

**Determinación de las fracciones de destilación al vacío (371°C - 565°C+) de crudos a partir de modelos predictivos obtenidos por Espectroscopía FTIR-ATR y Quimiometría**

**Marcela Vargas Maldonado**

**Trabajo de grado para optar el título de Magister en Química**

**Director**

**Dr. Enrique Mejía Opino**

**Doctor en Química**

**Co-Directora**

**Dra. Diana Alexandra Torres Sánchez**

**Doctora en Química**

**Universidad de Pamplona**

**Facultad de Ciencias Básicas**

**Escuela de Química**

**Pamplona, Norte de Santander**

**2018**

**Dedicatoria**

A Dios por sostenerme en todo momento.

A mis abuelos José Maldonado y Raquel Puentes, a mi mami Ana Diva Maldonado por su ejemplo y apoyo.

A mi hermosa familia: mi esposo Fernando Rincón, mis hijos Santiago y Sophia quienes me motivan a diario por lograr mis metas.

### **Agradecimientos**

A los profesores Alexandra Torres, Enrique Mejía, Jáder Guerrero y Rafael Cabanzo por las enseñanzas, aportes y confianza.

A la Universidad de Pamplona, la Universidad Industrial de Santander y a los investigadores del Laboratorio de Espectroscopía Atómica y Molecular (LEAM) sede Guatiguará por sus aportes en esta investigación.

### Resumen

Esta investigación describe el desarrollo y evaluación de modelos predictivos para determinar los rendimientos de las fracciones de destilación al vacío: destilado liviano (371°C - 427°C), destilado medio (427°C - 482°C), destilado pesado (482°C - 565°C) y fondo de vacío (565°C+); a partir de los espectros infrarrojos obtenidos por reflectancia total atenuada (FTIR-ATR) de crudos colombianos, empleando métodos quimiométricos de análisis; mínimos cuadrados parciales (PLS por sus siglas en inglés). El objetivo principal de esta investigación es desarrollar técnicas alternas a la destilación al vacío de crudos ASTM D 5236 y ASTM D 7119 Destilación a Alta Temperatura (GC) que permita realizar análisis rápidos, no destructivos y amigables con el ambiente, minimizando la cantidad de muestra y eliminando el uso de reactivos en la preparación de las mismas. Los resultados de la investigación demuestran que los modelos PLS obtenidos son robustos para predecir los rendimientos de crudos colombianos.

### Abstract

This research describes the development and evaluation of predictive models to determine the yields of the vacuum distillation fractions: light distillate (371 ° C - 427 ° C), medium distillate (427 ° C -482 ° C), heavy distillate (482 ° C -565 ° C) and vacuum bottom (565 ° C +); from the infrared spectra obtained by attenuated total reflectance (FTIR-ATR) of Colombian crude oils, using chemometric methods of analysis (partial least squares - PLS) the aim of this research was to develop alternative techniques to the vacuum distillation of crude ASTM D 5236 and ASTM D 7119 High Temperature Distillation (GC) and thus achieve quick, non-destructive and environmentally friendly analyzes, minimizing the amount of sample and eliminating the use of reagents in its preparation.

**Contenido**

	<b>Pág.</b>
Introducción .....	12
1. Marco teórico .....	16
1.1 Destilación del crudo .....	16
1.2 Esquemas de destilación piloto tipo i, ii y iii de crudos.....	19
1.3 Espectroscopia y espectroscopia infrarroja con reflectancia total atenuada (FTIR-ATR) .....	22
1.4 Quimiometría .....	23
1.4.1 Calibración multivariable.....	24
2. Metodología experimental .....	29
2.1 Selección de las muestras.....	29
2.2 Caracterización de las muestras de calibración .....	29
2.3 Toma de espectros FTIR-ATR.....	29
2.3.1 Método directo .....	29
2.3.2 Equipos instrumentales y parámetros de adquisición de espectros .....	30
2.3.3. Adquisición de espectros FIR-ATR.....	31
2.4 Pre-Tratamiento de espectros FTIR-ATR.....	32
2.5 Desarrollo de los modelos de predicción .....	32
3. Resultados y análisis .....	36
3.1 Selección de muestras .....	36

---

3.2 Caracterización de las muestras de crudos colombianos .....	36
3.3 Desarrollo del modelo de predicción PLS para el destilado liviano 371 °C -427 °C .....	37
3.4 Desarrollo del modelo de predicción PLS para el destilado medio 427° -482 °C .....	42
3.5 Desarrollo del modelo de predicción PLS para el destilado pesado 482°C-565 °C .....	47
3.6 Desarrollo del modelo de predicción PLS para el fondo de vacío 565°C+ .....	51
4. Conclusiones .....	58
5. Recomendaciones .....	59
Referencias bibliográficas.....	60
Apéndices.....	65

## Lista de figuras

	<b>Pág.</b>
<i>Figura 1.</i> Aumento de pesos moleculares y estructuras en función de la temperatura de ebullición en °C y °F.....	18
<i>Figura 2.</i> Esquema de destilación atmosférica y al vacío, con evaluación tipo II .....	21
<i>Figura 3.</i> Esquema de destilación atmosférica y al vacío con evaluación tipo III .....	21
<i>Figura 4.</i> Notación matricial para la descomposición de la matriz de regresores .....	26
<i>Figura 5.</i> Notación matricial para la descomposición de la matriz de predictores.....	27
<i>Figura 6.</i> Equipo Infrarrojo .....	30
<i>Figura 7.</i> Celda ATR .....	30
<i>Figura 8.</i> Espectros FTIR-ATR de los crudos colombianos .....	32
<i>Figura 9.</i> Gráfica del modelo PLS para el destilado liviano 371°C -427°C. ....	38
<i>Figura 10.</i> Medidas de errores vs el número de componentes (variables latentes) para el modelo de predicción del destilado liviano 371°C -427°C.....	39
<i>Figura 11.</i> Medida del error cuadrático medio de validación cruzada (SECV) vs error medio de calibración (SEC) para la determinar el # de variables latentes del modelo de destilado liviano 371°C -427°C.....	40
<i>Figura 12.</i> Residuales para el modelo del destilado liviano 371°C -427°C.....	41
<i>Figura 13.</i> Gráfica del modelo PLS para el destilado medio 427°C -482°C.....	43

*Figura 14.* Medidas de errores vs el número de componentes (variables latentes) para el modelo de predicción del destilado medio 427°C-482°C..... 44

*Figura 15.* Medida del error cuadrático medio de validación cruzada (SECV) vs error medio de calibración (SEC) para la determinar el # de variables latentes del modelo de destilado medio 427°C-482°C..... 45

*Figura 16.* Residuales para el modelo del destilado medio 427°C-482°C..... 46

*Figura 17.* Gráfica del modelo PLS para el destilado pesado 482°C -565°C ..... 47

*Figura 18.* Medidas de errores vs el número de componentes (variables latentes) para el modelo de predicción del destilado pesado 482°C -565°C..... 48

*Figura 19.* Medida del error cuadrático medio de validación cruzada (SECV) vs error medio de calibración (SEC) para la determinar el # de variables latentes del modelo de destilado pesado 482°C -565°C..... 49

*Figura 20.* Residuales para el modelo del destilado pesado 482°C -565°C..... 50

*Figura 21.* Gráfica del modelo PLS para el fondo de vacío 565°C+..... 52

*Figura22.* Medidas de errores vs el número de componentes (variables latentes) para el modelo de predicción del fondo de vacío 565°C+..... 53

*Figura23.* Medida del error cuadrático medio de validación cruzada (SECV) vs error medio de calibración (SEC) para la determinar el # de variables latentes del modelo del fondo de vacío 565°C+ ..... 54

*Figura24.* Residuales para el modelo del fondo de vacío 565°C+ ..... 55

*Figura 25.* Diagrama comparativo entre los rendimientos de referencia y los predichos por modelo del destilado medio 427-482°C+ ..... 57

## Lista de tablas

	<b>Pág.</b>
Tabla 1. <i>Propiedades de crudos colombianos con esquema de destilación tipo I</i> .....	19
Tabla 2. <i>Rangos de rendimientos de destilación ASTM D5236-13 de las muestras de crudos</i> ....	37
Tabla 3. <i>Parámetros estadísticos del modelo del destilado liviano 371°C -427°C</i> .....	42
Tabla 4. <i>Parámetros estadísticos del modelo del destilado medio 427°C -482°C</i> .....	46
Tabla 5. <i>Parámetros estadísticos del modelo del destilado pesado 482°C -565°C</i> .....	50
Tabla 6. <i>Parámetros estadísticos del modelo del fondo de vacío 565°C+</i> .....	55
Tabla 7. <i>Parámetros estadísticos de los modelos PLS para predecir los rendimientos de las fracciones de vacío de crudos colombianos 371°C -565°C +</i> .....	56

**Lista de apéndices**

	<b>Pág.</b>
Apéndice A. Caracterización de los rendimientos de los crudos colombianos ASTM D5236-1365	
Apéndice B. Gravedad API de los crudos colombianos .....	68

## Introducción

La industria del petróleo emplea la espectroscopia para la caracterización de muestras de crudo y sus derivados como naftas, kerosene, gasolinas (Kelly & Callis, 1990, págs. 1451:1444–1451), ACPM, destilados, crudo reducido y fondo de vacío, específicamente la técnica de espectroscopia infrarroja (IR) permite la identificación de los grupos funcionales de un compuesto, esto es debido a que, la frecuencia de vibración intramolecular normalmente está en el rango de frecuencias de radiación infrarroja y por tanto una molécula puede absorber radiación en esta región y originar una señal inequívoca de cada una modo vibracional. (Abbas, Rebufa, Dupuy, A., & J., 2012, págs. 98:5–14. )

La espectroscopia FTIR-ATR, presenta ventajas en el análisis de muestras de hidrocarburos porque es una técnica analítica no destructiva, requiere poca cantidad de muestra (< 1g), no requiere procesos complejos de preparación y por tanto el análisis se puede llevar a cabo en tiempos cortos (5 y 10 min) (Diem, 1993); (Nakamoto, 2001); (Rubinson & Rubinson, 2011); Brown J.M., 2003). Adicionalmente, FTIR ofrece información espectral que correlacionan muy bien con las propiedades fisicoquímicas (Laxalde, y otros, 2011, págs. 227–234), en este trabajo en particular se usó el punto de ebullición de fracciones moleculares que se diferencian por su complejidad estructural.

El desarrollo de modelos predictivos para caracterizar propiedades fisicoquímicas del petróleo es una actividad científica que lleva varios años. En particular para determinación de fracciones, Chung y colaboradores determinaron las propiedades de destilación en el keroseno

utilizando Espectroscopía en el infrarrojo cercano (NIR, por su nombre en inglés Near Infrared) y el infrarrojo medio (MIR, por su nombre en inglés Medium Infrared), acoplado a metodologías quimiométricas de análisis multivariados de tipo mínimos cuadrados parciales (PLS, por su nombre en inglés Partial Least Squares) (Chung, Ku, & Lee, 1999, págs. 155-161). Còcco y colaboradores encontraron correlaciones entre las propiedades fisicoquímicas de la gasolina como densidad, curva de destilación y Presión de Vapor Reid (RVP) mediante redes neuronales artificiales y Cromatografía de gases (GC por sus siglas en inglés Gas Chromatography) (Còcco & Yamamoto, 2004, págs. 55-63. ). Aleme y colaboradores predicen la composición de etanol y el peso específico de gasolinas brasileñas usando destilación ASTM D-86 y regresión por mínimos cuadrados parciales (PLS por sus siglas en inglés) (Aleme, Costa, & J.S., 2009, págs. 78:1422-1428). Bao y colaboradores aplicando regresión por mínimos cuadrados parciales (PLS) realizaron la predicción de número de Octano (RON por sus siglas en inglés) y el contenido de aromáticos totales empleando en gasolinas empleando espectroscopía en el infrarrojo cercano (NIR por sus siglas en inglés) y Raman ( Bao & Dai, 2009, págs. 88: 1216-1222). Kardamakis y colaboradores determinaron un modelo para predecir número de octano (RON por sus siglas en inglés) a partir de los espectros de gasolinas obtenidos en el infrarrojo cercano (NIR por sus siglas en inglés), el modelo combina la codificación de predicción lineal múltiple (LPC por sus siglas en inglés) y la regresión lineal múltiple (MLR por sus siglas en inglés) como una técnica de estimación integrada (Kardamakis Andreas, 2010, págs. 89:158–161). Flôres y colaboradores determinaron punto de ignición, gravedad específica y azufre en mezclas de biodiesel/diesel utilizando espectros FTIR y HATR y PLS, iPLS, siPLS, para este trabajo se emplearon 57 muestras para calibrar y 28 muestras para realizar validación externa. (Flôres, y otros, 2011, págs. 90: 701-706) Aleme y colaboradores predicen el punto de ignición y el índice de cetano en diésel empleando curvas de destilación

ASTM D86 y Regresión por mínimos cuadrados parciales (PLS por sus siglas en inglés), obteniendo bajos valores en los errores medios de predicción RMSEP (Aleme & Barbeira, 2012, págs. 102:129-134). Mendes y colaboradores determinaron el índice de octano para motores (MON por sus siglas en inglés) y el índice de octano (RON) en gasolinas utilizando curvas de destilación ASTM D 86 relacionando con regresión por mínimos cuadrados parciales (PLS) (Mendes, Aleme, & Barbeira, 2012, págs. 97:131-136). Chu y colaboradores emiten un review donde recopilan las bases de datos espectrales para materiales complejos (como el suelo, materiales de pruebas forenses, productos farmacéuticos, aceites, etc.) basados en la espectroscopía molecular (UV, Infrarrojo, NIR, Raman y fluorescencia) apalancando el desarrollo en la producción industrial, agrícola y la investigación científica (Chu X., Chen, & Xu, 2014, págs. 1379-1386). Fakayode y colaboradores determinan los puntos de ebullición de gasolinas empleando GC-MS y PLS (Fakayode, Mitchell, & Pollard, 2014, págs. 126:151-156). Orrego y colaboradores presentan un método para predecir densidad (gravedad API), SARA (saturados, Aromáticos, Resinas y Asfaltenos) y Residuo de carbón micro (MCR) empleando Análisis de Componentes Principales (PCA por sus siglas en inglés) y análisis de clúster jerárquico (HCA) junto con FTIR-ATR y RMN en fondos de vacío colombianos. (Orrego-Ruiz, 2014) Han participado otros autores en la determinación de SARA (Aske & Kallevik, 2001, págs. 15:1304–12. ) y SAR (Filgueiras P. R., 2016, págs. 1972 – 1978), Metodologías similares se han publicado para predecir propiedades del Kero (Gómez-Carracedo, Andrade, Calviño, Fernández, & Prada, 2003, págs. 82:1211–8), de la gasolina (Kardamakis Andreas, 2010, págs. 89:158–161) y del diesel. (Al-Ghouti & Al-Degs, 2010, págs. 89:193 – 201.)

Lo anterior demuestra que la espectroscopia infrarroja ha sido utilizada junto con modelos multivariados para la predicción de varios parámetros en muestras de fracciones de hidrocarburos con resultados satisfactorios. En conclusión, el montaje de modelos predictivos para la determinación de las fracciones de destilación al vacío 371,1°C hasta 565°C+ de crudos, basados en el uso de la espectroscopia FIR-ATR, es una alternativa innovadora para los procesos de refinación porque permite conocer los rendimientos (porcentajes) de las fracciones al vacío: destilado liviano, medio, pesado y fondo de vacío de manera rápida, precisa y con relaciones señal/ruido (S/N) elevadas.

Para el desarrollo de los modelos se usaron 50 muestras de crudos colombianos y en el laboratorio de espectroscopía de la Universidad Industrial de Santander (LEAM- UIS) se tomaron los espectros FTIR-ATR de los crudos, así como los datos de destilación. La información espectral de los crudos se procesó y correlacionó con los resultados de destilación de los crudos, empleando análisis multivariado hasta lograr el mínimo error en los modelos de predicción. El proceso de evaluación de los modelos desarrollados se realizó mediante validación cruzada. La robustez de los modelos se atestiguó mediante los parámetros de error de calibración y error de validación.

## 1. Marco teórico

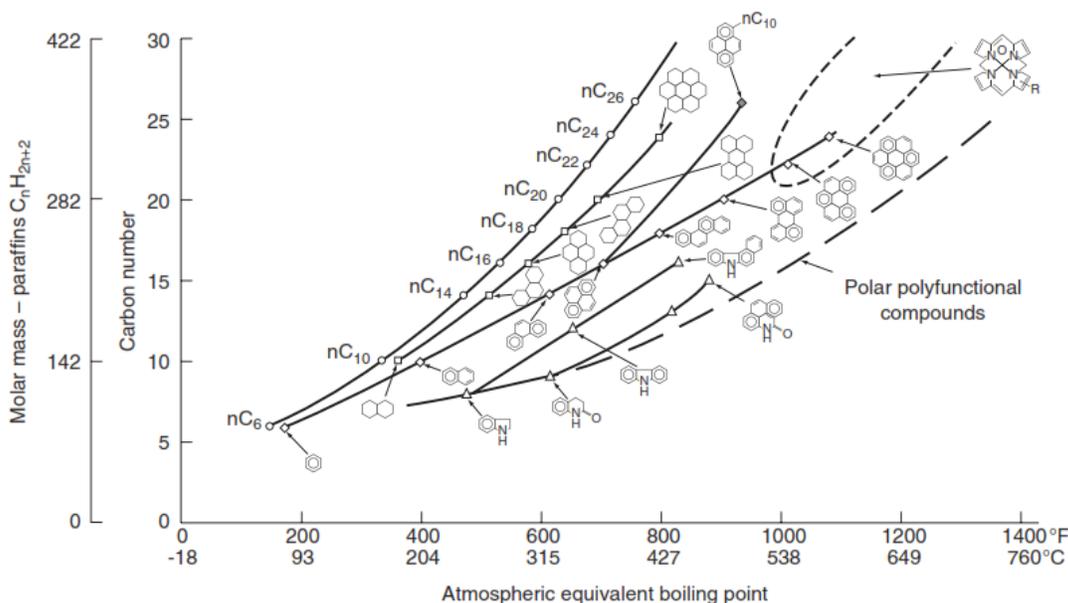
### 1.1 Destilación del crudo

La refinación del petróleo inicia con la separación de sus componentes mediante el proceso de destilación; que representa el porcentaje (%) de fracción (destilados) del crudo con respecto a los puntos de ebullición (cortes de temperatura); estas fracciones están directamente relacionadas con las propiedades químicas del crudo procesado y se diferencian por sus puntos de ebullición (propiedades físicas). Cada fracción obtenida de la unidad de destilación del crudo la constituyen un número de distintos compuestos químicos y se caracteriza por tener un punto de ebullición único. Estas fracciones incluyen gases livianos, naftas, destilados, gasóleos, crudo reducido y fondos. Cada una de estas fracciones pasa por un proceso de refinación diferente para continuar con sus subprocesos específicos. Por criterios de mercadeo y comercialización, los derivados de mayor importancia hacen referencia a los combustibles líquidos, cuyo uso principal es el de proveer la energía para el sector industrial.

Dentro de los destilados líquidos se encuentran, según el número de carbonos de menor a mayor, los siguientes componentes: naftas, kerosene (diésel), ACPM, destilados, crudo reducido y fondo de vacío.

Las naftas son materiales combustibles líquidos con punto de ebullición (IBP por sus siglas en inglés Initial Boiling Point) 15°C a 199°C (FBP por sus siglas en inglés Final Boiling Point). Se envían a la unidad para mejorar el octano, controlar el azufre, etc. y posteriormente a la mezcla de combustibles. El keroseno y el ACPM con punto de ebullición entre 199°C a 371°C, pasan por más tratamientos y luego se incorporan al combustible pesado, diésel y gasóleo. El gasóleo pasa por las unidades de conversión, donde se fraccionan en flujos más livianos (gasolina). Posteriormente, el crudo reducido es llevado a otras unidades de conversión (de alto vacío: 0.1 mmHg) generando los destilados de vacío con punto de ebullición entre 371°C a 565°C. Finalmente, el fondo de vacío con punto de ebullición de 565+°C se mezcla con el combustible industrial pesado y/o asfalto. Los residuos tienen relativamente escaso valor económico, en realidad un valor más bajo que el petróleo crudo del cual se extraen.

Básicamente los hidrocarburos pesados son los que no se destilan a presión atmosférica, presentan un punto de ebullición por arriba de 350°C, su densidad API es inferior a 20 g/cm<sup>3</sup> y sus fracciones se caracterizan por que sus moléculas contienen más de 25 átomos de Carbono (C), presentan complejidad estructural que aumenta con el punto de ebullición (ver figura 1), así como el peso molecular, la densidad, la viscosidad, el índice de refracción (Aromaticidad), polaridad (contenido de heteroátomos y metales).



**Figura. 1.** Aumento de pesos moleculares y estructuras en función de la temperatura de ebullición en °C y °F. Adaptada de Altgelt, K.H.; Boduszynski, M.M. (1994). *Composition and analysis of heavy peroleum fractions*. New York, Marcel Dekker

Para cualquier tipo de combustible, alta gravedad específica se asocia con hidrocarburos aromáticos o nafténicos y baja gravedad específica se asocia con hidrocarburos parafínicos. El calor de combustión del combustible es influenciado por el estado de oxidación, es decir la relación Carbono-Hidrógeno. Un combustible tendrá mayor contenido de cetano si aumenta la proporción de parafinas normal a isoparafinas. El punto de nube de un combustible disminuye al aumentar el de isoparafinas a normal parafinas. Así estas propiedades están determinadas por la ramificación de hidrocarburos saturados, la relación de isoparafinas normales a isoparafinas, el contenido de hidrocarburos aromáticos incluido el tipo y grado de sustitución en los anillos aromáticos y la relación de estos diversos constituyentes. Todas estas características químicas se reflejan en cierto grado en los espectros IR del compuesto. (Fodor & Kohl, 1993, págs. 7: 598 – 601. )

## 1.2 Esquemas de destilación piloto tipo i, ii y iii de crudos

Las muestras de hidrocarburos, son cargadas en las torres destiladoras (pilotos) en el laboratorio de caracterización y evaluación de crudos (ECP-ICP), allí se destila el crudo y sus fracciones.

A estos productos de hidrocarburos se les realiza tres tipos de esquemas de destilación que evalúan el crudo carga; el esquema de destilación tipo I, permite realizar la caracterización fisicoquímica del crudo, evaluando los siguientes parámetros: gravedad API, factor K, carbón Conradson, densidad, metales, destilación (%fracción), número de neutralización, viscosidad, cenizas, RVP, Contenido de agua, cera, punto de fluidez, azufre, sal, punto de inflamación, entre otros. En la tabla 1 se presentan ejemplos de caracterización fisicoquímica de algunos crudos colombianos.

**Tabla 1. Propiedades de crudos colombianos con esquema de destilación tipo I**

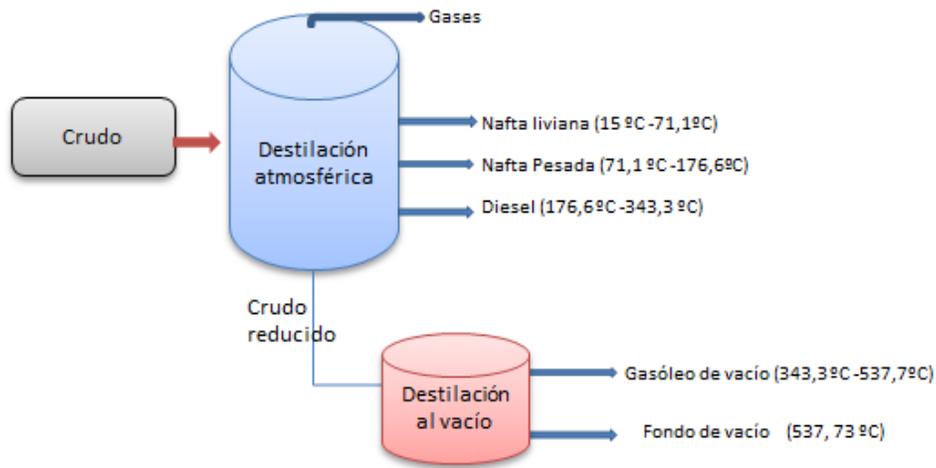
Propiedades	Cupiagua	Cusiana	Caño Limón	LCT	Castilla
Densidad a 15°C (g/ml)	0,801	0,8103	0,8809	0,9106	0,9836
Gravedad API	45,1	43	29	23.8	12.3
Número ácido (mg KOH/g)	<0,05	<0,05	0,154	2,034	0,1
Azufre, %m	0,0450	0,129	0,534	0,808	2,185
Punto de inflamación, °C	<-15.0	<-45	5	8	21
Punto de fluidez, °C	9	6	-9	<-33.0	3
Nitrogeno total, mg/kg	179	264	2477	2520	4856

Tabla 1. *Continuación*

Propiedades	Cupiagua	Cusiana	Caño Limón	LCT	Castilla
Ceras (%m)	15,84	7,78	6,57	0,18	1,58
R carbon micro ,%m	0,23	0,77	5,69	5,18	15,85
Insolubles en nC7, %m	0	0,14	2,16	1,29	14,58
Insolubles en nC5, %m	0,03	0,32	5,75	3,21	18,72
Vanadio, mg/kg	0,151	3,56	12,84	36,01	421,80
Niquel, mg/kg	0,087	1,05	40,41	20,71	82,60
Peso molecular promedio	154,41	165,92	246,92	222,99	256,22
MeABP, 50%SimDis	237,4	252,5	368,9	423,4	488,8
Factor KUOP	12,14	12,12	11,91	11,84	11,30
Indice de correlación	17,93	19,60	36,26	44,40	72,98
Viscosidad a 40°C, cSt	2,06	2,22	15,89	1,98	3510,57
CCAI	786,41	810,3	880,9	910,6	983,6

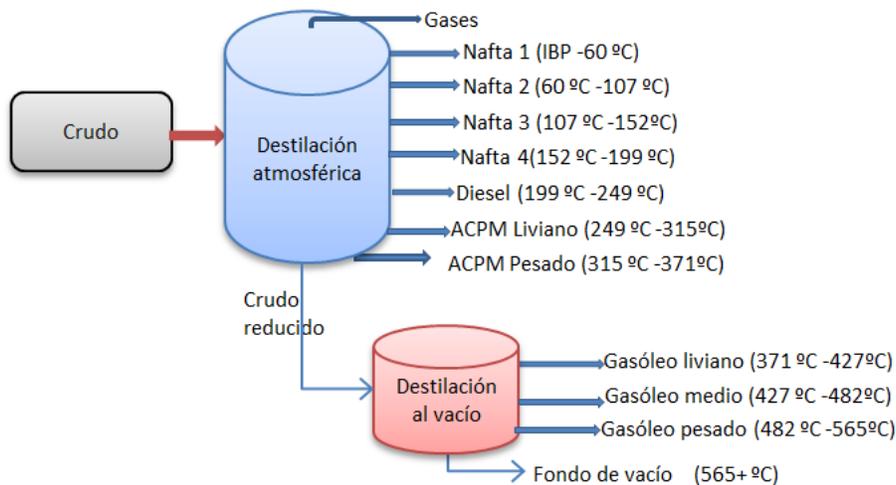
Fuente: Navarro, Uriel. (2002). Laboratorio de Caracterización de Crudos, ECP –ICP. PhD. Curso Química del Petróleo.

En el esquema de destilación tipo II (ver figura 2), se le realiza al crudo un proceso a condiciones atmosféricas (ASTM D2892-13) obteniendo gases, naftas, destilados y crudo reducido, este último -se lleva a destilación - vacío (ASTM D1160), obteniendo gasóleo y fondo de vacío.



**Figura 2. Esquema de destilación atmosférica y al vacío, con evaluación tipo II**

Y finalmente en el esquema de destilación tipo III (ver figura 3), se realiza la evaluación más completa del crudo a condiciones atmosféricas; se obtienen gases, naftas, diésel, ACPM y crudo reducido. El crudo reducido se lleva a destilación al vacío (ASTM D5236-13), para obtener gasóleo liviano, gasóleo medio, gasóleo pesado y fondo de vacío.



**Figura 3. Esquema de destilación atmosférica y al vacío con evaluación tipo III**

### 1.3 Espectroscopia y espectroscopia infrarroja con reflectancia total atenuada (FTIR-ATR)

La espectroscopia infrarroja es una técnica instrumental que puede adaptarse de muchas maneras para extraer la información necesaria como energías de estados vibracionales de las moléculas (Requena & Zuñiga, 2004) y se aplica en la industria petroquímica en la caracterización e identificación de materiales, síntesis, análisis de contaminantes, entre otros, esta técnica analítica es de fácil aplicación, es precisa y generalmente no requiere preparación de las muestras.

Específicamente las aplicaciones analíticas de la espectroscopia infrarroja se basan en el uso del infrarrojo medio (4000 - 600  $\text{cm}^{-1}$ ) y el infrarrojo cercano (2800 - 400  $\text{cm}^{-1}$ ), que proporciona la posibilidad de convertir esta metodología en una técnica cuantitativa (Hoeil & Min-Sik, 2000, págs. 54: 239 – 245. ). Se emplea para caracterizar una amplia variedad de materiales. En muchos casos, los tratamientos utilizados para la preparación de las muestras pueden resultar perjudiciales o impracticables, como por ejemplo cuando se intenta registrar espectros IR de crudo pesado, polímeros de elevada masa molecular, gomas, por lo que en estos casos la espectroscopia infrarroja media acoplada a celda con reflectancia total atenuada (ATR) constituyen una alternativa versátil y viable para este tipo de matrices complejas. (Requena & Zuñiga, 2004).

La reflectancia total atenuada (ATR) es un fenómeno en el que la radiación electromagnética que se propaga a través de un medio ópticamente denso llega a una interfase con un medio enrarecido (de menor índice de refracción), a un ángulo de incidencia mayor que el ángulo crítico, la radiación no escapa del medio denso, sino que es totalmente reflejada internamente. Debido a la naturaleza ondulatoria de la radiación electromagnética, la reflexión no sucede directamente en

la interfase de los dos medios; la radiación se emite del medio más denso y se extiende a una distancia corta del otro lado de la interfase en el medio enrarecido antes de volver. La extensión con la cual la radiación penetra el medio enrarecido depende de la longitud de onda de la radiación, los índices de refracción de los dos medios y el ángulo de incidencia a la interfase. (Skoog, Holler, & Nieman, 2002)

#### **1.4 Quimiometría**

La quimiometría es una herramienta de análisis, clasificación y calibración multivariable que permite obtener modelos para el cálculo de concentraciones y otras propiedades a partir de datos instrumentales mediante métodos matemáticos y estadísticos. (Hopke, 2003, págs. 365–377. )

Las primeras aplicaciones en quimiometría se dieron en el área de la química analítica en el año 1960, con el objetivo de analizar datos dependientes de más de una variable simultánea esto permitió observar el comportamiento de un experimento mediante la estadística multivariada. Las técnicas espectroscópicas suministran resultados de cada muestra, la quimiometría facilita el análisis de la información espectroscópica y su correlación con propiedades de interés de muestras complejas como el crudo y sus fracciones ( (Melendez, Lache, Orrego, Pachon, & Mejia Ospino, 2012, págs. 90-91:56–60)

Existen diversos métodos multivariable, depende del objetivo del trabajo, de la complejidad del sistema y de las características de las muestras. (Lavine & Workman, 2013, págs. 705–714.)

La quimiometría incorpora otras áreas entre estas están el diseño de experimentos, el reconocimiento de patrones y la calibración multivariable. En el área de diseño experimental se buscan las variables que más afectan a determinado proceso, así como la interacción entre ellos. En reconocimiento de patrones, a partir de una amplia gama de información (medidas espectrales, por ejemplo) en una serie de objetos, se pretende encontrar grupos de muestras (objetos) que son similares entre sí y por lo tanto detectar las tendencias de datos y finalmente en la calibración multivariable el objetivo es el desarrollo de modelos predictivos de propiedades fisicoquímicas con base en propiedades medidas del sistema químico (tales como datos espectrales, concentraciones, etc. ), que determinen los mismos resultados que los encontrados con los procedimientos estándar para dichas propiedades, obteniendo métodos alternos a los existentes, confiables, de corto tiempo y que no impliquen preparación de la muestras con reactivos químicos. (Ferreira da Silva, y otros, 2013, págs. 116:151–157. )

A partir de la correlación de los espectros FTIR-ATR con las propiedades obtenidas por métodos estandarizados de referencia (ASTM D5236-13) Rendimientos de las fracciones al vacío (%V): Destilado liviano, medio, pesado y fondo de vacío, se pueden construir modelos mediante la regresión de mínimos cuadrados parciales (PLS) que puedan predecir los rendimientos de las fracciones al vacío de nuevos crudos.

#### **1.4.1 Calibración multivariable**

El desarrollo de la quimiometría en química analítica se debe fundamentalmente a la existencia de herramientas informáticas acopladas a la moderna instrumentación analítica. Las medidas con gran cantidad de información, por ejemplo, medidas espectrales (multivariable), se

pueden transformar matemáticamente a la forma más adecuada para la obtención de un modelo de calibración, que permita predecir el valor de la propiedad fisicoquímica estudiada a partir de la relación entre la matriz de datos  $X$ , que contiene los datos espectroscópicos y la matriz de datos  $Y$  que representan los valores de las propiedades de las muestras (rendimientos %V/%V). Una vez obtenido el modelo de calibración la propiedad se puede predecir por medio de la ecuación:

$$Y = X B + E \quad (1)$$

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

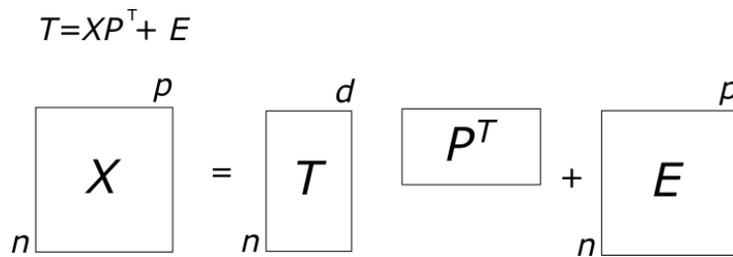
Dentro de las técnicas más usadas tenemos la regresión por componentes principales(PCR), regresión por mínimos cuadrados parciales(PLSR) y la regresión lineal múltiple(MLR). (Matthias, 2007)

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

**2.4.1.1 Regresión parcial por mínimos cuadrados (PLS)**

La técnica PLS busca hacer la descomposición de la matriz X independientemente de la matriz Y (figuras 4 y 5). Usa la información contenida en la matriz X (datos espectroscópicos) y la información en la matriz Y (propiedad a estimar), y obtiene unas variables auxiliares llamadas variables latentes. Cada dirección de la variable latente de la matriz X es modificada de manera que la covarianza entre esta y la matriz Y es maximizada. Las matrices X y Y se descomponen simultáneamente en matrices más pequeñas. (Matthias, 2007)

$$X = TP^T + E = \sum_{i=1}^d t_i p_i^T + E \tag{2}$$



**Figura 4. Notación matricial para la descomposición de la matriz de regresores**

$$Y = UQ^T + E = \sum_{i=1}^d u_i q_i^T + F \tag{3}$$

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

**Figura 5.** Notación matricial para la descomposición de la matriz de predictores

Donde  $T_{n \times d}$  es la matriz de resultados (scores),  $P_{d \times p}$  es la matriz de pesos, y  $E$  los residuales para la matriz de datos  $X$ ;  $U_{n \times d}$  es la matriz de resultados (scores),  $Q_{d \times m}$  de pesos y  $F$  de residuales, para la matriz de la propiedad a determinar  $Y$ . De la misma manera  $n$  es el número de muestras,  $d$  es el número de componentes,  $p$  son las variables originales (espectros) y  $m$  las variables dependientes (rendimientos %V/%V). (Matthias, 2007)

La descomposición de las matrices se realiza simultáneamente, estableciendo una relación lineal interna entre los resultados (scores) de las matrices  $X$  y  $Y$ , de manera que se cumpla para cada componente  $p$ :

$$u_d = b_d t_d \quad (4)$$

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

$$Y = TBQ^T + F \quad (5)$$

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

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

$$y_i = bX_i + e_i \quad (2)$$

Siendo  $X_i$  vector que contiene la señal de la muestra,  $y_i$  el vector de propiedades a determinar y b un vector que permite determinar la predicción de la propiedad de una muestra sin necesidad de descomponer su matriz en matrices T y P. (Matthias, 2007)

## 2. Metodología experimental

Se siguió la metodología establecida en la norma ASTM E-1655, para la construcción de modelos de calibración multivariable a partir de mediciones espectrales de muestras de interés y que propone las siguientes etapas.

### 2.1 Selección de las muestras

Se utilizó un grupo de aproximadamente 50 muestras de crudos de diferentes campos de Colombia suministradas por la Universidad Industrial de Santander (UIS).

### 2.2 Caracterización de las muestras de calibración

Todas las muestras se caracterizaron por el método ASTM D 5236-13 que es el método de referencia previamente establecido, evaluado estadísticamente y totalmente confiable. Los datos se encuentran en el Apéndice A.

### 2.3 Toma de espectros FTIR-ATR

#### 2.3.1 Método directo

Cada muestra de crudo, aproximadamente 1 gramo, fue depositada con la ayuda de una espátula directamente en la ventana de trayectoria del haz de radiación, sin tratamiento. Después

de tomar cada espectro se limpiaba la ventana de la celda ATR y se realizaba un blanco vacío para garantizar que no quedara residuo del crudo anterior y garantizar los resultados.

### 2.3.2 Equipos instrumentales y parámetros de adquisición de espectros

La toma de los espectros FTIR-ATR se realizó en el laboratorio de espectroscopía atómica y molecular (LEAM) de la Universidad Industrial de Santander (UIS), sede Guatiguará, en el equipo marca Bruker modelo Tensor 27 (ver figura 6), con celda Platinum ATR marca Bruker, (Figura 7), que cuenta con un cristal de diamante de reflexión simple, plano, con distancia de trabajo 20 nm (altura de la muestra) e índice de refracción de 2,4.



**Figura 6. Equipo Infrarrojo**



**Figura 7. Celda ATR**

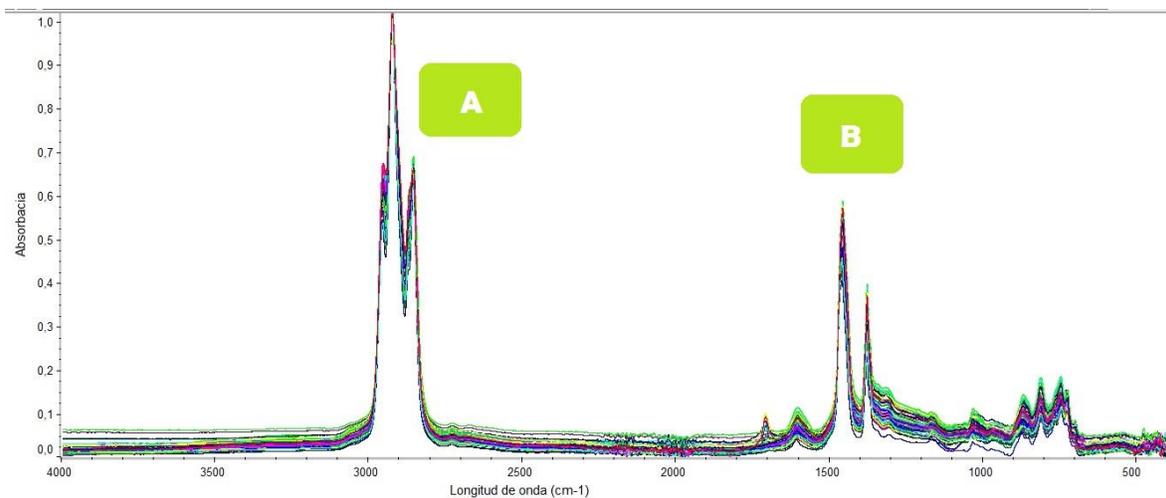
Antes de tomar los espectros se acondicionó el equipo IR a la celda de medida, alineando el sistema óptico del equipo infrarrojo.

Los parámetros después de optimizar el FTIR-ATR fueron: número de barridos: 32, velocidad del espejo del interferómetro: 0.4747 cm/s, resolución: 4 cm<sup>-1</sup> y el ángulo de incidencia del haz IR 45°.

Los espectros infrarrojos se tomaron en un rango MIR (región infrarroja media) de 400-4000 cm<sup>-1</sup> (Requeña & Zúñiga, 2004). Las muestras de crudos se colocaron directamente en la trayectoria del haz de radiación, sin tratamiento.

### **2.3.3. Adquisición de espectros FIR-ATR**

Los espectros infrarrojos se adquirieron en la región infrarroja media (MIR) que corresponde al rango de 400-4000 cm<sup>-1</sup>, en la figura 8., se evidencian dos regiones de interés, **A** que va de 3100-2500 cm<sup>-1</sup> que corresponden en su mayoría a bandas de grupos alifáticos y **B** que va de 1800-600 cm<sup>-1</sup> que corresponden en su mayoría a bandas de grupos aromáticos. (Kuptsov & Arbusova, 2011, págs. 203–211. )



**Figura 8.** Espectros FTIR-ATR de los crudos colombianos

#### 2.4 Pre-Tratamiento de espectros FTIR-ATR

Antes de calcular los modelos predictivos se realizó un pre-tratamiento de los espectros para minimizar errores por fallas espectrales. A los espectros se les realizó normalización máxima, ajuste a la misma escala (1), corrección de la línea base off set, para minimizar el ruido en las señales. La normalización máxima tiene como objetivo hacer que los datos espectroscópicos de todas las muestras estén a la misma escala y consiste en dividir cada punto del espectro por el valor de máxima absorbancia.

#### 2.5 Desarrollo de los modelos de predicción

A continuación, se describe el protocolo usado para el desarrollo y evaluación de los modelos de predicción de las fracciones de destilación al vacío, destilado liviano 371°C -427°C (%v/v), destilado medio 427°C - 482 °C (%v/v), destilado pesado 482°C - 565 °C (%v/v) y fondo de vacío

565°C+ (%v/v), a partir de los espectros FTIR-ATR de las muestras de crudos colombianos utilizando el algoritmo NIPALS programado en Matlab.

A partir de los espectros de los crudos previamente pretratados y empleando el algoritmo matemático de Regresión de Mínimos Cuadrados Parciales (PLS) y el programa MatLab se evaluó la capacidad de la espectroscopía FTIR-ATR para predecir los rendimientos de las fracciones al vacío de los crudos colombianos. PLSR es una herramienta matemática ampliamente utilizada en análisis quimiométrico de datos espectrales con alta colinealidad. El algoritmo implementado en Matlab para tal fin fue el NIPALS (Non-linear Iterative Partial Least Squares).

La técnica PLS usa la información contenida en los datos espectroscópicos y la información de la propiedad que se quiere predecir (rendimientos) y obtiene unas variables auxiliares llamadas variables latentes o componentes que reducen la dimensionalidad de los datos. Estas nuevas variables contienen los datos en X (datos espectroscópicos) que mejor predicen la propiedad. (Abbas, Rebufa, Dupuy, A., & J., 2012)

El desarrollo del modelo se realizó en dos etapas, calibración y validación. Para esto el grupo total de muestras se separan en dos grupos; el grupo de calibración que está compuesto por las muestras usadas para construir el modelo, sus variables X y sus propiedades de referencia Y, y el grupo de validación que contiene a las muestras restantes para las cuales los valores de las variables X se introducen al modelo y se obtienen los valores predichos Y. (Bezerra de Lira, y otros, 2010, págs. 89:405 – 409)

La técnica de validación cruzada de K iteraciones o K-fold (K= 10) fue usada sobre la población de calibración para determinar el número de variables latentes y consiste en dividir aleatoriamente el grupo de muestras en k subconjuntos. Los K-1 subconjuntos se utilizan como datos de entrenamiento o calibración y el subconjunto restante se utiliza como datos de validación. El proceso de validación cruzada es repetido durante k iteraciones, con cada uno de los posibles subconjuntos de datos de prueba. Para encontrar los mejores grupos de entrenamiento y validación, se ejecuta el K-fold varias veces (multi cross-validation) hasta que se obtiene el mejor modelo con un número óptimo de variables latentes, de acuerdo con el mayor valor de  $R^2_{cv}$  (el más cercano a 1) y el error cuadrático medio de validación cruzada (SECV) más bajo, calculado por la ecuación 7,

$$SECV = \sqrt{\frac{\sum_{i=1}^{N_v} (y_i - \hat{y}_i)^2}{N_v}} \quad (73)$$

En la ecuación (7),  $y_i$  es la propiedad medida,  $\hat{y}_i$  es el valor de la propiedad predicha por el modelo y  $N_v$  es el número de muestras usadas en la validación.

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

$$\hat{y}_i = bX + e \quad (8)$$

En la ecuación 8,  $X$  es el espectro de la muestra a la cual se le quiere predecir la propiedad,  $e$  es el residual y  $b$  es vector de regresión que considera la contribución de cada variable al modelo PLS, esto significa que entre más alto sea el valor de  $b$  más importante es la variable para el modelo de calibración.

Una vez obtenido el modelo de calibración se evalúa el desempeño del modelo para predecir la propiedad con el grupo de validación. Los resultados obtenidos se comparan con los observados produciendo un residuo que se puede utilizar para calcular un error cuadrático medio de calibración (SEC por sus siglas en inglés Square Error of calibration), que proporciona información sobre el ajuste del modelo a los datos de calibración, evaluando el desempeño del modelo, calculado por la ecuación 9.

$$SEC = \sqrt{\frac{\sum_{i=1}^{N_c} (y_i - \hat{y}_i)^2}{N_c}} \quad (94)$$

En donde,  $y_i$  es la propiedad medida,  $\hat{y}_i$  es el valor de la propiedad predicha por el modelo y  $N_c$  es el número total de muestras en el conjunto de calibración. (Hongfu, Xiaoli, Haoram, & Yupeng, 2006, págs. 1720–1728.)

Los modelos de calibración se evaluaron de acuerdo a los parámetros estadísticos: coeficiente de determinación  $R^2$ , el error cuadrático medio de validación cruzada (SECV) y el error cuadrático medio de calibración (SEC).

### **3. Resultados y análisis**

#### **3.1 Selección de muestras**

Las muestras de crudos se dividieron aleatoriamente en dos grupos: 33 muestras para calibración y 16 muestras para validación cruzada. Para obtener los modelos de los rendimientos de las fracciones al vacío de crudos colombianos, información importante para el proceso de refinación y preparación del crudo (dietas) para mejorar productos.

Las 50 muestras de crudos fueron suministradas por la Universidad Industrial de Santander (UIS) y tienen su origen en diferentes regiones de Colombia.

#### **3.2 Caracterización de las muestras de crudos colombianos**

Todas las muestras de crudos contaron con su respectivo análisis de destilación ASTM D5236-13, esta información fue suministrada por el laboratorio de espectroscopia (LEAM) de la Universidad Industrial de Santander (UIS).

**Tabla 2. Rangos de rendimientos de destilación ASTM D5236-13 de las muestras de crudos**

Propiedad	Valor Mínimo (V %)	Valor Máximo (V %)	Valor Promedio (V %)
<b>Destilado Liviano (371°C-427°C)</b>	3,60	14,40	7,43
<b>Destilado Medio (427°C-482°C)</b>	6,89	17,00	11,06
<b>Destilado Pesado (482°C-565°C)</b>	3,72	16,90	9,72
<b>Fondo de Vacío (565 +°C)</b>	9,52	60,15	35,48

Los modelos generados son correlativos, se requiere que las nuevas muestras de crudos que se usen para aplicar en estos modelos, tengan las mismas características (rendimientos %v/v de las fracciones de destilación al vacío) de las usadas en esta investigación como se observa en la tabla 2.

### 3.3 Desarrollo del modelo de predicción PLS para el destilado liviano 371°C -427°C

Los valores medidos de los rendimientos de los crudos, se relacionaron linealmente con las intensidades espectrales (Fodor & Kohl, 1993, págs. 598-601) y el desarrollo del modelo dependió del valor del coeficiente de determinación  $R^2$ , el cual obtuvo el mayor valor 0,98 empleando 8 componentes (variables latentes), determinando de esta manera que 8 es el número de componentes óptimo para construir el modelo PLSR de predicción del destilado liviano de vacío 371°C -427°C mostrando que la recta de regresión ajusta bien los datos de calibración con los de validación cruzada (ver figura 9), comportamiento que se confirma con los bajos valores de error cuadrático medio de calibración (SEC) 0.40% y de error cuadrático medio de validación cruzada (SECV) 0.41%.

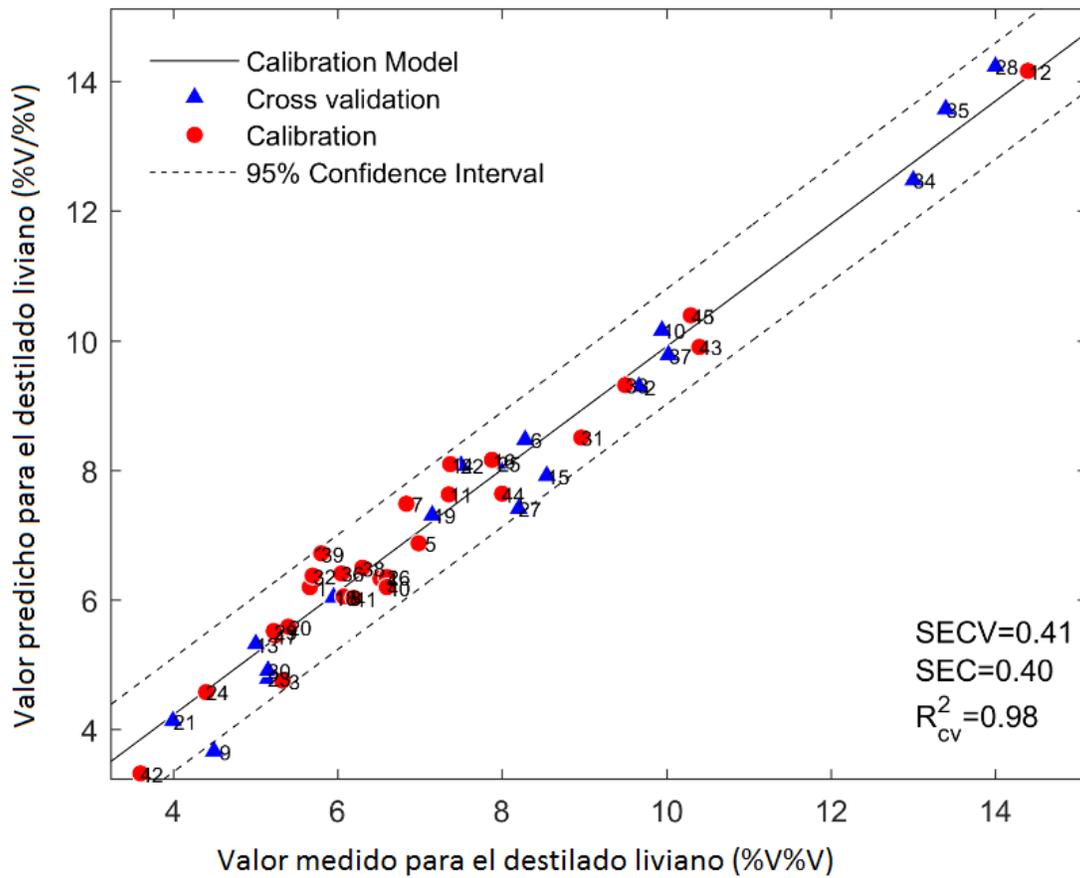
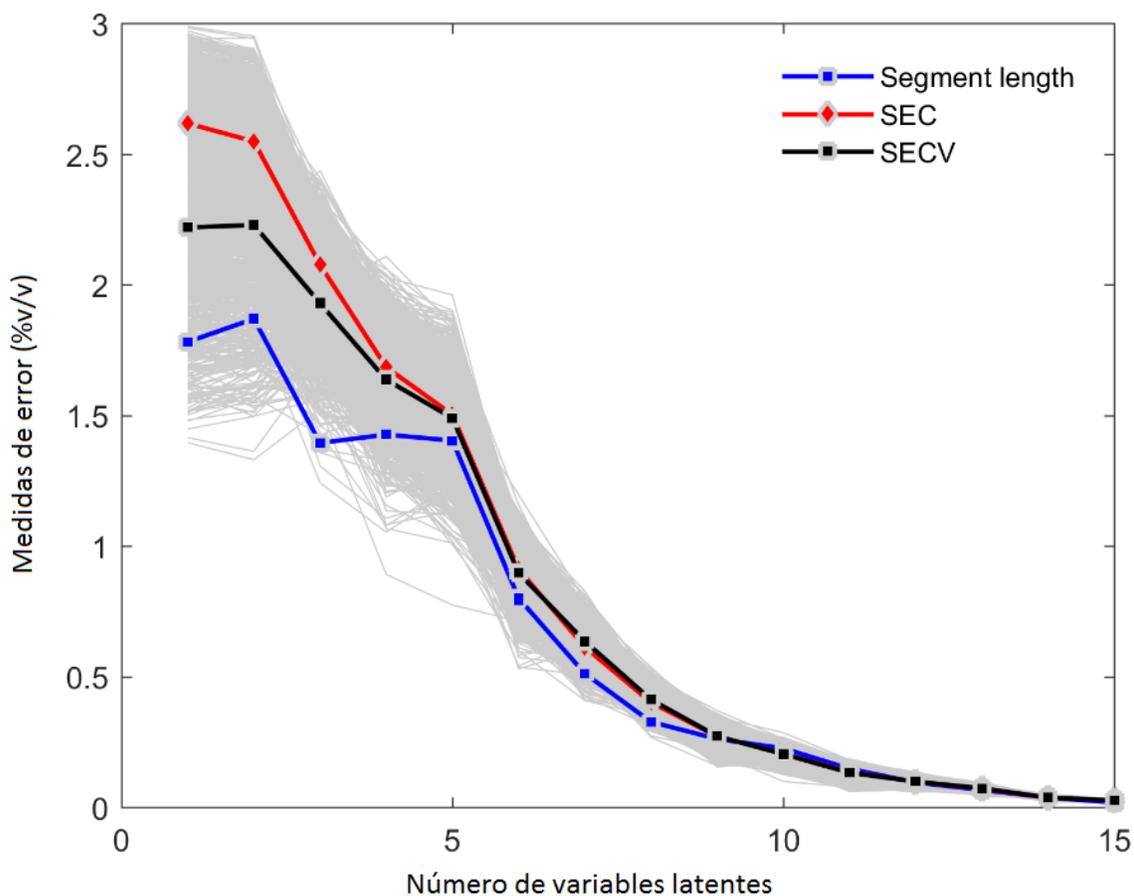


Figura 9. Gráfica del modelo PLS para el destilado liviano 371°C -427°C.

Una vez obtenido el modelo óptimo PLS de calibración para la predicción del destilado liviano que presentaba los menores valores de SECV, SEC y la mayor linealidad ( $R^2$ ), se encontró que los parámetros estadísticos SECV y SEP varían con el número de componentes (variables latentes).

En la figura 10 se observa como disminuye el comportamiento del error cuadrático medio de validación cruzada (SECV) y el error medio de calibración (SEC) a medida que aumenta el

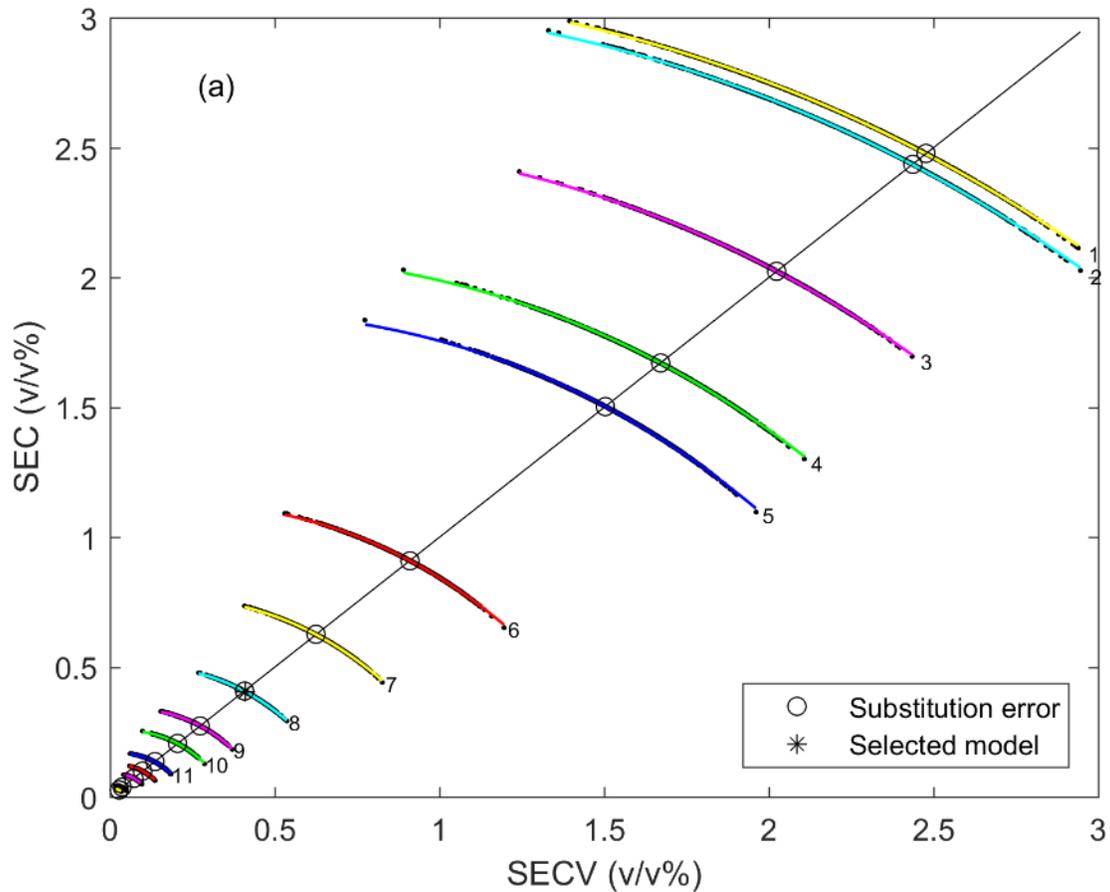
número de componentes (variables latentes), obteniéndose los valores de errores más bajos SECV 0.41 % y SEC 0.40%, para 8 componentes, estos parámetros estadísticos dan información de la medida de la dispersión de las muestras de calibración alrededor de la línea de regresión cuando se emplea la validación cruzada.



**Figura10.** Medidas de errores vs el número de componentes (variables latentes) para el modelo de predicción del destilado liviano 371°C -427°C.

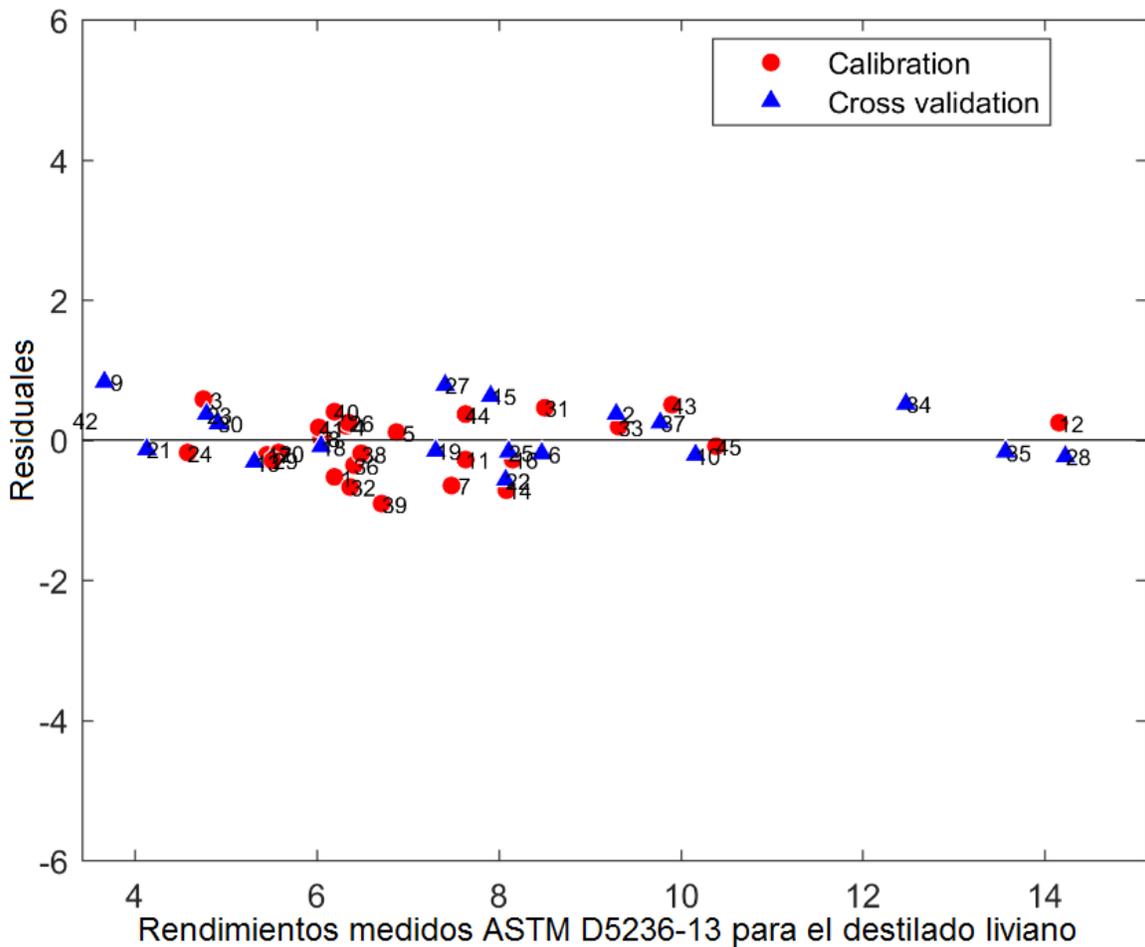
Nuevamente la selección del número de componentes (variables latentes) dependió estrechamente de la relación con los errores, en la figura 11 se observa como al disminuir las

variables latentes (componentes) aumenta el error, obteniéndose un bajo error al diseñar el modelo con 8 componentes, finalmente se seleccionó este modelo como el óptimo de predicción del destilado liviano de vacío  $371^{\circ}\text{C}$  - $427^{\circ}\text{C}$  (PLS).



**Figura 11.** Medida del error cuadrático medio de validación cruzada (SECV) vs error medio de calibración (SEC) para la determinar el # de variables latentes del modelo de destilado liviano  $371^{\circ}\text{C}$  - $427^{\circ}\text{C}$ .

En la figura 12 se observa un comportamiento aleatorio de los residuales alrededor de cero y no presenta tendencias, entendiéndose que los residuales son las desviaciones entre los puntos experimentales y los valores teóricos (Mongay Fernandez, 2005) para el modelo del gasóleo liviano 371°C -427°C.



**Figura12.** Residuales para el modelo del destilado liviano 371°C -427°C.

Los parámetros estadísticos obtenidos para el modelo predictivo PLS del destilado liviano 371°C -427°C se observan en la tabla 3.

**Tabla 3. Parámetros estadísticos del modelo del destilado liviano 371°C -427°C**

<b>Rendimiento</b>	<b>Número de componentes</b>	<b>SECV ( %v/v)</b>	<b>SEC ( %v/v)</b>	<b>R<sup>2</sup><sub>CV</sub></b>	<b>R<sup>2</sup><sub>C</sub></b>
<b>Destilado liviano</b>	8	0,41	0,40	0,98	0,97

### 3.4 Desarrollo del modelo de predicción PLS para el destilado medio 427° -482 °C

En la figura 13 se observa el modelo predictivo del destilado medio 427°C -482°C calibrado con los valores de destilación de los crudos obtenidos por la metodología estándar ASTM-5236 y bajo un intervalo del 95% de confianza, obteniendo unos bajos valores de los errores medio de calibración (SEC) de 0.13%, de validación cruzada de 0.13%, y un coeficiente de determinación (R<sup>2</sup>)de 0.99, muy cercano a 1 lo que garantiza la linealidad del modelo y alta certeza en los datos predichos, siendo este el modelo más robusto en esta investigación.

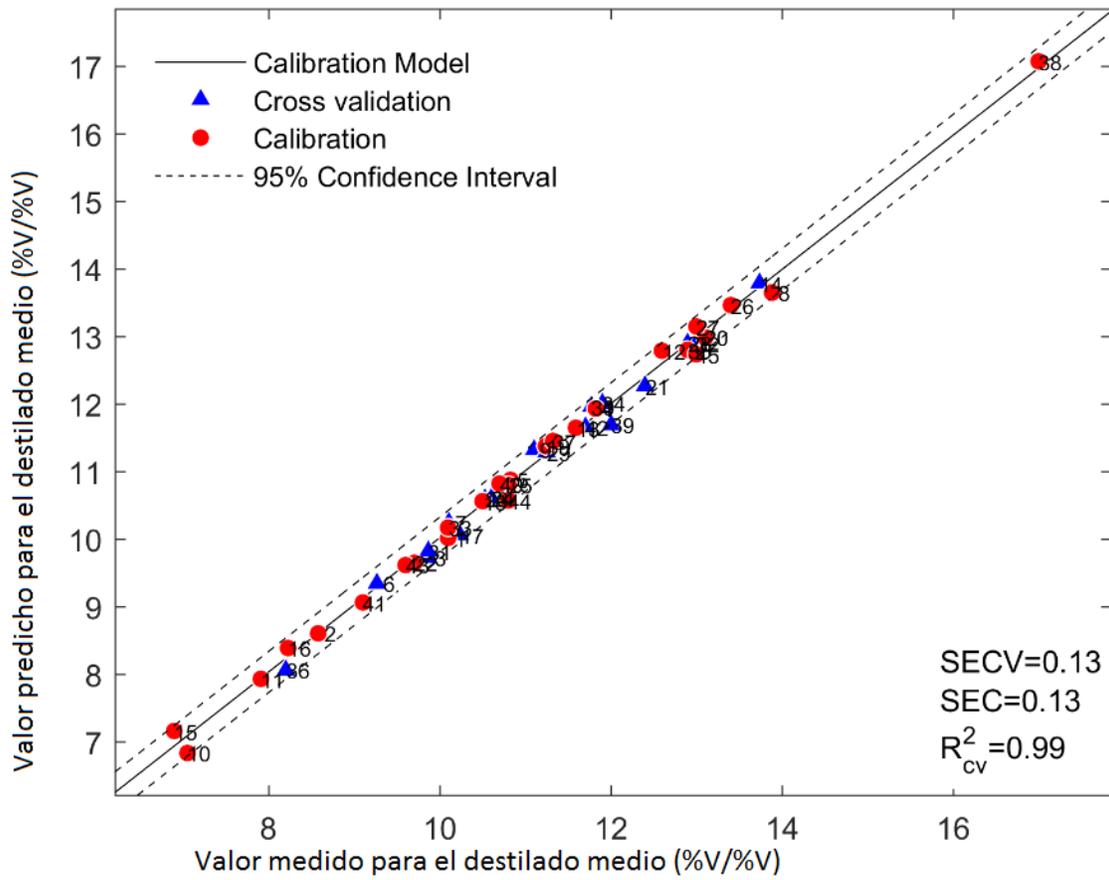
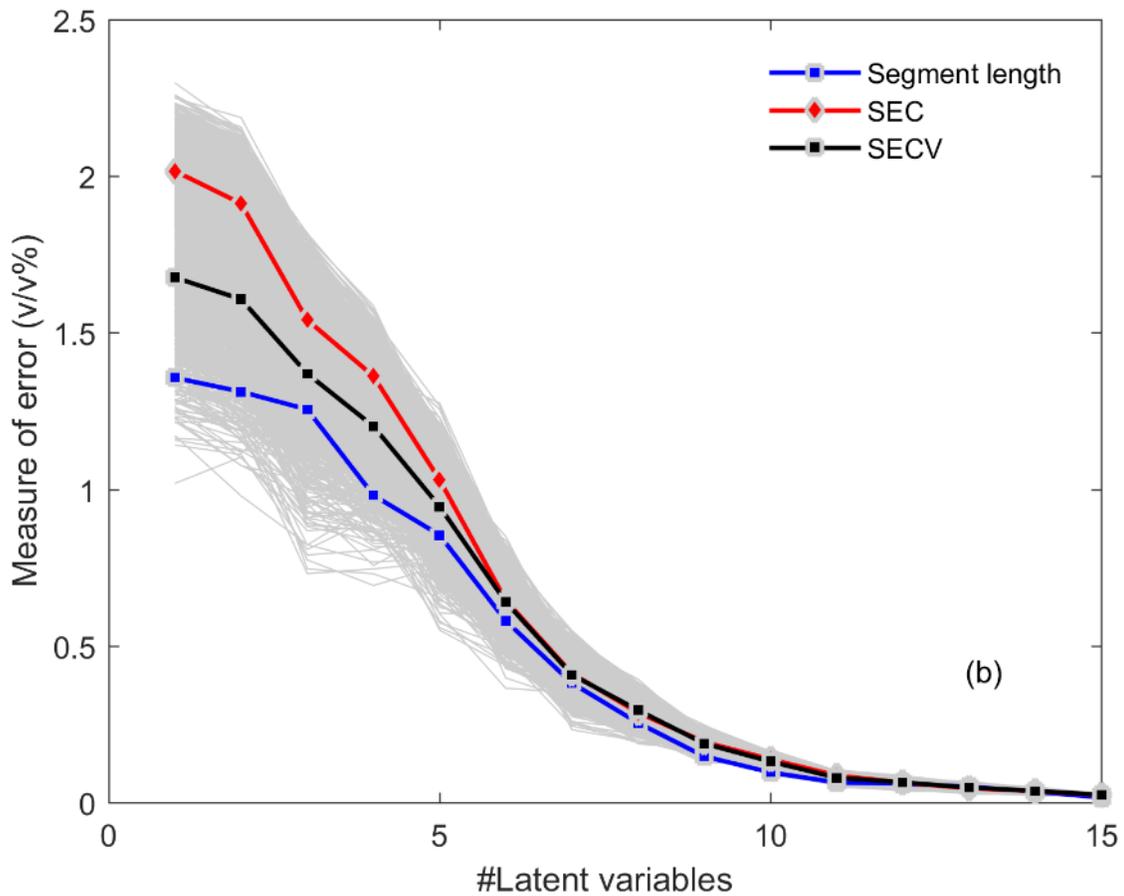


Figura 13. Gráfica del modelo PLS para el destilado medio 427°C -482°C

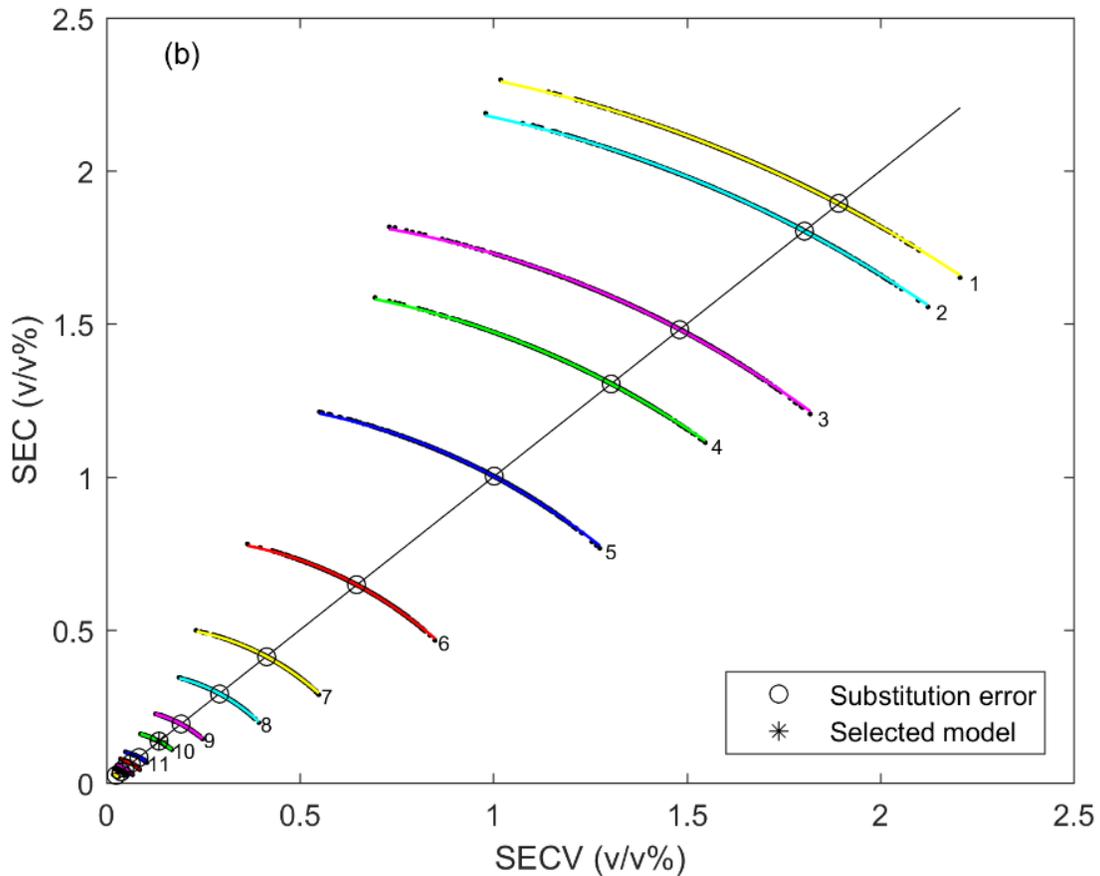


**Figura14.** Medidas de errores vs el número de componentes (variables latentes) para el modelo de predicción del destilado medio 427°C-482°C

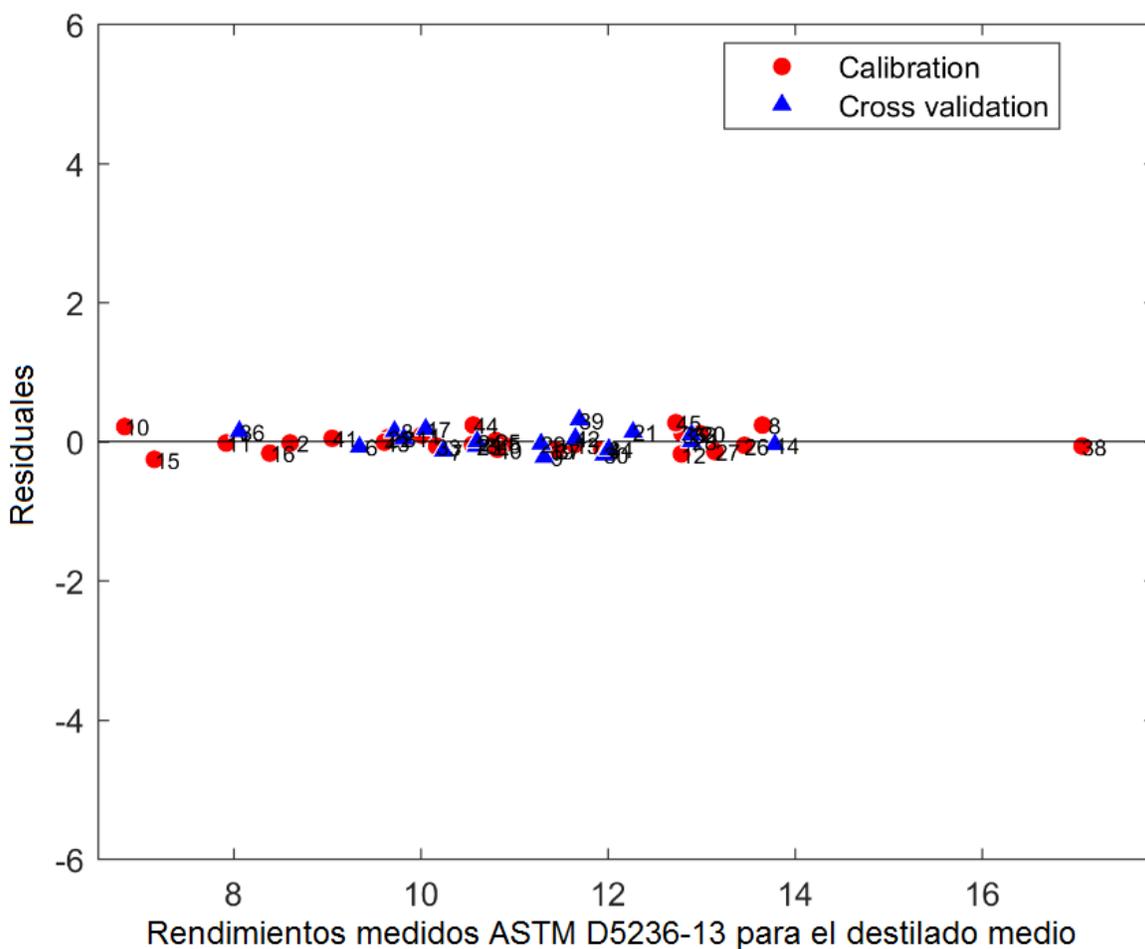
Siguiendo el mismo protocolo empleado para obtener los modelos para el destilado liviano se determinaron los demás rendimientos de los destilados de vacío.

En las figuras 14 y 15 se observa que el número óptimo de variables latentes del modelo del destilado medio 427°C-482°C es de 10 componentes, porque en este valor se obtienen bajos valores de errores 0.13% (SEC y SEVC) y alto coeficiente de determinación ( $R^2$ ) 0.99.

La baja dispersión de los residuales, indica poca variabilidad entre los datos experimentales y predictivos para el modelo del destilado medio 427°C-482°C, garantizando un mínimo error en los resultados, como se observa en la figura 16.



**Figura15.** Medida del error cuadrático medio de validación cruzada (SECV) vs error medio de calibración (SEC) para la determinar el # de variables latentes del modelo de destilado medio 427°C-482°C



**Figura16. Residuales para el modelo del destilado medio 427°C-482°C.**

Los parámetros estadísticos obtenidos para el modelo predictivo PLS del destilado medio 427°C-482°C se observan en la tabla 4.

**Tabla 4. Parámetros estadísticos del modelo del destilado medio 427°C -482°C**

Rendimiento	Número de componentes	SECV (%v/v)	SEC (%v/v)	$R_{CV}^2$	$R_C^2$
Destilado medio	10	0,13	0,13	0,99	1,00

3.5 Desarrollo del modelo de predicción PLS para el destilado pesado 482°C-565 °C

El modelo predictivo del destilado pesado 482°C -565°C calibrado con los valores de destilación de los crudos obtenidos por la metodología estándar ASTDM-5236-13 y bajo un intervalo del 95% de confianza, se obtiene: un valor de error medio de calibración (SEC) de 0.92%, de error medio de validación cruzada (SECV) de 0.94%, y un coeficiente de determinación ( $R^2$ ) de 0.90 como se observa en la figura 17.

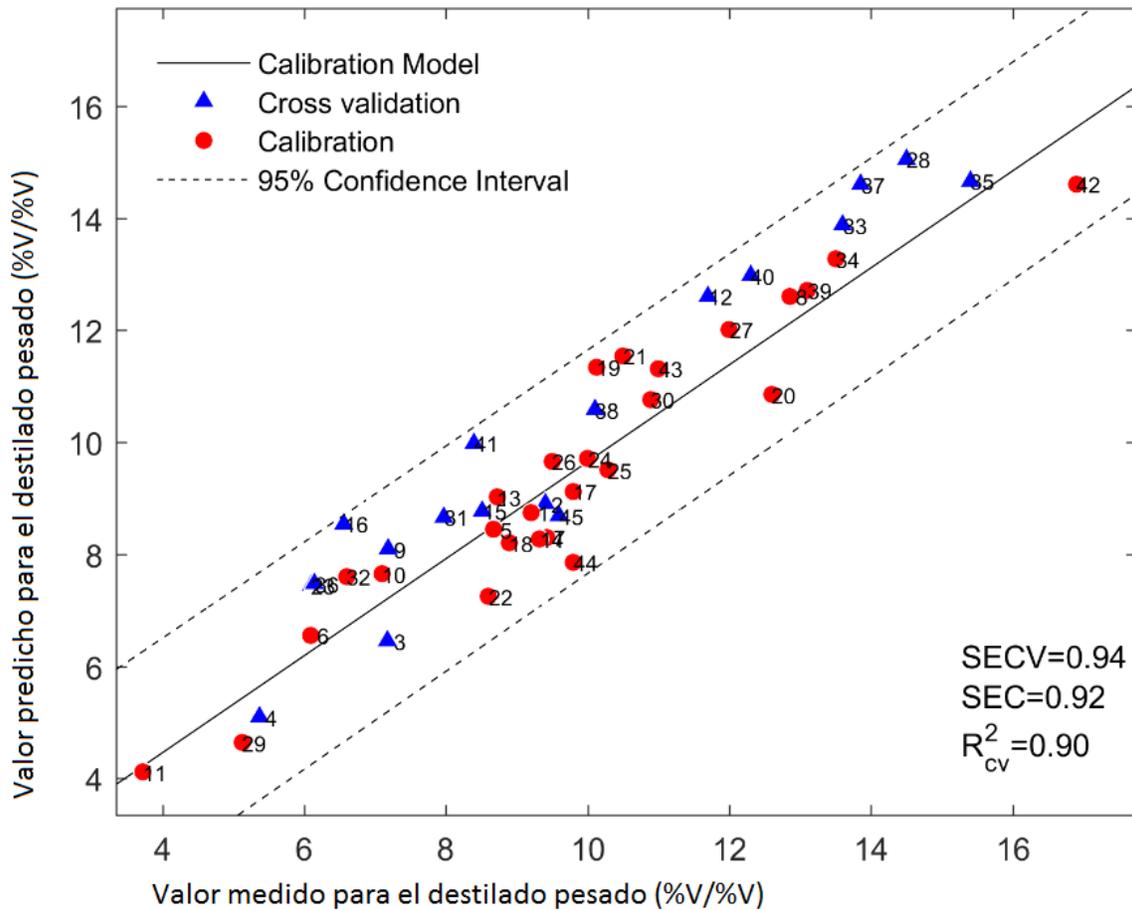
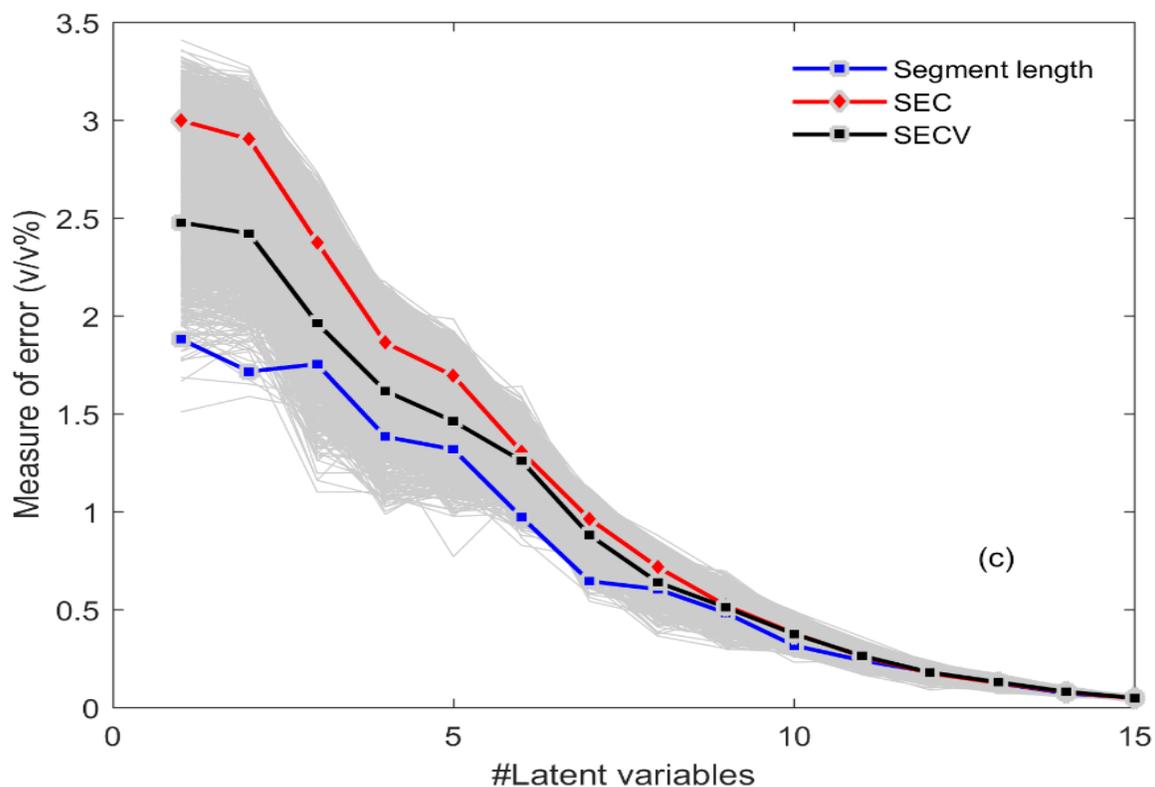


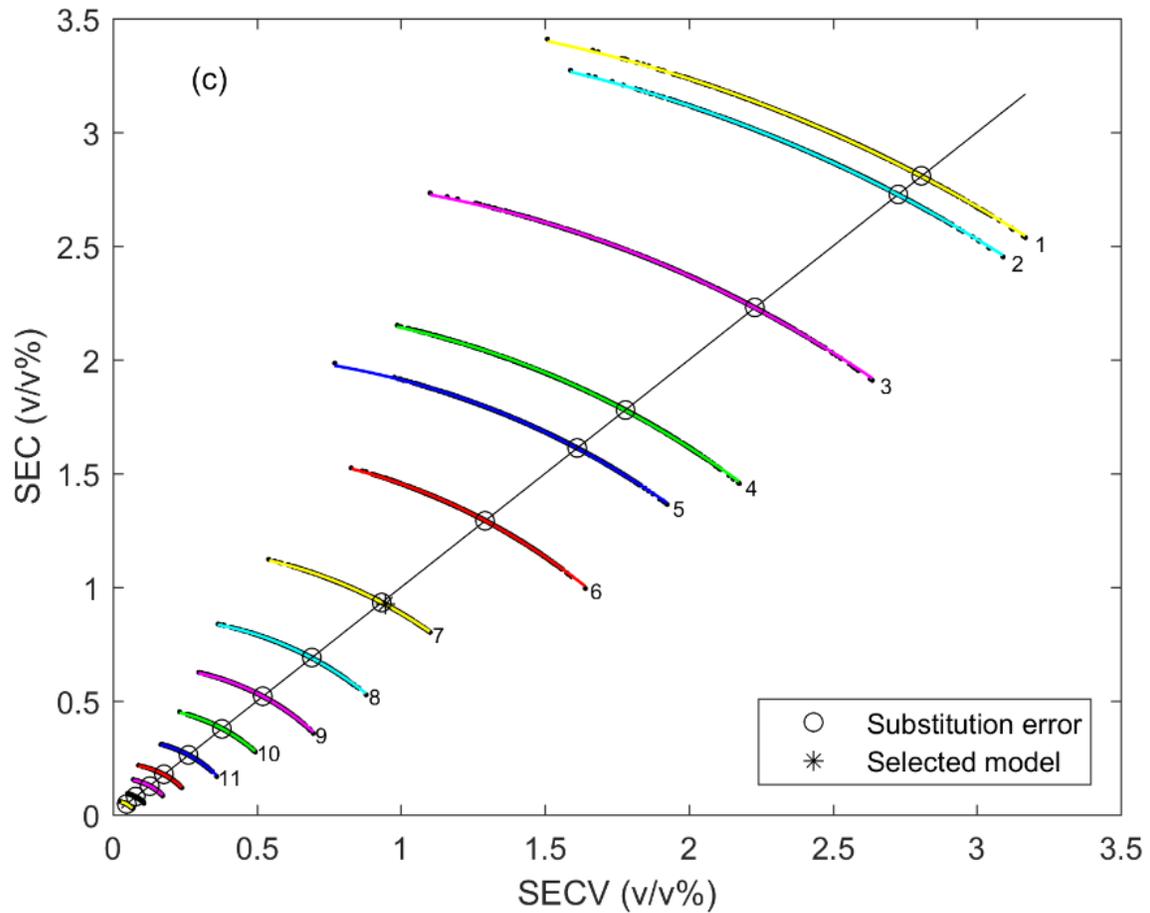
Figura 27. Gráfica del modelo PLS para el destilado pesado 482°C -565°C

El número óptimo de variables latentes para el modelo del destilado pesado corte 482°C-565°C es de 7 componentes como se observa en las figuras 18 y 19, obteniendo unos valores de error  $<1$ .

La dispersión de los residuales presenta un comportamiento homocedástico es decir variación de los residuales es constante y sin tendencias (Mongay-Fernández. 2005), indicando la baja variabilidad entre los datos experimentales y los predictivos para el modelo del destilado pesado 482°C -565°C, como se observa en la figura 20.



**Figura18.** Medidas de errores vs el número de componentes (variables latentes) para el modelo de predicción del destilado pesado 482°C -565°C



**Figura19.** Medida del error cuadrático medio de validación cruzada (SECV) vs error medio de calibración (SEC) para la determinar el # de variables latentes del modelo de destilado pesado 482°C -565°C

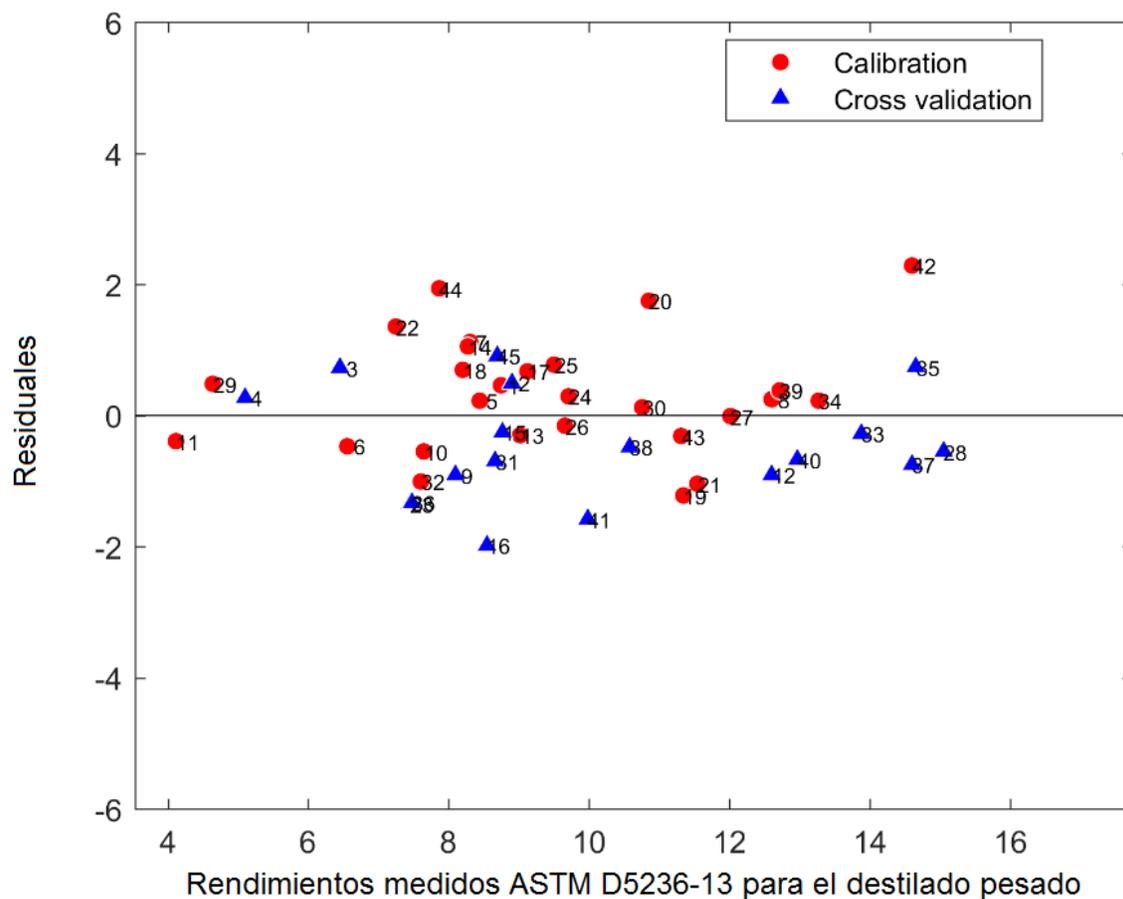


Figura 20. Residuales para el modelo del destilado pesado 482°C -565°C

Los parámetros estadísticos obtenidos para el modelo predictivo PLS del destilado pesado 482°C-565°C se observan en la tabla 5.

Tabla 5. Parámetros estadísticos del modelo del destilado pesado 482°C -565°C

Rendimiento	Número de componentes	SECV (%v/v)	SEC (%v/v)	$R^2_{CV}$	$R^2_C$
Destilado pesado	7	0,94	0,92	0,90	0,88

### 3.6 Desarrollo del modelo de predicción PLS para el fondo de vacío 565°C+

El modelo óptimo de predicción para el fondo de vacío 565°C+ se observa en la figura 21, con un error medio de validación cruzada (SECV) de 4.13, un error medio de calibración (SEC) de 3.70 y un coeficiente de determinación ( $R^2$ ) de 0.92 manteniendo un intervalo de confianza del 95%.

En las figuras 22 y 23 se observa como disminuye el comportamiento de los errores SECV y SEC a medida que aumenta el número de componentes (variables latentes). Para este modelo los valores óptimos de errores y coeficiente de determinación ( $R^2$ ) se obtuvieron con un número de 4 componentes (variables latentes).

Y finalmente se observa en la figura 24 que los residuales presentan un comportamiento homocedástico alrededor del cero, en un rango entre -5 % v y 5 %v.

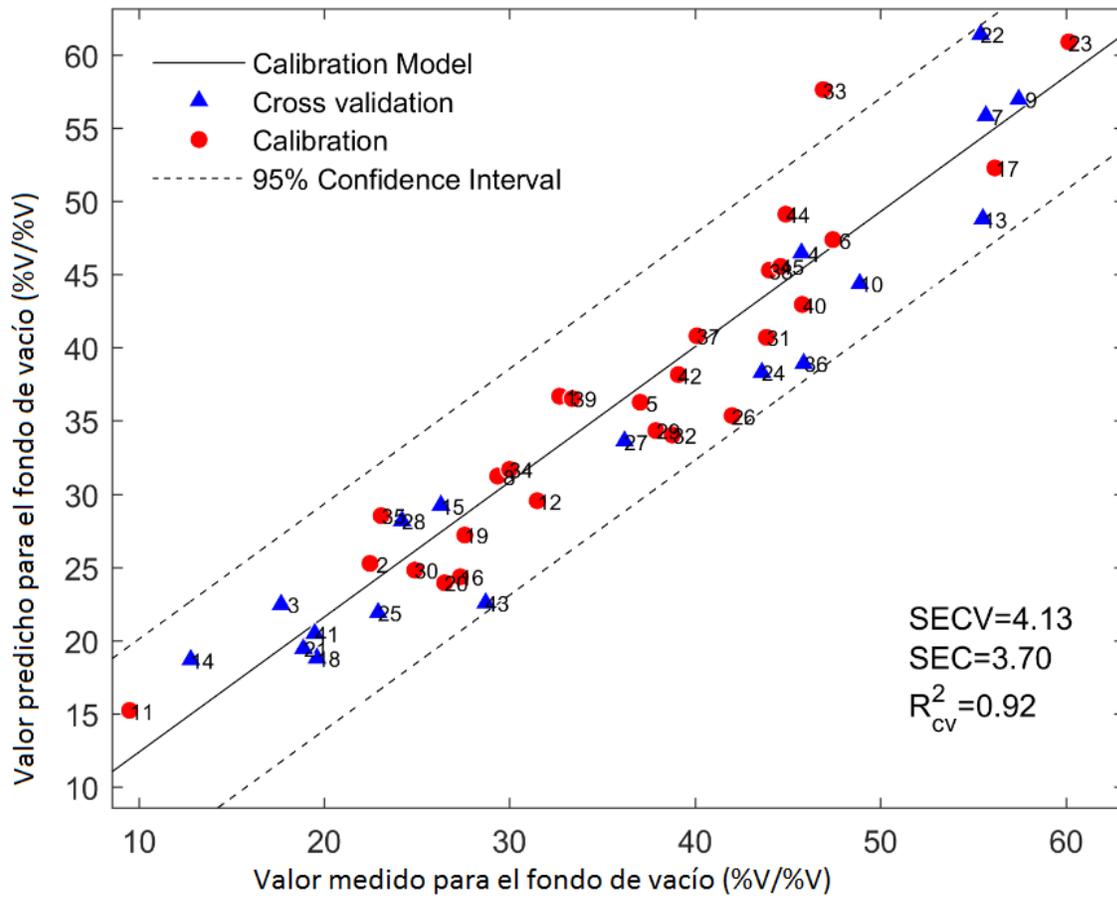
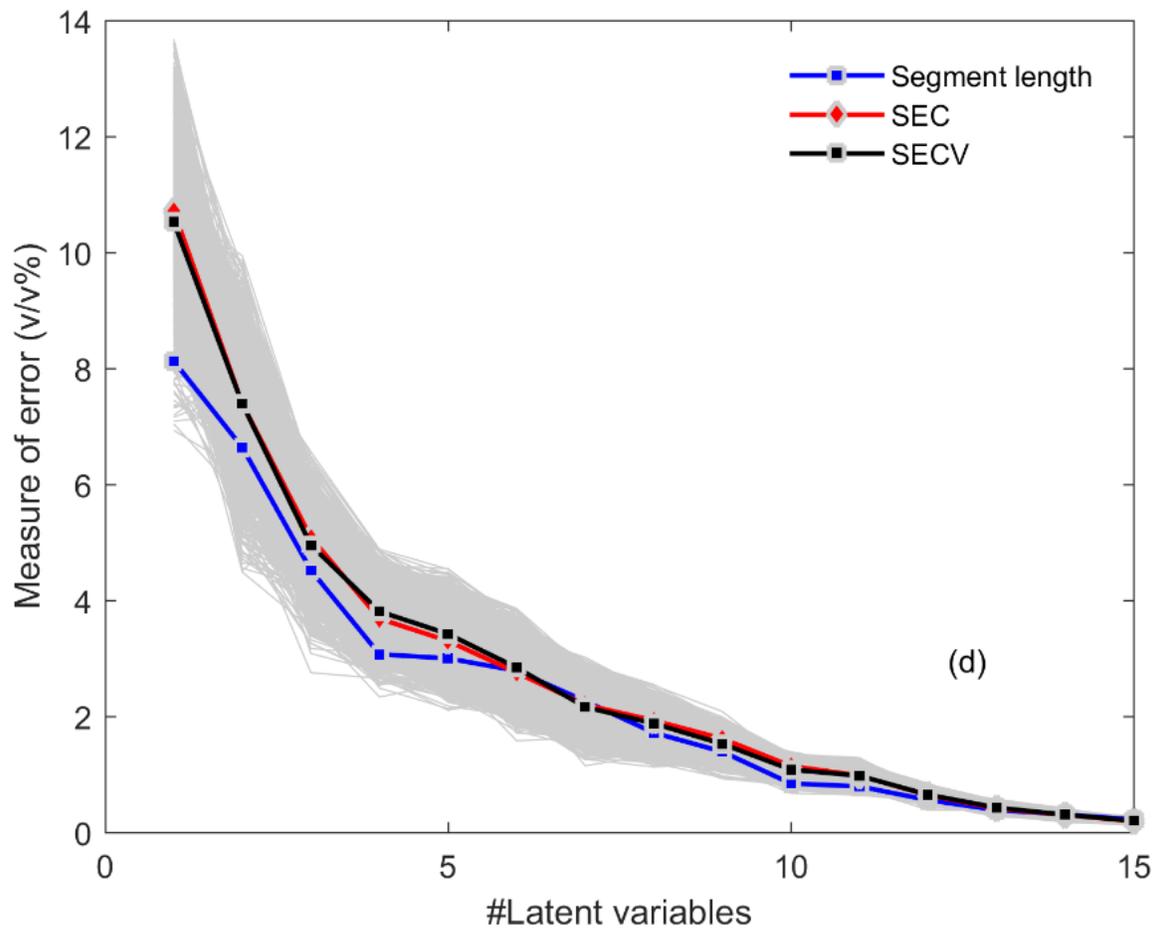
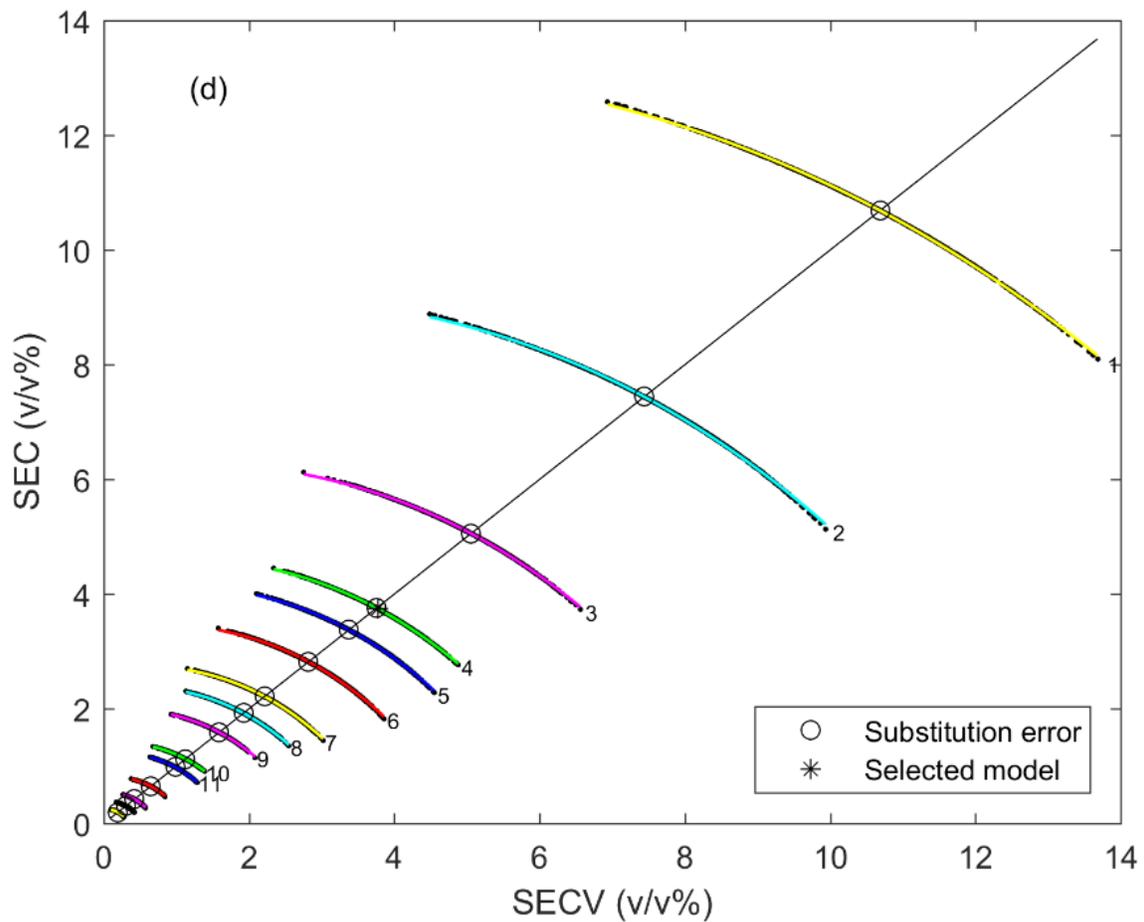


Figura 23. Gráfica del modelo PLS para el fondo de vacío 565°C+



**Figura 22.** Medidas de errores vs el número de componentes (variables latentes) para el modelo de predicción del fondo de vacío 565°C+



**Figura 23.** Medida del error cuadrático medio de validación cruzada (SECV) vs error medio de calibración (SEC) para la determinar el # de variables latentes del modelo del fondo de vacío 565°C+

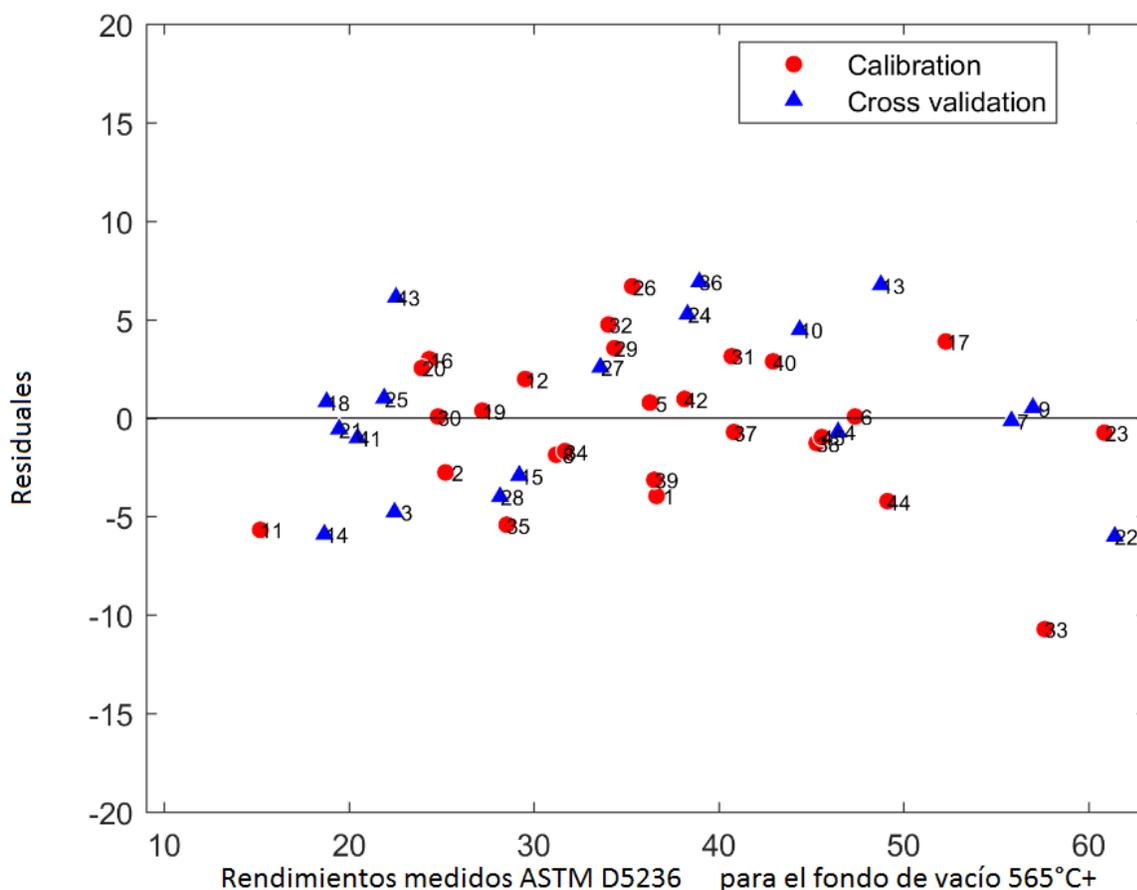


Figura 24. Residuales para el modelo del fondo de vacío 565°C+

Los parámetros estadísticos obtenidos para el modelo predictivo PLS del fondo de vacío 565°C+ se observan en la tabla 6.

Tabla 6. Parámetros estadísticos del modelo del fondo de vacío 565°C+

Rendimiento	Número de componentes	SECV (%v/v)	SEC (%v/v)	$R_{CV}^2$	$R_C^2$
Fondo de vacío	4	4,13	3,70	0,92	0,88

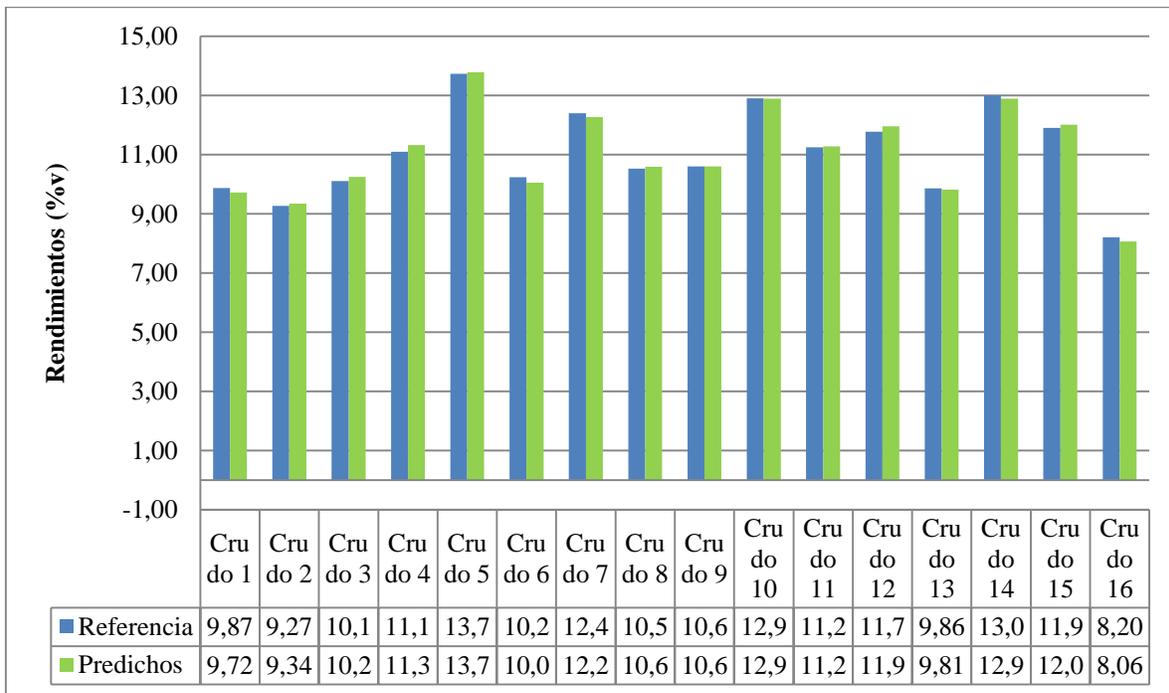
La validación de un modelo basado en datos experimentales busca estimar la incertidumbre de predicciones futuras, es decir evalúa el desempeño del modelo con nuevas muestras de la misma naturaleza de las que se usaron para su construcción. Si la incertidumbre es razonablemente baja el modelo es válido (Murcia, 2016, págs. 38-137)

El desempeño de los modelos PLS para predecir el rendimiento de las fracciones de vacío se evaluó de dos formas. La primera consiste en una validación cruzada (K-fold), con múltiples iteraciones que se realizó sobre el grupo de muestras de calibración resultando un error medio de validación cruzada de SECV entre el rango 0.13% y 4.13 %v. Por otra parte, ya que los modelos de regresión están diseñados para hacer predicciones futuras, se realizó una validación con muestras que no participaron en la construcción del modelo obteniéndose unos errores medios de calibración entre 0.40 y 3.70 %v, el número de variables latentes para los modelos predictivos de las fracciones al vacío se encuentran en el rango de 4 a 10 y el coeficiente de determinación  $R^2$  para los modelos estuvo entre 0.90 y 0.98 (ver tabla 7).

**Tabla 1. Parámetros estadísticos de los modelos PLS para predecir los rendimientos de las fracciones de vacío de crudos colombianos 371°C -565°C +**

Propiedad	Número de componentes	SECV (v %)	SEC (v %)	$R_{CV}^2$	$R_C^2$
Destilado liviano 371°C -427°C	8	0,41	0,40	0,98	0,97
Destilado medio 427°C -482°C	10	0,13	3,40	0,99	1,00
Destilado pesado 482°C -565°C	7	0,94	1,12	0,90	0,88
Fondo de vacío 565°C+	4	4,13	3,70	0,92	0,88

Al comparar los resultados entre los datos de referencia y los predichos por el modelo de la fracción de destilado medio 427° C- 482 °C (ver figura 25), se obtuvo una variación menor del 5%, lo que demuestra que el modelo tiene un buen rendimiento para la predicción de las fracciones de destilación al vacío en nuevas muestras de crudo.



**Figura 25. Diagrama comparativo entre los rendimientos de referencia y los predichos por modelo del destilado medio 427-482°C+**

#### 4. Conclusiones

La investigación demostró que la espectroscopía FTIR-ATR genera información espectral robusta que junto con la aplicación de métodos quimiométricos de análisis (PLS) logra determinar los rendimientos de las fracciones de destilación al vacío de crudos en menor tiempo y a bajo costo sin requerir el análisis de destilación.

Con los modelos obtenidos en esta investigación se logra predecir los rendimientos de las fracciones al vacío (destilado liviano, destilado medio, destilado pesado y fondo de vacío de crudos colombianos, siendo un método alternativo y versátil para obtener resultados previos de manera confiable, los modelos se desarrollaron con un intervalo de confianza del 95%.

La aplicabilidad de los modelos se evaluó a partir de los parámetros estadísticos de error cuadrático medio de calibración y error cuadrático medio de validación cruzada (SEC y SECV) obteniéndose resultados estadísticos satisfactorios.

### **5. Recomendaciones**

Para una mayor aplicación de los modelos, se recomienda ampliar el número de muestras, si es posible incluir muestras de crudos extranjeros o nuevas mezclas de los crudos colombianos, para que el rango cubra nuevas especificaciones en los rendimientos de las fracciones al vacío.

Ampliar el número de muestras de crudos colombianos para calibrar y realizar la validación cruzada, mantendría actualizado cada modelo predictivo.

**Referencias bibliográficas**

- Abbas, O.; Rebufa, C.; Dupuy, N.; Permanyer A. & Kister J. (2012). PLS regression on spectroscopic data for the prediction of crude oil quality: API gravity and aliphatic/aromatic ratio. *Fuel*; 98:5–14. DOI: 10.1016/j.Fuel.2012.03.045.
- Aleme, H.G.; Costa, L.M.; Barbeira J.S. (2009). Determination of ethanol and specific gravity in gasoline by distillation curves and multivariate analysis. *Talanta*.78:1422-1428.
- Aleme, H. G.; Barbeira, J.S. (2012). Determination of flash point and cetano index in diesel using distillation curves and multivariate calibration. *Fuel*.102:129-134.
- Al-Ghouti, M.; Al-Degs, Y. & Mustafa, M. (2010). Determination of hydrogen content, gross heat of combustion, and net heat of combustion of diesel fuel using FTIR spectroscopy and multivariate calibration. *Fuel* ;89:193 – 201. DOI:10.1016/j.fuel.2009.08.044.
- Altgelt, K.H.; Boduszynski, M.M. (1994). *Composition and analysis of heavy peroleum fractions*. New York, Marcel Dekker.
- Aske, N.; Kallevik, H. & Sjöblom, J. (2001). Determination of Saturate, Aromatic, Resin and Asphaltenic (SARA) Components in Crude Oils by Means of Infrared and Near-infrared Spectroscopy. *Energy & Fuels*. 15:1304–12. DOI 10.1021/ef010088h.
- ASTM E1655. Standard Practices for Infrared Multivariate Quantitative Analysis. 1 2012;05:29. DOI:10.1520/E1655-05R12.
- ASTM. ASTM D1160-15. Standard Test Method for Distillation of Petroleum Products at Reduced Pressure. 1 2013:1–20. DOI:10.1520/D1160-15.

- ASTM. ASTM D2892-13. Standard Test Method for Distillation of Crude Petroleum (15-Theoretical Plate) 2013:34. DOI:10.1520/D2892-13.
- ASTM. ASTM D5236-13. Standard Test Method for Distillation of Heavy Hydrocarbon Mixtures (Vacuum Potstill Method)1 2013:18. DOI:10.1520/D5236-17.
- Bao, X.; Dai, L.(2009). Partial least squares with outlier detection in spectral analysis: A tool to predict gasoline properties. *Fuel*.88: 1216-1222.
- Bezerra de Lira, L.de F.; Vera Cruz de Vasconcelos. F.; Fernandes Pereira. C.; Silveira Paim, A.P.; Stragevitch, L. & Pimentel, M.F. (2010). Prediction of properties of diesel/biodiesel blends by infrared spectroscopy and multivariate calibration. *Fuel*; 89:405 – 409.
- Chu X., LI J.; Chen, P.; Xu, Y. (2014). Review, Algorithms, Strategies and Application Progress of Spectral Searching Methods. *Chin J. Anal Chem.* 42(9):1379-1386.
- Chung, H.; Ku, M.; Lee, J. (1999). Comparison of near-infrared and mid-infrared spectroscopy for the determination of distillation property of kerosene. *Vibrational Spectroscopy*;20: 155-161.
- Côcco, L.C.; Yamamoto, C. I., Meien, O.F. (2004). Study of correlations for physicochemical properties of Brazilian gasoline. *Chemometrics and intelligent laboratory systems*;76: 55-63. doi.org/10.1016/j.chemolab.2004.09.004
- Diem, M. (1993) Introduction to modern vibrational spectroscopy. Hunter College. New York. Wiley-VCH. 1993.
- Fakayode, S.; Mitchell, B.; Pollard, D. (2014). Determination of boiling point of petrochemical by gas chromatography-mass spectrometry and multivariate regression analysis of structural activity relationship. *Talanta*.126:151-156.

- Ferreira da Silva, M. P.; Rodrigues e Brito, L.; Araújo Honorato, F.; Silveira Paim, A. P.; Pasquini, C. & Pimentel, M. F. (2013). Classification of Gasoline as with or without dispersant and detergent additives using infrared spectroscopy and multivariate classification. *Fuel*; 116:151–157. DOI:10.1016/j.fuel.2013.07.110.
- Filgueiras, P.R. (2016). Determination of Saturates, Aromatics, and Polars in Crude Oil by <sup>13</sup>C NMR and Support Vector Regression with variable Selection by Genetic Algorithm. *Energy & Fuel*. V30, pp 1972 – 1978.
- Flôres, M.; Viera, M.; Panta, R.E.; Fachini, D.; Gerbase, A.E.; Marder, L. (2011). Simultaneous determination of quality parameters of biodiesel/diesel blends using HATR-FTIR spectra and PLS, iPLS or siPLS regressions. *Fuel*.2011;90: 701-706.
- Fodor, G.E.; Kohl, KB. (1993). Analysis of Middle Distillate Fuels by Midband Infrared Spectroscopy. *Energy & Fuels*; V7: 598 – 601. DOI: 10.1021/ef00041a006
- Gómez-Carracedo, M.P.; Andrade, J.M.; Calviño, M.; Fernández, E.; Prada, D. (2003). Multivariate Prediction of Eight Kerosene Properties Employing Vapour-Phase Mid-Infrared Spectrometry. *Fuel*; V82:1211–8.
- Hoeil, C.; Min-Sik, K. (2000). Comparison of Near - Infrared, Infrared, and Raman Spectroscopy for the Analysis of Heavy Petroleum Products. *Applied Spectroscopy*.; V54: 239 – 245. DOI: 10.1366/0003702001949168
- Hongfu, Y.; Xiaoli, Ch.; Haoram, L.; Yupeng, X. (2006). Determination of Multi-properties of Residual Oils Using Mid-infrared Attenuated Total Reflection Spectroscopy. *Fuel*; 85:1720–1728. DOI: 10.1016/j.fuel.2006.02.003
- Hopke, P.K. (2003). The evolution of chemometrics. *Analytica Chimica Acta*. 500, 365–377. DOI:10.1016/S0003-2670(03)00944-9.

- Kardamakis Andreas, A. NP. (2010). Autoregressive modeling of near-IR spectra and MLR to predict RON values of gasolines. *Fuel*; 89:158–161.
- Kelly, JJ.; Callis, JB. (1990). Nondestructive Analytical Procedure for Simultaneous Estimation of the Major Classes of Hydrocarbon Constituents of Finished Gasolines. *Analytical Chemistry*; 1451:1444–1451.
- Kuptsov, A. K.; Arbuzova, T. V. (2011). A study of heavy oil fractions by Fourier-transform near-infrared Raman spectroscopy. *Petroleum Chemistry*; V51:203–211. doi:10.1134/S0965544111020083.
- Lavine, B.K.; Workman, J. (2013). Chemometrics. *Analytical Chemistry*; V85:705–714.
- Laxalde, J.; Ruckebusch, C.; Devos, O.; Caillol, N.; Wahl, F.; Duponchel, L. (2011). Characterisation of Heavy Oils Using Near-Infrared Spectroscopy: Optimisation of Pre-Processing Methods and Variable Selection. *Analytical Chimica Acta* 705: pp 227–234. doi:10.1016/j.aca.2011.05.048.
- Matthias, O. (2007). Chemometrics. Statistics and computer application in analytical chemistry. 3<sup>th</sup> Edition. Freiberg University of Mining and technology. Leipziger Germany: Wiley-VCH.
- Mendes, G.; Aleme, H.; Barbeira, J.S. (2012). Determination of octane numbers in gasoline by distillation curves and partial least squares regression. *Fuel*. 97:131-136.
- Mongay Fernandez, C.; Quimiometría (2005). Universitat de València. Educación. Materials. Primera edición.
- Murcia Celis, B. (2016). Predicción de propiedades fisicoquímicas de destilados atmosféricos (199-371°C+) a partir de sus crudos madre mediante FTIR-ATR y análisis multivariado. M.Sc. Thesis. Escuela de Química. Facultad de Ciencias Básicas. Universidad Industrial de Santander, Bucaramanga. Pp 38-137.

- Nakamoto, K. Infrared and Raman Spectra of inorganic and Coordination Compounds. Part A. Theory and applications in inorganic chemistry. 6th edition. Marquette University. Wiley-VCH. Chap. 1. 2001
- Orrego-Ruiz, J.A.(2014). Fraccionamiento por destilación molecular y caracterización de fondos de vacío de crudos colombianos. Tesis doctoral. Escuela de Química. Facultad de Ciencias. Postgrado en química. Universidad Industrial de Santander.
- Requena, A.; Zuñiga, J. (2004). Espectroscopía. Madrid. Pearson Educación S.A.
- Rubinson, K. A.; Rubinson, J. F. (2011). Análisis instrumental. Traductor: Ros Luis, 3ª edición. Nueva Jersey. Estados Unidos. Prentice Hall.
- Skoog, D.; Holler, F.J.; Nieman, T. (2002). Principios de análisis instrumental. 5ta edición. McGraw Hill. Cap. 16,17.

## Apéndices

## Apéndice A. Caracterización de los rendimientos de los crudos colombianos ASTM D5236-

13

Muestra LEAM	Identificación de la muestra	Destilado Liviano (371°- 427°C)	Destilado Medio (427°- 482°C)	Destilado Pesado (482°- 565°C)	FONDO DE VACÍO (565+)
1-1Norm	Crudo 1	5,66	10,10	9,20	32,70
1Norm	Crudo 2	5,66	10,10	9,20	32,70
2Norm	Crudo 3	9,66	8,58	9,40	22,48
3Norm	Crudo 4	5,33	9,87	7,18	17,71
4Norm	Crudo 5	6,52	11,82	5,37	45,76
5Norm	Crudo 6	6,99	10,83	8,67	37,05
6Norm	CRUDO 7	8,28	9,27	6,09	47,45
7Norm	Crudo 8	6,84	10,11	9,43	55,68
8Norm	Crudo 9	6,07	13,89	12,85	29,36
10Norm	Crudo 10	4,49	11,10	7,18	57,48
11Norm	Crudo 11	9,95	7,05	7,10	48,87
13Norm	Crudo 12	7,35	7,90	3,72	9,52
14Norm	Crudo 13	14,40	12,60	11,70	31,50
14- 1Norm	Crudo 14	5,00	11,59	8,72	55,54
16Norm	Crudo 15	7,37	13,74	9,32	12,80

<b>Muestra LEAM</b>	<b>Identificación de la muestra</b>	<b>Destilado Liviano (371°- 427°C)</b>	<b>Destilado Medio (427°- 482°C)</b>	<b>Destilado Pesado (482°- 565°C)</b>	<b>FONDO DE VACÍO (565+)</b>
16- 1Norm	Crudo 16	7,37	13,74	9,32	12,80
18Norm	Crudo 17	8,54	6,89	8,52	26,30
19Norm	Crudo 18	7,88	8,22	6,56	27,35
20Norm	Crudo 19	5,24	10,23	9,80	56,18
22Norm	Crudo 20	5,95	10,50	8,89	19,61
23Norm	Crudo 21	7,16	11,24	10,12	27,58
24Norm	Crudo 22	5,40	13,10	12,60	26,50
25Norm	Crudo 23	4,00	12,40	10,50	18,90
26Norm	Crudo 24	7,50	9,70	8,60	55,40
27Norm	Crudo 25	5,15	10,53	6,10	60,15
29Norm	Crudo 26	4,40	10,60	10,00	43,60
33Norm	Crudo 27	7,94	10,81	10,28	22,92
34Norm	Crudo 28	6,60	13,40	9,50	42,00
35Norm	Crudo 29	8,20	13,00	12,00	36,20
36Norm	Crudo 30	14,00	12,90	14,50	24,20
37Norm	Crudo 31	5,22	11,25	5,12	37,90
39Norm	Crudo 32	5,15	11,77	10,89	24,88
40Norm	Crudo 33	8,96	9,86	7,97	43,85
41Norm	Crudo 34	5,70	13,00	6,60	38,80
42Norm	Crudo 35	9,50	10,10	13,60	46,90
43Norm	Crudo 36	13,00	11,90	13,50	30,00
44Norm	Crudo 37	13,40	12,90	15,40	23,10

<b>Muestra LEAM</b>	<b>Identificación de la muestra</b>	<b>Destilado Liviano (371°- 427°C)</b>	<b>Destilado Medio (427°- 482°C)</b>	<b>Destilado Pesado (482°- 565°C)</b>	<b>FONDO DE VACÍO (565+)</b>
45Norm	Crudo 38	6,05	8,20	6,15	45,87
46Norm	Crudo 39	10,03	11,32	13,85	40,09
47Norm	Crudo 40	6,30	17,00	10,10	44,00
49Norm	Crudo 41	5,80	12,00	13,10	33,40
50Norm	Crudo 42	6,60	10,70	12,30	45,80
51Norm	Crudo 43	6,20	9,10	8,40	19,50
52Norm	Crudo 44	3,60	11,70	16,90	39,10
53Norm	Crudo 48	10,40	9,60	11,00	28,70
54Norm	Crudo 49	8,00	10,80	9,80	44,90
55Norm	Crudo 50	10,30	13,00	9,60	44,60

## Apéndice B. Gravedad API de los crudos colombianos

Muestra LEAM	Identificación de la muestra	Gravedad API
1-1Norm	Crudo 1	26,20
1Norm	Crudo 2	28,50
2Norm	Crudo 3	31,80
3Norm	Crudo 4	23,00
4Norm	Crudo 5	24,50
5Norm	Crudo 6	12,50
6Norm	CRUDO 7	8,50
7Norm	Crudo 8	25,0
8Norm	Crudo 9	8,2
10Norm	Crudo 10	11,9
11Norm	Crudo 11	42,0
13Norm	Crudo 12	14,1
14Norm	Crudo 13	9,40
14-1Norm	Crudo 14	33,8
16Norm	Crudo 15	12,80
16-1Norm	Crudo 16	12,80
18Norm	Crudo 17	26,30
19Norm	Crudo 18	27,35
20Norm	Crudo 19	56,18
22Norm	Crudo 20	19,61
23Norm	Crudo 21	27,58
24Norm	Crudo 22	26,50
25Norm	Crudo 23	18,90

Muestra LEAM	Identificación de la muestra	Gravedad API
26Norm	Crudo 24	55,40
27Norm	Crudo 25	60,15
29Norm	Crudo 26	43,60
33Norm	Crudo 27	22,92
34Norm	Crudo 28	42,00
35Norm	Crudo 29	36,20
36Norm	Crudo 30	24,20
37Norm	Crudo 31	37,90
39Norm	Crudo 32	24,88
40Norm	Crudo 33	43,85
41Norm	Crudo 34	38,80
42Norm	Crudo 35	46,90
43Norm	Crudo 36	30,00
44Norm	Crudo 37	23,10
45Norm	Crudo 38	45,87
46Norm	Crudo 39	40,09
47Norm	Crudo 40	44,00
49Norm	Crudo 41	33,40
50Norm	Crudo 42	45,80
51Norm	Crudo 43	19,50
52Norm	Crudo 44	39,10
53Norm	Crudo 48	28,70
54Norm	Crudo 49	44,90
55Norm	Crudo 50	44,60