bajo. Los valores predichos y los de referencia de la propiedad modelada se presentan en la Tabla 7.

Tabla 7.

Validación cruzada del modelo de predicción del %p/p de azufre

No.	Muestra	Referencia (%p/p de S)	Predicho (%p/p de S)	Residual	Error Relativo (%)
1	L60P40	1,23	1,20	-0,03	2,74
2	L55P45	1,23	1,23	0,00	0,34
3	L75P25	1,19	1,13	-0,06	4,70
4	L100	1,16	1,15	-0,01	0,76
5	L50P20C30	1,01	1,05	0,04	3,78
6	L55P15C30	1,02	1,03	0,01	1,12
7	L75C25	1,09	1,04	-0,05	4,44
8	L55C45	1,00	1,03	0,04	3,75
9	L55P45	1,25	1,22	-0,02	1,99
10	L55P40A5	1,21	1,23	0,02	1,48
11	L55P35A10	1,21	1,23	0,02	1,71
12	L60P25A15	1,19	1,22	0,03	2,83
13	L65P30A5	1,17	1,19	0,02	1,76
14	L65P25A10	1,19	1,20	0,02	1,46
15	L75C10A15	1,10	1,08	-0,02	1,75
16	L85A15	1,14	1,14	-0,01	0,55
17	L90A10	1,14	1,12	-0,01	0,96
18	L65P25A10	1,14	1,18	0,04	3,14
19	L40P45A15	1,27	1,23	-0,04	2,81
20	L45P40C15	1,11	1,13	0,02	1,36
21	L45P40C5A10	1,21	1,21	0,00	0,32

No.	Muestra	Referencia (%p/p de S)	Predicho (%p/p de S)	Residual	Error Relativo (%)
22	L45P45A10	1,26	1,28	0,02	1,75
23	L50P40A10	1,25	1,24	-0,01	0,77
24	L85A15	1,15	1,15	0,00	0,31
25	P30LCT70	1,06	1,05	-0,02	1,53
26	L35LCT65	1,03	1,02	0,00	0,18
27	P50LCT50	1,12	1,10	-0,02	1,56
28	P40LCT60	1,07	1,04	-0,03	2,91
29	A15LCT85	0,96	0,99	0,04	4,11
30	A20LCT80	0,99	1,02	0,03	3,17
31	A25LCT75	1,01	1,00	-0,01	0,59
32	P25LCT75	1,02	1,05	0,02	2,36
<i>Desviación Estándar</i>					1,29

4. Conclusiones

La espectroscopia NIR combinada con técnicas de calibración multivariable como el PCA y la regresión PLS, permitió clasificar las muestras de mezclas de crudos y la obtención de un modelo para la predicción de la gravedad API y otro para la predicción del %p/p de azufre, ambos con desempeño satisfactorio; la validación cruzada aplicada a cada modelo desarrollado en la región espectral determinada según este trabajo, presentó un valor del RMSECV aceptable por comparación con la reproducibilidad del método de ensayo (Norma ASTM) para las muestras en el laboratorio de la refinería, estos modelos pueden ser una alternativa a las técnicas convencionales empleadas comúnmente e incrementar la reproducibilidad de los métodos en un menor tiempo de análisis.

5. Recomendaciones

- Para una posterior aplicación del modelo desarrollado en este trabajo, se deben emplear más muestras de la misma naturaleza. Además, las señales espectroscópicas deben ser obtenidas bajo las mismas condiciones y parámetros instrumentales con el fin de construir modelos de predicción más robustos.
- Es necesario evaluar la reproducibilidad en el tiempo del modelo utilizando nuevas muestras y contribuir así, en la actualización de la predicción de parámetros referentes al control de calidad de los productos de la GRB.
- Los resultados de este trabajo muestran el gran potencial de la espectroscopia NIR para la caracterización de la gravedad API, el porcentaje en peso de azufre y posibilita la implementación para la predicción de otras propiedades de mezclas de crudos colombianos.

Referencias Bibliográficas

ABB, 2016. *Espectrómetro Versátil FT-IR de laboratorio*. [Online] Available at: <http://new.abb.com/products/measurement-products/analytical/ft-ir-and-ft-nir-analyzers/laboratory-spectrometers/mb3000>

Abbas, O. a., 2008. *Assessing petroleum oils biodegradation by chemometric analysis of spectroscopic data*. s.l.:s.n.

Alves, J. C. H. C. B. P. R. J., 2012. *Determination of diesel quality parameter using support vector regression and near infrared spectroscopy for an in – line blending optimizer system*. s.l.:s.n.

American Society for Testing and Materials., 2005. *Standard practices for infrared multivariate quantitative analysis*. Philadelphia: ASTM.

Ancheyta, J. & Speight, J. G., 2007. *Hydroprocessing of heavy oils and residua (en inglés)*. CRC Press. p. 125.. Boca Ratón, Florida, EUA: CRC Press.

Ancheyta, J. & Speight, J. G., 2007. *Wikipedia*. [Online] Available at: <http://es.wikipedia.org/wiki/Petróleo>

AS, C. S., 2016. [Online] Available at: <http://www.camo.com/rt/Products/Unscrambler/unscrambler.html> [Accessed 25 06 2016].

Aske, N., Kallevik, H., Johnsen, E. E. & Sjöblom, J., 2002. Asphaltene Aggregation from Curde Oils and Model Systems Studied by High-Pressure {NIR} Spectroscopy. *Energy Fuels*, 16(7), pp. 1287 -1295.

ASTM International, 2010. *ASTM D4294-10*. [Online] Available at: <http://www.astm.org/Standards/D4294.htm>

ASTM International, 2013. *ASTM D5002-13*. [Online]
Available at: <http://www.astm.org/search/fullsite-search.html?query=D5002&>

ASTM International, 2016. *http://www.astm.org*. [Online]
Available at: <http://www.astm.org/search/fullsite-search.html?query=e1655&resStart=0&resLength=10&>

ASTM, 2013. *D5002 -13*. [Online]
Available at: <http://www.astm.org/search/fullsite-search.html?query=D5002&>

ASTM, 2015. *ASTM International*. [Online]
Available at: http://www.astm.org/ABOUT/full_overview.html

ASTM, 2015. *Form and style for ASTM standards [Forma y estilo de las normas ASTM] (en inglés)*. West Conshohocken, PA, EUA: ASTM International.

ASTM, 2015. *Standard Guide for Proficiency Testing by Interlaboratory Comparisons*. s.l.:s.n.

Balabin, R. M. & Safieva, R. Z., 2008. Gasoline classification by source and type based on near infrared (NIR) spectroscopy data. *Fuel*, Volume 87, pp. 1096 -1101.

Baldrich, C. & Novoa, L. A., 2005. Infrared spectrophotometry, a rapid and effective tool for characterization of directdistillation naphthas. *Ciencia Tecnología y Futuro*, 3(1), p. 26.

Beebe, K. R., Pell, R. J. & Seasholtz., M. B., 1998. *Chemometrics. A practical guide*. New York: John Wiley & sons.

Blanco, M. A. M., 2009. Multivariate calibration for quantitative analysis. In: *Infrared spectroscopy for food quality analysis and control*. New York: Academic Press, Elsevier, pp. 51-82.

Bueno Velandia, A., 2008. *Tesis: Caracterización de corrientes de la planta de Aromátios de la GCB mediante espectroscopía NIR y métodos quimiométricos de análisis*. Bucaramanga, Colombia: UIS.

- Bueno Velandia, A., 2008. *Trabajo de grado de maestría (Magister en Química). Universidad Industrial de Santander. Caracterización de corrientes de la planta de aromáticos de la GCB mediante espectroscopia NIR y métodos quimiométricos de análisis*. Bucaramanga: s.n.
- Camo Software , 2012. *Multivariate Data Analysis for Dummies*. s.l.: John Wiley & sons , Ltd. .
- Carvajal Escobar, Y. & Marco Segura, J. B., 2005. Análisis de variabilidad de datos ambientales aplicando funciones ortogonales empíricas o componentes principales. *Meteorología Colombiana* , pp. 1-6.
- Casal Martínez, C., 2013. *Implantación de la tecnología NIRS en aplicaciones in-situ como herramienta de apoyo a las decisiones en el sector primario*. Oviedo, España: Universidad de Oviedo.
- Castillo Martínez, M., 2007. *Aplicación de la Espectroscopia NIR al control analítico de procesos de la industria química*. , . *Tesis doctoral (Programa de doctorado de química). Universidad Autónoma de Barcelona..* Bellaterra: s.n.
- Castillo Martínez, M., 2007. *Aplicación de la Espectroscopia NIR al Control Analítico de Procesos de la Industria Química Tesis Doctoral*. Bellaterra: Universidad Autónoma de Barcelona.
- Castro Zaldúa, A., 2010. *Análisis de capacidad y carga laboral del puesto de cromatografía en el laboratorio de la Refinería de Barrancabermeja - Ecopetrol S.A.*, Barrancabermeja: Ecopetrol S.A..
- Ciurczak, E., 2001. *Handbook of near-infrared analysis*. New York: Taylor & Francis Group.
- Coello, J. & MasPOCH., S., 2007. *Calibración Multivariable. En Temas avanzados de Quimiometría*". Palma de Mallorca, España: M. Blanco, V. Cerdà Editores.
- Davies, T., 1998. The history of near infrared spectroscopic analysis: past, present and future - "From sleeping technique to the morning star of spectroscopy". *Analisis Magazine*, 26(4), pp. M17-M19.

Ecopetrol S.A., 2009. *DESARROLLO DE MODELOS MULTIVARIABLES CON BASE EN THE UNSCRAMBLER*, Bucaramanga: Ecopetrol S.A..

Ecopetrol S.A., 2009. *Manual de Descripción del Proceso de la Unidad de Materias Primas y Productos*, Barrancabermeja: RWD Technologies.

Ecopetrol S.A., 2014. *Nuestra Historia*, Bogotá: Ecopetrol S.A..

Ecopetrol, 2006. *PMRB- ECP-1263*, Barrancabermeja: s.n.

Ecopetrol, 2013. *Plantilla para elaborar documentos técnicos - 81131_ECP-VIN-P-MBC-MT-008-MMH*, Barrancabermeja: Ecopetrol S.A..

Ecopetrol, n.d. *Ecopetrol. El petróleo y su mundo.* [Online] Available at: <http://www.ecopetrol.com.co/especiales/elpetroleoysumundo/refinacion3.htm> [Accessed 16 02 2016].

Empresa Colombiana de Petróleos., 2014. *Acuerdo para la definición de parámetros de calidad para el despacho de productos refinados desde y hacia la Refinería de Barrancabermeja (GRB) en los poliductos y oleoductos de la Vicepresidencia de Transporte y Logística.* s.l.:s.n.

Empresa Colombiana de Petróleos, 2009. *Manual de operación de área externa. Refinería de Barrancabermeja.* s.l.:s.n.

Eriksson, L. J. E. K.-W. N. T. J. W. C. W. S., 2006. *Multi- and megavariate data analysis. Part I, Basic principles and applications.* 2a ed. Umea: Umetrics.

Falla, F. S. et al., 2006. Characterization of crude petroleum by NIR. *Journal of Petroleum Science and Engineering*, Volume 51, pp. 127 -137.

Ganduglia, F., 2009. *Manual de biocombustibles. Instituto Interamericano de Cooperación para la Agricultura.* s.l.:s.n.

Geladi., P., 1988.. Notes on the history and nature of partial least squares (PLS) modelling. *Journal. Chemometry*, Issue 2, pp. 231-246..

Gemperline, P., 2006. *Practical guide to chemometrics*. s.l.:Taylor y Francis Grupo.

Gustato, T., 2003. *Análisis de Crudo*. s.l.:s.n.

Hannisdal, A., Hemmingsen, P. V. & Sjöblom, J., 2005. Group-Type Analysis of Heavy Crude Oils Using Vibrational Spectroscopy in Combination with Multivariate Analysis. *Industrial and Engineering Chemistry Research*, 19 January, 44(5), pp. 1349 - 1357.

Hermes Muller, A. L. e. a., 2011. Determinación total de azufre en los residuos de la destilación del petróleo crudo mediante ft-ir/atr y métodos de selección de variables.

Hillaraza Brito, P. E., 2011. *Tesis: Aplicación de los métodos de ensayos estandarizados según normas astm, para la certificación del crudo en el laboratorio del fso nabarima, ubicado en el campo corocoro de pdvsa petrosucre*. Barcelona, Anzoátegui, Venezuela: Universidad del Oriente.

Horlick, G., 1972. Digital data handling of spectra utilizing Fourier transformations. *Analytical Chemistry*, Issue 44, pp. 943-947..

Jackson, J. E., 1980. Principal components and factor analysis: part I – principal components. *Journal of Quality Technologies*, Issue 12 , pp. 201-213.

Kister, J. et al., 1996. Effects of preheating and oxidation on two bituminous coals assessed by synchronous UV fluorescence and FTIR spectroscopy. *Energ Fuels*, pp. 948-957.

Kowalski., B. R., 1975. Chemometrics: views and propositions.. *Journal of Chemical Information and Computer Sciences*, pp. 201-203.

Lamus Blanco, C. R., 2010. *Uso de análisis multivariado en la determinación sara de crudos mediante la espectroscopia NIR*. Bucaramanga: UIS.

Lamusa, C. et al., 2011. Uso de Análisis Multivariado En La Determinación SARA De Crudos Por Espectroscopia NIR. *Revista Colombiana de Física*, Volume 3, p. 43.

Lamus, C. et al., 2011. Uso de análisis multivariado en la determinación SARA De crudos por Espectroscopia NIR. *Revista Colombiana de Física*, 43(3), pp. 635 -642.

Macho Aparicio, S., 2012. *Metodologías analíticas basadas en espectroscopia de infrarrojo y calibración multivariante Aplicación a la industria petroquímica*., Tesis doctoral (Programa de doctorado de química). Tarragona: s.n.

Manual, U. U., 1998.. Trondheim: CAMO A/S.

Martens, H. & Naes, T., 1989. Multivariate calibration. p. 348.

Martens, H. & Naes, T., 1989. *Multivariate calibration*. England: John Wiley & Sons.

Massart, D. L. et al., 1997. *Handbook of chemometrics and qualimetrics*. Amsterdam: Elsevier.

Metrohm, N., n.d. Petroleum product analysis by NIR spectroscopy (e.g., cetane index, TAN, aromatics, and sulfur). *Near-Infrared Spectroscopy Application Note NIR-22*.

milller, J. & MILLER, J., 202. Analisis Multivariante. In: *Estadística y Quimiometría para Química Analítica*. 4a ed. Madrid: Pearson Educación S.A., pp. 221-238.

Ministerio de Minas y Energía y Ministerio de Ambiente y Desarrollo Sostenible., 2014. *Resolución No. 90963*. Bogotá.: s.n.

Monfort Fraga, C., 2014. *Validación y transferencia de métodos analíticos mediante espectroscopia NIR*. Barcelona, España: Universidad Autónoma de Barcelona.

OTT, D., 1992. *Manual de Laboratorio de Ensayos*. Zaragoza, España: s.n.

Palou García, A., 2014. *Desarrollo de nuevas metodologías espectrales para el control analítico de productos y procesos petroquímicos y farmacéuticos* Tesis doctoral (Programa de doctorado de química). Universidad Autónoma de Barcelona. Bellaterra: s.n.

Peguero Gutiérrez, A., 2010. *Tesis Doctoral: La espectroscopia NIR en la determinación de propiedades físicas y composición química de intermedios de producción y productos acabado*. Barcelona, España: Universidad Autónoma de Barcelona.

Peguero Gutiérrez, A., 2010. *Tesis Doctoral: La espectroscopia NIR en la determinación de propiedades físicas y composición química de intermedios de producción y productos acabado*. Barcelona: Universidad Autónoma de Barcelona.

Pinedo Salcedo, O., n.d. Origen del petróleo. In: *La química aplicada y la elaboración del petróleo colombiano*. Bogotá: s.n., pp. 378-381.

P, P., 2004. *Dean's Analytical Chemistry Handbook*.

Quintero, K., López, L. & Lima., L. D., 2014. Tesis: Espectroscopía infrarroja con transformadas de Fourier - Reflectancia total atenuada (IRTF/RTA) aplicada a la caracterización de crudos y su relación con la gravedad API. 2014. Caracas, Venezuela. Univers. *Revista Facultad de Ingeniería UCV*.

Rubinson, K. & Rubinson, J., 2001. *Análisis Instrumental*. Madrid: Prentice Hall Hispanoamericana S.A..

Savitsky, S. & G., 1964. Smoothing and differentiation of data by simplified. *Analytical Chemistry*, Volume 36, pp. 1627-1639.

Schrader, B., 1995. *Infrared and Raman Spectroscopy, VCH, New York, 1995, p.2 y ss.*, New York: VCH.

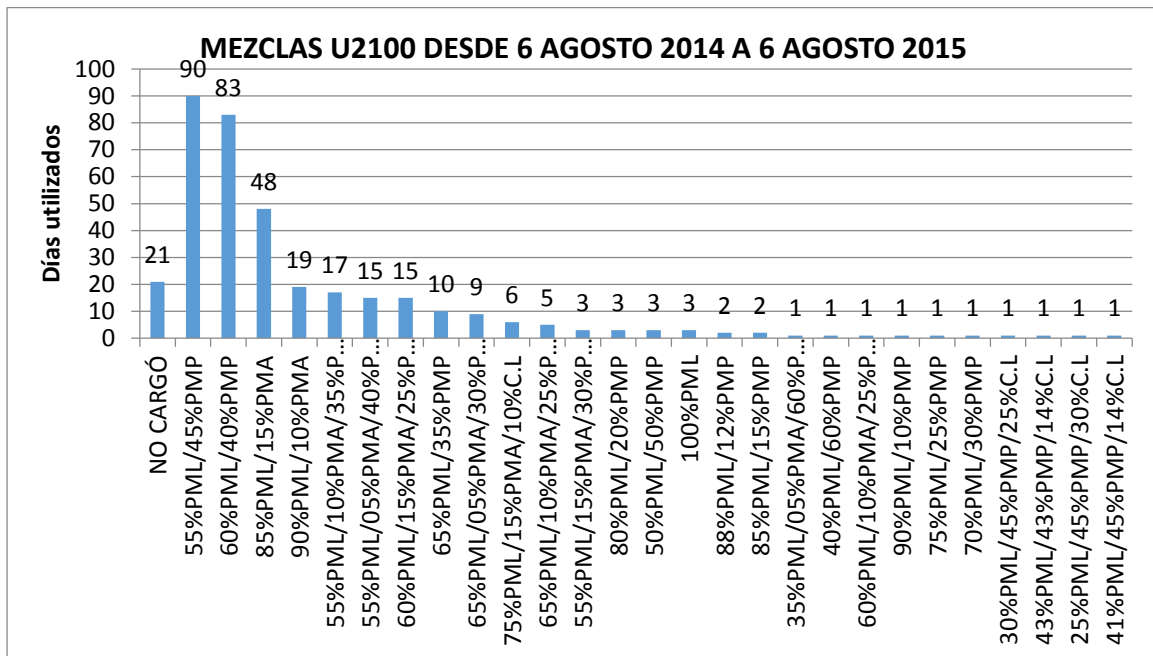
Skoog, D. A., Holler, F. J. & Nieman., T. A., 2001. *Principios de análisis instrumental*. s.l.:McGraw-Hill..

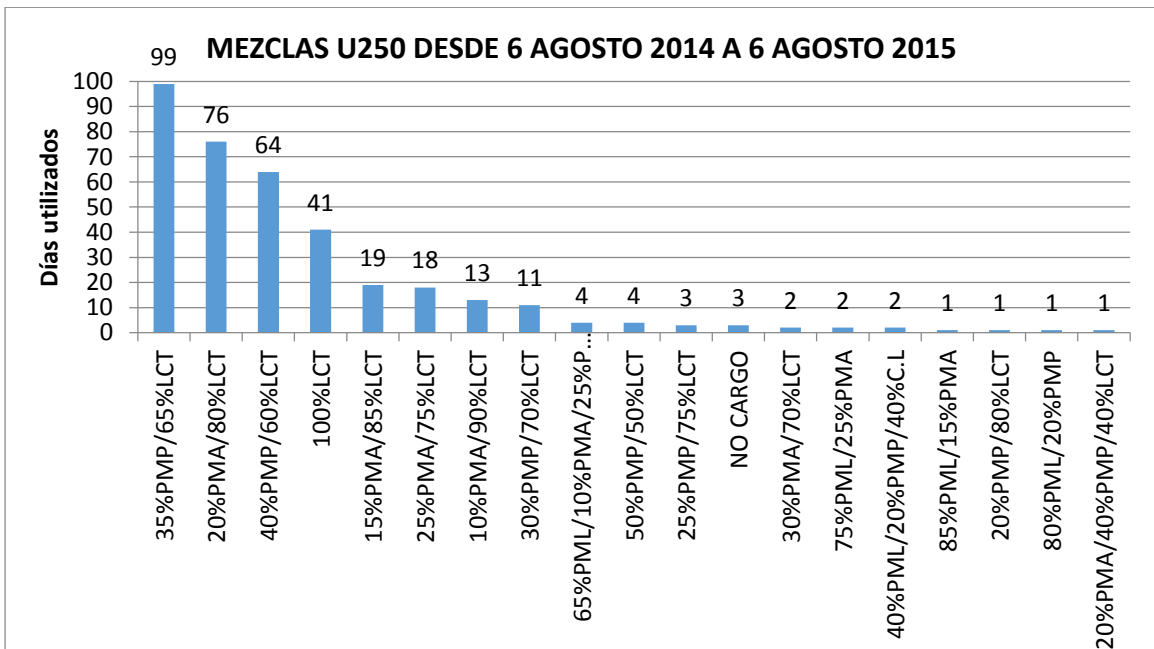
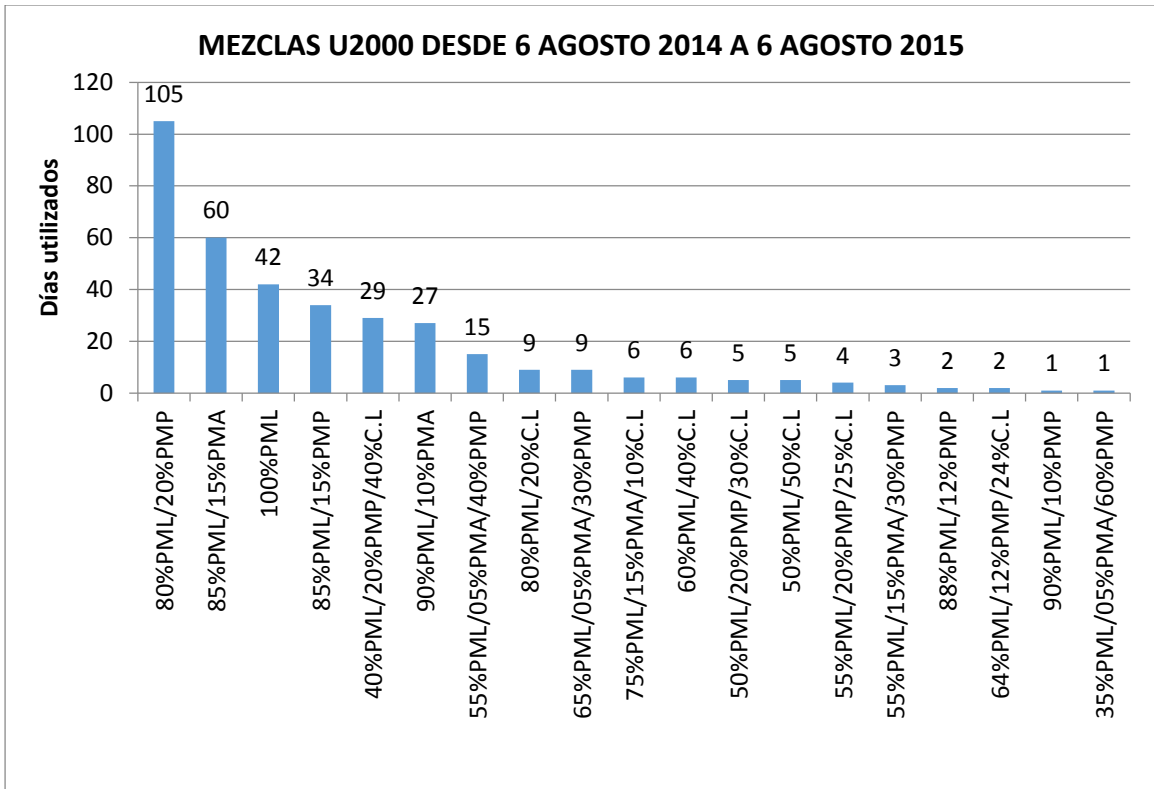
Skoog, D., Holler, F. & Nieman, T., 2001. *Principios de Análisis Instrumental*. 5a ed. Madrid: s.n.

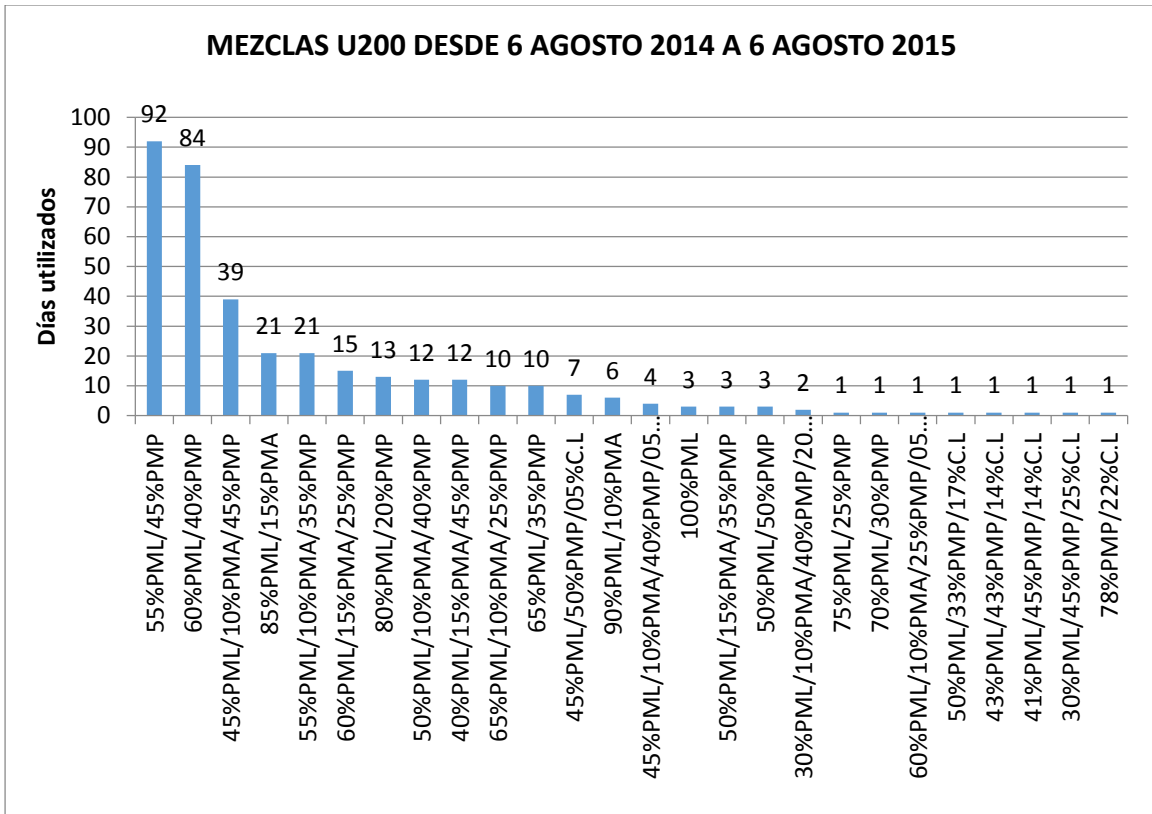
- Stark, E., 1996. Near Infrared Spectroscopy: The future Waves. *NIR Publications*, p. 704.
- Stone, M. J., 1974. Cross-validatory choice and assessment of statistical. *Journal of the Royal Statistical Society*, Volume B36, pp. 111-147.
- Stuart, B. H., 2004. *Infrared Spectroscopy: Fundamentals and Applications*. s.l.:John Wiley & Sons, Ltd..
- Vásquez C, H., 1994. La historia del petróleo en Colombia. *Revista Universidad EAFIT*, 30(93), pp. 99 - 109.
- Wikström, C. A. C. E. L. F. H. J. E. N. Å. S. S. M. K.-W. N. W. S., n.d. Multivariate process and quality monitoring applied to an electrolysis process: Part I. Process supervision with multivariate control charts.. *Chemometr. Intell. Lab.*, Volume 42, pp. 221-231.
- Wold., H. & Jöreskog, K. G., 1982. *Soft modelling. The basic design and some extensions. In systems under indirect observation.* Amsterdam: K. G. Jöreskog and H. Wold.
- Wold, S., Esbensen, K. H. & Geladi, P., 1987. Principal component analysis. *Chemom. Intell. Lab. Syst.*, Issue 2, pp. 37-52..
- Zamora Zamora, D. & Blanco Romia, M., 2012. *Desarrollo de nuevas metodologías analíticas para el control de procesos y productos industriales.* Barcelona, España: UAB.

Apéndice


Apéndice A. Estadística Unidades de Mezcla







Apéndice B. Identificación según Anexo 8 del procedimiento CID-CID-P-001 (ver Pág. 25)

	PROCEDIMIENTO PARA TRANSPORTE, RECEPCION, DISTRIBUCIÓN, MANEJO, RECHAZO Y DISPOSICIÓN DE MUESTRAS DE ENSAYO EN EL LABORATORIO		
	SOPORTE TECNICO COORDINACIÓN INSPECCION DE CALIDAD		
	CODIGO CID-CID-P-001	Elaborado 09/05/2013	Versión: 10

ANEXO 8

Requerimientos para el etiquetado de Muestras de ensayo

1. IDENTIFICACIÓN DE MUESTRAS

Todas las muestras que se envíen al laboratorio de la GRB deben ser rotuladas con al menos uno de los dos rótulos que se muestran en las Figuras 1 y 2; estos rótulos se deben diligenciar en su totalidad y con letra legible. Como medio de control se establece que las muestras que lleguen al Laboratorio y no estén identificadas con uno de estos rótulos, o que no contengan toda la información solicitada en el rótulo, se rechazarán y se generarán las fallas de control correspondiente a la unidad operativa de donde provienen las muestras.

2. INFORMACION CONTENIDA EN LOS ROTULOS.

A continuación se describe la información que se debe diligenciar en cada uno de los rótulos

2.1. Rotulo para diligenciamiento manual.

 GERENCIA REFINERIA BARRANCA BERMEJA ROTULO PARA IDENTIFICACION DE MUESTRAS	
PRODUCTO: _____	Ej.: Gasolina, Gasóleo, ACPM, etc.
UBICACIÓN: _____	Punto de Muestreo y Unidad
TIPO: _____ <small>(SUPERFICIE, MITAD, FONDO, CORRIDA, EN LÍNEA)</small>	Superficie, Mitad, Fondo, Corrida, en Línea.
TOMADA POR: _____	Número de Registro del Funcionario de Ecopetrol que Toma la Muestra.
FECHA: _____	Corresponden a la fecha y hora en que se realizó la toma de la
HORA: _____	Este espacio se debe llenar solo a las muestras de VoBo.
ID SILAB: _____ <small>ABRIL PARA 14</small>	Se debe valorar el nivel de cada ítem según el producto.
TOXICO <input type="checkbox"/> INFLAMABLE <input type="checkbox"/> CORROSIVO <input type="checkbox"/>	Se debe consignar cualquier información adicional que se considere relevante.
OBSERVACIONES: _____	

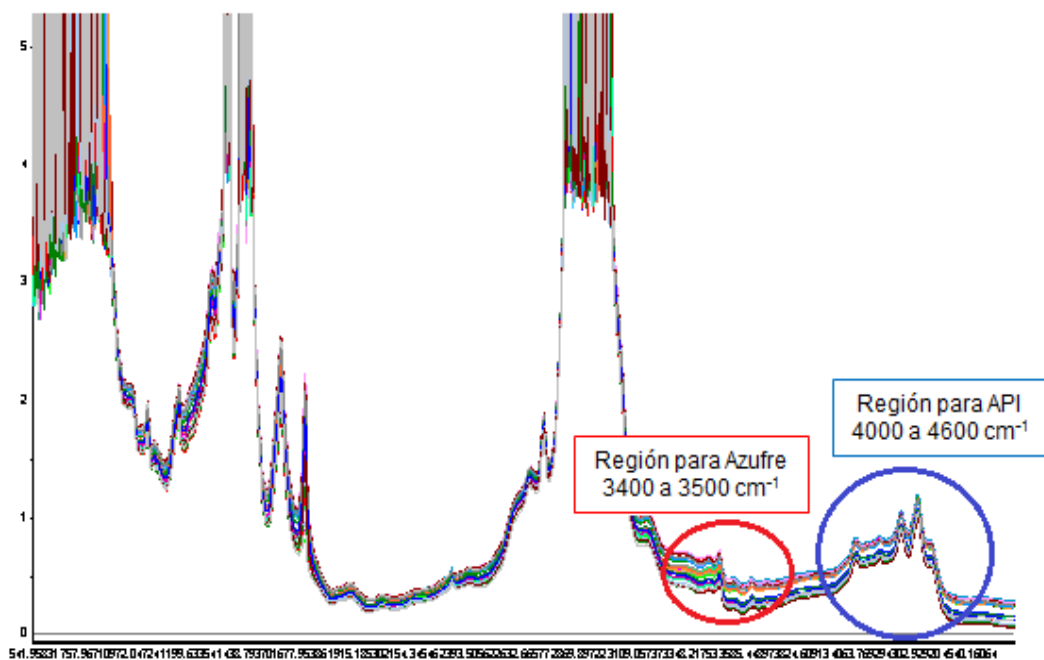
Figura 1. Rotulo para diligenciamiento manual.

Plantilla 010

Todos los derechos reservados para Ecopetrol S.A. Ninguna reproducción externa copia o transmisión digital de esta publicación puede ser hecha sin permiso escrito. Ningún párrafo de esta publicación puede ser reproducido, copiado o transmitido digitalmente sin un consentimiento escrito o de acuerdo con las leyes que regulan los derechos de autor y con base en la legislación vigente.

25/28

Apéndice C. Región espectral total del conjunto de mezclas de crudo con selección de rangos espectrales



Apéndice D. Gráfico de influencia en PLS en el modelo de predicción de la gravedad API

