

Desarrollo de un modelo multivariado predictivo de propiedades fisicoquímicas de crudos mejorados en procesos térmicos de recobro usando espectroscopia infrarroja FTIR-ATR y Quimiometría

José Miguel Coneo Vergara
Marcos Felipe Carvajal Antolinez

Trabajo de Grado para Optar al Título de Ingeniero de Petróleos

Director:
Adan Yovani León Bermúdez
Ingeniero Químico, Ph.D

Codirectores:
Julian Ernesto Sanabria Becerra
Ingeniero de Petróleos, M.Sc (C)

Universidad Industrial de Santander
Facultad de Ingeniería Fisicoquímicas
Escuela de Ingeniería de Petróleos
Bucaramanga
2025

Dedicatoria

En primer lugar, quiero dedicar este logro a Dios, porque a pesar de este camino tan largo no me ha dejado solo y me ha permitido vivir grandes cosas, logros, vivencias y muchas experiencias en toda la carrera.

A mi padre, porque gracias a su fortaleza mental y física no hubiese podido cumplir todos mis sueños y metas, gracias por esforzarte por mí y por mi hermana, porque sé que todo lo que hiciste en este camino fue con mucho esfuerzo, gracias a ti padre porque con tus enseñanzas, con tu cariño y con cada uno de esos consejos siempre los atesoro en mi ser y aunque no los tomaba todos siempre terminabas teniendo la razón, gracias por orientarme en la palabra de Dios y por permitirme ver más allá de lo físico, gracias por tus oraciones a mi favor porque sé que sin ellas no hubiese podido llegar hasta este momento. Te amo padre.

A mi madre, a ella dedico esto porque sé que las madrugadas y traspasadas han llegado a la meta que tanto anheló, también porque sé que oraste a Dios junto con papá para que me fuese bien, gracias por ser esa inspiración en mi vida, por ser esa persona que siempre creía en mis capacidades y me enseñaba a ser mejor, gracias infinitas por enseñarme a ser un hombre mejor. Este logro es tuyo madre mía

A mi hermana, porque desde primer semestre me escuchaste y me apoyaste en esos momentos en los que me quería rendir, este logro es para ti porque estuviste en cada uno de los logros, te dedico esto “negrita”, porque eres mi inspiración más sincera para ser el gran ingeniero y hermano de la futura doctora, que algún día llegaré a ser. Te amo.

. A la memoria de mi nonita Martha que, aunque ya no estas con nosotros y no pueda disfrutar este logro al lado tuyo, siempre llevaré en mi corazón tus risas, tu cariño y el amor que me demostrabas siempre en cada visita, me enseñaste tanto que no sabría cuánto agradecerte, te amo nonita, más que a nadie, esto va dedicado a ti...Lo logramos nonita.

A mis abuelos Marco y Victoria, les agradezco con todo mi corazón por su amor incondicional, por cada gesto de cariño y por enseñarme, con su ejemplo, la verdadera esencia de la familia, gracias por estar siempre presentes, por sus consejos llenos de sabiduría y por el inmenso afecto que me han brindado a lo largo de mi vida. Su amor ha sido un refugio y una inspiración constante.

A mis amigos, Coneo cuchito, que con sus consejos dentro y fuera del campo de softbol y su dedicación en este trabajo de grado no hubiese sido posible este logro, a mi prima Dani porque me aconsejaste en los momentos duros de la carrera y me guiaste a través de ellos, a Ronaldo que desde segundo semestre ha sido participe de muchas experiencias dentro y fuera del país, a Mario porque a pesar de la distancia ha sido de gran apoyo, a Garri por su amistad y su guía en caminos internacionales y por su valentía, a Dani S, que aunque ya no estes en este país, me apoyaste y me brindaste una mano amiga durante el tiempo que estudiamos juntos, a Leidy C porque me brindaste tu amistad incondicional en esta etapa, a Yoha, a Nataly lo logramos enana, a Dianita, al Coste, a Chachan, y a cada uno de mis compañeros de universidad que no nombré, porque el camino ha sido demasiado largo y culebrero, gracias por ser partícipes de este logro y por ser quienes son.

Dedicado a toda mi familia que de una u otra forma me apoyo en este sueño, lo logramos señores, el primer ingeniero de la familia.

Marcos Felipe Carvajal Antolinez

Dedicatoria

Primeramente, agradecerle a Dios en este trayecto que no ha sido nada fácil, que con el pude refugiarme en ocasiones las cuales no podía, pude desahogarme y nunca me ha dejado solo.

Agradecerle a mi mamá, que gracias a ella pude lograr todo lo que soy ahora, con sus palabras, consejos, regaños y sacrificio pudo sacarme adelante y culminar conmigo esta etapa. Sé que no ha sido nada fácil para ella, que fue un camino lleno de obstáculos para los dos pudimos salir adelante en todo lo que se nos iba presentando, gracias mamita por tu sacrificio, por tus oraciones, por doblar rodilla mañana tarde y noche, este logro es tanto mío como tuyo, te amo.

Darle gracias a mi papá, que ha sido ese hombre de la casa, que siempre ha soportado todo por vernos felices, que cada día se esfuerza en el trabajo para que no nos falte absolutamente nada, que ha sabido mantener la calma cuando no hemos podido más, ese bastón que Dios me dio, gracias papito, este logro también es tuyo, gracias por esas llamadas, esos consejos, esos te quiero y esos “tu puedes hijo, estoy orgulloso de ti” que en algún momento me ayudaron a levantarme, no te lo digo casi, pero eres el mejor padre, te amo.

A mi hermanita, que ha sido fundamental en todo este recorrido, gracias pulga, por ser esa alegría de la casa, por sacarnos canas de las buenas, te amo con toda mi alma, por fin tu hermanita logrará el sueño que tanto quieres, ser in ingeniero.

Un agradecimiento especial a ti Angélica y a toda tu familia, gracias por ayudarme, por acogerme como un hijo, por siempre darme apoyo y ‘pullarme’ para salir adelante, por nunca dejarme solo, este logro también va dirigido a ustedes que siempre estuvieron ahí desde el principio y hasta ahora, gracias por todo, los quiero como una familia.

Una dedicatoria especial en el cielo, a mi abuelita Rochi, que a pesar de que no está conmigo en cuerpo, siempre me ha acompañado en alma, sé que ella está orgullosa allá en el cielo por ver a su negro convertirse en un ingeniero, era el sueño que tenía y por fin se lo cumpliré, te amo abuelita, gracias por absolutamente todo.

A mi amigo/hermano de otra madre, Keismer por ser ese apoyo incondicional que tanto necesite, por esos consejos que tanto me diste, esas peleas que en algún momento tuvimos, gracias, hermano lo logramos!

A todos mis amigos, Marcos, gracias por acompañarme en este proceso y poder culminar juntos esta etapa, Ronaldo, por ser ese amigo que me llamaba y me sacabas de la casa a despejar la mente a jugar, salir, tomarse algo, aunque nunca te lo dije, necesitaba esos espacios, a María José, que ayudo tanto en la etapa final de este logro y ya no está en el país, esto también es tuyo, tus ocurrencias y risas lograron ser más llevadero cada clase vista juntas y a cada uno de las personas que me acompañaron a lo largo de este camino, les doy las gracias por ser siempre un apoyo en cada momento de esta etapa.

A toda mi familia, gracias por nunca desampararme y por siempre velar por mi bienestar, les agradezco por estar siempre presente, los quiero a todos.

JOSÉ MIGUEL CONEO VERGARA

Agradecimientos

Los autores de este trabajo expresan agradecimientos a:

La Universidad Industrial de Santander.

La Escuela de Ingeniería de Petróleos.

Grupo de Investigación de Corrosión y el doctor Darío Peña.

El doctor Adán León y el ingeniero Julián Sanabria por su paciencia, rigor y apoyo durante este proyecto.

Asociación Centro de Estudios (ACEIP) por sus espacios de conocimiento y esparcimiento en el desarrollo de este proyecto.

Tabla de Contenido

INTRODUCCIÓN.....	19
1 OBJETIVOS.....	21
1.1 OBJETIVO GENERAL.....	21
1.2 OBJETIVOS ESPECÍFICOS.....	21
2 FUNDAMENTOS TEÓRICOS Y ESTADO DEL ARTE	22
2.1 CLASIFICACIÓN DE CRUDOS.....	22
2.2 CRUDOS PESADOS	22
2.3 MÉTODOS DE RECOBRO	23
2.3.1 <i>Inyección de vapor.</i>	25
2.3.2 <i>Inyección Cíclica con Vapor.</i>	26
2.3.3 <i>Drenaje Gravitacional asistido por vapor.</i>	27
2.4 ACUATERMÓLISIS CATALÍTICA.	28
2.5 ESPECTROSCOPIA INFRARROJA	30
2.5.1 <i>Tipos de espectroscopía infrarroja</i>	30
2.6 QUIMIOMETRÍA.....	32
2.6.1 <i>Resolución de curvas por automodelado (SMCR)</i>	33
2.6.2 <i>Análisis por componentes principales</i>	33
2.6.3 <i>Regresión a regresión en componentes principales (PCR)</i>	34
2.6.4 <i>Regresión de mínimos cuadrados parciales (PLS)</i>	35
NOTA: CREADA Y ADAPTADA POR AUTORES.....	39
3 METODOLOGÍA EXPERIMENTAL.....	39

NOTA: CREADA Y ADAPTADA POR AUTORES	40
3.1 SELECCIÓN DE LAS MUESTRAS	40
3.2 CARACTERIZACIÓN DE LAS MUESTRAS.....	41
3.2.1 <i>API y Densidad.</i>	41
3.2.2 <i>Viscosidad</i>	41
3.2.3 <i>Contenido de asfaltenos</i>	45
3.3 ANÁLISIS MULTIVARIABLE	46
3.3.1 <i>Espectroscopia Infrarroja</i>	46
3.3.2 <i>Modelos PLS</i>	48
3.4 ANÁLISIS ESPECTROSCOPÍA INFRARROJA FTIR-ATR.....	49
4 RESULTADOS Y ANÁLISIS.	53
4.1 MODELOS DE REGRESIÓN LINEAL MÚLTIPLE	59
4.1.1 <i>Modelo asfaltenos por RLM</i>	59
4.1.2 <i>Modelo densidad RLM</i>	62
4.1.3 <i>Modelos de viscosidad</i>	64
4.2 MODELOS PLS.....	71
4.2.1 <i>Modelos Asfaltenos C7+</i>	72
4.2.2 <i>Modelo PLS densidad</i>	80
4.2.3 <i>Modelo PLS viscosidad</i>	83
5 ANÁLISIS DE RESULTADOS.....	107
6 CONCLUSIONES.....	109
7 RECOMENDACIONES.....	110
REFERENCIAS BIBLIOGRÁFICAS.....	111

APÉNDICES..... 117

Lista de Tablas

<i>Tabla 1. Clasificación de crudos.</i>	22
<i>Tabla 2. Asignación de Grupos Funcionales en Espectroscopia Infrarroja.</i>	31
<i>Tabla 3. Resumen de investigaciones espectroscopia infrarroja y/o quimiometría.</i>	35
<i>Tabla 4. Promedio de Temperaturas y viscosidades en muestra.</i>	43
<i>Tabla 5. Correlaciones con la Ecuación de Andrade.</i>	43
<i>Tabla 6.</i>	44
<i>Tabla 7. Algoritmo de la regresión PLS.</i>	49
<i>Tabla 8. Parámetros moleculares promedio.</i>	52
<i>Tabla 9. Valores obtenidos de viscosidad de crudos provenientes de pruebas de AQ.</i>	54
<i>Tabla 10. Estadística de Parámetros Moleculares Promedio Asfaltenos.</i>	59
<i>Tabla 11. Análisis de varianza, ANOVA.</i>	60
<i>Tabla 12. Variables y R2 de modelamiento para el contenido de asfaltenos.</i>	60
<i>Tabla 13. Estadística de Parámetros Moleculares Promedio Densidad.</i>	62
<i>Tabla 14. Análisis de varianza, ANOVA</i>	62
<i>Tabla 15. Variables y R2 de modelamiento de la densidad .</i>	62
<i>Tabla 16. Estadística de Parámetros Moleculares Promedio para el modelo de viscosidad @ 30°C (LnVisc30).</i>	64
<i>Tabla 17. Análisis de varianza, ANOVA.</i>	64
<i>Tabla 18. Variables y R2 de modelamiento para la viscosidad @ 30°C (LnVis30).</i>	65
<i>Tabla 19. Estadística de Parámetros Moleculares Promedio para el modelo de viscosidad @ 45°C (LnVis45).</i>	65
<i>Tabla 20. Análisis de varianza, ANOVA.</i>	66
<i>Tabla 21. Variables y R2 de modelamiento para el modelo de viscosidad @ 45°C (LnVis45).</i>	66

<i>Tabla 22. Estadística de Parámetros Moleculares Promedio para el modelo de viscosidad @ 60°C (LnVis60).</i>	68
<i>Tabla 23. Análisis de varianza, ANOVA.</i>	68
<i>Tabla 24. Variables y R2 de modelamiento para el modelo de viscosidad @ 60°C (LnVis60).</i>	69
<i>Tabla 25. Parámetros de cálculo de propiedades.</i>	70
<i>Tabla 26. Anchos de bandas de los Espectros usados.</i>	71
<i>Tabla 27. Modelos representativos Asfaltenos < 5%.</i>	73
<i>Tabla 28. Modelos representativos asfaltenos < 5%.</i>	76
<i>Tabla 29. Modelos representativos.</i>	80
<i>Tabla 30. Modelos para el comportamiento en la viscosidad de crudos usando espectro 2.</i>	83
<i>Tabla 31. Modelos para el comportamiento en la viscosidad de crudos usando espectro 4.</i>	84
<i>Tabla 32. Modelos para el comportamiento en la viscosidad de crudos usando tratamiento del espectro 2.</i>	84
<i>Tabla 33. Modelos para el comportamiento en la viscosidad de crudos usando tratamiento del espectro 4.</i>	85
<i>Tabla 34. Modelos para viscosidades mayores a 1000 cP usando espectro 2.</i>	86
<i>Tabla 35. Modelos viscosidades mayores a 1000 cP usando espectro 3.</i>	86
<i>Tabla 36. Modelo Lnvisc @ 30°C menor a 1000 cP.</i>	87
<i>Tabla 37. Modelo visc @ 30°C > 1000 cP.</i>	90
<i>Tabla 38. Modelo Lmvisc @ 45°C < 1000 cP</i>	93
<i>Tabla 39. Modelo Lnvisc @ 45°C para viscosidades mayores a 1000 cP</i>	97
<i>Tabla 40. Modelo Lnvisc @ 60°C < 1000 cP</i>	100
<i>Tabla 41. Modelo Lnvisc @ 60°C > 1000 cP</i>	103

<i>Tabla 42. Asfaltenos > 5% Primera regresión.....</i>	<i>117</i>
<i>Tabla 43. Asfaltenos > 5% Análisis Final.</i>	<i>117</i>
<i>Tabla 44. Asfaltenos > 5% línea base primera regresión.</i>	<i>118</i>
<i>Tabla 45. Asfaltenos > 5% línea base última regresión.....</i>	<i>119</i>
<i>Tabla 47. Asfaltenos >5% normalizado última regresión</i>	<i>124</i>
<i>Tabla 48. Asfaltenos < 5%. base primera regresión.....</i>	<i>127</i>
<i>Tabla 49. Asfaltenos< 5%. base última regresión.....</i>	<i>127</i>
<i>Tabla 50. Asfaltenos normalizado primera regresión< 5%.....</i>	<i>128</i>
<i>Tabla 51. Asfaltenos normalizado última regresión < 5%.</i>	<i>129</i>
<i>Tabla 52. Asfaltenos línea base primera regresión < 5%.</i>	<i>132</i>
<i>Tabla 53. Asfaltenos línea base última regresión < 5%.</i>	<i>132</i>
<i>Tabla 54. Modelo Densidad Primera regresión.....</i>	<i>135</i>
<i>Tabla 55. Modelo densidad última regresión.....</i>	<i>135</i>
<i>Tabla 56. Modelo densidad línea base, primera regresión.....</i>	<i>138</i>
<i>Tabla 5757. Modelo densidad línea base, última regresión.</i>	<i>138</i>
<i>Tabla 5858. Modelo densidad normalizado, primera regresión.....</i>	<i>138</i>
<i>Tabla 59. Modelo densidad normalizado, última regresión.</i>	<i>139</i>

Lista de Figuras

<i>Figura 1 Métodos de Recobro Mejorado.....</i>	<i>24</i>
<i>Figura 2 Inyección de Vapor.....</i>	<i>27</i>
<i>Figura 3. Ventana Acuatermolisis</i>	<i>29</i>
<i>Figura 4. Metodología experimental de la investigación.</i>	<i>40</i>
<i>Figura 5. Viscosímetro rotacional B-one plus Lamy Rheology.</i>	<i>42</i>
<i>Figura 6. Husillo MK-SV418 – Camisa de Medición.....</i>	<i>44</i>
<i>Figura 7. Lavado Soxhlet de los asfaltenos.....</i>	<i>45</i>
<i>Figura 8. Imagen de referencia Asfaltenos después de lavado y secado, tomada del Laboratorio Grupo de Investigación en Corrosión, UIS Guatiguará (2025).....</i>	<i>46</i>
<i>Figura 9. Reflexión de la Muestra.</i>	<i>47</i>
<i>Figura 10. Equipo Nicolet Summit X. FTIR Spectrometer.....</i>	<i>47</i>
<i>Figura 11. Repetibilidad Espectro General Muestra CR58.</i>	<i>50</i>
<i>Figura 12. Vista 3D Axial de la Repetibilidad.</i>	<i>51</i>
<i>Figura 13. Gráfico Viscosidad vs Temperatura.</i>	<i>54</i>
<i>Figura 14. Paso a paso para realización del modelo.....</i>	<i>59</i>
<i>Figura 15. Cantidad de coeficientes principales para el contenido de asfaltenos.....</i>	<i>61</i>
<i>Figura 16. Cantidad de coeficientes principales para la densidad.....</i>	<i>63</i>
<i>Figura 17. Cantidad componentes principales para el modelo de viscosidad @ 45°C (LnVisc45). 66</i>	
<i>Figura 18. Coeficientes principales para el modelo de viscosidad @ 60°C (LnVisc60)</i>	<i>70</i>
<i>Figura 19. Valores medidos vs predichos para el modelo sin alteración para el contenido de asfaltenos (>5% peso).....</i>	<i>74</i>

<i>Figura 20. Cantidad Componentes Principales Modelo seleccionado sin alteración para el contenido de asfaltenos (% peso)</i>	<i>75</i>
<i>Figura 21. Valores medidos vs predichos para el modelo de asfaltenos < 5%.</i>	<i>78</i>
<i>Figura 22. Gráfica Componentes Principales Asfaltenos < 5%.</i>	<i>79</i>
<i>Figura 23. Valores medidos vs predichos para el modelo línea base seleccionado.</i>	<i>81</i>
<i>Figura 24. Componentes principales modelo escogido.</i>	<i>82</i>
<i>Figura 25. Valores medidos vs predichos para el modelo LnVisc @ 30°C menor a 1000 cP.</i>	<i>88</i>
<i>Figura 26. Componentes principales modelo escogido.</i>	<i>89</i>
<i>Figura 27. Valores medidos vs predichos para el modelo Lnvisc @ 30°C mayores a 1000 cP.</i>	<i>91</i>
<i>Figura 28. Componentes principales modelo escogido.</i>	<i>92</i>
<i>Figura 29. Valores medidos vs predichos para el Modelo Lnvisc @ 45°C menores a 1000 cP.</i>	<i>95</i>
<i>Figura 30. Componentes principales modelo escogido.</i>	<i>96</i>
<i>Figura 31. Modelo Lnvisc @ 45°C > 1000 cP.</i>	<i>98</i>
<i>Figura 32. Componentes principales modelo escogido.</i>	<i>99</i>
<i>Figura 33. Modelo Lnvisc @ 60°C < 1000 cP.</i>	<i>101</i>
<i>Figura 34. Componentes principales modelo escogido.</i>	<i>102</i>
<i>Figura 35. Modelo Ln visc @ 60°C > 1000 cP.</i>	<i>105</i>
<i>Figura 36. Componentes principales modelo escogido</i>	<i>106</i>
<i>Figura 37. Valores medidos vs predichos para el modelo de contenido de asfaltenos (% peso). ..</i>	<i>121</i>
<i>Figura 38. Cantidad Componentes Principales Modelo escogido, línea base.</i>	<i>122</i>
<i>Figura 39. Valores medidos vs predichos para el modelo normalizado de Asfaltenos >5%.</i>	<i>125</i>
<i>Figura 40. Cantidad Componentes Principales Modelo escogido, normalizado.</i>	<i>126</i>

<i>Figura 41. Valores medidos vs predichos para el modelo de Asfaltenos < 5% en peso normalizado.</i>	130
<i>Figura 42. Gráfico Componentes Principales Modelo escogido.</i>	131
<i>Figura 43. Valores medidos vs predichos para el modelo escogido, línea base asfaltenos < 5%.</i>	133
<i>Figura 44. Componentes Principales del modelo escogido.</i>	134
<i>Figura 45. Valores medidos vs predichos para el modelo seleccionado para la densidad.</i>	136
<i>Figura 46. Componentes principales modelo escogido.</i>	137
<i>Figura 47. Valores medidos vs predichos para el modelo densidad normalizado.</i>	140
<i>Figura 48. Componentes principales modelo seleccionado.</i>	141
<i>Figura 49. Vista análisis grafico general The Unscrambler Modelo Lnvisc @ 30°C < 1000 cP.</i>	142
<i>Figura 50. Vista análisis grafico general The Unscrambler para Modelo Lnvisc @ 30°C > 1000 cP.</i>	143
<i>Figura 51. Vista análisis grafico general The Unscrambler</i>	144
<i>Figura 52. Vista analisis grafico general The Unscrambler</i>	145
<i>Figura 53. Vista analisis grafico general The Unscrambler para Modelo Lnvisc @ 60°C mayor a 1000 cP.</i>	146

Lista de apéndices.

<i>Apéndice A. Modelos evaluados para asfaltenos > 5 (% peso).....</i>	<i>117</i>
<i>Apéndice B. Modelos evaluados para asfaltenos < 5 (% peso).....</i>	<i>127</i>
<i>Apéndice C. Modelos evaluados para densidades.....</i>	<i>135</i>
<i>Apéndice D. Vistas análisis gráfico de modelos.....</i>	<i>142</i>

Resumen

Título: Desarrollo de un modelo multivariado predictivo de propiedades fisicoquímicas de crudos mejorados en procesos térmicos de recobro usando espectroscopia infrarroja FTIR-ATR y Quimiometría *

Autor: José Miguel Coneo Vergara, Marcos Felipe Carvajal Antolinez**

Palabras Clave: FTIR-ATR, viscosidad, densidad, asfaltenos, crudos mejorados.

Descripción: El presente estudio se centra en la estimación de la viscosidad en rangos inferiores y superiores a 1000 cP a temperaturas de 30 °C, 45 °C y 60 °C, el contenido de asfaltenos, así como parámetros de densidad y gravedad API en 65 crudos mejorados, empleando espectros obtenidos por FTIR ATR mediante el uso del equipo Nicolet Summit X. Los modelos generados presentaron valores de R^2 de calibración superiores a 0,98 y R^2 de validación mayores a 0,94. Los errores cuadráticos medios por sus siglas RMSE se mantuvieron bajos en relación con los rangos experimentales, evidenciando una alta capacidad predictiva. Para el caso de la viscosidad, se observó un mejor ajuste en muestras menores a 1000 cP, mientras que en rangos más elevados se mantuvo una correlación adecuada pese a una ligera dispersión. Los modelos de asfaltenos mostraron sensibilidad a las regiones espectrales asociadas a enlaces aromáticos, mientras que los de densidad y API reflejaron una estrecha relación con la absorción en zonas características de las cadenas hidrocarbonadas. Finalmente, la aplicación de PLS a partir de datos espectroscópicos constituye una estrategia eficaz y no destructiva para la estimación de la viscosidad en hidrocarburos. Los resultados obtenidos respaldan la viabilidad de implementar este enfoque en laboratorios y entornos industriales, optimizando el control de procesos y reduciendo tiempos de análisis en comparación con los métodos tradicionales.

* Trabajo de Grado

** Facultad de Ingenierías Fisicoquímicas. Escuela de Ingeniería de Petróleos. Director: Adán Yovani León Bermúdez. Ph.D. Codirector: Julian Ernesto Sanabria Becerra. Ingeniero de Petróleos

Abstract

Title: Development of a Multivariate Predictive Model for the Physicochemical Properties of Upgraded Crude Oils in Thermal Recovery Processes Using FTIR-ATR Spectroscopy and Chemometrics*

Author: José Miguel Coneo Vergara, Marcos Felipe Carvajal Antolinez¹

Key Words: FTIR-ATR, viscosity, density, asphaltenes, upgraded crude oils.

Description: The present study focuses on the estimation of viscosity in ranges below and above 1000 cP at temperatures of 30 °C, 45 °C, and 60 °C, the asphaltene content, as well as density and API gravity parameters in 65 upgraded crude oils, using spectra obtained by FTIR ATR with a Nicolet Summit X instrument. The generated models exhibited calibration R^2 values above 0.98 and validation R^2 values greater than 0.94. Root Mean Square Errors (RMSE) remained low relative to the experimental ranges, evidencing a high predictive capacity. For viscosity, a better fit was observed in samples below 1000 cP, while in higher ranges the models maintained an adequate correlation despite slight dispersion. The asphaltene models showed sensitivity to spectral regions associated with aromatic bonds, whereas the density and API models reflected a close relationship with absorption in characteristic regions of hydrocarbon chains. Finally, the application of PLS to spectroscopic data constitutes an effective and non-destructive strategy for estimating viscosity in hydrocarbons. The results obtained support the feasibility of implementing this approach in laboratory and industrial environments, optimizing process control and reducing analysis time compared to traditional methods.

* Degree Work

¹ Faculty of Physicochemical Engineering. School of Petroleum Engineering. Advisor: Adán Yovani León Bermúdez, Ph.D. Co-advisor: Julian Ernesto Sanabria Becerra, Petroleum Engineer.

Introducción

La espectroscopía infrarroja (IR) es una herramienta muy útil para analizar los crudos mejorados mediante procesos térmicos en la industria petrolera (Smith, B. C., et al. 2011). Permite examinar de forma rápida y precisa la composición molecular en las muestras. En este estudio, se emplea un enfoque quimiométrico para establecer relaciones cuantitativas entre los espectros infrarrojos y las propiedades fisicoquímicas relevantes en crudos mejorados obtenidos bajo diferentes condiciones térmicas. Esto facilita la interpretación de los resultados y la toma de decisiones en la optimización de los procesos térmicos. La quimiometría por otro lado, consiste en utilizar métodos matemáticos y estadísticos para analizar datos químicos, esta nos permite extraer información valiosa de conjuntos complejos de datos provenientes de espectros infrarrojos. A través del desarrollo de modelos quimiométricos, es posible predecir y relacionar las propiedades fisicoquímicas del crudo con sus características espectrales. (Pedersen, L., et al. 2010). Este estudio resulta significativo debido a su potencial para mejorar la eficacia y rentabilidad a los procesos de producción del crudo mejorado, a partir de esto será posible desarrollar una metodología precisa y no invasiva para caracterizar las propiedades fisicoquímicas claves en un hidrocarburo.

Al comprender mejor cómo influyen los procesos térmicos en estas propiedades, se pueden tomar decisiones fundamentadas para optimizar dichos procesos. En la industria petrolera, son ampliamente utilizados para mejorar las propiedades de los crudos, lo que permite aumentar su calidad y valor comercial. Sin embargo, la eficacia y la optimización de estos dependen en gran medida de la comprensión detallada de las propiedades fisicoquímicas de los crudos antes, durante y después del tratamiento térmico. La espectroscopia infrarroja (IR) ha demostrado ser una herramienta sustancial para el análisis de compuestos orgánicos en muestras de petróleo y sus derivados. La técnica ofrece la capacidad de identificar y cuantificar diferentes componentes presentes en los crudos, así como

monitorear los cambios que experimentan durante los procesos térmicos. Sin embargo, el análisis de los datos IR de manera directa puede resultar complejo debido a la gran cantidad de información obtenida. En este contexto, la aplicación de modelos quimiométricos se presenta como una solución prometedora para el análisis de datos espectroscópicos, permitiendo extraer información relevante y obtener una caracterización detallada de las propiedades fisicoquímicas de crudos mejorados sometidos a procesos térmicos.

Por lo tanto, el presente estudio se centra en la caracterización de propiedades fisicoquímicas de crudos mejorados en procesos térmicos mediante el uso de espectroscopia infrarroja y modelado quimiométrico. Específicamente, se pretende desarrollar un modelo el cual permita aprovechar al máximo la información proporcionada por la espectroscopia IR, mediante la aplicación de técnicas de modelado quimiométrico, con el fin de identificar y cuantificar de manera precisa los cambios en las propiedades fisicoquímicas de los crudos como la viscosidad, API, asfaltenos, entre otros.

1 Objetivos

1.1 Objetivo General

Desarrollar un modelo multivariado predictivo de propiedades fisicoquímicas de crudos mejorados en procesos térmicos de recobro usando espectroscopia infrarroja FTIR-ATR y quimiometría.

1.2 Objetivos Específicos

Caracterizar crudos mejorados en procesos térmicos de recobro a partir de propiedades fisicoquímicas contempladas como viscosidad, contenido de asfaltenos (C7+) y gravedad API, entre otras.

Desarrollar la metodología de caracterización de crudos mejorados en procesos térmicos de recobro mediante espectroscopia FTIR-ATR.

Construir modelos multivariados para la predicción de propiedades fisicoquímicas de crudos mejorados en procesos térmicos de recobro usando espectroscopia infrarroja FTIR-ATR y la técnica de quimiometría.

2 Fundamentos teóricos y estado del arte

2.1 Clasificación de crudos

La Tabla 1 muestra la clasificación de los crudos según las 3 variables más importantes, en la cual, la gravedad específica es correlacionada en función de la densidad, la viscosidad medida en centipoises (cP) que refiere a la resistencia del crudo a fluir por un medio y de la misma manera la densidad equivalente a la cantidad de masa que se puede concentrar en un volumen (Cardenas, 2015).

Tabla 1.

Clasificación de crudos.

Crudo	Gravedad API (°)	Viscosidad (cP)	Densidad (kg/m ³)
<i>Extra Pesado</i>	<i>< 10</i>	<i>10000 - 100000</i>	<i>> 1000</i>
<i>Pesado</i>	<i>10 – 22.3</i>	<i>1000 - 10000</i>	<i>966 - 1000</i>
<i>Mediano</i>	<i>22.3 – 31.1</i>	<i>100 - 1000</i>	<i>934 – 966</i>
<i>Liviano</i>	<i>> 31.1</i>	<i>< 100</i>	<i>< 934</i>

Nota: Adaptado de Cárdenas, B. (2015). *Análisis del número ácido que presentan crudos de diferente gravedad API y sus fracciones típicas* [Trabajo de grado]. Instituto Politécnico Nacional.

En este caso se estará trabajando con diferentes muestras entre 10 a 22.3 °API correspondientes a crudos pesados colombianos.

2.2 Crudos pesados

Un crudo con una viscosidad superior a 1000 cP o una gravedad API que oscile entre 10° y 20° se clasifica como petróleo pesado; estos crudos se caracterizan por sus problemas de bajo flujo en pozos de producción a condiciones normales. El término “pesado” esta dado gracias a que la densidad y la gravedad específica es más alta que los aceites de tipo liviano, esto ejemplificado por la tabla 1 además, la proporción de componentes pesados (resinas y asfaltenos) suele ser mayor que en

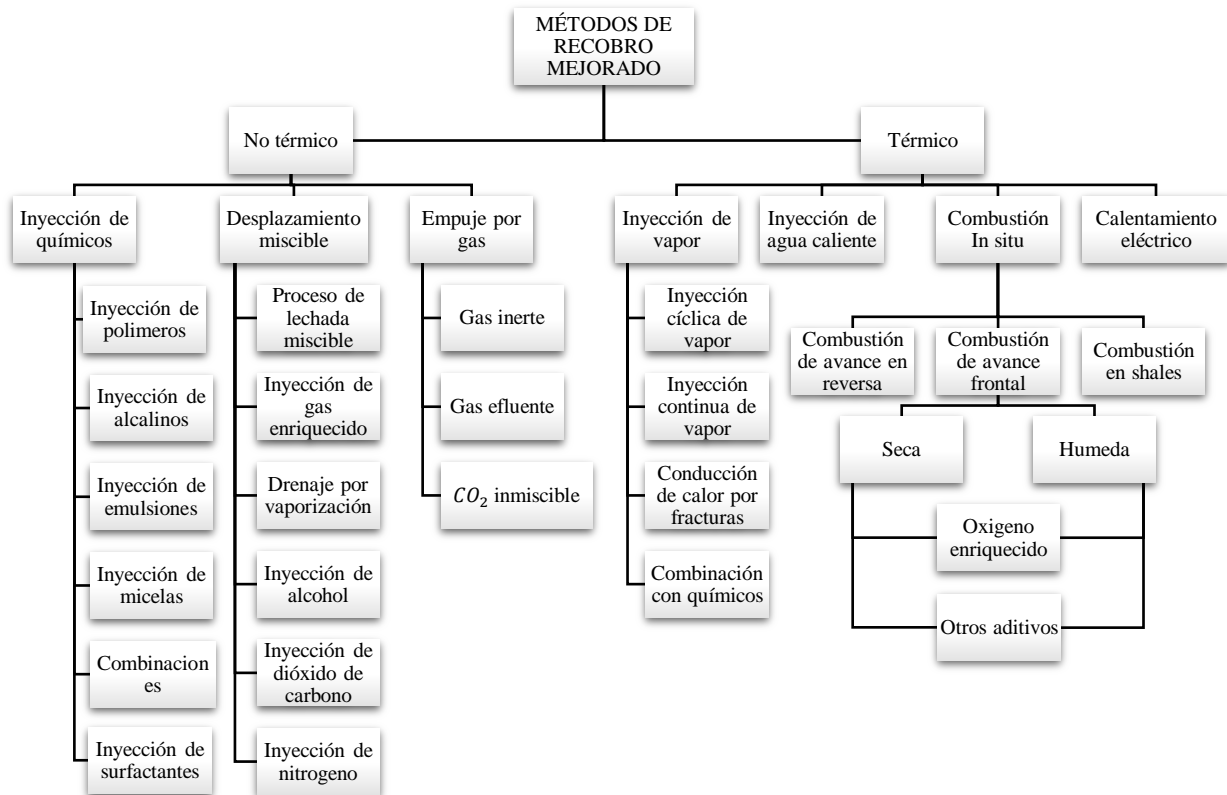
el crudo común (Guo et al., 2016; Speight, 2009). También, la explotación de yacimientos de crudo pesado mediante métodos de producción primaria y secundaria se ve restringida e ineficiente debido a las características de composición y movilidad que presentan este tipo de crudo. Por lo tanto, se hace necesario implementar métodos de recobro mejorado para la extracción de estos recursos (Aydin et al., 2022).

2.3 Métodos de recobro

Son técnicas utilizadas en la industria petrolera con el objetivo de incrementar la cantidad de petróleo extraíble de un yacimiento una vez que las etapas de producción primaria y secundaria han sido insuficientes para lograr extraer el petróleo original en sitio (OOIP) (Mogensen & Masalmeh, 2020). Los métodos de recobro mejorado se clasifican en tres categorías acorde al mecanismo que se llegue a emplear:

- Térmicos.
- Químicos
- Inyección de gases.

Según la figura 1 se resumen algunos métodos que se pueden presentar en los yacimientos convencionales y no convencionales, esto de acuerdo con lo que se requiera para su extracción.

Figura 1*Métodos de Recobro Mejorado.*

Nota: Adaptado de Farouq, A. (1996). The promise and Problems of Enhanced Oil Recovery Methods. The Journal of Canadian Petroleum Technology, Vol. 35, No. 7. University of Alberta.

Nota: Tomado de Sanabria, J. (2023). Caracterización fisicoquímica de un crudo pesado sometido a acuatermolisis catalítica mediante métodos convencionales y técnicas analíticas. [Trabajo de grado]. Universidad Industrial de Santander.

Entre los más comunes o usados en yacimientos no convencionales (pesados, extrapesados) son los métodos térmicos, debido a la alta eficiencia en la reducción de la viscosidad donde este es el principal inconveniente de este tipo de yacimientos. A su vez, estos se han desarrollado desde 1950 hasta la actualidad siendo así los más avanzados en el desarrollo experimental y tecnológico,

mayormente aplicados en distintos países con registros de crudos pesados y extrapesados como Canadá, Estados Unidos, Venezuela, Indonesia y demás (Thomas, 2008).

Algunos de los tratamientos térmicos son:

- Inyección cíclica de vapor.
- Inyección continua de vapor.
- Drenaje gravitacional asistido por vapor (SAGD).
- Combustión In Situ.

2.3.1 Inyección de vapor.

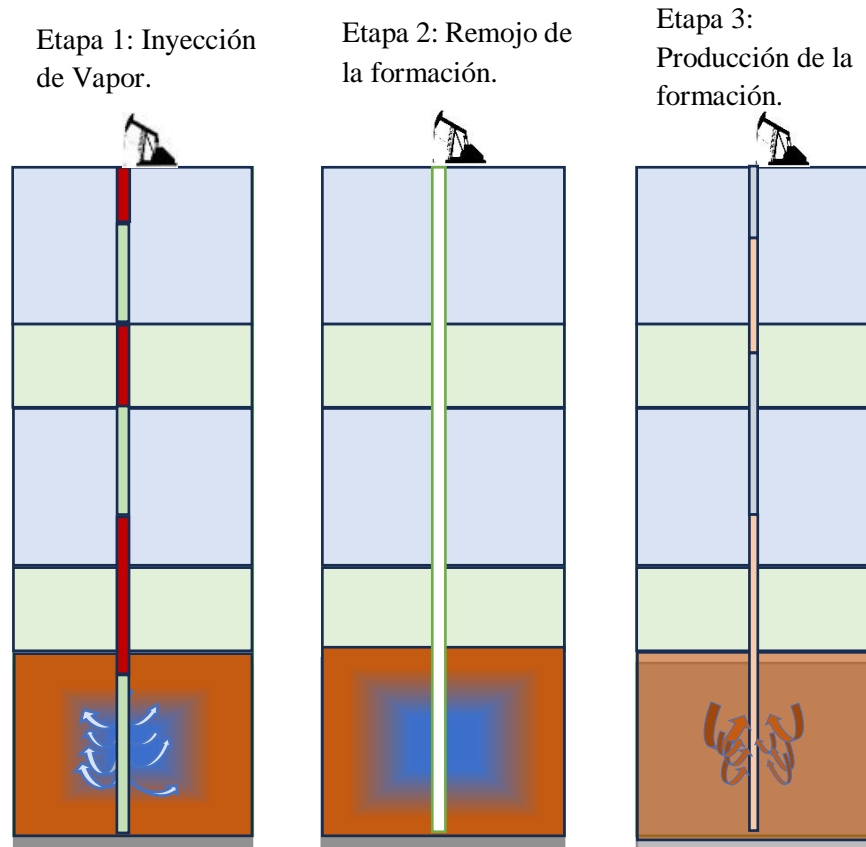
Esta es una técnica térmica de recobro mejorado empleada en yacimientos de petróleo pesado y extrapesado, en donde su aplicación comienza para los años de 1960 en California, EE. UU en donde fue usado en gran manera para suplir el uso de la estimulación cíclica de vapor (Hanzlik & Mims, 2003). El mecanismo principal es la reducción de la viscosidad debido a la transferencia de calor, esta afecta positivamente la relación de movilidades (movilidad del fluido desplazante sobre el desplazado) por medio de patrones similares a la inyección de agua en donde a su vez depende de estos patrones y de la geología, dado a esto se disminuye la saturación de aceite en el yacimiento, permitiendo que el factor de recuperación en el rango de un 50% a 60% del OOIP (Thomas, 2008).

Existen otros mecanismos que están presentes en la inyección de vapor, estos son:

- Expansión térmica e hinchamiento del aceite
- Destilación al vapor
- Empuje por gas en solución
- Desplazamiento miscible.

2.3.2 Inyección Cíclica con Vapor.

Esta técnica está basada en un proceso de pozo singular para yacimientos casi específicos con baja viscosidad, un API demasiado bajo, la operación de inyección consiste en 3 etapas, en donde la primera de estas es la inyección de vapor de manera continua por aproximadamente un mes, luego de esto el pozo se cierra para que el vapor inyectado en manera de calor se distribuya en todas las caras de la formación estudiada; a este paso se le tiene por nombre remojo y por último se da marcha a la producción en donde la rata de flujo de aceite incrementa de manera significativa pero solo por un tiempo muy corto. Todo lo anterior se repite en ciclos hasta que este proceso se vuelve económicamente poco rentable para la operación (Luo et al., 2020).

Figura 2*Inyección de Vapor*

Nota: Adaptado de Sanabria (2023) Caracterización fisicoquímica de un crudo pesado sometido a acuatermolisis catalítica mediante métodos convencionales y técnicas analíticas. [Trabajo de grado]. Universidad Industrial de Santander.

2.3.3 *Drenaje Gravitacional asistido por vapor.*

El proceso se basa en la segregación gravitacional del vapor, utilizando un par de pozos horizontales paralelos, ubicados a 5 m de distancia (en el caso de las arenas bituminosas) en el mismo plano vertical. Esto nos da una reducción de viscosidad la cual permite la movilización de bitumen drenado y capturado por otro pozo productor localizado en ocasiones justamente en la mitad del reservorio (Thomas, 2008).

2.4 Acuatermolisis catalítica.

Se denomina como ‘Acuatermolisis’, a un conjunto de reacciones químicas generadas entre los 200 y 320°C (Hyne, 1986). Se define como la interacción química del vapor de agua y compuestos orgánicos del petróleo crudo en presencia de calor y ausencia de oxígeno, en el que se logra un cambio en la estructura molecular de los compuestos orgánicos. Las reacciones que componen a la acuatermolisis son: pirolisis, hidrólisis, reacción del gas al agua (W.G.S.), hidrodesulfuración (HDS), estas dos últimas consisten en la reducción del contenido de heteroátomos de oxígeno y azufre, donde se produce dióxido de carbono CO_2 , monóxido de carbono CO , ácido sulfhídrico H_2S , hidrógeno H_2 e hidrocarburos livianos (Arboleda et al., 2018).

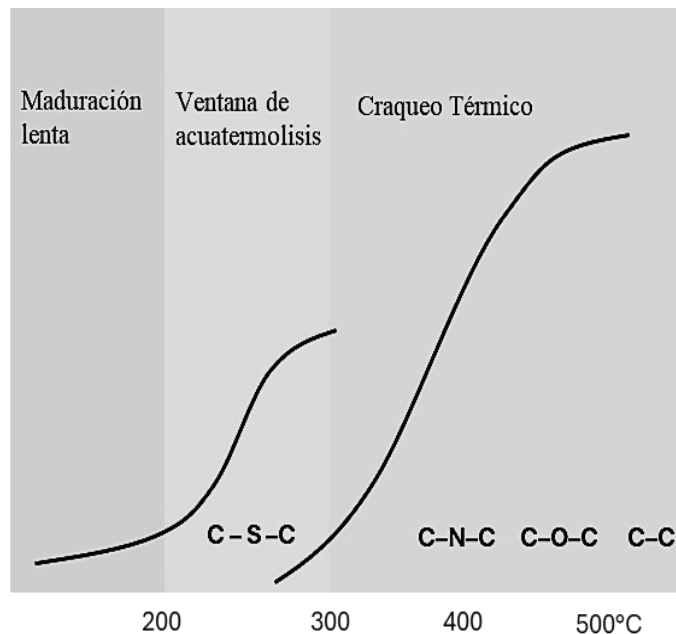
Por lo general, las reacciones de acuatermolisis ocurren en periodos extensos a altas temperaturas y no poseen grandes efectos sobre el crudo a condiciones típicas de su aplicación, sin embargo, en presencia de minerales de yacimiento o iones metálicos hay un incentivo/aceleración de las reacciones de acuatermolisis ya que se reduce la energía de activación a la que se genera, y por lo tanto el mejoramiento del crudo es mayor. Así mismo, esto permite una mejor descomposición y conversión térmica en crudos sulfurados, propiciando un rendimiento químico aún mayor de gases típicos de estas reacciones como lo son Hidrogeno (H_2), metano (CH_4), monóxido de carbono (CO), dióxido de carbono (CO_2) y la reducción de gases ácidos como ácido sulfhídrico (H_2S), que por su alta toxicidad y corrosividad ha desmeritado la técnica de inyección de vapor convencional en el pasado (Arboleda et al., 2018).

Diferentes autores han evaluado diversos esquemas reactivos de la acuatermolisis evaluando la evolución de sus productos y la composición de las muestras.

Hyne (1986) logró identificar que la temperatura es una de las condiciones críticas para que las reacciones ocurran, por ende, se definió como una ventana de acuatermolisis, la cual se encuentra entre los 200°C y 325°C, donde la eficiencia de la acuatermolisis aumenta considerablemente, todo ocurre gracias a que la producción de residuos sólidos (coque) y de gas hidrocarburos es muy poca o nada. Por encima de este rango de temperaturas, comienzan a primar las reacciones de craqueo térmico donde ocurre rompimiento de enlaces C-N-C, C-O-C y C-C generando un aumento en la producción de coque y gas.

Figura 3.

Ventana Acuatermolisis



Nota: Adaptado de Hyne (1986) Aquathermolysis: a synopsis of work on the chemical reaction between water (steam) and heavy oil sands during simulated steam stimulation.

Nota: AQ corresponde a la Acuatermolisis.

La acuatermolisis catalítica está compuesta por dos procesos: W.G.S e Hidrotratamiento, los cuales son llevados a cabo por las reacciones de pirolisis, hidrolisis, W.G.S, Hidrodesulfuración

(HDS), Hidrodesoxigenación (HDO), cuyos productos son dióxido de carbono (CO_2), monóxido de carbono (CO), ácido sulfhídrico (H_2S), Hidrogeno (H_2), hidrocarburos livianos, amoniaco (NH_3) y agua (H_2O) (Arboleda J, Castillo A, Muñoz S, 2018).

Muraza (2015) plantea que la aplicación de catalizadores durante el proceso de la acuatermólisis mejora el grado de ruptura de enlaces y una mayor reducción de la viscosidad. Por lo tanto, la acuatermólisis catalítica implica la introducción de catalizadores adecuados para lograr una reducción eficiente de la viscosidad. Por otro lado, estudios han demostrado que la presencia de catalizadores en fase acuosa reduce significativamente la viscosidad del petróleo, dado a esto, el aumento neto en la movilidad del petróleo se atribuye a la reducción de resinas y asfaltenos de alta viscosidad.

2.5 Espectroscopía infrarroja

La espectroscopia examina cómo la radiación electromagnética interactúa con una sustancia. Esta disparidad energética, conocida como energía de interacción o absorción, puede provocar cambios a nivel atómico o molecular cuando interactúa con la muestra. Además, el efecto molecular resultante depende de la frecuencia de la radiación electromagnética involucrada en la interacción. En el caso del análisis de muestras mediante espectroscopia infrarroja, la región de radiación infrarroja se divide comúnmente en tres zonas: infrarrojo cercano (1.0-2.5 μm), infrarrojo medio (2.5-50 μm) e infrarrojo lejano (50-1000 μm).

2.5.1 Tipos de espectroscopía infrarroja

2.5.1.1 Reflectancia total atenuada

La reflectancia total atenuada (ATR por sus siglas en inglés), 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 menos 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 de 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 (Maldonado, 2018).

Tabla 2.

Asignación de Grupos Funcionales en Espectroscopia Infrarroja.

Pico	Longitud de Onda cm^{-1}	Estructura encontrada
1	720	Enlaces aromáticos CH lineales a la molécula
2	812	Enlaces aromáticos CH perpendiculares a la molécula
3	870	Condensación angular de los anillos aromáticos
4	1030	Grupos Sulfoxidos
5	1380	Grupo metil CH_3
6	1465	Grupo metileno CH_2CH_3
7	1600	Estructuras aromáticas $\text{C}=\text{C}$
8	1710	Grupos carbonilo en los ácidos
9	2920	Grupos CH_2
10	2955	Grupos CH_3

Nota: Tomado y adaptado de Márquez & Sánchez (2024).

2.6 Quimiometría

La quimiometría es una disciplina interdisciplinaria que combina principios de la química con métodos matemáticos y estadísticos para analizar e interpretar datos químicos y sensoriales. Se utiliza principalmente en campos como la química analítica, la espectroscopia, la cromatografía, la investigación de alimentos, la farmacología y la ingeniería química entre otras (Mongay C, 2011).

Para predecir las propiedades fisicoquímicas de los crudos con precisión y en períodos cortos, se han utilizado varios modelos como modelos de análisis multivariado por regresión lineal múltiple y componentes principales (PCAMLR), así como regresión de mínimos cuadrados parciales (PLS) (Campillo N, 2011).

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 (Maldonado, 2018)

Esta técnica es de gran importancia en estudios de investigación y caracterización de materiales, lo que permite la optimización de métodos de preprocesamiento y la selección de variables en ensayos a escala de laboratorio. Para construir modelos quimiométricos cuantitativos con técnicas analíticas, se requieren tres pasos, primero, el pretratamiento de los espectros, segundo. la selección del intervalo de longitud de onda y por último la calibración multivariada (Mongay C, 2011).

2.6.1 Resolución de curvas por automodelado (SMCR)

La resolución de curvas por automodelado (SMCR, por sus siglas en inglés) es un enfoque quimiométrico compuesto por un conjunto de métodos matemáticos diseñados para descomponer señales bidimensionales provenientes de mezclas multicomponentes que no pueden ser separadas directamente mediante técnicas instrumentales convencionales. Este proceso permite aislar los perfiles correspondientes a cada componente individual en forma de factores.

Tal como sugiere el término "automodelado", esta técnica no requiere información previa detallada sobre los datos para extraer las variables puras. En cambio, se basa en un modelo bilineal que describe adecuadamente la estructura de los datos, junto con supuestos generales acerca de las características de las variables puras, tales como los valores iguales o mayores a cero y la presencia de un único máximo de datos, estadísticamente llamado unimodalidad. Estas condiciones suelen cumplirse de manera natural en los datos bidireccionales generados a partir de análisis multivariantes de mezclas con variaciones en su composición (Tamiru Tefera et al., 2017).

2.6.2 Análisis por componentes principales

El Análisis por Componentes Principales (ACP), es una técnica para reducir la cantidad de variables cuando está presente la correlación. Resulta interesante afirmar que no es una técnica útil cuando las variables no están correlacionadas. La idea que se encuentra detrás del ACP es encontrar componentes principales Z_1, Z_2, \dots, Z_n partiendo de una matriz en donde estas sean combinaciones lineales de las variables originales X_1, X_2, \dots, X_n , (Miller James N. & Miller Jane C., 2002) que describen de tal manera que cada fila es una muestra y cada columna una variable presentando de así m muestras y k variables, por lo cual la matriz será expresada como $m \times k$. En la ecuación 1 se puede observar la referencia de solución de un componente principal.

$$Z_1 = a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + \dots + a_{1n}X_n \quad Z_2 = a_{21}X_1 + a_{22}X_2 + a_{23}X_3 + \dots + a_{2n}X_n \quad (1)$$

Una decisión que se debe tomar antes de llevar a cabo un PCA, es el tener en cuenta si se emplean los datos originales o, por el contrario, se deben estandarizar cada una de las variables a media cero y varianza unidad. Teniendo en cuenta de que, si no se estandarizan las variables y una de estas tiene una varianza mucho más grande, entonces esta variable controlará el primer componente principal, por lo tanto, la estandarización hace que todas las variables tengan la misma varianza. Para datos estandarizados cada una de las variables originales tiene una desviación estándar, y por tanto una varianza, de 1, así que, si esto pasa, la varianza total del conjunto de datos y la suma de las componentes principales son iguales al número de variables (Rodríguez C, 2014).

2.6.3 Regresión a regresión en componentes principales (PCR)

Para la estadística, la regresión por componentes principales (*PCR por sus siglas en inglés*), es una técnica de análisis que se basa en los componentes principales de una muestra, donde el tamaño de esta tiene un rango más reducido (Bair et al., 2006), específicamente, se usa para estimar los coeficientes de regresión desconocidos en un modelo de regresión lineal estándar.

Normalmente para esto solo se llega a usar un conjunto de los componentes para la regresión, lo que convierte a la PCR en un método de ajuste y a su vez en un estimador de variación en las muestras, ya que se seleccionan como regresores los componentes con menores varianzas para dar una estimación menor o igual a 1 (Frank & Friedman, 1993).

2.6.4 Regresión de mínimos cuadrados parciales (PLS)

El método de Mínimos Cuadrados Parciales (PLS) es una técnica de calibración fundamental en Quimiometría. Fue desarrollado a mediados de los años 60 por Herman Wold en Suecia, inicialmente con un enfoque heurístico, que luego evolucionó hacia procedimientos iterativos más formales. En sus inicios, las principales aplicaciones del PLS se encontraban en la economía y las ciencias sociales.

La aplicación del PLS al análisis de datos químicos recibió un impulso significativo en los años 80 gracias a Svante Wold, hijo de Herman Wold. Desde entonces, el PLS se ha convertido en un método de uso rutinario en Quimiometría. Los avances en informática y software han permitido la implementación de estos métodos, destacando su gran potencial de modelización, versatilidad y capacidad analítica (Guzmán & Vega-Vilca, 2010). En la tabla 3, se presentan varios trabajos realizados por modelos PLS en la industria.

Tabla 3.

Resumen de investigaciones espectroscopia infrarroja y/o quimiometría.

Autores	Título	Condiciones	Técnica	Resultados
(Aldana-Jiménez et al., 2018)	Aplicación de la espectroscopia infrarroja y técnicas de reconocimiento de patrones en la caracterización de aditivos de la industria del petróleo.	Aditivo antiestático. Aditivo Inhibidor de corrosión.	Espectroscopia infrarroja con transformada de Fourier (FT – MIR). Análisis por componentes principales (PCA). Modelo suave independiente de analogías de clase (SIMCA).	Antiestático: PCA explicó el 89% de la variabilidad. SIMCA clasificó correctamente las 15 muestras dentro de especificaciones. Inhibidor de corrosión: PCA explicó el 98% (CP1: 98,5%; CP2: 0,58%). SIMCA mostró clasificación correcta en una sola clase

(Sáenz Paéz, 2018)	Avances en la aplicación de resonancia magnética nuclear – RMN de bajo y alto campo – y quimiometría para caracterizar crudos pesados y fondos de vacío.	Crudos pesados y fondos de vacío del petróleo. Asfáltenos y gasóleos de vacío. EQUIPOS. Espectrómetros RMN de bajo (1H) y alto campo (13C). FT-IR (Espectroscopia infrarroja).	RMN de bajo campo (con protones). RMN de alto campo (con carbono). Regresión de Mínimos Cuadrados Parciales (PLS). Regresión Lineal Múltiple (MLR). Análisis componentes principales (PCA). Análisis de agrupamiento jerárquico (HCA).	RMN más quimiometría, permite predecir propiedades fisicoquímicas (viscosidad, azufre, API), reducir reactivos, tiempo y muestra Crudos pesados: alta correlación con °API, viscosidad y SARA Fondos de vacío: Predice destilación, estabilidad y SARA
--------------------	--	--	---	--

(Camperos, 2017)	Aplicación de la espectroscopia de infrarrojo cercano para la determinación de curvas de destilación en crudos de carga en la refinería de Barrancabermeja	89 crudos de carga (refinería de Barrancabermeja) almacenadas a <10°C EQUIPOS. Cromatógrafo de gases Agilent 6890 Espectrómetro Infrarrojo ABB MB3000 Solo® versión 8.2.1 (Eigenvector), para modelamiento quimiométrico.	Espectroscopia de infrarrojo cercano (NIR), aplicada a 4000 y 4700 cm ⁻¹ . Regresión por mínimos cuadrados parciales (PLS). Análisis de componentes principales (PCA)	Los modelos generados por PLS, mostraron altos coeficientes de correlación (R ² >0.98) entre valores inferidos y medidos por ASTM D7169. Reducción significativa del tiempo de análisis, se obtuvo una disminución de más o menos 5 horas, la espectroscopía NIR permite obtener resultados en minutos. Se construyeron modelos predictivos para curvas de destilación: IBP, 5%, 10%... hasta el 70% de volumen destilado.
------------------	--	--	--	---

(Comesaña García et al., 2009)	Comparación de dos métodos supervisados de reconocimiento de patrones para la clasificación de destilados medios de petróleo mediante Espectroscopia Infrarroja.	60 muestras de kerosinas. 10 de turbocombustibles. Espectrofotómetro infrarrojo FT-MIR Avatar 360Esp. Cromatografía gaseosa acoplada a espectrometría de masa.	Espectroscopia Infrarroja por transformada de Fourier (FT-IR); Infrarrojo Medio (FT-MIR). Análisis por componentes principales (PCA). Modelo suave independiente de analogías de clase (SIMCA). Regresión por mínimos cuadrados parciales (PLS).	Resultados de R^2 de calibración y validación, respectivamente. Clase C: 0.991; 0.986 Clase V: 0.995; 0.994 Las clases C y V corresponden a los dos tipos de kerosinas detectados en el análisis por PCA.
(Lamus et al., 2011).	Uso de análisis multivariado en la determinación sara de crudos mediante la espectroscopia NIR	Crudos petróleo de diferentes regiones de Colombia. Fracciones SARA para cada muestra. Espectrofotómetro NIR.	Espectroscopía infrarrojo cercano (NIR). Regresión PLS. Corrección línea base, normalizado, derivadas.	Se construyeron modelos PLS para predecir las fracciones SARA a partir de espectros NIR. R^2 calibración y validación respectivamente. Saturados: 0.99; 0.976. Aromáticos: 0.9924; 0.9793. Resinas: 0.9757; 0.9483. Asfáltenos: 0.9951; 0.9699.
(Pichardo, 2011)	Caracterización de hidrocarburos de diferente gravedad api mediante espectroscopia de infrarrojo (FTIR)	12 muestras de crudo. Espectrofotómetro FTIR.	Regresión mínimos cuadrados parciales (PLS). Espectroscopía Infrarroja por Transformada de Fourier (FTIR).	R^2 calibración y validación respectivamente. API: 0.988; 0.956. Viscosidad: 0.985; 0.960. Densidad: 0.996; 0.981.

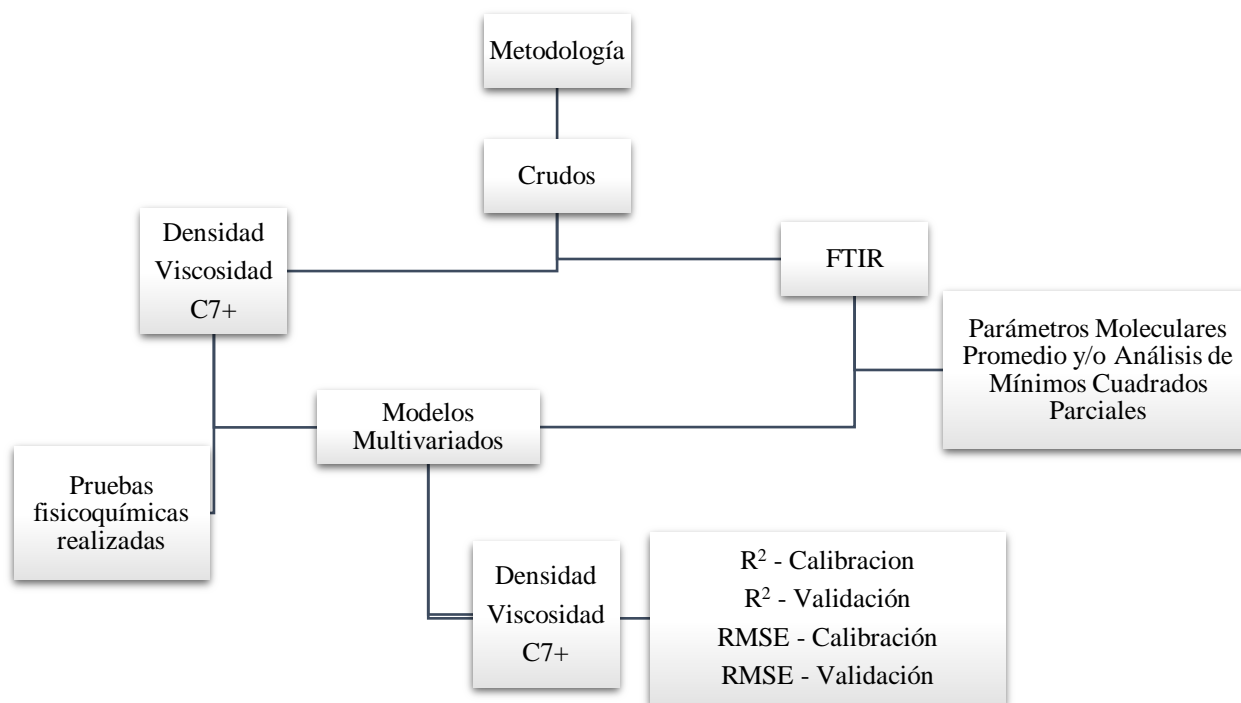
(Casal Trigo, 2024)	Desarrollo de métodos para la caracterización inteligente y automatizada de productos derivados del petróleo mediante machine learning y técnicas espectroscópicas.	360 mezclas + 45 puras, con un total de 405 muestras. Espectrómetro.	Quimiometría Análisis de componentes parciales (PCA). Análisis de agrupamiento jerárquico (HCA). Regresión de mínimos cuadrados parciales (PLS).	R ² calibración y validación Gasolina: 0.998; 0.996 Metanol: 0.999; 0.997 Queroseno: 0.997; 0.996 Modelo global (todas las mezclas): 0.997; 0.995.
(Ballesteros-Lancheros et al., 2011)	Determinación De La Composición De Fondos De Vacío De Petróleos Colombianos Por Espectroscopia De Fluorescencia Inducida Por Láser (LIF).	27 muestras de fondo de vacío. Espectrómetro.	Quimiometría. Regresión de mínimos cuadrados parciales (PLS). Análisis de componentes parciales (PCA). Espectroscopía de fluorescencia inducida por lase (LIF)	Correlación de R ² Saturados: 0.926 Aromáticos: 0.738 Resinas: 0.897 Asfáltenos: 0.734
(Mena Sánchez, 2021)	Modelos de clasificación y cuantificación quimiométrica de hidrocarburos fracción pesada durante la biorremediación de un histosol contaminado con petróleo crudo.	70 muestras de histosol contaminadas por petróleo crudo. Espectroscopía Cromatografía de gases.	Espectroscopía de reflexión difusa. FTIR Quimiometria. Análisis de componentes principales (PCA) Regresión de mínimos cuadrados parciales (PLS)	Hidrocarburos de fracción pesada con un R ² de calibración del 0.9998. Suelo contaminado con un R ² de calibración del 0.9998.

(Guzmán-Osorio, 2022)	Determinación de tipo y concentración de petróleo utilizando espectrometría FTIR.	22 muestras de crudo superligeros, ligeros, medianos, pesados y extrapesados.	Espectrometría infrarroja por transformada de Fourier (FTIR).	Grados API R ² calibración: 0.9632 R ² validación: 0.9024
		EQUIPOS.	Análisis de componentes principales (PCA)	Curvas de calibración los R ² varían entre 0.9496 a 0.9946 según el tipo de petróleo.
		Espectrómetro FTIR con ATR. Espectrómetro de banda.	Regresión de mínimos cuadrados parciales (PLS)	Espectroscopía de banda fija los R ² < 0.98.

Nota: Creada y adaptada por autores

3 Metodología Experimental

La metodología mostrada en la Figura 4 refleja el desarrollo de la investigación, en donde se inicia con la caracterización fisicoquímica de los crudos, evaluando primero la densidad, seguido de la °API y, por último, la viscosidad a diferentes temperaturas (30, 45 y 60°C), a su vez se encontró el contenido de asfaltenos que tenía cada muestra de hidrocarburo. Finalmente se realizó la medición del FTIR-ATR garantizando una repetibilidad en el muestreo. Continuando con la segunda etapa, en donde se coteja y se evalúa la información obtenida de la primera fase, dando paso al diseño y construcción de modelos quimiométricos mediante PLS.

Figura 4.*Metodología experimental de la investigación.*

Nota: Creada y adaptada por autores

3.1 Selección de las muestras

Se seleccionaron en total 65 muestras de crudo de diferente naturaleza, provenientes de diversas regiones del país y la mayoría de estos son productos de investigaciones termocatalíticas en crudos pesados que se han realizado previamente en el laboratorio, todo esto con el fin de garantizar un amplio rango de variabilidad en el muestreo, obteniendo de esta manera modelos confiables al tener una base de datos amplia.

3.2 Caracterización de las muestras.

3.2.1 API y Densidad.

La toma de muestras y caracterización para la °API y densidad, se realizaron en el laboratorio usando tres elementos básicos como fueron, suficiente muestra de crudo, un picnómetro previamente calibrado y una balanza analítica de 4 cifras significativas. Así mismo para la densidad en donde se pudo calcular a cada una de las muestras por medio de las fórmulas a continuación:

Gravedad Específica.

$$GE = \frac{141.5}{\text{°API} + 131.5} \quad (2)$$

Densidad.

$$\rho = GE * \rho_{H_2O} \quad (3)$$

3.2.2 Viscosidad

Las pruebas de viscosidad se realizaron con un viscosímetro rotacional de referencia B-one Plus @Lamy Rheology como se puede evidenciar en la figura 5, el rango de torque de este equipo va desde 0,05 a 13 mNm y tiene en su función un número limitados de velocidades que van entre 0,3 y 300 rpm. Cuenta con 7 diferentes husillos rotatorios, para este caso se utilizó específicamente el husillo MK-SV418 que está diseñado para mediciones reológicas de pequeño volumen, ofreciendo así una alta precisión y reproducibilidad (ver figura 6). Las dimensiones del viscosímetro son 780mm de alto, 280mm de ancho y 200mm de profundidad, con un peso aproximado de 6,7kg. Cumple certificación bajo las normas ASTM D789, ASTM D2196, ISO 1652 (Sanabria, 2023).

Figura 5.

Viscosímetro rotacional B-one plus Lamy Rheology.



Nota: tomadas del laboratorio del Grupo de Investigaciones en Corrosión (GIC)

Cada crudo se midió a 3 temperaturas diferentes, y luego haciendo corrección por la ecuación de Andrade (4) para tener una precisión mejor de cada toma realizada, como lo muestra la Tabla 5.

$$\mu = A * e^{\frac{b}{T}} \quad (4)$$

$$\ln(\mu) = \ln(A) * \frac{b}{T} \quad (5)$$

En donde:

μ = viscosidad dinámica del líquido [cP]

T = temperatura absoluta (K)

A = constantes empíricas que dependen del líquido

$b = m$ = Pendiente calculada en X,Y mostrado en la tabla 4.

Tabla 4.

Promedio de Temperaturas y viscosidades en muestra.

Temp, °C	Visc, cP
34.3	6374.5
42.4	2929.2
61.1	447.5

Nota: Promedio de temperaturas trabajado en el laboratorio

Tabla 5.

Correlaciones con la Ecuación de Andrade.

Temp, °K	X= 1/T	Y=ln(visc)
307.5	0.0032526	8.7600609
315.6	0.0031687	7.9824732
334.2	0.0029924	6.1036207
a	-24.64577	
A	1.979E-11	
m	10281.1	

Nota: Correlaciones en las temperaturas con la ecuación de Andrade

Tabla 6.

Viscosidad con la Ecuación de Andrade.

Temp, °C	Temp, °°K	Visc, cP
30	303.2	10597.925
45	318.2	2141.8861
60	333.2	499.9246

Nota: Viscosidad modificada con la ecuación de Andrade.

Para el husillo MK-SV418 se necesita un volumen de muestra de 7,5 mL y la velocidad de cizalla se va variando según la viscosidad del crudo y teniendo como punto de referencia que el par esté dentro del rango de 6500 ± 100 mNm/m.

Figura 6.

Husillo MK-SV418 – Camisa de Medición.



Nota: tomadas del laboratorio del Grupo de Investigaciones en corrosión (GIC)

3.2.3 *Contenido de asfaltenos*

Para el cálculo se garantizó muestras mínimas de 1gr para cada crudo y así tener un porcentaje realmente significativo de asfaltenos. Para ello se sigue la metodología establecida por la norma ASTM D6560 - 17 en la cual se disuelve el hidrocarburo con n-heptano, posteriormente se hace el proceso de filtración usando papel filtro Whatman, luego se procede al lavado soxhlet, para así poder limpiar los asfaltenos ya que al filtrar puede que aun queden trazas de maltenos en ellos, esta limpieza se realiza con el fin de obtener la mayor cantidad de asfaltenos “limpios” o completamente separados de los demás componentes del SARA. Luego de esto, se procede a secar cada una de las muestras o sedimentos en un horno a 60°C por 24 horas y así finalmente pesar y calcular la cantidad de asfaltenos que han resultado de este proceso.

Figura 7.

Lavado Soxhlet de los asfaltenos.



Nota: tomadas del laboratorio del Grupo de Investigaciones en corrosión (GIC)

Figura 8.

Imagen de referencia Asfaltenos después de lavado y secado, tomada del Laboratorio Grupo de Investigación en Corrosión, UIS Guatiguará (2025)



Nota: tomasas del laboratorio del Grupo de Investigaciones en corrosión (GIC)

3.3 Análisis multivariable

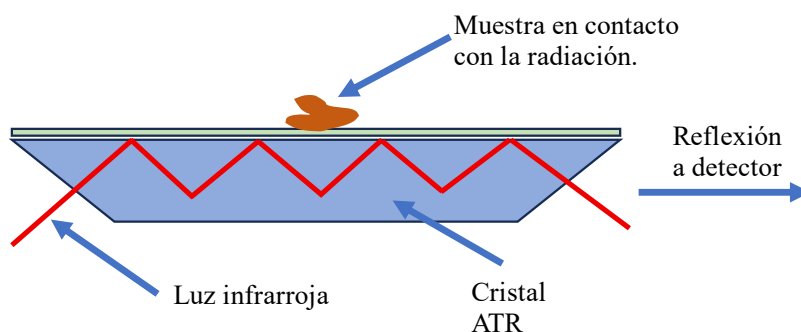
3.3.1 Espectroscopia Infrarroja

Cada uno de los espectros tomados en las muestras, fue hecho en el equipo Thermo Scientific NICOLET Summit X el cual se muestra en la Figura 11, este está dentro del rango desde 400 cm^{-1} hasta los 4000 cm^{-1} correspondiente al rango espectral infrarrojo medio (MIR) anteriormente mencionado, en donde el procedimiento de uso de este equipo, empieza por el encendido del equipo cerciorándose de que todo esté en buen funcionamiento, luego se procede a hacer la calibración del equipo mediante un background previo a cada muestra con la finalidad de que el muestreo sin errores de calibración, se sugiere limpiar el cristal con un paño suave sin pelusa humedecido en isopropanol o el solvente recomendado por el fabricante. Después de cada

background, se procede a preparar cada una de las muestras, con las cuales se sugiere poner una mínima gota directamente sobre el cristal de reflexión de diamante de un solo paso mostrado en la Figura 10, el cual tiene un ángulo de incidencia de 45° ,

Figura 9.

Reflexión de la Muestra.



Nota: Tomado y adaptado de Apasionados por la Metrología Guía (2015).

Figura 10.

Equipo Nicolet Summit X. FTIR Spectrometer.



Nota: Imagen de Thermo Fischer Scientific (2023) [User Guide]. Madison, Wisconsin USA.

3.3.2 Modelos PLS

Los datos iniciales son previamente estandarizados por columnas, tras lo cual se da inicio al procedimiento de estimación del primer componente PLS. En esta etapa se determina el vector $w = (w_1, \dots, w_p)$, cuyos elementos corresponden a la covarianza entre la variable de respuesta y cada una de las variables predictoras; finalmente, dicho vector es normalizado con el fin de garantizar que posea norma unitaria.

A partir de este resultado, se obtiene el componente PLS definido como:

$$T_h = X_w(X_1, \dots, X_p)(W_1, \dots, W_p)' \quad (6)$$

Seguidamente se estima el coeficiente de regresión simple de la variable de respuesta sobre dicho componente y finalmente, se calcula el vector $b = (b_1, \dots, b_p)$, cuyos elementos representan los coeficientes de regresión simple de cada variable predictora X_i respecto al componente T_h .

Una vez obtenidos estos resultados, se procede a la actualización de la matriz de variables predictoras, así como del vector de respuestas, con el fin de eliminar la información ya explicada por el componente calculado. Posteriormente, a partir del procedimiento descrito cuando se empieza a encontrar el vector w , se estima el siguiente componente PLS, repitiendo de manera iterativa el proceso hasta alcanzar el número de componentes requeridos.

Un paso a paso general es mostrado en la tabla 7.

Tabla 7.

Algoritmo de la regresión PLS.

	<i>Entrada</i>	<i>X (0), Y (0)</i>
1	Para h	1 hasta p
2	cov (Y, X)	normalizar w
3	T_h	X_w
4	v	$(T_h'Y) / (T_h'T_h)$
5	b	$(T_h'Y) / (T_h'T_h)$
6	X(h)	$X_{(h-1)} - T_h b$
7	Y(h)	$Y_{(h-1)} - T_h v$
8	Próximo h	

Nota: (Guzmán & Vega–Vilca, 2010)

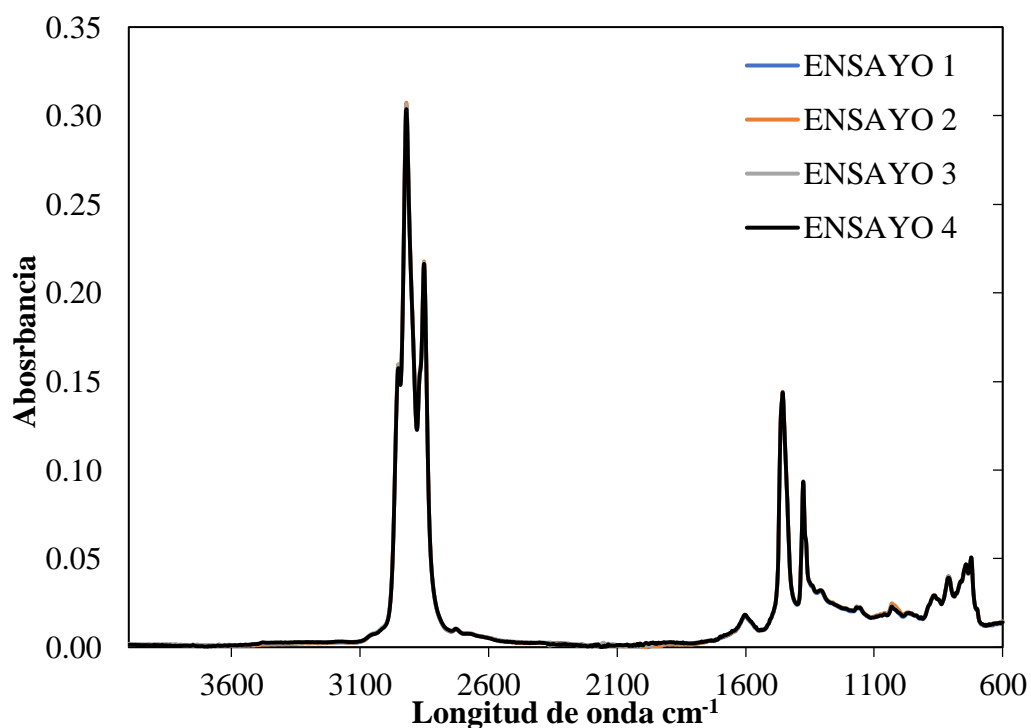
3.4 Análisis Espectroscopía Infrarroja FTIR-ATR

Cada espectro tomado, fue adquirido mediante el programa OMNIC con el cual, los resultados mostrados son exportados en formato TXT y así mediante este, poder llevar a cabo una respectiva normalización, ajuste, modelamiento y realización de las gráficas en el programa base, el cual va a ser The Unscrambler, tomando como ajuste previo a 32 scans y de esta manera al momento de tomar el infrarrojo, no tendrá una cantidad alta de ruido y de esta manera se obtendrán resultados positivos con una intensidad de señal adecuada con un tiempo de adquisición corto.

Antes de iniciar los análisis, primero se evaluó la repetibilidad de los infrarrojos, a partir de esto fueron tomados 4 espectros de la muestra CR58 bajo las condiciones ambientales y con características fisicoquímicas representativas a la investigación siendo este un crudo pesado. Las gráficas mostradas en la figura 13 fueron recopiladas en el software The Unscrambler para dar un análisis más detallado de la muestra, teniendo en cuenta que la desviación estándar de estas señales fue demasiado baja, por consiguiente, se asegura y se comprueba la buena repetibilidad en los espectros.

Figura 11.

Repetibilidad Espectro General Muestra CR58.

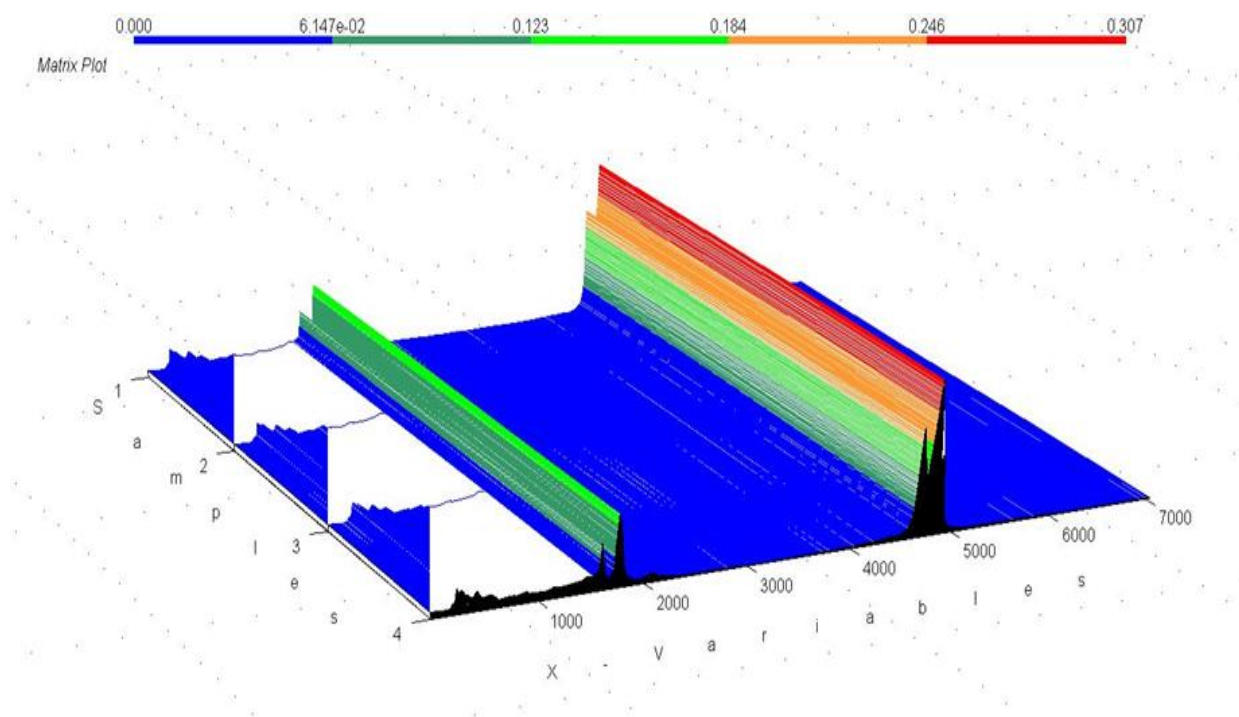


Nota: El análisis del crudo CR58 mediante espectroscopia infrarroja ATR, mostró espectros altamente consistentes entre las repeticiones que se hicieron, coinciden en varias zonas la posición y forma de las bandas principales, destacando las señales de estiramientos C-H

alifáticos en $2850 - 2950 \text{ Cm}^{-1}$ y las deformaciones CH_2/CH_3 en $1450 - 1375 \text{ Cm}^{-1}$ como se presencia en la figura 12 y 13.

Figura 12.

Vista 3D Axial de la Repetibilidad.



Nota: La desviación estándar general es de 0.03838, demostrando que el sistema tiene una alta precisión y buena repetibilidad, ya que la dispersión entre las medidas es mínima. Esto indicando que las diferencias observadas entre los espectros no provienen de fallas instrumentales o errores de muestreo significativos, por el contrario, provienen del ruido inherente al método y de pequeñas variaciones en el contacto con el cristal ATR.

Siguiendo con la metodología, a partir de las investigaciones reportadas por Vakhin et al., (2018) el análisis en la estructura química es evaluado con información obtenida del espectro de

infrarrojo FTIR-ATR mediante la intensidad asignada a cada longitud de onda como se muestra en la tabla 10. Para realizar un mejor análisis sobre los cambios estructurales en los crudos mejorados se utilizaron los parámetros moleculares promedio descritos en la tabla 10 . Donde D representa la absorbancia a una longitud de onda dada.

Tabla 8.

Parámetros moleculares promedio.

Parámetro	Nomenclatura	Ecuación
Aromaticidad	C ₁	$C_1 = \frac{D_{1600}}{D_{720}}$
Oxidación	C ₂	$C_2 = \frac{D_{1710}}{D_{1465}}$
Ramificación	C ₃	$C_3 = \frac{D_{1380}}{D_{1465}}$
Parafinicidad	C ₄	$C_4 = \frac{D_{1720} + D_{1380}}{D_{1600}}$
Sulfuración	C ₅	$C_5 = \frac{D_{1030}}{D_{1465}}$
Longitud de cadenas alifáticas	C ₆	$C_6 = \frac{D_{2920} + D_{2850}}{D_{2955} + D_{2870}}$
Grado de condensación aromática	C ₇	$C_7 = \frac{D_{835} + D_{899}}{D_{733} + D_{775}}$

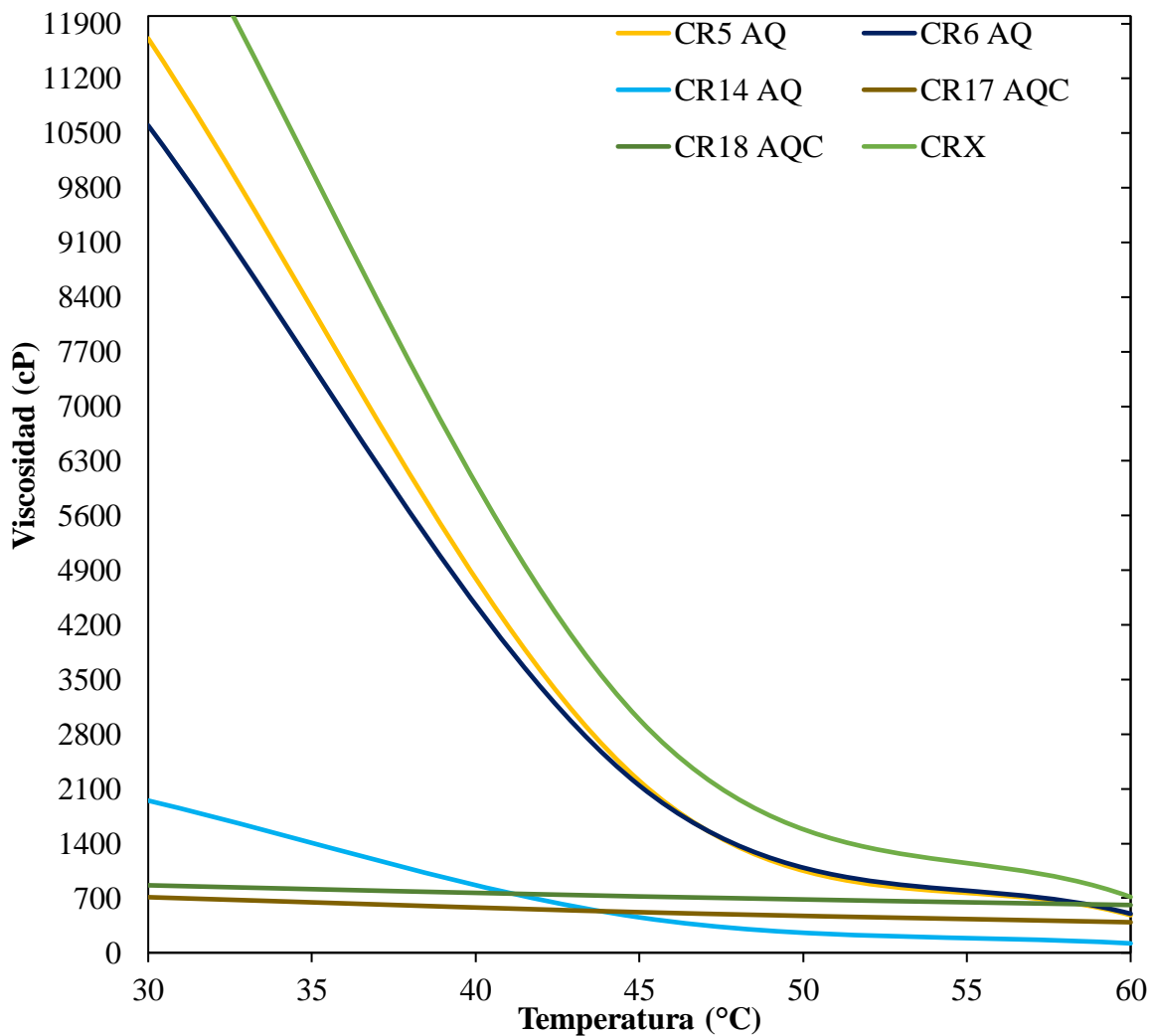
Nota: Parámetros moleculares promedio con su respectiva ecuación.

4 Resultados y análisis.

Para analizar la capacidad de cambio que tiene la acuatermólisis catalítica en la viscosidad de un crudo, se desarrolló una gráfica en la que se pueden observar los cambios que vienen desde un crudo base hasta un crudo con acuatermólisis catalítica (figura 7), este referenciado con el seudónimo AQC.

Siendo así:

- CRX, crudo base.
- CR5, CR6 y CR14 crudos con catalizadores a 100 y 200 ppm a temperatura de 270°C.
- CR18 AQC y CR17 AQC con proceso de acuatermólisis catalítica a 30h y 60h respectivamente.

Figura 13.*Gráfico Viscosidad vs Temperatura.***Tabla 9.***Valores obtenidos de viscosidad de crudos provenientes de pruebas de AQ.*

	CRX	CR5	CR6	CR14	CR18 AQ	CR17 AQ
Visc, 30°C	14015.3	11712.3	10597.9	1949.0	864.3	711.7
Visc, 45°C	2988.17	2201.76	2141.89	453.16	721.15	519.97
Visc, 60°C	714.37	481.13	499.92	120.16	611.56	390.79

Nota: De esta manera vemos en la tabla 9 una comparativa de los valores de viscosidad, comparando el crudo base con los demás, en donde, se logra percibir un gran mejoramiento de la viscosidad de los crudos, de esta manera, se procede a calcular la reducción de viscosidad con la ecuación 5:

$$\mu_{red} [\%] = \frac{\mu_i - \mu_m}{\mu_i} \times 100 \quad (7)$$

En donde:

μ_{red} = viscosidad reducida.

μ_i = viscosidad inicial

μ_m = viscosidad mejorada

En este caso, la viscosidad reducida a 30° C, 45°C y 60 °C es la siguiente:

$$\mu_{red} [\%] = \frac{14015.3 - 711.7}{14015.3} \times 100 \quad (8)$$

$$\mu_{red} @ 30^{\circ}C = 94.9 \%$$

$$\mu_{red} @ 45^{\circ}C = 82.5 \%$$

$$\mu_{red} @ 60^{\circ}C = 45.29\%$$

De este modo contrastamos que para la temperatura de 60°C no hubo una mejoría tan amplia respecto a la viscosidad del crudo base CRX, aun así, un 45.3% es una reducción optima, también da reflejo que a 30°C y 45°C. La acuatermolisis catalítica es altamente eficaz reduciendo desde casi entre un 82.5% a un 94.9% de viscosidad original de la muestra y así sucesivamente con los diferentes crudos muestreados en la investigación.

Para la comprensión de los datos arrojados en los resultados, los parámetros estadísticos se pueden definir como:

Coeficiente de determinación (R^2)

El parámetro R^2 representa la proporción de la variabilidad total de la variable respuesta explicada por el modelo. Matemáticamente, se define como aquel valor que es cercano a la unidad (1), refleja un alto poder explicativo y, por tanto, una adecuada capacidad del modelo para representar la relación lineal entre las variables latentes predictoras y la respuesta. En términos prácticos, $R^2 > 0.9$ se considera evidencia de un ajuste robusto en modelamientos quimiométricos.

Componentes Principales (Cp)

El número de componentes principales (Cp) seleccionados determina la dimensionalidad efectiva del espacio latente. Este parámetro, equilibra la varianza explicada y la complejidad del modelo, evitando el sobreajuste. Un modelo óptimo se caracteriza por un número reducido de Cp que capture la mayor parte de la varianza sistemática sin incorporar ruido experimental.

Suma de Cuadrados

Cuantifica la variabilidad en los datos. Se descompone en tres fuentes principales:

Variabilidad total observada.

Variabilidad explicada por el modelo (efecto sistemático).

Variabilidad residual no explicada (error aleatorio).

La relación entre estos constituye la base de la estimación de R^2 y por tanto, del poder explicativo global del modelo.

Grados de libertad (GL)

Los grados de libertad reflejan el número de observaciones independientes disponibles para la estimación de cada parámetro. En regresión, el GL asociado al modelo corresponde al número de parámetros ajustados, mientras que el GL residual equivale a la diferencia entre el número total de observaciones y los parámetros estimados. Este valor es esencial para la inferencia estadística y el cálculo de varianzas ajustadas.

Cuadrado Medio

El cuadrado medio es la varianza promedio asociada a una fuente de variabilidad, y se calcula como la razón entre la suma de cuadrados y sus grados de libertad correspondientes:

$$MS = \frac{SS}{GL} \quad (9)$$

Se utiliza para comparar la magnitud relativa de la variabilidad explicada (modelo) frente a la no explicada (error), constituyendo la base para el cálculo del estadístico F.

Estadístico F (F-Ratio)

El F-Ratio evalúa la hipótesis nula de que el modelo no explica significativamente más variabilidad que el error aleatorio. En donde los valores elevados de F implican que la variabilidad explicada por el modelo es sustancialmente mayor que la atribuible al error, sugiriendo una relación estadísticamente significativa entre las variables predictoras y la respuesta. Su

interpretación se apoya en los valores críticos de la distribución F para los grados de libertad respectivos.

Valor-P (P-Value)

El valor-P expresa la probabilidad de obtener un estadístico igual o más extremo que el observado, bajo la hipótesis nula de ausencia de relación significativa, $p < 0.05$ se considera evidencia suficiente para rechazar la hipótesis nula y aceptar la significancia del modelo o del parámetro evaluado. En modelos quimiométricos, valores de p extremadamente bajos corroboran la fiabilidad estadística del ajuste y la no aleatoriedad de las correlaciones observadas.

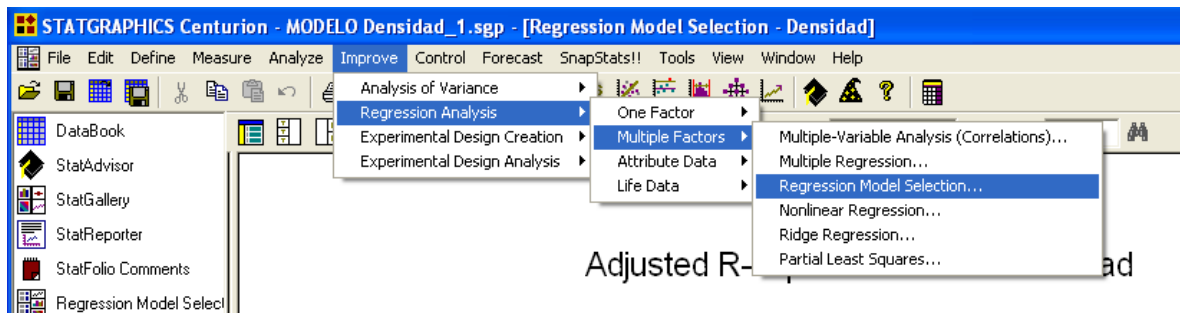
Si el análisis de los parámetros estadísticos evidencia que los modelos desarrollados presentan un ajuste sólido y confiable, los valores altos de R^2 nos indican una buena representación de los datos experimentales, a su vez, si se tiene un número adecuado de componentes principales (C_p), nos reflejarán un modelo eficiente sin sobreajuste. Los resultados de F-Ratio y p-value confirman la significancia estadística de las relaciones obtenidas, validando la capacidad predictiva de los modelos. En conjunto, estos indicadores pueden demostrar un ajuste robusto y adecuado para estimar propiedades fisicoquímicas de crudos y fracciones pesadas.

4.1 Modelos de Regresión Lineal Múltiple

En primer lugar, se desarrollan los modelos de regresión lineal múltiple en el software Statgraphics ingresando a este tal y como muestra la figura 14.

Figura 14.

Paso a paso para realización del modelo.



Nota: A partir del ingreso y con los parámetros moleculares promedios definidos, se lleva a cabo la realización de cada modelo de asfaltenos, densidad y viscosidad, teniendo como resultados lo siguiente.

4.1.1 Modelo asfaltenos por RLM

En la tabla 11 se observa la estadística dada a partir del modelo de asfaltenos a partir de los parámetros moleculares promedio que en este caso son las 4 variables independientes o y_i .

Tabla 10.

Estadística de Parámetros Moleculares Promedio Asfaltenos.

Parámetros	Estimado	Error estándar	T Estadístico	P-Value
Constante	-16.066	3.853	-4.169	0.0001
C2	-93.887	6.579	-14.270	0.0000
C3	27.443	4.862	5.644	0.0000

C4	-0.3748	0.223	-1.678	0.0990
C5	77.063	8.458	9.110	0.0000

Tabla 11.

Análisis de varianza, ANOVA.

	Suma de cuadrados	Grados de libertad	Cuadrado Medio	F-Ratio	P-Value
Modelo	2324.35	4	581.087	149.81	0.0000
Residual	213.343	55	3.878		
Total (Corr.)	2537.69	59			

Nota: Dado que el P-Value en la tabla ANOVA es menor que 0,05 se puede indicar que existe una relación estadísticamente significativa entre las variables al nivel de confianza del 95,0%.

Nótese que, en la tabla 13, las variables incluidas corresponden a la siguiente clasificación: A, corresponde a C₁, B corresponde a C₂, C corresponde a C₃, y así sucesivamente hasta G que corresponde a C₇.

Tabla 12.

Variables y R² de modelamiento para el contenido de asfaltenos.

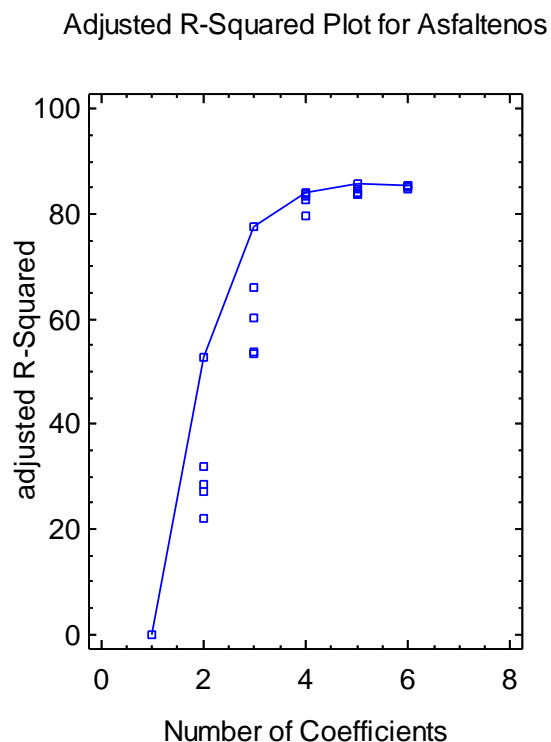
MSE	R ²	R ² Ajustada	Cp	Variables incluidas
3.878	91.593	90.982	3.262	BCDE
3.895	91.710	90.942	4.522	BCDEG
3.896	91.555	90.941	3.504	BCEG
3.906	91.688	90.918	4.661	ABCEF
3.9127	91.52	90.903	3.725	ABCE

Nota: El coeficiente de determinación R^2 evidencia que el modelo ajustado explica el 91,593% de la variabilidad observada en el contenido de asfaltenos. Por su parte, el coeficiente de determinación R^2 ajustado, considerado un estimador más robusto para la comparación entre modelos con distinto número de variables independientes, alcanza un valor de 90,9816%

Gráficamente, el modelo da como resultante lo siguiente.

Figura 15.

Cantidad de coeficientes principales para el contenido de asfaltenos.



Nota: La figura 15 muestra que el modelo alcanza una capacidad explicativa óptima con 3 a 4 coeficientes, ya que a partir de este número el valor de R^2 ajustado se estabiliza alrededor del 80–85%. Esto indica que incluir más predictores no aporta mejoras sustanciales en la calidad del ajuste, y podría incluso introducir redundancia o sobreajuste.

4.1.2 Modelo densidad RLM

Para el modelo de densidad, se desarrollan las tablas 14, 15 y 16.

Tabla 13.

Estadística de Parámetros Moleculares Promedio Densidad.

Parámetros	Estimado	Error estándar	T Estadístico	P-Value
Constante	-1.147	0.176	-6.532	0.000
C1	2.179	0.146	14.959	0.000
C2	-1.377	0.099	-13.885	0.000
C4	0.036	0.004	9.495	0.000
C6	0.592	0.060	9.887	0.000

Tabla 14.

Análisis de varianza, ANOVA

	Suma de cuadrados	Grados de libertad	Cuadrado Medio	F-Ratio	P-Value
Modelo	0.297383	4	0.0743458	102.87	0.0000
Residual	0.0383055	53	0.000722745		
Total (Corr.)	0.335689	57			

Nota: Dado que el valor P en la tabla ANOVA es menor que 0,05, existe una relación estadísticamente significativa entre las variables al nivel de confianza del 95,0%.

Tabla 15.

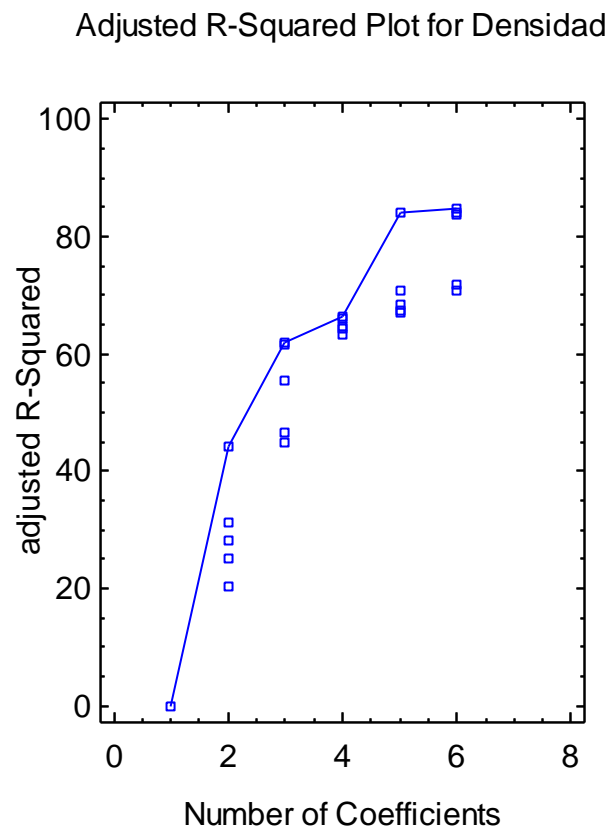
Variables y R2 de modelamiento de la densidad .

MSE	R ²	R ² Ajustada	Cp	Variables incluidas
3.87896	91.593	90.982	3.262	BCDE
3.89589	91.710	90.942	4.522	BCDEG
3.89656	91.555	90.941	3.504	BCEG
3.90622	91.688	90.918	4.661	ABCEF
3.91266	91.52	90.903	3.725	ABCE

El coeficiente de determinación R^2 ajustado, considerado más adecuado para la comparación entre modelos con diferente número de variables independientes, es de 91.59%. El error estándar de la estimación muestra que la desviación estándar de los residuos es 0,0268839. Este valor puede emplearse para la construcción de intervalos de predicción.

Figura 16.

Cantidad de coeficientes principales para la densidad.



Nota: La figura 16 nos muestra que el valor de R^2 ajustado aumenta de forma considerable con la inclusión de las primeras variables, alcanzando cerca del 60% con 3 coeficientes y superando el 80% alrededor de 5–6 coeficientes. A partir de este punto, el incremento se estabiliza, lo que indica que la adición de más variables no mejora significativamente la capacidad explicativa

del modelo, es decir, el modelo logra un nivel de ajuste adecuado con aproximadamente 5–6 variables independientes, alcanzando un R^2 ajustado cercano al 82%, lo cual refleja una alta capacidad explicativa sin necesidad de incorporar predictores adicionales que podrían introducir redundancia o riesgo de sobreajuste.

4.1.3 Modelos de viscosidad

Para el modelo de viscosidad @ 30°C se obtienen las tablas 17, 18 y 19

Tabla 16.

Estadística de Parámetros Moleculares Promedio para el modelo de viscosidad @ 30°C

(LnVisc30).

Parámetros	Estimado	Error estándar	T Estadístico	P-Value
Constante	-54.611	9.943	-5.492	0.000
C4	1.263	0.205	6.174	0.000
C5	23.548	6.520	3.612	0.001
C6	7.703	3.253	2.368	0.021
C7	61.010	9.118	6.691	0.000

Tabla 17.

Análisis de varianza, ANOVA.

	Suma de cuadrados	Grados de libertad	Cuadrado Medio	F-Ratio	P-Value
Modelo	538.364	4	134.591	33.26	0.000
Residual	238.717	59	4.046		
Total (Corr.)	777.081	63			

Nota: La tabla 17 muestra los resultados del ajuste de un modelo de regresión lineal múltiple para describir la relación entre LnVis30 y 4 variables independientes, dado que el valor

P en la tabla ANOVA es menor que 0,05, existe una relación estadísticamente significativa entre las variables al nivel de confianza del 95,0%.

Tabla 18.

Variables y R² de modelamiento para la viscosidad @ 30°C (LnVis30).

MSE	R ²	R ² Ajustada	Cp	Variables incluidas
2.487	75.182	73.042	7.965	BDEFG
2.488	75.175	73.034	7.982	CDEFG
2.501	74.612	72.891	7.340	DEFG
2.531	74.746	72.569	9.017	ABCDF
2.534	74.278	72.5341	8.148	BDEG

Nota: El coeficiente de determinación R² indica que el modelo ajustado explica el 69,2803% de la variabilidad en LnVIS30. El coeficiente de determinación R² ajustado, considerado más adecuado para la comparación entre modelos con diferente número de variables independientes, es de 67,1976%. El error estándar de la estimación muestra que la desviación estándar de los residuos es de 2,01148.

Para el modelo de viscosidad @ 45°C se obtienen las tablas 19, 20 y 21.

Tabla 19.

Estadística de Parámetros Moleculares Promedio para el modelo de viscosidad @ 45°C (LnVis45).

Parámetros	Estimado	Error estándar	T Estadístico	P-Value
Constante	-36.176	8.320	-4.348	0.000
C2	-6.215	5.388	-1.153	0.254
C4	0.764	0.170	4.506	0.000
C5	21.124	5.125	4.122	0.000
C6	3.902	2.685	1.453	0.152
C7	44.864	7.231	6.204	0.000

La salida presenta los resultados del ajuste de un modelo de regresión lineal múltiple para describir la relación entre LnVIS45 y 5 variables independientes.

Tabla 20.

Análisis de varianza, ANOVA.

	Suma de cuadrados	Grados de libertad	Cuadrado Medio	F-Ratio	P-Value
Model	436.978	5	87.396	35.14	0.000
Residual	144.252	58	2.4871		
Total (Corr.)	581.23	63			

Nota: Dado que el valor P en la tabla ANOVA es menor que 0,05, se establece una relación estadísticamente significativa entre las variables al nivel de confianza del 95,0%.

Tabla 21.

Variabes y R2 de modelamiento para el modelo de viscosidad @ 45°C (LnVis45).

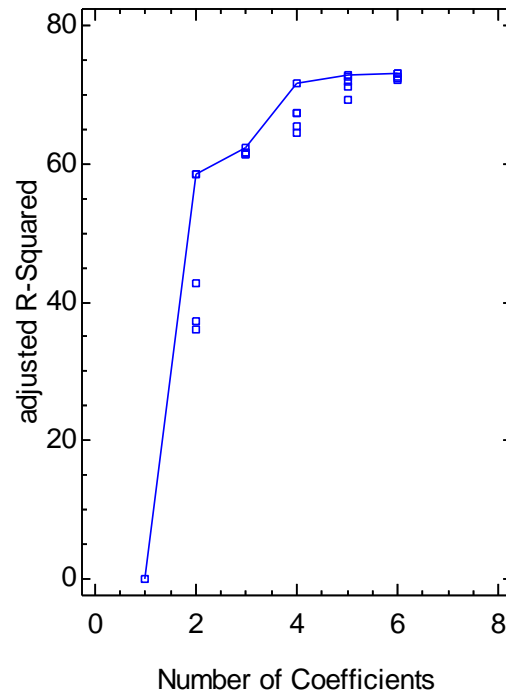
MSE	R ²	R ² Ajustada	Cp	Variabes incluidas
2.033	78.190	76.208	8.219	ABCDF
2.108	77.387	75.332	10.326	BDEFG
2.139	77.061	74.976	11.181	CDEFG
2.161	76.402	74.717	10.910	DEFG
2.110	76.407	74.262	12.899	ADEFG

El coeficiente de determinación R² ajustado, considerado más apropiado para la comparación entre modelos con distinto número de variables independientes, es de 76.208. El error estándar de la estimación muestra que la desviación estándar de los residuos es de 2.033

Figura 17.

Cantidad componentes principales para el modelo de viscosidad @ 45°C (LnVisc45).

Adjusted R-Squared Plot for LnVIS45



Nota: La figura 17 evidencia que el modelo alcanza una capacidad adecuada con alrededor de 4 a 5 variables independientes, logrando un R^2 ajustado en torno 75%. En este sentido, el incluir un mayor número de variables no incrementa de manera relevante la varianza explicada y podría introducir redundancia o riesgo de sobreajuste. Por lo tanto, el modelo óptimo se encuentra en el rango de 4 a 5 coeficientes.

Para el modelo de viscosidad @ 60°C se obtienen las tablas 22, 23 y 24.

Tabla 22.

Estadística de Parámetros Moleculares Promedio para el modelo de viscosidad @ 60°C (LnVis60).

Parámetros	Estimado	Error estándar	T Estadístico	P-Value
Constante	-25.263	3.157	-8.003	0.000
C4	0.818	0.161	5.074	0.000
C5	20.900	5.456	3.831	0.000
C7	35.126	3.867	9.083	0.000

La salida presenta los resultados del ajuste de un modelo de regresión lineal múltiple para describir la relación entre LnVIS60 y 3 variables independientes.

Tabla 23.

Análisis de varianza, ANOVA.

	Suma de cuadrados	Grados de libertad	Cuadrado Medio	F-Ratio	P-Value
Model	376.106	3	125.369	52.29	0.000
Residual	136.669	57	2.398		
Total (Corr.)	512.775	60			

Dado que el valor P en la tabla ANOVA es menor que 0.05, se establece una relación estadísticamente significativa entre las variables al nivel de confianza del 95.0%.

Tabla 24.

Variables y R² de modelamiento para el modelo de viscosidad @ 60°C (LnVis60).

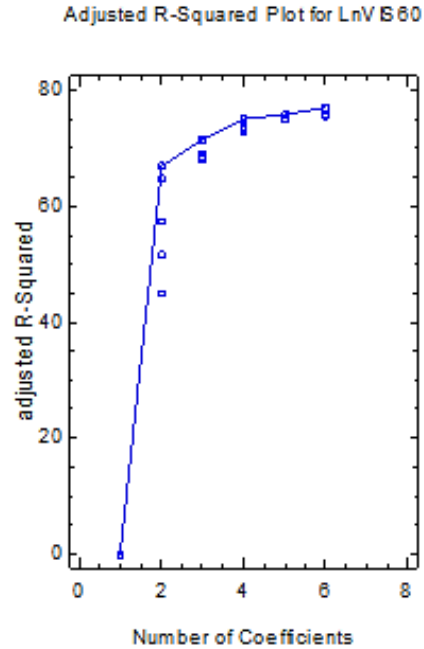
MSE	R ²	R ² Ajustada	Cp	Variables incluidas
2.033	78.190	76.208	8.2189	ABCDF
2.108	77.387	75.332	10.326	BDEFG
2.139	77.061	74.976	11.181	CDEFG
2.161	76.402	74.717	10.910	DEFG
2.199	76.407	74.262	12.899	ADEFG

Nota: El coeficiente de determinación R² indica que el modelo ajustado explica el 78.190% de la variabilidad en LnVis60. El coeficiente de determinación R² ajustado, considerado más apropiado para la comparación entre modelos con diferente número de variables independientes, es de 76.208%. El error estándar de la estimación muestra que la desviación estándar de los residuos es de 2.033%

La figura 18 evidencia que el R² ajustado aumenta de manera rápida con la inclusión de las primeras variables, alcanzando cerca del 70% al considerar 3 coeficientes. Posteriormente, el valor se estabiliza en torno al 78% a partir de 4 variables, sin presentar mejoras significativas al añadir más predictores.

Figura 18.

Coefficientes principales para el modelo de viscosidad @ 60°C (LnVisc60)



Nota: Imagen tomada del software statgraphics

Para el cálculo de cada componente y_i , se tiene la ecuación 7 asumida como la sumatoria de la multiplicación de un componente B_o y un C_n , la cual es desarrollada mediante la tabla 26.

$$y_i = \beta_o + \sum_{i=1}^{C_7} \beta_i * C_i \quad (10)$$

Tabla 25.

Parámetros de cálculo de propiedades.

Propiedad y_i	Parámetros β_o								R^2_{cal}	R^2_{val}	RMSE
	β_o	C_1	C_2	C_3	C_4	C_5	C_6	C_7			
Asfaltenos	-16.066	-	-93.887	27.443	-0.375	77.063	-	-	91.593	90.982	1.970
Densidad	-1.147	2.179	-1.377	-	0.036		0.592	-	88.589	87.728	0.003

LnVisc@30	-54.610	-	-	-	1.263	23.548	7.703	61.008	69.280	67.198	2.012
LnVisc@45	-36.176	-	-6.215	-	0.764	21.124	3.902	44.864	75.182	73.042	1.260
LnVisc@60	-25.264	-	-	-	0.818	20.899	-	35.126	73.347	71.944	1.548

Nota: Después del desarrollo de los modelos de regresión múltiple por parámetros moleculares promedio, se logra evidenciar que, aunque estos modelos presentan un ajuste con valores de correlación de calibración en el intervalo de 69.28 a 91.593. Por tanto, las variables de estudio un grado de confiabilidad estadístico en el orden decreciente para contenido de asfaltenos > densidad > viscosidad 45 °C > viscosidad 60 °C > viscosidad 30 °C. Considerando lo anterior, para el mejoramiento de los modelos se plantea el desarrollo de modelos por PLS usando gran parte de la información de los espectros de infrarrojo, dado que puede presentar una mejor correlación entre las propiedades dependientes y la composición química de los crudos.

4.2 Modelos PLS

Para los modelos multivariados PLS se tendrán en cuenta 4 tipos de bandas las cuales pueden corroborar en la tabla 26,

Tabla 26.

Anchos de bandas de los Espectros usados.

Espectros	Tamaño	Banda [cm^{-1}]
Espectro 1	6431	500 – 3600
Espectro 2	2697	500 - 1800
Espectro 3	1246	2600 - 3200
Espectro 4	3943	500 - 1800 ; 2600 - 3200

Nota: Luego de obtener los anchos de banda de cada espectro, se puede asegurar la captura de señales asociadas a vibraciones características de los grupos funcionales relevantes en los crudos, a partir de ellas se desarrollan cada una de las regresiones sin hacer ningún cambio, dando como resultado los datos presentados en las tablas siguientes en el documento, para su respectivo análisis las regresiones se manejarán de la siguiente manera:

- Sin corrección de banda.
- Normalización, para lograr un ajuste a escala y de esta manera poder comparar los espectros entre sí.
- Transformación por línea base para lograr un ajuste en el área bajo la curva de cada espectro en el que se asemejen cada uno de ellos.

4.2.1 Modelos Asfaltenos C7+

Para desarrollar el modelo de asfaltenos, se clasifica la data en dos secciones, primero, los que tengan un contenido mayor al 5% de peso y menor del 5%, ya que al desarrollar el modelo manteniendo todos los datos en conjunto y hacer una división de los datos, se genera espacio muestral muy significativo entre los valores, generando un R^2 menor a 0,95. Por lo cual se decidió hacer esta discriminación de datos en 2 categorías.

En primer lugar, se tiene el modelo estándar de los asfaltenos mayores a 5%, con los anchos de banda mostrados en la tabla 26.

Para detallar mejor los datos obtenidos, se organizaron en 2 tablas, la primera será llamada primera regresión, correspondiente a aquella que se hace al inicio sin extraer ningún dato original, y la segunda será llamada regresión final, esta será en donde se termina de realizar la regresión con un número de muestras extraídas correspondientes a menos del 10% de ellas (Anexos).

Seguido a esto, se presentan los modelos que son más representativos para las muestras de asfaltenos < 5% son presentados en la tabla 27.

Tabla 27.

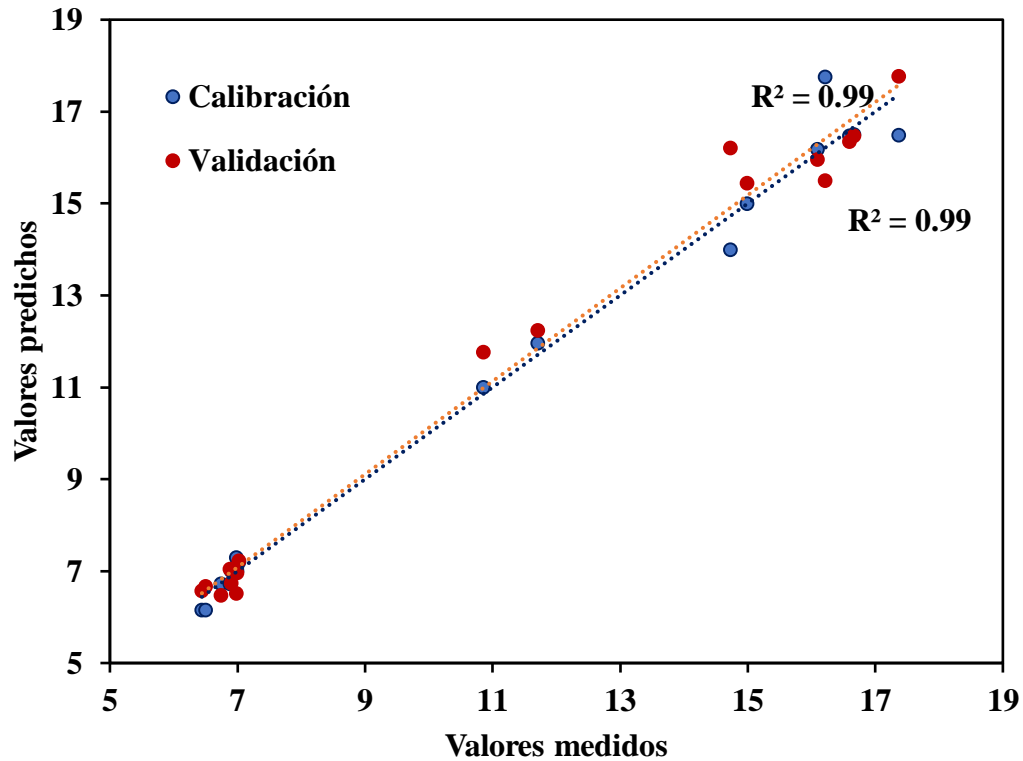
Modelos representativos Asfaltenos < 5%.

Parámetro estadístico	Modelos representativos			
	Tipo de Estudio	Modelo base	Modelo línea base	Modelo normalizado
Espectro	3	4	4	4
Rcal	0.999	0.995	0.987	0.987
Rval	0.990	0.981	0.963	0.963
RMSEcal	0.074	0.285	0.495	0.495
RMSEval	0.502	0.608	0.901	0.901
CP	10	8	5	5

Nota: Al comparar los tres modelos representativos, se observa que, aunque el modelo sin ningún cambio en el espectro requiere un mayor número de componentes principales, es el que ofrece un mejor desempeño global. Sus valores de correlación en calibración y validación son más altos y los errores son menores solo con el descarte de 4 muestras equivalentes a un 8% de la muestra, lo que indica que logra describir con mayor precisión la relación entre los espectros y la variable de interés (figura 19) . En contraste, los modelos con correcciones como línea base o normalización reducen el número de componentes necesarios, pero sacrifican capacidad predictiva, mostrando un ajuste menos confiable.

Figura 19.

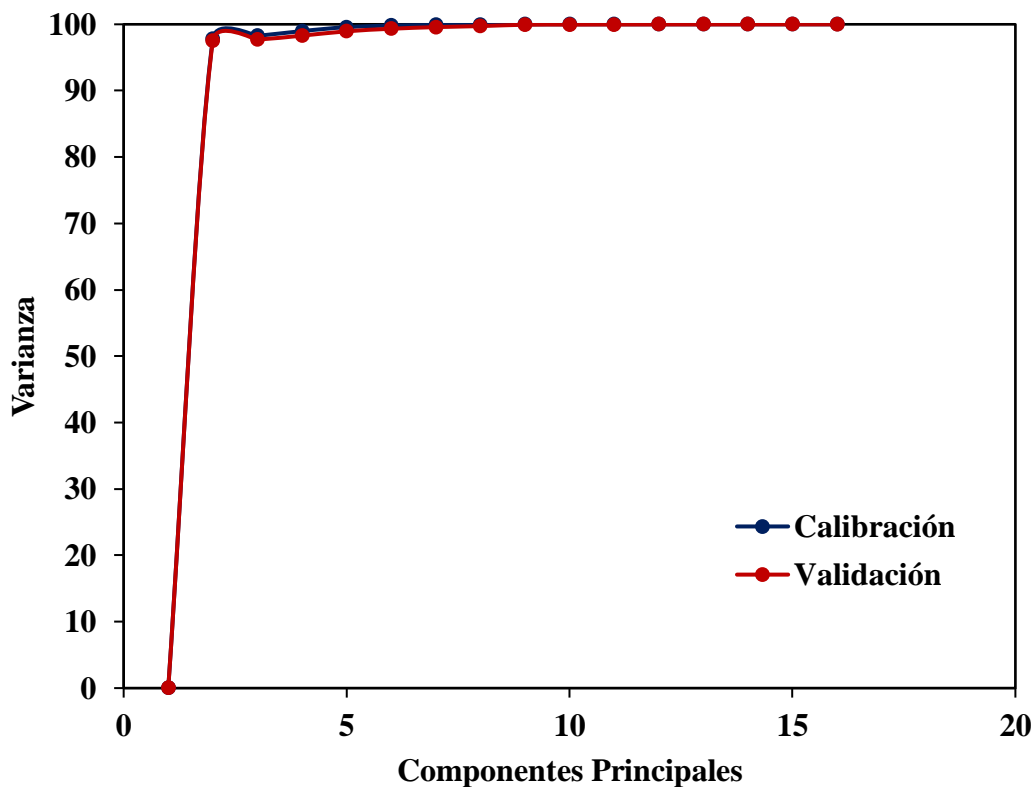
Valores medidos vs predichos para el modelo sin alteración para el contenido de asfaltenos (>5% peso).



Nota: En la figura 19 se presenta la regresión PLS donde muestra los valores predichos contra los valores medidos, de esta manera las dos líneas (azul y roja) corresponden a ajustes de calibración y validación cruzada respectivamente, donde se puede visualizar la poca diferencia entre los puntos con una alta precisión y capacidad para predecir con una mínima desviación respecto a los valores medidos.

Figura 20.

Cantidad Componentes Principales Modelo seleccionado sin alteración para el contenido de asfaltenos (% peso)



Nota: la figura 20, Se observa que PC1 ya explica alrededor del 100% de la varianza (la curva sube casi verticalmente y se aplana enseguida), esto indica que los datos tienen una estructura altamente correlacionada y que un solo componente principal es suficiente para describirlos casi en su totalidad. A partir de PC2 en adelante, la contribución adicional es mínima (la línea se vuelve plana). En pocas palabras: este gráfico confirma que la mayor parte de la información de las variables originales se concentra en el primer componente principal.

Ahora se desarrollan las muestras menores a un 5% de valores de asfaltenos, iniciando de la misma manera que los crudos mayores a 5 con los infrarrojos base y luego con las modificaciones por línea base y normalizado respectivamente (Anexos).

En la tabla 28 se presentan los modelos representativos para aquellos asfaltenos menores a 5% en peso.

Tabla 28.

Modelos representativos asfaltenos < 5%.

Modelos representativos			
Tipo de Estudio	Modelo base	Modelo línea base	Modelo normalizado
Espectro	2	4	4
Rcal	0.988957	0.982092	0.990458
Rval	0.983770	0.978384	0.982613
RMSEcal	0.138534	0.177837	0.130469
RMSEval	0.174182	0.202359	0.182404
CP	4	2	5

Nota: El análisis comparativo de los tres modelos revela que, si bien todos exhiben métricas de ajuste satisfactorias de R^2 cal y R^2 val > 0.97 , existen diferencias en términos de estabilidad. El modelo con corrección de línea base, a pesar de su simplicidad estructural con 2 CP, muestra un deterioro en capacidad predictiva, reflejado en los valores relativamente más altos de RMSE, lo que indica un sobreajuste insuficiente para capturar la variabilidad sistemática de los datos.

El modelo normalizado presenta un ajuste sobresaliente en calibración, con el menor error de entrenamiento y la correlación más alta; sin embargo, esta ventaja se atenúa en la etapa de validación, donde el incremento del error y los 5 CP que, aunque siendo el mayor número de componentes principales, de sugieren una complejidad innecesaria y potencial riesgo de sobreajuste.

En contraste, el modelo base evidencia un comportamiento más robusto: mantiene un equilibrio entre precisión predictiva y complejidad, alcanzando errores reducidos tanto en calibración como en validación, con un número de 4 CP equivalente a un número moderado de componentes. Este balance lo posiciona como el modelo más equilibrado y confiable, optimizando simultáneamente ajuste y capacidad de generalización.

Lo anterior se refleja en la Figura 21, donde el modelo seleccionado muestra un alineamiento casi perfecto entre los valores medidos y predichos, y en la Figura 22, donde el análisis de componentes principales evidencia que la mayor parte de la varianza se concentra en los primeros factores, reafirmando la solidez del modelo obtenido para este rango de concentración de asfaltenos.

Figura 21.

Valores medidos vs predichos para el modelo de asfaltenos < 5%.

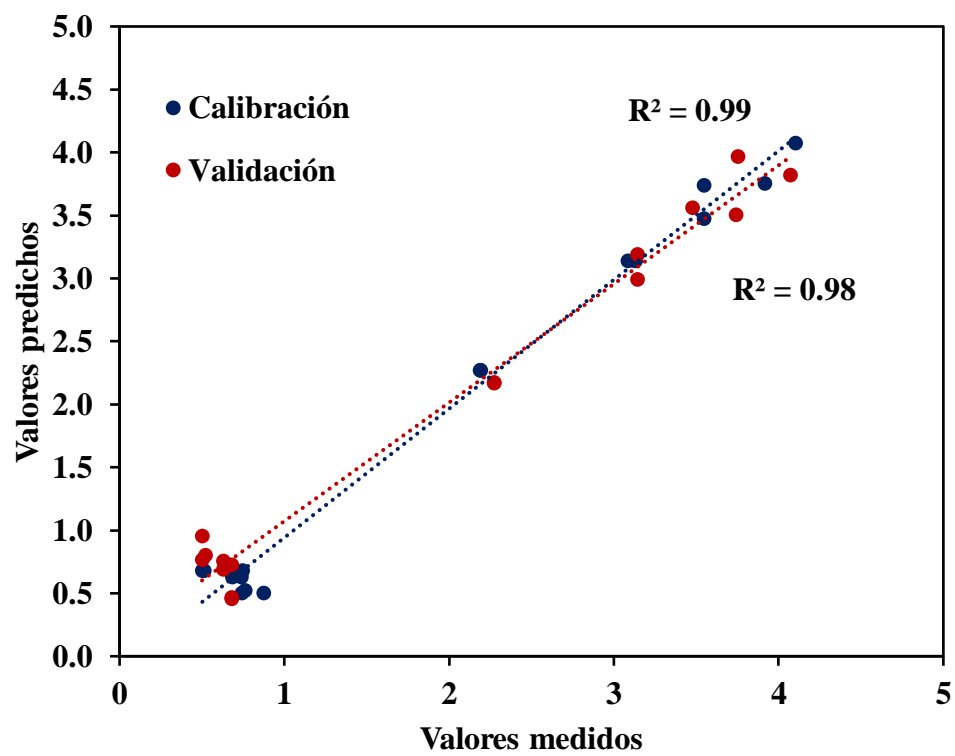
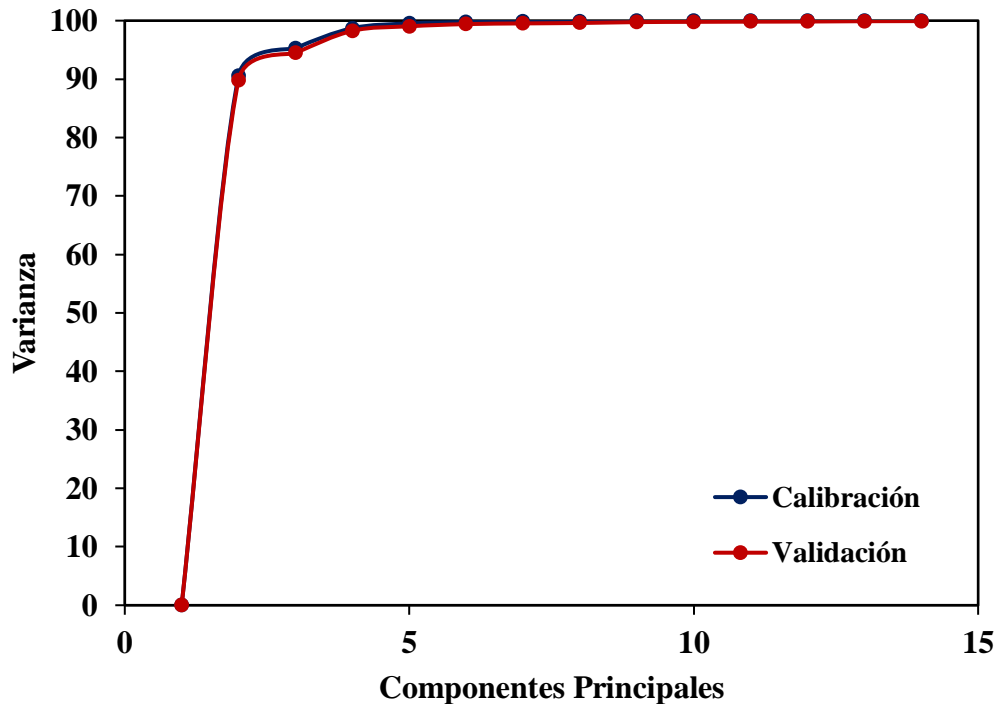


Figura 22.

Gráfica Componentes Principales Asfaltenos < 5%.



Nota: los modelos base para asfaltenos < 5% demostraron tener un mayor grado de robustez. Aunque el ajuste inicial fue bueno, la eliminación de unas pocas muestras permitió alcanzar un R^2 de hasta 0.99 en calibración y validación, con errores mínimos. El espectro 2 (figura 21), se destacó como el modelo más representativo, mientras que el análisis de componentes principales (figura 22) confirma que la estructura de los datos es simple y altamente correlacionada.

En términos estadísticos, el modelo base constituye la solución más apropiada, al maximizar la varianza explicada con un número reducido de factores y preservar la estabilidad predictiva.

4.2.2 *Modelo PLS densidad*

Procedemos al modelo de densidad, mostrado también con 3 diferentes cambios (Anexos). dando como resultados los modelos presentados en la tabla 29. De esta manera se presentan los 3 modelos representativos de la densidad.

Tabla 29.

Modelos representativos.

Modelos representativos			
Tipo de Estudio	Modelo base	Modelo línea base	Modelo normalizado
Espectro	2 primera regresión	2 primera regresión	1
Rcal	0.996	0.998	0.991
Rval	0.997	0.998	0.984
RMSEcal	0.005	0.003	0.007
RMSEval	0.006	0.004	0.009
CP	5	5	5

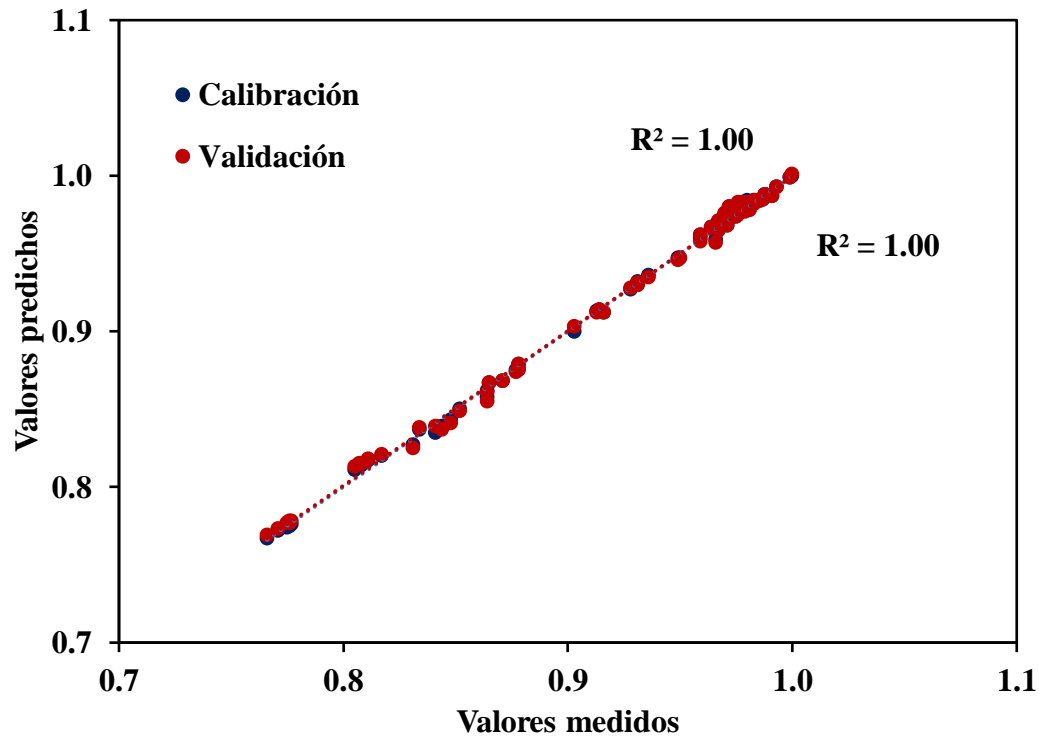
Nota: En el análisis comparativo de los modelos reflejan que, aunque todos presentan desempeños sobresalientes, la corrección de línea base ofrece la mejor combinación entre precisión y estabilidad. Sus coeficientes de correlación se mantienen en niveles casi perfectos, mientras que los errores de calibración y validación son los más bajos observados, lo cual indica una mayor capacidad de generalización sin incrementar la complejidad estructural del modelo.

En términos prácticos, esto significa que una corrección estadísticamente sencilla, como el ajuste de la línea base, aporta un beneficio significativo en donde no solo optimiza la predicción,

sino que lo hace con un mismo número de componentes principales, evitando así caer en modelos innecesariamente complicados.

Figura 23.

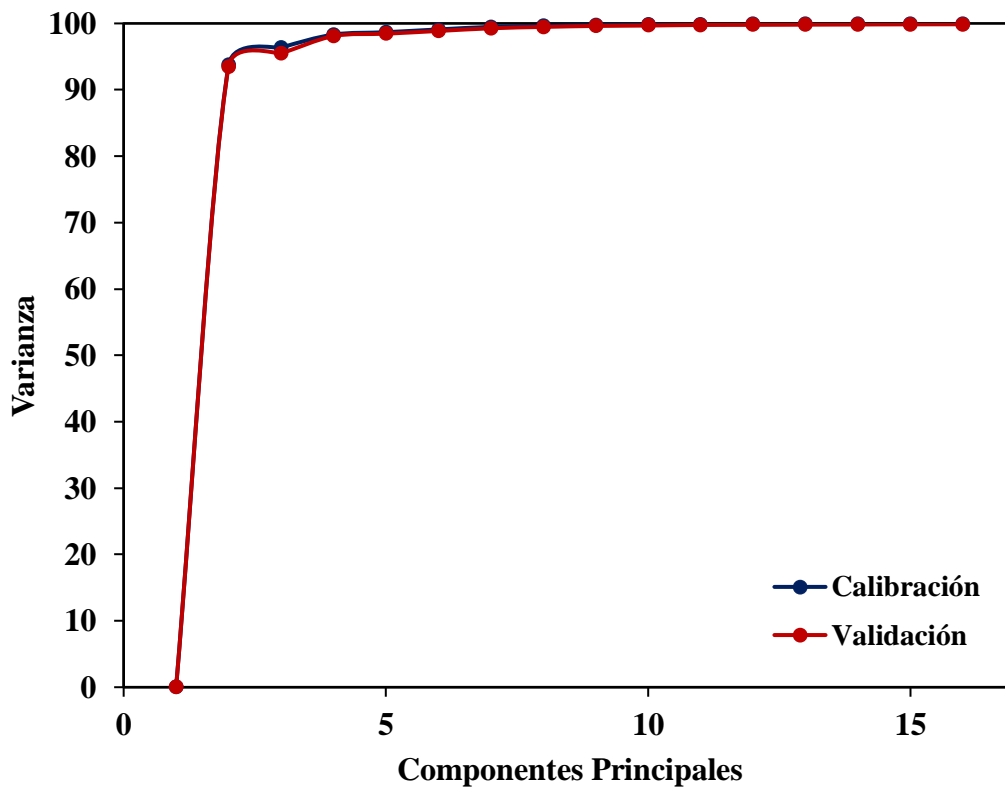
Valores medidos vs predichos para el modelo línea base seleccionado.



Nota: La figura 23 muestra la dispersión entre los valores medidos y predichos tras aplicar la corrección de línea base al modelo de densidad, en ella se aprecia un alineamiento estrecho de los puntos con respecto a la recta de calibración, lo que confirma la alta precisión del ajuste. Dichos resultados podemos compararlos ya que coinciden con lo que se evidencia en la tabla 44, donde se pueden observar los coeficientes de correlación alcanzan valores cercanos a 0.99 y los errores de predicción se reducen, validando la robustez del modelo en la última regresión.

Figura 24.

Componentes principales modelo escogido.



Nota: Por otra parte, si observamos la Figura 24, podemos notar que nos está representando la varianza explicada por los componentes principales del modelo corregido por línea base. La curva nos muestra que los primeros factores concentran prácticamente toda la información relevante, estabilizando rápidamente.

En general, la evidencia respalda que el modelo con corrección de línea base es la alternativa más sólida y confiable, no solo por su rendimiento numérico, sino porque logra un equilibrio entre rigor estadístico y aplicabilidad práctica, lo cual es fundamental en esta investigación.

4.2.3 *Modelo PLS viscosidad*

Para las viscosidades se desarrollan 2 matrices de modelos usando los crudos con viscosidades inferiores a 1000 cP.

La primera se desarrolla respecto al espectro 2, con viscosidades a 30°C , 45°C y 60°C y a su vez se incluyen los logaritmos naturales de cada viscosidad para poder tener un mejor coeficiente de correlación R^2 de calibración y validación, así como un RMSE menor a 1, la segunda se efectúa con el espectro 4 a las mismas condiciones, esta información se ve reflejada en las tablas 30 y 31

Tabla 30.

Modelos para el comportamiento en la viscosidad de crudos usando espectro 2.

Modelo Base					
	R^2_{cal}	R^2_{val}	RMSE _{cal}	RMSE _{val}	CP
Visc @ 30°C	0.89	0.818	79.21	108.52	4
Visc @ 45°C	0.936	0.856	51.5	79.6	4
Visc @ 60°C	0.897	0.71	59.2	91.6	4
LnVisc @ 30°C	0.962	0.93	0.48	0.65	4
LnVisc @ 45°C	0.955	0.923	0.47	0.645	4
LnVisc @ 60°C	0.905	0.847	0.58	0.993	4

Tabla 31.*Modelos para el comportamiento en la viscosidad de crudos usando espectro 4*

Modelo Base					
	R ² cal	R ² val	RMSEcal	RMSEval	CP
Visc @ 30°C	0.781	0.741	82.3	109.35	4
Visc @ 45°C	0.956	0.676	40.4	100.	6
Visc @ 60°C	0.774	0.666	86.7	108.7	4
LnVisc @ 30°C	0.978	0.954	0.373	0.554	5
LnVisc @ 45°C	0.963	0.926	0.939	0.633	5
LnVisc @ 60°C	0.81	0.815	0.866	1.083	3

Nota: Luego de obtener estos resultados, en donde el procesamiento de las viscosidades por logaritmo natural arroja R's más acordes a un modelamiento y errores medios relativamente bajos, se procede a hacer una transformación de los espectros, con los cuales se decide partir hacia los modelos pls por medio de línea base y normalización por área de espectros, en donde los resultados de cada espectro son reflejados en las tablas 32 y 33.

Tabla 32.*Modelos para el comportamiento en la viscosidad de crudos usando tratamiento del espectro 2.*

Espectro 2						
	Transformación	R ² cal	R ² val	RMSEcal	RMSEval	CP
LnVisc @ 30°C	Base line	0.951	0.817	0.561	1.116	6
LnVisc @ 45°C	Base line	0.926	0.882	0.61	0.802	3
LnVisc @ 60°C	Base line	0.913	0.865	0.723	0.131	3

LnVisc @ 30°C	Área	0.981	0.998	0.34	0.54	5
LnVisc @ 45°C	Área	0.934	0.901	0.55	0.3	4
LnVisc @ 60°C	Área	0.878	0.815	0.85	1.09	4

Tabla 33.

Modelos para el comportamiento en la viscosidad de crudos usando tratamiento del espectro 4.

Espectro 4						
	Transformación	R ² cal	R ² val	RMSEcal	RMSEval	CP
LnVisc @ 30°C	Base line	0.957	0.801	0.52	1.16	7
LnVisc @ 45°C	Base line	0.948	0.879	0.516	0.809	5
LnVisc @ 60°C	Base line	0.898	0.851	0.785	0.983	3
LnVisc @ 30°C	Área	0.983	0.963	0.327	0.503	5
LnVisc @ 45°C	Área	0.96	0.928	0.431	0.626	5
LnVisc @ 60°C	Área	0.815	0.713	1.058	1.153	2

Después de hacer cada uno de los modelos, se opta por la normalización de área del espectro 4, donde respecto al espectro 2 arroja mejores resultantes de validación R²val que van de 0.815 a 0.98, un RMSE para LnVisc @ 30°C y LnVisc @ 45°C poco menor al base line y menor cantidad de componentes principales.

Para las viscosidades mayores a 1000 cP se desarrollan de igual manera 2 matrices de modelos las cuales se muestran en las tablas 34 y 35 por medio de los espectros 3 y 4.

Tabla 34.*Modelos para viscosidades mayores a 1000 cP usando espectro 2.*

Modelos Espectro 2						
	Transformación	R ² cal	R ² val	RMSEcal	RMSEval	CP
LnVisc @ 30°C	Normal	0.985	0.856	0.242	0.729	11
LnVisc @ 45°C	Normal	0.981	0.903	0.219	0.502	10
LnVisc @ 60°C	Normal	0.91	0.894	0.249	0.924	9
LnVisc @ 30°C	Área	0.986	0.882	0.209	0.639	11
LnVisc @ 45°C	Área	0.981	0.912	0.211	0.479	10
LnVisc @ 60°C	Área	0.963	0.876	0.281	0.581	9

Tabla 35.*Modelos viscosidades mayores a 1000 cP usando espectro 3.*

Modelos Espectro 4						
	Transformación	R ² cal	R ² val	RMSEcal	RMSEval	CP
LnVisc @ 30°C	Normal	0.99	0.791	0.0069	0.842	17
LnVisc @ 45°C	Normal	0.967	0.867	0.77	0.591	8
LnVisc @ 60°C	Normal	0.965	0.897	0.23	0.48	8
LnVisc @ 30°C	Área	0.982	0.781	0.242	0.812	12
LnVisc @ 45°C	Área	0.975	0.86	0.241	0.592	10
LnVisc @ 60°C	Área	0.960	0.873	0.293	0.541	8

Nota: A partir de estos modelos seleccionados, se proceden a desarrollar los modelos finales, en donde las resultantes son variables respecto al espectro 2 ya que presenta valores mejores con la transformación por área.

De esta manera se van a desarrollar los modelos de manera individual para viscosidades y temperaturas de la siguiente manera,

En primer lugar, se desarrollan las regresiones correspondientes a $\mu < 1000$ cP, seguido a esto se toman las regresiones para $\mu > 1000$ cP, cada una de estas evaluadas a temperaturas de 30°C, 45°C y 60°C respectivamente, teniendo en cuenta que para las $\mu > 1000$ cP se tiene un tamaño muestral de 32 crudos distintos y para $\mu < 1000$ cP un muestreo de 33 crudos

Para el modelo de viscosidad a 30 °C para muestras menores a 1000 cP Se usa el espectro 4, ya que este se conforma por el espectro 2 y 3, a su vez tiene mucho menos ruido que el primer espectro. Seguido a esto se le hizo tratamiento con normalización por área y se eliminan las 3 muestras como se muestra en la tabla 36.

Tabla 36.

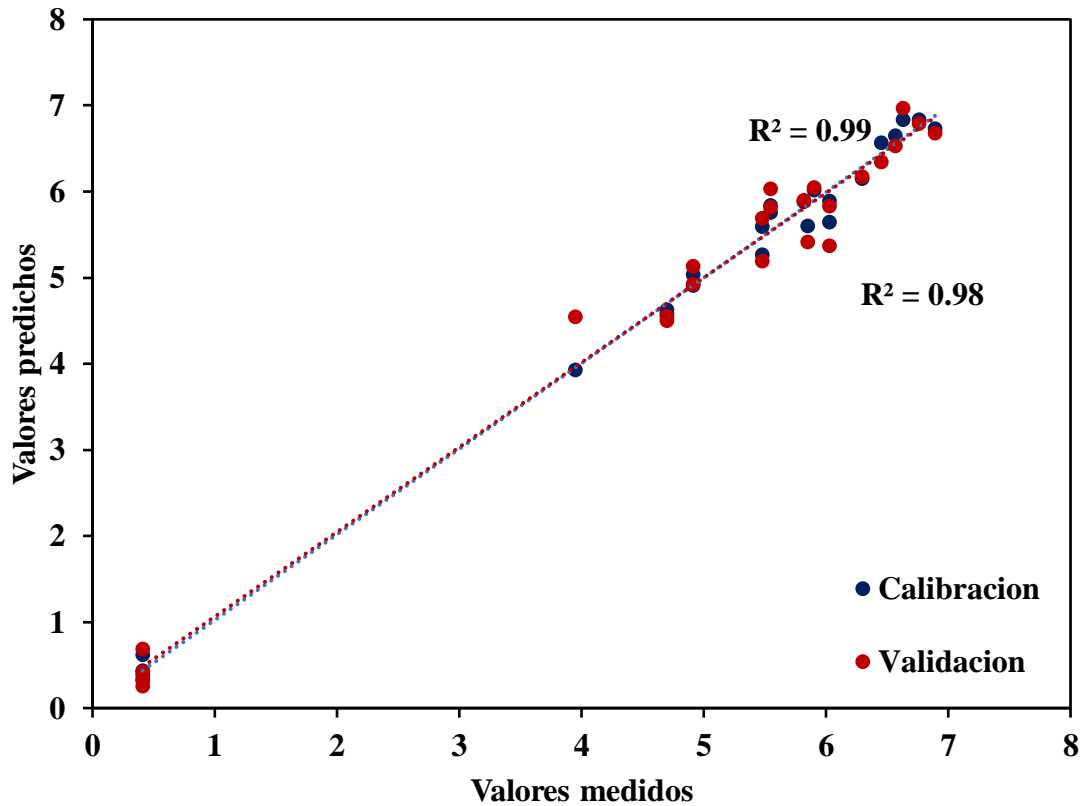
Modelo Lnvisc @ 30°C menor a 1000 cP.

Normalización por área espectro 4.					
Muestras eliminadas	R^2_{cal}	R^2_{val}	RMSEcal	RMSEval	CP
-	0.982	0.941	0.264	0.498	6
1	0.988	0.954	0.216	0.448	6
2	0.984	0.959	0.254	0.433	5
3	0.9938	0.983	0.162	0.280	6

Nota: El análisis de varianza indica que se explica el 82,13% de la varianza en X y el 16.76% en Y, los valores son suficientes considerando la naturaleza no lineal de la viscosidad. El grafico de scores permite evidenciar que las muestras seleccionadas están distribuidas homogéneamente dentro de los límites de confianza, sin presencia de valores atípicos significativos, esto nos aporta una mayor robustez en el modelo.

Figura 25.

Valores medidos vs predichos para el modelo LnVisc @ 30°C menor a 1000 cP.

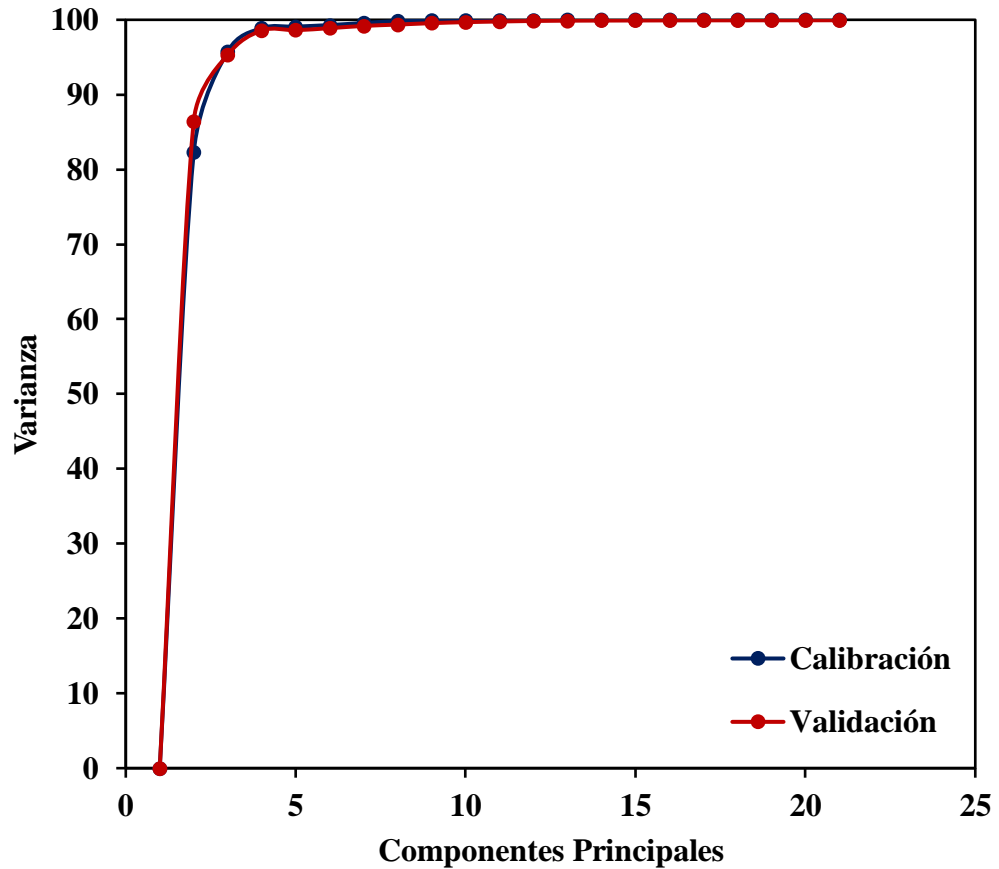


Nota: El modelo alcanzó un R^2 de calibración de 0.9939 y un R^2 de validación de 0.9836, Figura 25 con lo cual deducimos que es una correlación casi perfecta entre valores medidos y

predichos. Así mismo los valores de RMSE bajos (0.168 y 0.280) y las pendientes cercanas a la unidad confirman la ausencia de sesgos significativos.

Figura 26.

Componentes principales modelo escogido.



Nota: En la figura 26, el modelo muestra que es bastante confiable, preciso y adecuado para la estimación de viscosidades en el rango estudiado, pudiendo ser utilizado como herramienta predictiva en aplicaciones de laboratorio.

Para el modelo de viscosidad a 30 °C para muestras mayores a 1000 cP, se evalúan diferentes muestras de la tabla 57 para $L_{\text{visc}} @ 30 \text{ °C} > 1000 \text{ cP}$ con el fin de obtener el mejor resultado,

En la tabla 37 se refleja en resultado inicial del modelo, y el final después de hacer la eliminación de 6 muestras, lo cual significa para el estudio, uno de los modelos más difíciles de llevar a cabo en donde se toma una buena parte de la programación para su desarrollo.

Tabla 37.

Modelo visc @ 30°C > 1000 cP.

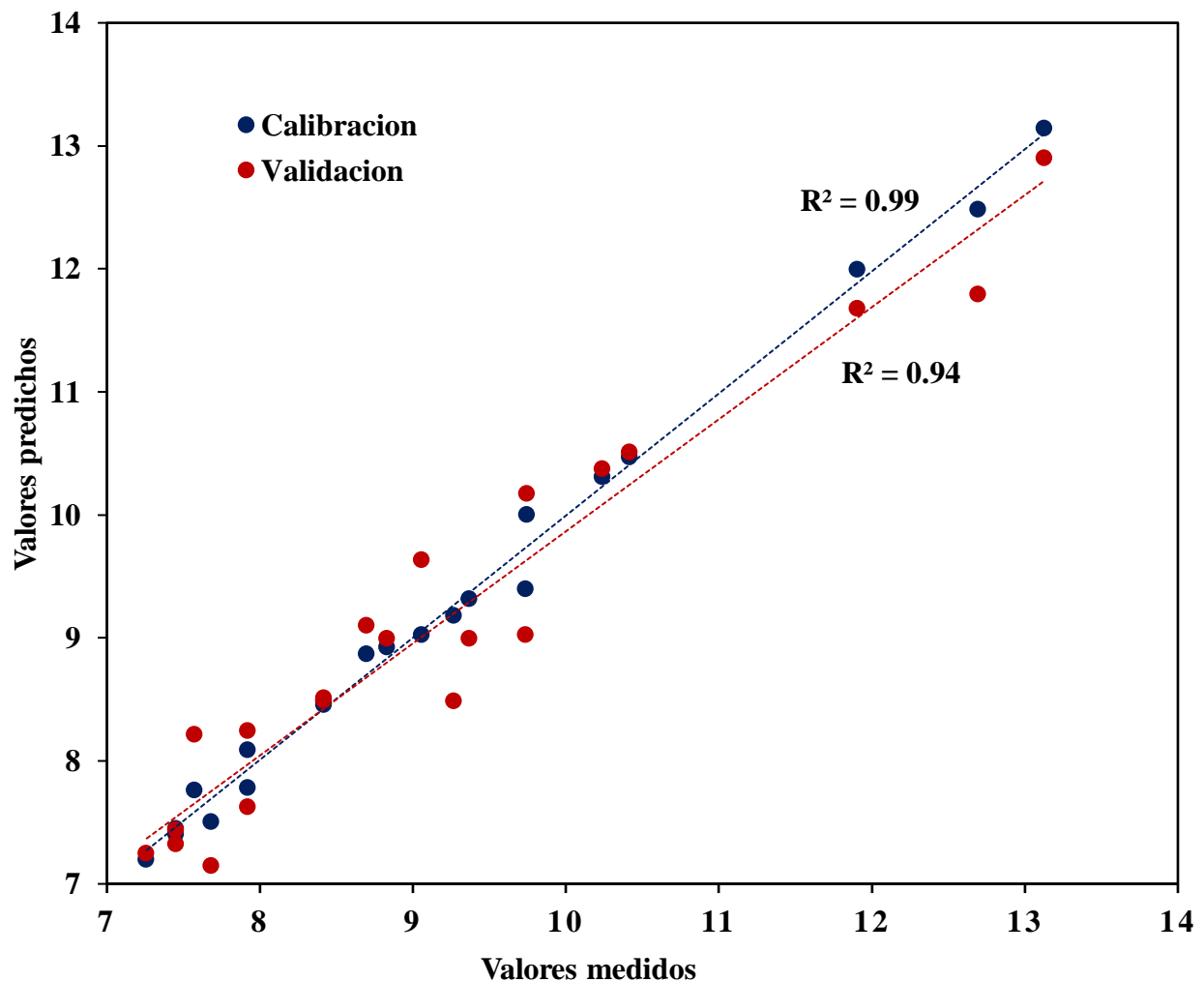
Normalización por área espectro 4.					
Muestras eliminadas	R^2_{cal}	R^2_{val}	RMSEcal	RMSEval	CP
-	0.982	0.789	0.21	0.857	12
6	0.992	0.940	0.1409	0.268	10

Nota: La tabla 37 muestra un desarrollo donde el modelo inicial presenta un R^2_{cal} de 0,982 y un R^2_{val} de 0,789, con un error de predicción relativamente alto, tras la eliminación de las 6 muestras, el modelo mejora de manera significativa en donde el R^2_{val} aumenta a 0,940 y el RMSEval disminuye a 0,268, lo que indica una capacidad predictiva mucho más robusta y estable. Los resultados permiten indicar que, la exclusión de las muestras atípicas optimiza el desempeño del modelo, reduciendo errores y aumentando la correlación en la validación, aunque son un gran número de muestras quitadas, hay una gran mejoría en el modelo, en donde inicialmente no tenía gran validez de demostración, y como finalidad sugiere que dichas muestras actuaban como outliers que deterioraban la calidad del ajuste y la capacidad de predicción.

El desempeño del predictivo, evaluado mediante los estadísticos de calibración y validación muestra en la Figura 42 unos R^2 superiores a 0.97, acompañados de RMSE bajos en relación con el rango de viscosidad evaluado. Las pendientes cercanas a la unidad y los offsets reducidos confirman que no se observan sesgos sistemáticos.

Figura 27.

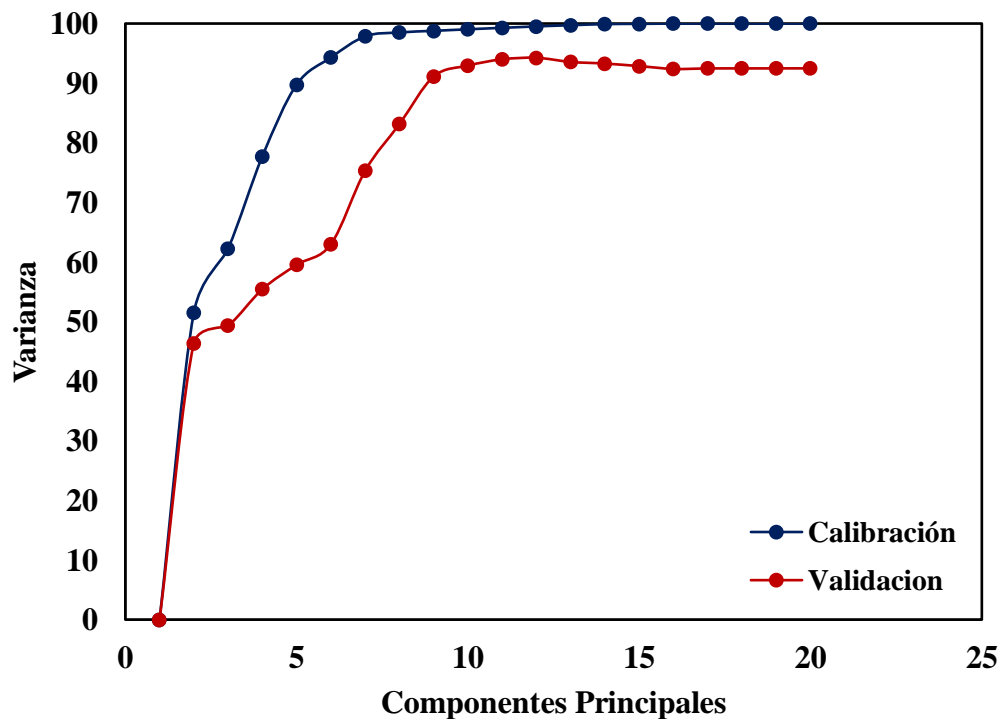
Valores medidos vs predichos para el modelo Lnvisc @ 30°C mayores a 1000 cP.



Nota: Como se puede observar en la Figura 27, el modelo presenta un ajuste estadísticamente sólido, aunque con mayor complejidad debido a la naturaleza altamente no lineal de fluidos en este rango. Los gráficos de scores arrojan que las muestras se atribuyen dentro del intervalo de confianza, lo que indica estabilidad en el modelo y ausencia de valores atípicos significativos.

Figura 28.

Componentes principales modelo escogido.



Nota: En la figura 28 se observa que el modelo explica un porcentaje elevado en X y un valor representativo en Y, los que refleja que, aunque la predicción de viscosidades altas presenta más dispersión, la relación espectral es capturada de manera adecuada. Los coeficientes de

regresión evidencian regiones espectrales con influencia marcada, se refiere a la sensibilidad del modelo frente a cambios estructurales en las muestras

Para el modelo de viscosidad a 45 °C para muestras menores a 1000 cP, se tiene en cuenta también el espectro 4 en donde con la sustracción de 3 muestras con las cuales se alcanza un R^2 bastante significativo dado a que la variación de viscosidad en este punto es visiblemente variable, visible en la tabla 38.

Tabla 38.

Modelo Lmvisc @ 45°C < 1000 cP

Normalización por área espectro 4.					
Muestras eliminadas	R^2_{cal}	R^2_{val}	RMSEcal	RMSEval	CP
-	0.927	0.984	0.541	0.754	4
1	0.984	0.956	0.251	0.4439	5
3	0.989	0.977	0.145	0.301	5

Nota: La tabla 38, permiten evidenciar que, al inicio, el modelo parece bueno en validación $R^2_{val} = 0,984$, pero sus errores son altos con un $RMSE_{val} = 0,754$, lo que indica que no es muy preciso al hacer predicciones, luego de esto, cuando se quita la muestra 7, el modelo se vuelve más estable, es decir, aumenta su capacidad de ajuste ya que el R^2_{cal} sube a 0,98 y los errores disminuyen tanto en calibración como en validación. Finalmente, al eliminar las 3 muestras, el modelo alcanza su mejor desempeño y logra explicar casi toda la variabilidad de los datos con un $R^2_{cal} = 0,989$ y $R^2_{val} = 0,977$ y los errores se reducen al mínimo donde el $RMSE_{val} = 0,301$.

El modelo presenta un desempeño altamente confiable en termino de calibración y validación. Las muestras en el gráfico de scores visto en la Figura 43 podemos notar que se encuentran bien distribuidos dentro de la elipse de Hotelling (95%), sin presencia de outliers significativos, lo que respalda la homogeneidad del conjunto de datos y la estabilidad estructural del modelo.

Los indicadores estadísticos mostrados en la figura 30, reflejan valores de R^2 de calibración a 0.99 y de validación mayores a 0.98 como se puede ver en la figura 45, esto nos confirma un ajuste casi perfecto entre viscosidades medidas y predichas. Asimismo, los errores cuadráticos medios RMSE permanecen en niveles bajos siendo menores o iguales a 0.3 en unidades normalizadas, y los offsets próximos a cero evidencian que el modelo no presenta sesgos sistemáticos, reforzando su aplicabilidad práctica.

Figura 29.

Valores medidos vs predichos para el Modelo Lnvisc @ 45°C menores a 1000 cP.

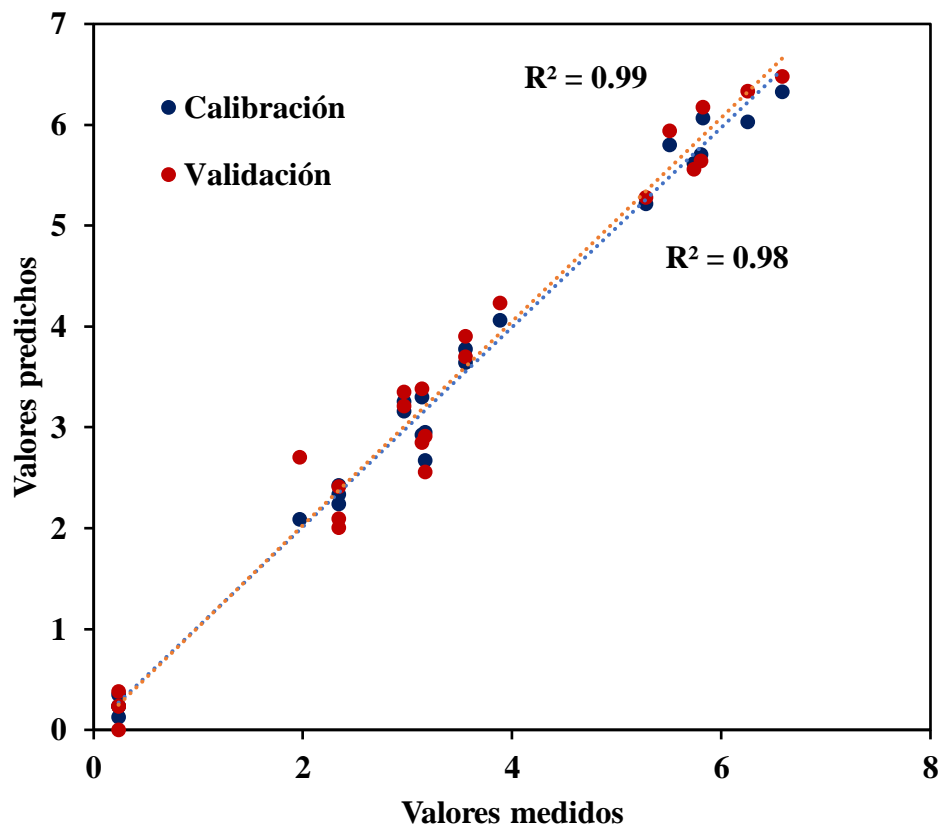
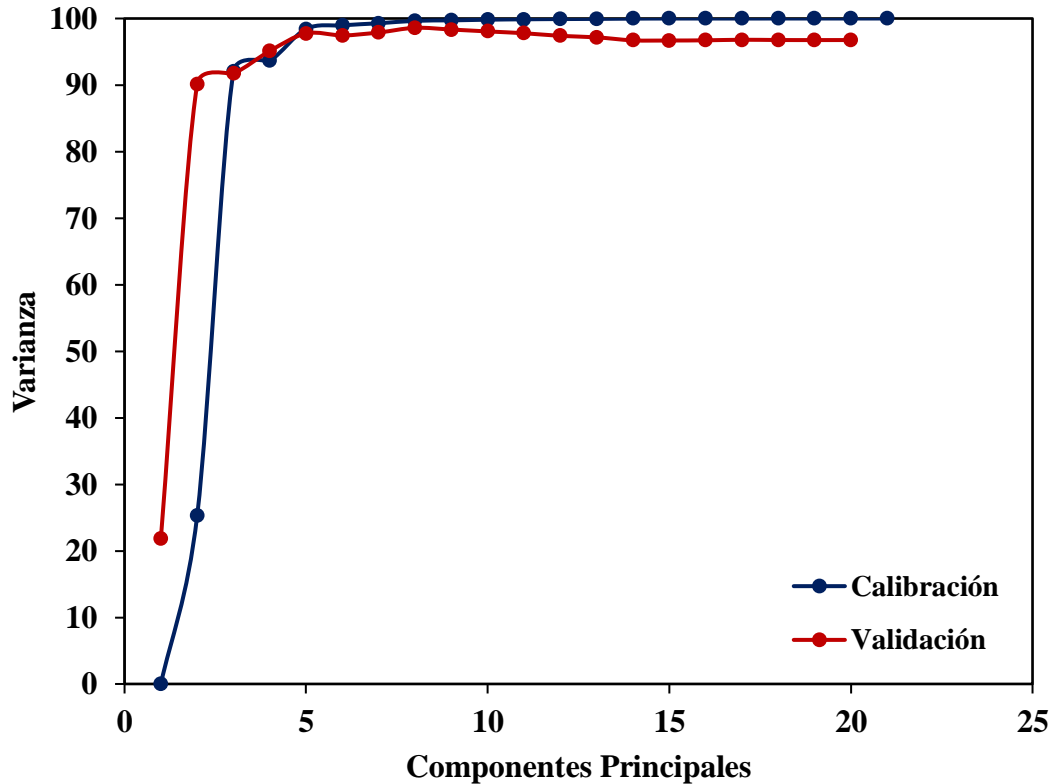


Figura 30.*Componentes principales modelo escogido*

Nota: Podemos observar que la varianza explicada en la figura 30, concentra un porcentaje elevado en X (superior al 80%), mientras que la variabilidad capturada en Y se mantiene en un rango adecuado (15% y 20%), lo que indica que la información espectral es suficiente para describir la respuesta de viscosidad en este rango.

Modelo de viscosidad a 45 °C para muestras mayores a 1000 cP

Para este modelo, se tiene en cuenta el espectro 4, para esto se cuenta con la sustracción de 5 muestras con las cuales se alcanza un R^2 bastante significativo dado a que la variación de viscosidad en este punto es visiblemente variable.

A partir del desarrollo del modelo, se muestran resultados en la tabla 39.

Tabla 39.

Modelo Lnvisc @ 45°C para viscosidades mayores a 1000 cP

Normalización por área espectro 4.					
Muestras eliminadas	R ² cal	R ² val	RMSEcal	RMSEval	CP
-	0.975	0.781	0.292	0.812	12
5	0.99	0.959	0.068	0.3206	13

Nota: Sin eliminar muestras: el modelo parece ajustarse bien a los datos de calibración con un $R^2_{cal} = 0,975$, pero su capacidad de validación es débil con un $R^2_{val} = 0,781$ y nos presenta un error elevado $RMSE_{val} = 0,812$. Esto indica sobreajuste que quiere decir que el modelo “aprende” los datos de entrenamiento, pero falla al predecir nuevos casos, eliminando las 5 muestras, el ajuste mejora tanto en calibración como en validación. El modelo logra un $R^2_{val} = 0,959$ el cual es mucho más alto, con errores reducidos $RMSE_{val} = 0,321$. Además, el R^2_{cal} aumenta a 0,99, lo que confirma una mejor consistencia en el modelo.

El modelo alcanzó un R2 de calibración de 0.998 y un R2 de validación de 0.953 como se puede observar en la figura 31, lo que nos determina una correlación fuerte entre los valores medidos y los estimados. Los errores cuadráticos medios fueron reducidos (0.083 en calibración y 0.378 en validación), lo que representa una incertidumbre baja en relación con el rango de viscosidad evaluado.

El modelo PLS de viscosidad a 45°C para valores superiores a 1000cP observado en la figura 31, se puede considerar confiable, robusto y aplicable en contextos de laboratorio e

industriales. A pesar de la ligera dispersión observada en las segundas componentes principales, el modelo mantiene una capacidad predictiva elevada, se convierte en una herramienta útil para la estimación no destructiva de viscosidad en fluidos de alta densidad.

Figura 31.

Modelo Lnvisc @ 45°C > 1000 cP

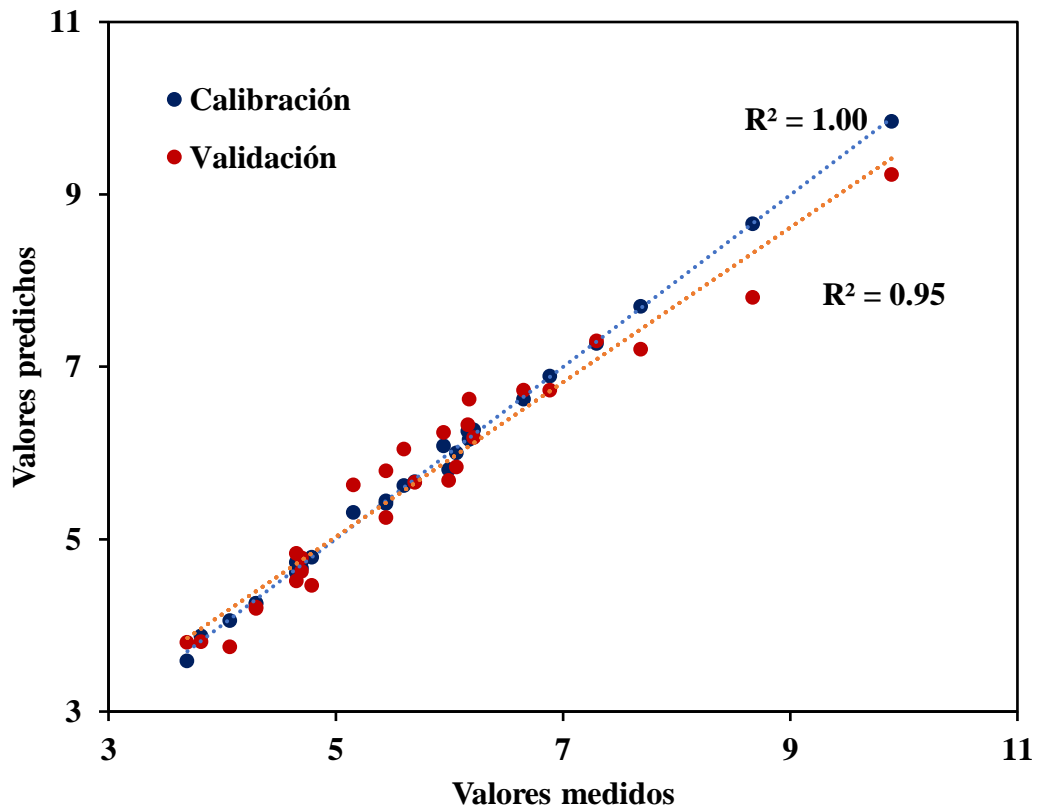
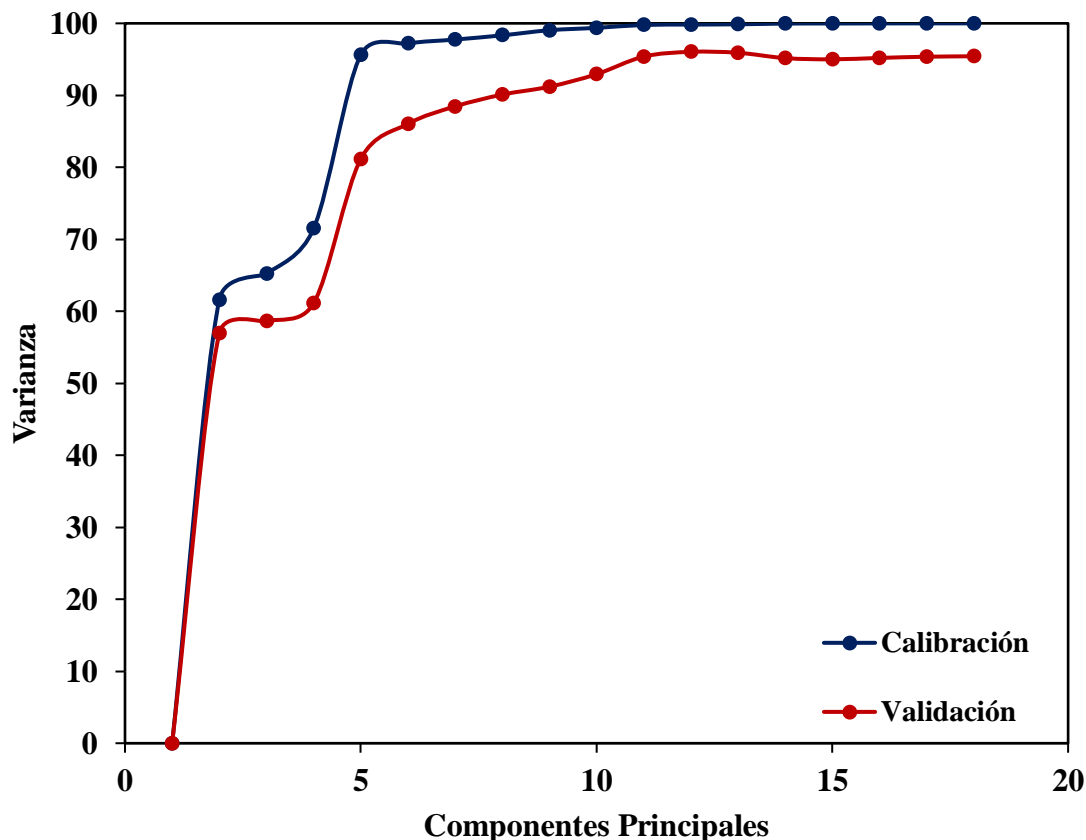


Figura 32.*Componentes principales modelo escogido*

La varianza analizada en la Figura 32, alcanzó un 88,9% en X y un 63,2% en Y, lo cual confirma que los espectros contienen la mayor parte de la información relevante para describir la propiedad de interés, aun cuando la viscosidad en este rango presenta una elevada complejidad molecular.

Para este modelo de viscosidad a 60 °C para muestras menores a 1000 Cp, se tiene en cuenta el espectro 4, para esto se cuenta con la sustracción de 7 muestras con las cuales se alcanza un R^2 con mayor robustez como se indica en la tabla 40

Tabla 40.*Modelo Lnvisc @ 60°C < 1000 cP*

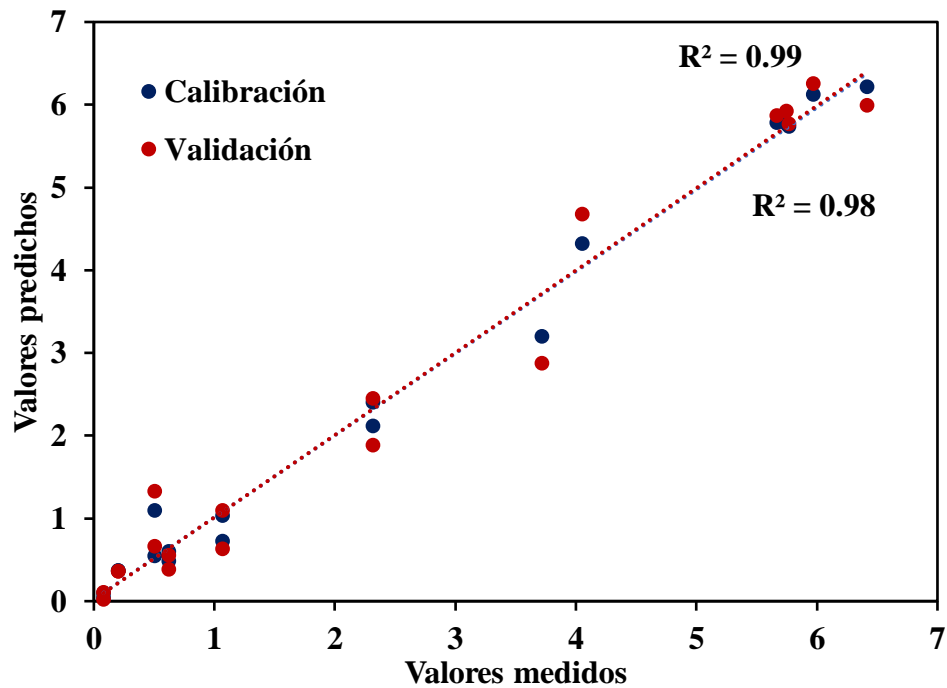
Normalización por área espectro 4.					
Muestras eliminadas	R ² cal	R ² val	RMSEcal	RMSEval	CP
-	0.933	0.876	0.628	0.891	4
7	0.991	0.978	0.225	0.364	5

Nota: Al analizar los resultados se observa que, cuando se incluyen todas las muestras, el modelo tiene un ajuste aceptable, pero los errores son altos y la capacidad predictiva se ve limitada. Al eliminar las muestras, los valores de R² mejoran notablemente tanto en calibración como en validación, y los errores RMSE disminuyen de forma importante, esto quiere decir que las muestras estaban actuando como un punto atípico que afectaba el desempeño del modelo. Su exclusión permitió obtener un modelo mucho más confiable y preciso, con una mejor correlación y menor error en la predicción de la viscosidad de los crudos pesados. En otras palabras, al limpiar los datos de valores que distorsionan la tendencia general, el modelo se vuelve más robusto y representativo de la realidad.

El desempeño estadístico mostrado en la Figura 33, el modelo alcanzó un R² de calibración de 0.99 y un R² de validación de 0.98, lo cual refleja una correlación muy fuerte entre los valores experimentales y los estimados. Los errores de predicción fueron bajos RMSE: 0.225 en calibración y 0.364 en validación, con pendientes próximas a la unidad y offsets mínimos, lo que confirma la ausencia de sesgos sistemáticos relevantes.

Figura 33.

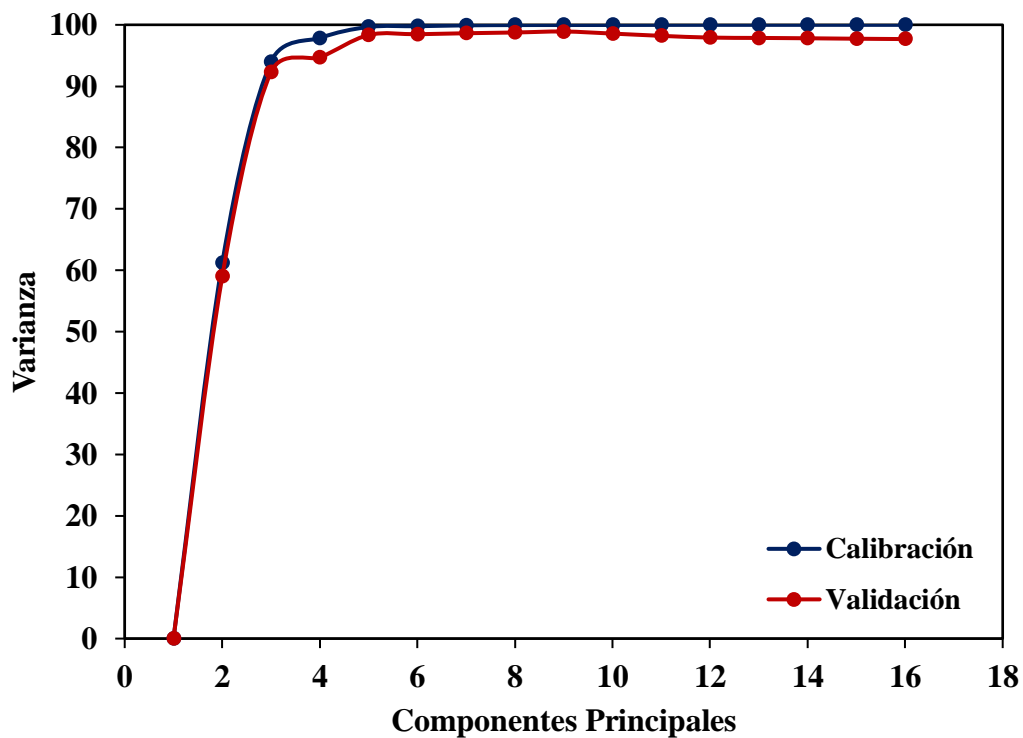
Modelo Lnvisc @ 60°C < 1000 cP.



Nota: En términos de varianza, el modelo explica que un 93,2% de la variabilidad en X y un 61,3% en Y, lo que demuestra la información contenida en los espectros es suficiente para describir de forma adecuada la respuesta de viscosidad en este rango como se observa en las figura 33 y 34.

Figura 34.

Componentes principales modelo escogido



Por el contrario, para el Modelo de viscosidad a 60 °C para muestras mayores a 1000 cP, se tiene en cuenta el espectro 2, en este caso la totalidad de muestras que se removieron respecto a la original fueron de tan solo 4, con las cuales se alcanza un R^2 óptimo.

A partir del desarrollo del modelo, se muestran resultados en la tabla 41.

Tabla 41.*Modelo Lnvisc @ 60°C > 1000 cP*

Normalización por área espectro 4.					
Muestras eliminadas	R ² cal	R ² val	RMSEcal	RMSEval	CP
-	0.982	0.781	0.22	0.812	12
4	0.997	0.959	0.0681	0.3206	13

Nota: En la Tabla 41 muestra un modelo inicial, que, sin eliminar muestras, presenta un buen ajuste en calibración, pero una validación baja, esta, siendo 0.781 llevando a tener un error de predicción relativamente alto $RMSE_{val} = 0.812$. Al sacar las 4 muestras, el desempeño del modelo no da una mejora de manera significativa en donde el coeficiente de determinación en validación aumenta a 0.959 y el error de predicción baja a 0.3206, Esto indica que estas 4 muestras actuaban como un valor atípico que afectaba la robustez del modelo y su respectiva eliminación permitió construir un modelo mucho más confiable, con alta capacidad de predicción y menor error, lo que refuerza la importancia de identificar y tratar datos que distorsionan el ajuste estadístico.

Respecto a la capacidad predictiva, el modelo alcanzó un R^2 de calibración de 0,998 y un R^2 de validación de 0,955 observado en la Figura 36, lo que demuestra una fuerte correlación entre los valores medidos y los estimados. Los errores de predicción fueron bajos en relación con el rango analizado ($RMSE = 0,068$ en calibración y 0,320 en validación), con pendientes cercanas a la unidad (0,998 y 0,897) y offsets mínimos en calibración. Cabe resaltar que, aunque en validación

la pendiente es menor (0,897), lo cual sugiere una ligera tendencia a subestimar viscosidades altas, el modelo conserva un desempeño sólido y aplicable.

Respecto a la capacidad predictiva, el modelo alcanzó un R^2 de calibración de 0,998 y un R^2 de validación de 0,95 observado en la Figura 35, lo que demuestra una fuerte correlación entre los valores medidos y los estimados. Los errores de predicción fueron bajos en relación con el rango analizado (RMSE = 0,068 en calibración y 0,320 en validación), con pendientes cercanas a la unidad (0,998 y 0,897) y offsets mínimos en calibración. Cabe resaltar que, aunque en validación la pendiente es menor (0,897), lo cual sugiere una ligera tendencia a subestimar viscosidades altas, el modelo conserva un desempeño sólido y aplicable.

Figura 35.

Modelo Ln visc @ 60°C > 1000 cP.

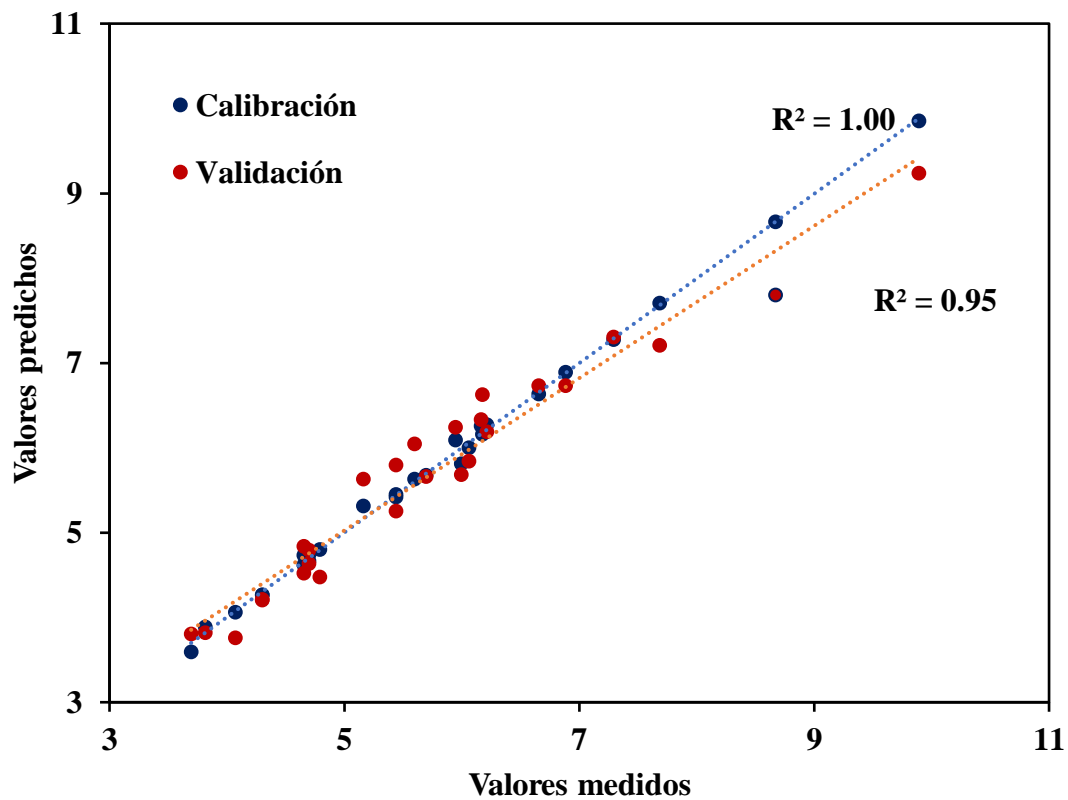
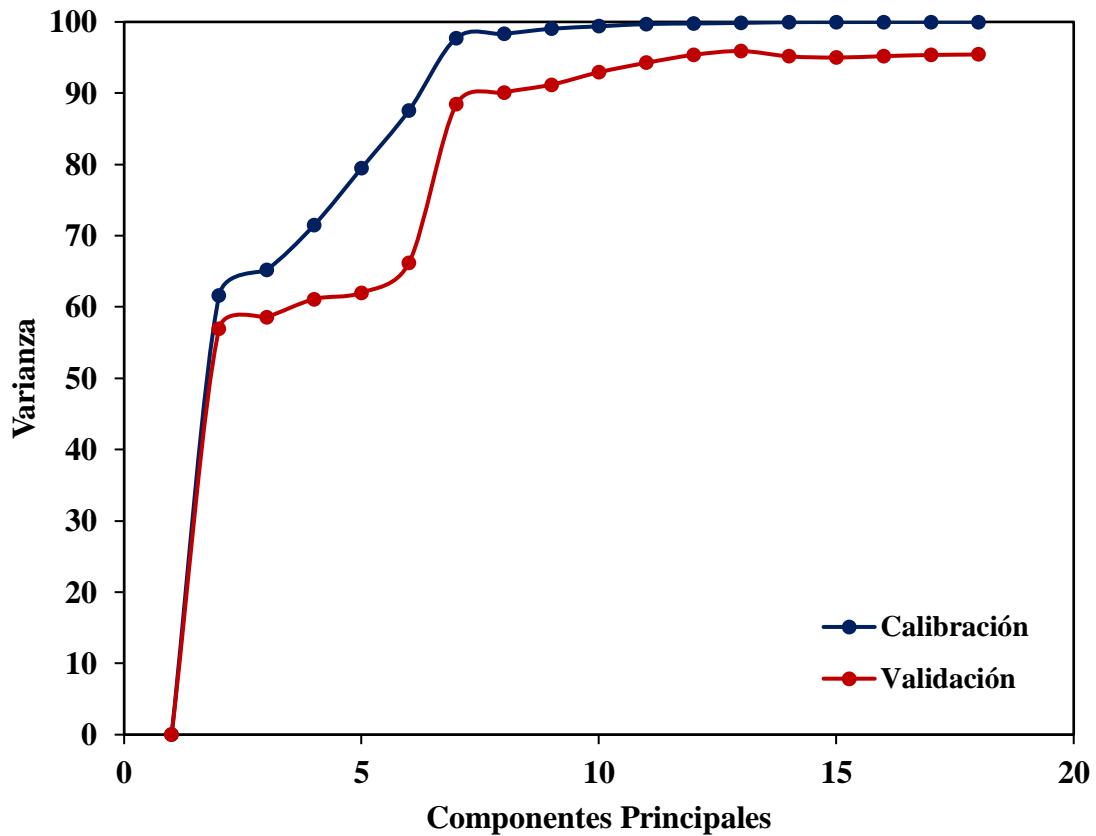


Figura 36.*Componentes principales modelo escogido*

En cuanto a la explicación de la varianza (figura 36), el modelo logró capturar un 86,1 % de la variabilidad en X y un 62,4 % en Y, lo cual refleja que la información espectral contiene patrones suficientes para describir la propiedad de viscosidad en este rango como lo muestran las figuras 35 y 36, también los coeficientes de regresión presentaron picos definidos en diferentes regiones espectrales, lo que evidencia la contribución de múltiples bandas a la predicción, consistente con la complejidad estructural de los fluidos altamente viscosos.

5 Análisis de Resultados

En primer lugar, el modelo de asfaltenos presentó un excelente ajuste tanto en los conjuntos de calibración como de validación, alcanzado un R^2 cercanos a 0.9 y errores de valores RMSE mínimos. Este modelo mostró alta sensibilidad frente a las regiones espectrales asociada a enlaces aromáticos y grupos funcionales representativos de compuestos pesados del crudo, especialmente en el rango de 1600 a 870 cm^{-1} . Los modelos diferenciados para concentraciones mayores y menores a 5% en peso confirmaron la estabilidad predictiva del modelo base, el cual logró equilibrar la precisión estadística y la complejidad computacional, destacándose como el más representativo para la estimación del contenido de asfaltenos en crudos mejorados.

Por su parte, el modelo de densidad reflejó un comportamiento igualmente robusto, con coeficientes de determinación entre 0.98 y 0.99, y un bajo errores cuadrático medio (RMSE <0.001). este modelo demostró una correlación directa entre la absorción espectral y la longitud de las cadenas hidrocarbonadas, particularmente en las bandas comprendidas entre 2920 y 2850 cm^{-1} , correspondientes a los estiramientos C-H alifáticos. Entre las diferentes configuraciones ensayadas, la corrección de línea base ofreció el mejor desempeño general, garantizando estabilidad, simplicidad y reproducibilidad en las predicciones. De este modo, el modelo de densidad se consolida como una herramienta confiable para la estimación no destructiva de esta propiedad en sistemas complejos de crudo.

Finalmente, el modelo de viscosidad permitió establecer relaciones directas entre la estructura molecular y el comportamiento reológico del crudo, mostrando resultados altamente satisfactorios en los tres rangos de temperatura analizados (30, 45, 60°C). Los valores de R^2 oscilaron entre 0.95 y 0.99, con errores de RMSE reducidos, confirmando la coherencia interna del modelo. Se observó una mayor precisión para viscosidades inferiores a 1000Cp, aunque los

modelos correspondientes a viscosidades más elevadas también mantuvieron un desempeño aceptable, con valores de R^2 altos. Además, se verificó la influencia significativa de los procesos de acuatermolisis catalítica sobre la reducción de la viscosidad, lo que demuestra la capacidad del modelo para capturar variaciones fisicoquímicas asociadas a transformaciones térmicas y catalíticas.

En conjunto, los tres modelos planteados consolidan una metodología predictiva eficiente, rápida y no destructiva para la caracterización de crudos pesados y mejorados. La combinación de espectroscopia FTIR-ATR y análisis multivariado PLS permitió obtener modelos con alta capacidad de generalización, reduciendo los tiempos de análisis y los requerimientos de laboratorio en comparación con las técnicas convencionales. De esta manera, se evidencia el potencial de las herramientas quimiométricas como soporte analítico en la optimización de procesos térmicos y en la evaluación de propiedades clave de los hidrocarburos.

6 Conclusiones

Se desarrollaron con éxito modelos multivariados predictivos mediante el uso de espectroscopia FTIR-ATR y técnicas quimiométricas, específicamente regresión lineal múltiple y PLS. Estos modelos demostraron alta capacidad predictiva con $R^2 > 0.98$ para estimar propiedades como el contenido de asfaltenos y la densidad que aunque los modelos resultantes y escogidos fueron modelos en donde los espectros no obtuvieron tratamiento, estos dieron resultados casi semejantes a aquellos que si lo tuvieron, lo anterior concluye a partir de la validación de la viabilidad de esta metodología como una herramienta rápida, precisa y no destructiva para la caracterización de crudos mejorados.

La investigación concluye con la combinación de la espectroscopia FTIR-ATR y métodos quimiométricos, en donde se pueden construir modelos de manera robusta y eficiente para el análisis predictivo de crudos, la técnica mostró una buena repetibilidad y sensibilidad para la cuantificación simultánea de múltiples propiedades fisicoquímicas a partir de un único espectro. Esto valida su potencial como una alternativa rápida y confiable frente a los métodos de caracterización convencionales, más lentos y demandantes de recursos.

7 Recomendaciones

Teniendo en cuenta las futuras investigaciones relacionadas en este tema, los autores de este documento recomiendan tener en cuenta lo siguiente:

Ampliar la caracterización para un mayor número de muestras de crudos mejorados y no mejorados, para así poder tener más robustez y una alta convergencia en los modelos de caracterización de propiedades fisicoquímicas de crudos de diferente naturaleza.

Analizar y caracterizar los crudos, mediante análisis SARA y reproducir su modelamiento quimiométrico por FTIR ATR en donde puedan someterse a crudos extrapesados y livianos para obtener una amplia gama de características fisicoquímicas.

Los modelos PLS de asfaltenos mostraron sensibilidad en regiones espectrales asociadas a enlaces aromáticos. Se recomienda incluir mayor diversidad de crudos con variaciones significativas en fracciones SARA, lo que permitiría un modelo más representativo y aplicable a diferentes tipos de yacimientos colombianos.

Referencias Bibliográficas

- Aldana-Jiménez, K., Comesaña-García, Y., & Alexis Ochoa-Valladares, F. (2018). Aplicación de la espectroscopia infrarroja y técnicas de reconocimiento de patrones en la caracterización de aditivos de la industria del petróleo. *Revista CENIC. Ciencias Químicas*, 49(1), 5–10.
<https://www.redalyc.org/articulo.oa?>
- Arboleda, J., Castillo, Á., & Muñoz, S. (2018). Estudio de la acuatermólisis catalítica en procesos de upgrading de crudos pesados como método complementario en el recobro térmico de hidrocarburos. *Revista Fuentes El Reventón Energético*, 16(1), 57–69.
<https://doi.org/10.18273/revfue.v16n1-2018005>
- Aydin, H., Nagabandi, N., Temizel, C., & Jamal, D. (2022). Heavy Oil Reservoir Management - Latest Technologies and Workflows. *SPE Western Regional Meeting Proceedings, 2022-April*.
<https://doi.org/10.2118/209328-MS>
- Bair, E., Hastie, T., Paul, D., & Tibshirani, R. (2006). Prediction by Supervised Principal Components. *Journal of the American Statistical Association*, 101(473), 119–137.
<https://doi.org/10.1198/0162145050000000628>
- Ballesteros-Lancheros, S. L., Orrego, J. A., Cabanzo, R., & Mejía-Ospino, E. (2011). Determinación De La Composición De Fondos De Vacío De Petróleos Colombianos Por Espectroscopia De Fluorescencia Inducida Por Láser (LIF). *Revista Colombiana de Física*, 43(3), 649–652.
https://www.researchgate.net/publication/279669317_Determinacion_De_La_Composicion_De_Fondos_De_Vacio_De_Petroleos_Colombianos_Por_Espectroscopia_De_Fluorescencia_Inducida_Por_Laser_LIF

- Guo, K., Li, H., & Yu, Z. (2016). In-situ heavy and extra-heavy oil recovery: A review. *Fuel*, 185, 886–902. <https://doi.org/10.1016/J.FUEL.2016.08.047>
- Guzmán, J., & Vega–Vilca, J. (2010). Regresion PLS y PCA como solución al problema de multicolinealidad en regresion múltiple . *Revista de Matematica: Teoría y Aplicaciones*, 11–17.
- Guzmán-Osorio, F. J. (2022). *DETERMINACIÓN DE TIPO Y CONCENTRACIÓN DE PETRÓLEO UTILIZANDO ESPECTROMETRÍA FTIR*. UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO.
- Hanzlik, E. J., & Mims, D. S. (2003). *Forty Years of Steam Injection in California – The Evolution of Heat Management*. <https://doi.org/10.2118/84848-MS>
- Hyne, J. B. (1986a). *Aquathermolysis: a synopsis of work on the chemical reaction between water (steam) and heavy oil sands during simulated steam stimulation*.
- Hyne, J. B. (1986b). *Aquathermolysis: a synopsis of work on the chemical reaction between water (steam) and heavy oil sands during simulated steam stimulation*.
- Lamus, C., Guzman, A., Murcia, B., Cabanzo, R., & Mejía-Ospino, E. (2011). Uso de Análisis Multivariado En La Determinación SARA De Crudos Por Espectroscopia NIR. *Revista Colombiana de Física*, 43(3), 638–641. https://www.researchgate.net/profile/Enrique-Ospino/publication/237045492_Uso_de_Analisis_Multivariado_En_La_Determinacion_SARA_De_Crudos_Por_Espectroscopia_NIR/links/0c96052eda028b4c88000000/Uso-de-Analisis-Multivariado-En-La-Determinacion-SARA-De-Crudos-Por-Espectroscopia-NIR.pdf
- Luo, E., Fan, Z., Hu, Y., Zhao, L., Bo, B., Yu, W., Liang, H., Liu, M., Liu, Y., He, C., & Wang, J. (2020). An efficient optimization framework of cyclic steam stimulation with experimental design

in extra heavy oil reservoirs. *Energy*, 192, 116601.

<https://doi.org/10.1016/J.ENERGY.2019.116601>

Maldonado, M. V. (2018). *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*

[Universidad de Pamplona].

http://repositoriodspace.unipamplona.edu.co:8080/jspui/bitstream/20.500.12744/2706/1/Vargas_2018_TG.pdf

Márquez, A., & Sánchez, J. D. (2024). *Análisis del efecto termo catalítico sobre la composición y las propiedades fisicoquímicas de crudos pesados colombianos*. [Universidad Industrial de

Santander]. <https://noesis.uis.edu.co/server/api/core/bitstreams/fe917531-378b-4db2-a567-7f16f35ff26b/content>

Mena Sánchez, C. U. (2021). *Modelos de clasificación y cuantificación quimiométrica de hidrocarburos fracción pesada durante la biorremediación de un histosol contaminado con*

petroleo crudo. [Colegio de Postgraduados, Campus Tabasco].

<https://www.sidalc.net/search/Record/oai:colposdigital.colpos.mx:10521-4783/Description>

Miller James N., & Miller Jane C. (2002). ESTADÍSTICA Y QUIMIOMETRÍA PARA QUÍMICA ANALÍTICA. In I. Capella & Clares José (Eds.), *ESTADÍSTICA Y QUIMIOMETRÍA PARA*

QUÍMICA ANALÍTICA (Cuarta edición, pp. 224–253). PEARSON EDUCACIÓN, S.A.

<https://sceqa.wordpress.com/wp-content/uploads/2012/05/quimica-y-quimiometria.pdf>

- Mogensen, K., & Masalmeh, S. (2020). A review of EOR techniques for carbonate reservoirs in challenging geological settings. *Journal of Petroleum Science and Engineering*, 195, 107889. <https://doi.org/10.1016/J.PETROL.2020.107889>
- Pichardo, D. (2011). *Caracterización de Hidrocarburos de Diferente Gravedad API Mediante Espectroscopía de Infrarrojo (FTIR)* [Instituto Politécnico Nacional]. <https://es.scribd.com/document/511444562/Caracterizacion-de-Hidrocarburos-de-Diferenrte-Gravedad-API-Mediante>
- Rodriguez C, S. (2014). PREDICCIÓN DE LA COMPOSICIÓN ELEMENTAL DE FRACCIONES DEL PROCESO DE DELAYED COKING DE FONDOS DE VACÍO DE CRUDOS COLOMBIANOS EMPLEANDO ESPECTROSCOPIA DE FLUORESCENCIA Y MÉTODOS QUIMIOMÉTRICOS. In *Universidad Industrial de Santander*. Universidad Industrial de Santander.
- Sáenz Paéz, A. Y. (2018). *AVANCES EN LA APLICACIÓN DE RESONANCIA MAGNÉTICA NUCLEAR-RMN DE BAJO Y ALTO CAMPO-Y QUIMIOMETRIA PARA CARACTERIZAR CRUDOS PESADOS Y FONDOS DE VACÍO*. UNIVERSIDAD NACIONAL ABIERTA Y A DISTANCIA.
- Sanabria, J. E. (2023). *Caracterización fisicoquímica de un crudo pesado sometido a acuatermólisis catalítica mediante métodos convencionales y técnicas analíticas* [Universidad Industrial de Santander]. <https://noesis.uis.edu.co/server/api/core/bitstreams/6bc0cad8-95b3-46a0-a55c-e3935cd5e533/content>

Speight, J. (2009). *Enhanced Recovery Methods for Heavy Oil and Tar Sands* - James G. Speight - Google Libros. Gulf Publishing Company.

https://books.google.com.co/books?hl=es&lr=&id=_bxAAQAAQBAJ&oi=fnd&pg=PP1&ots=ZWlqidnXk-&sig=Ii490764-uS6zASXjMEP5PX_g28&redir_esc=y#v=onepage&q&f=false

Tamiru Tefera, D., Agrawal, A., Maria Yan, L., Jaramillo, ez, de Klerk, A., & Prasad, V. (2017). *Self-Modeling Multivariate Curve Resolution Model for Online Monitoring of Bitumen Conversion Using Infrared Spectroscopy*. <https://doi.org/10.1021/acs.iecr.7b01849>

Thomas, S. (2008). Enhanced Oil Recovery-An Overview Molecular Structures of Heavy Oils and Coal Liquefaction Products Structure moléculaire des huiles lourdes et produits de liquéfaction du charbon. *Oil & Gas Science and Technology-Rev. IFP*, 63(1), 9–19. <https://doi.org/10.2516/ogst:2007060>

Vakhin, A. V., Aliev, F. A., Kudryashov, S. I., Afanasiev, I. S., Petrashov, O. V., Sitnov, S. A., Mukhamatdinov, I. I., Varfolomeev, M. A., & Nurgaliev, D. K. (2018). Aquathermolysis of heavy oil in reservoir conditions with the use of oil-soluble catalysts: part I—changes in composition of saturated hydrocarbons. *Petroleum Science and Technology*, 36(21), 1829–1836. <https://doi.org/10.1080/10916466.2018.1514411>

Apéndices

Apéndice A.

Modelos evaluados para asfaltenos > 5 (% peso)

Tabla 42.

Asfaltenos > 5% Primera regresión.

Parámetro estadístico	Asfaltenos primera regresión			
	Espectro1	Espectro 2	Espectro 3	Espectro 4
Rcal	0.988	0.997	0.976	0.988
Rval	0.888	0.904	0.834	0.912
RMSEcal	0.526	0.242	0.744	0.529
RMSEval	1.679	1.555	0.834	1.490
CP	6	7	7	5

Nota: En la tabla 42, se permite observar que todos los espectros muestran un ajuste aceptable, aunque con diferencias notorias en su desempeño, el espectro 2 es el que presenta el mejor comportamiento, ya que combina el valor de correlación más alto Rcal de 0.9974 y Rval de 0.9043 con los errores de calibración y validación más bajos respecto a los demás, lo que refleja una mayor capacidad de predicción y estabilidad.

Tabla 43.

Asfaltenos > 5% Análisis Final.

	Asfaltenos última regresión			
	Espectro 1	Espectro 2	Espectro 3	Espectro 4
Rcal	0.997	0.987	0.999	0.996
Rval	0.953	0.974	0.990	0.946

RMSEcal	0.208	0.484	0.074	0.449
RMSEval	0.953	0.974	0.502	1.141
CP	6	3	10	4

Nota: La tabla 43, muestra la comparación de los cuatro espectros, allí todos muestran buen desempeño en calibración, pero el espectro 3 se destaca al alcanzar el mayor ajuste con Rcal y Rval más altos y el menor error de calibración. Aunque requiere un número mayor de componentes principales (10) y aunque su error de validación no es el más bajo, sigue siendo el modelo más representativo y robusto en este conjunto de datos. Esto indica que, para la última regresión de asfaltenos, el Espectro 3 es la opción más confiable para describir la variabilidad de la muestra, aun a costa de mayor complejidad.

Ahora, hacemos correcciones a los espectros para tener una mejora, empezando por la línea base y seguido a esta hacemos la normalización a cada uno de los espectros, sin suprimir datos, los resultados de este análisis son mostrados.

Tabla 44.

Asfaltenos > 5% línea base primera regresión.

Parámetro estadístico	Asfaltenos línea base primera regresión			
	Espectro 1	Espectro 2	Espectro 3	Espectro 4
Rcal	0.988	0.995	0.972	0,990
Rval	0.907	0.888	0.862	0.918
RMSEcal	0.569	0.327	0.972	0.488
RMSEval	1.537	1.681	0.862	1.442

CP	5	7	6	6
-----------	---	---	---	---

Nota: En la regresión aplicada en la tabla 44, se observa que los modelos presentan buenos niveles de ajuste, aunque con variaciones importantes en su desempeño. El espectro 2 alcanza el mayor valor de correlación en calibración y el menor error de calibración. Sin embargo, en validación pierde estabilidad con un $R_{val} = 0.888$ y muestra el $RMSE_{val}$ más alto de 1.681, lo que evidencia un sobreajuste, a su vez, el espectro 1 y el espectro 4 ofrecen un comportamiento más balanceado, con buenos valores de R_{val} , lo que los hace más confiables que el Espectro 2 en términos de robustez. Por su parte, el Espectro 3 presenta el ajuste más débil, con la menor correlación en validación y un RMSE relativamente alto tanto en calibración como en validación.

Luego de este análisis, se calculan modelos descartando entre 3 y 4 muestras en cada uno, donde se obtienen los siguientes datos mostrados en la tabla 30.

Tabla 45.

Asfaltenos > 5% línea base última regresión.

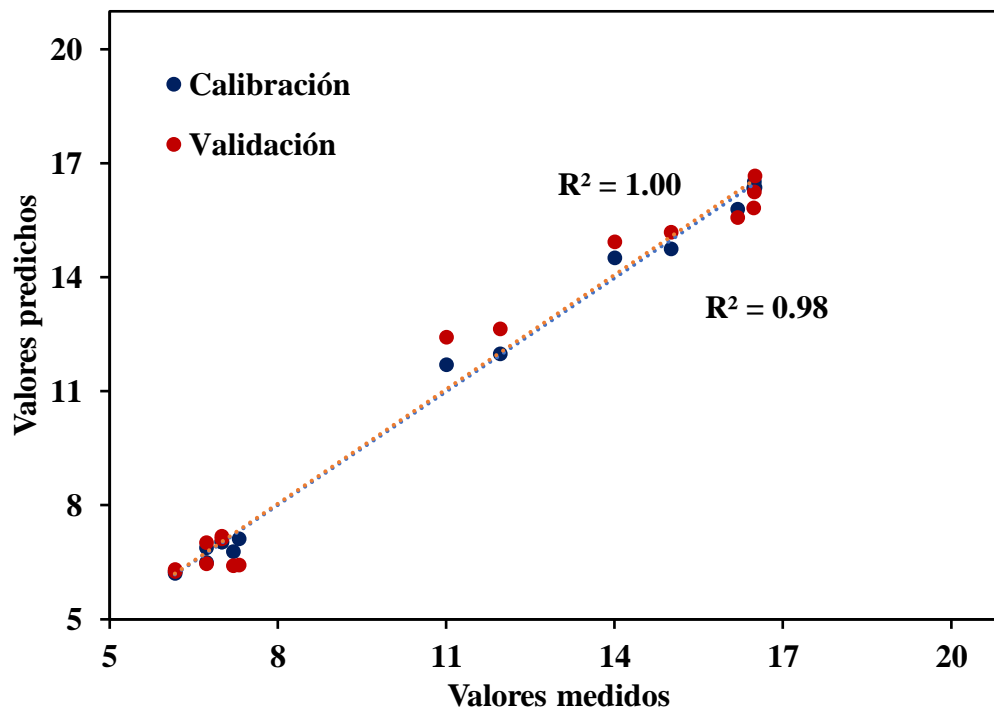
Parámetro estadístico	Asfaltenos línea base última regresión			
	Espectro 1	Espectro 2	Espectro 3	Espectro 4
R_{cal}	0.997	0.997	0.998	0.995
R_{val}	0.955	0.961	0.969	0.982
RMSE_{cal}	0.2915	0.262	0.191	0.286
RMSE_{val}	1.092	0.961	0.904	0.608

CP	6	6	8	8
-----------	---	---	---	---

Nota: En la tabla 45, los datos obtenidos muestran que todos los modelos presentan un ajuste sobresaliente en la etapa de calibración, con coeficientes R_{cal} superiores a 0.99 en los cuatro espectros evaluados. En cuanto a la validación, el mejor desempeño se observa en el Espectro 4, con un R_{val} de 0.982 y un $RMSE_{val}$ de 0.608, lo que indica un modelo robusto y con adecuada capacidad predictiva. Aunque los espectros 1, 2 y 3 también evidencian una buena capacidad de predicción, ellos presentan errores de validación ligeramente superiores. Por lo tanto, el espectro 4 se perfila como la opción más confiable para la estimación de asfaltenos, al combinar alta correlación en validación con un número adecuado de componentes principales ($CP = 8$), lo que garantiza precisión y estabilidad en el modelo.

Figura 37.

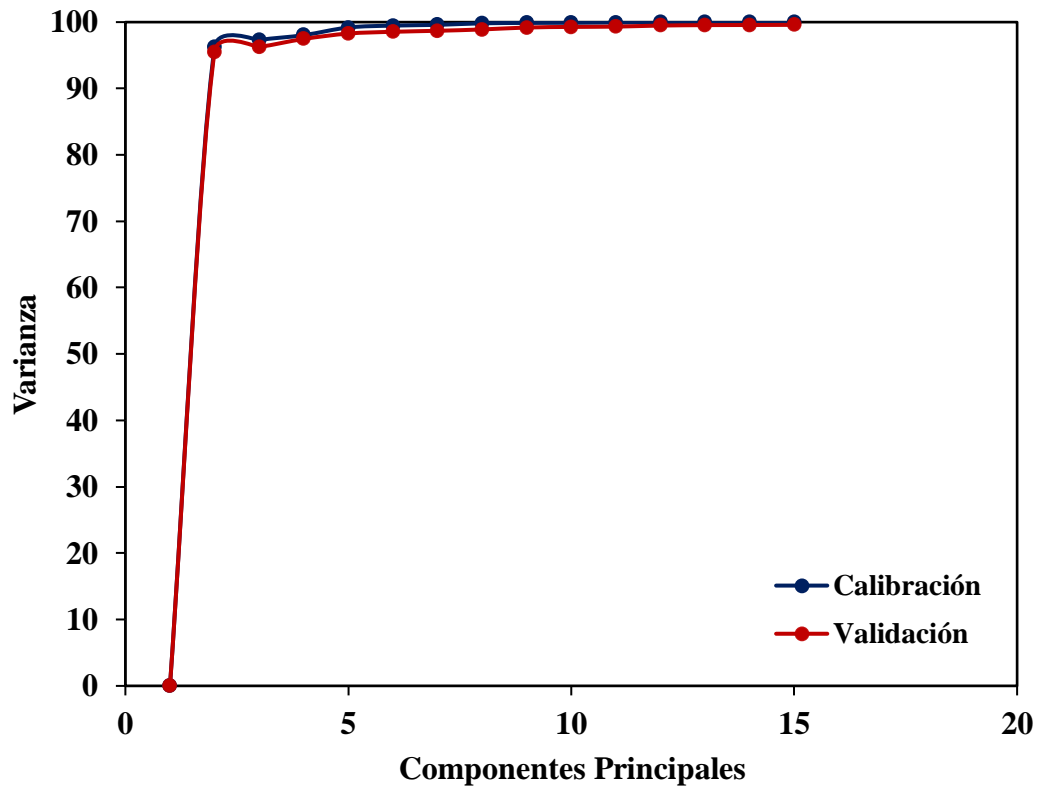
Valores medidos vs predichos para el modelo de contenido de asfaltenos (% peso).



Nota: La figura 37, refleja la selección del modelo ya que presenta un ajuste sobresaliente entre valores medidos y predichos de asfaltenos, con coeficientes de determinación superiores al 98 %. Las pendientes cercanas a la unidad y los interceptos próximos a cero evidencian ausencia de sesgo sistemático, mientras que los valores de RMSE reflejan un error de predicción bajo, especialmente en el modelo de calibración de color azul. En conjunto, los resultados confirman la alta confiabilidad del modelo, con ligeras desviaciones únicamente en el rango intermedio de concentraciones.

Figura 38.

Cantidad Componentes Principales Modelo escogido, línea base.



Nota: La figura 38 de varianza explicada, evidencia que entre el primer componente principal y el segundo concentran por sí solos cerca del 98 % de la variabilidad total del sistema, mientras que con el tercer componente se aproxima al 100 %. A partir del cuarto componente, la contribución marginal es casi nulo e imperceptible, lo que confirma que el modelo puede representarse de manera robusta y eficiente con únicamente 4 componentes principales. Esto implica una reducción sustancial de la dimensionalidad, conservando casi toda la información original y facilitando tanto la interpretación de los resultados como la construcción de modelos predictivos confiables.

Tabla 46.*Asfaltenos > 5% normalizado primera regresión.*

Parámetro estadístico	Asfaltenos normalizado primera regresión			
	Espectro 1	Espectro 2	Espectro 3	Espectro 4
Rcal	0.959	0.993	0.998	0.992
Rval	0.839	0.872	0.952	0.883
RMSEcal	0.959	0.399	1,121	0.431
RMSEval	2.014	1,799	2,447	1.172
CP	4	7	5	6

Nota: La tabla 46 muestra resultados de los cuales el espectro 1 tiene el mejor ajuste en la calibración junto con valores altos de Rcal, por otra parte, el espectro 3 presenta el mayor error en validación RMSEval, el espectro 2, aunque tiene un excelente ajuste en calibración, no generaliza tan bien a nuevos datos, el espectro 4, aunque tiene un buen desempeño en calibración, no es el mejor en términos de validación. En general, los espectros 1 y 3 destacan, pero la elección del modelo ideal depende de la importancia que se le dé al ajuste en calibración versus la capacidad de generalización a nuevos datos. Luego de ese análisis y de no obtener un buen resultado en la regresión, se procede a remover algunas muestras con el fin de optimizar el modelo, dando como resultado la tabla 47.

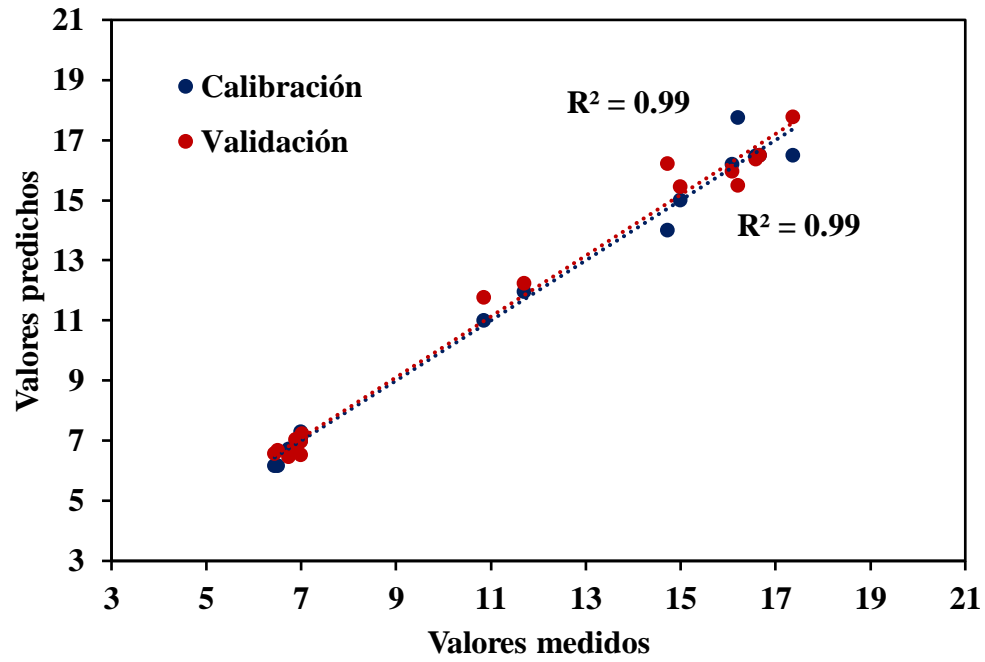
Tabla 46.*Asfaltenos >5% normalizado última regresión*

Parámetro estadístico	Asfaltenos normalizado última regresión			
	Espectro 1	Espectro 2	Espectro 3	Espectro 4
Rcal	0.995	0.992	0.998	0.987
Rval	0.938	0.930	0.952	0.987
RMSEcal	0.349	0.436	0.238	0.495
RMSEval	1,281	1,371	1,131	0.901
CP	7	6	10	5

Nota: En los modelos obtenidos con datos normalizados de la tabla 47, se observa un desempeño consistente en la calibración, con coeficientes Rcal muy altos que están entre 0.987 y 0.998, lo que indica un buen ajuste a los datos experimentales. En la etapa de validación, el Espectro 4 presenta el mejor balance, con un Rval de 0.987 y el menor RMSEval de 0.901, confirmando una adecuada capacidad predictiva con solo 5 componentes principales, lo cual lo hace más eficiente en comparación con los demás. Si bien el Espectro 3 alcanza el Rcal más alto de 0.998 y un buen Rval de 0.952, requiere un mayor número de componentes, lo que podría aumentar la complejidad del modelo. Siendo de esta manera, el espectro 4 es el más confiable y eficiente para la predicción de asfaltenos normalizados, al combinar buena correlación de validación con bajo error y simplicidad en el número de componentes.

Figura 39.

Valores medidos vs predichos para el modelo normalizado de Asfaltenos >5%.

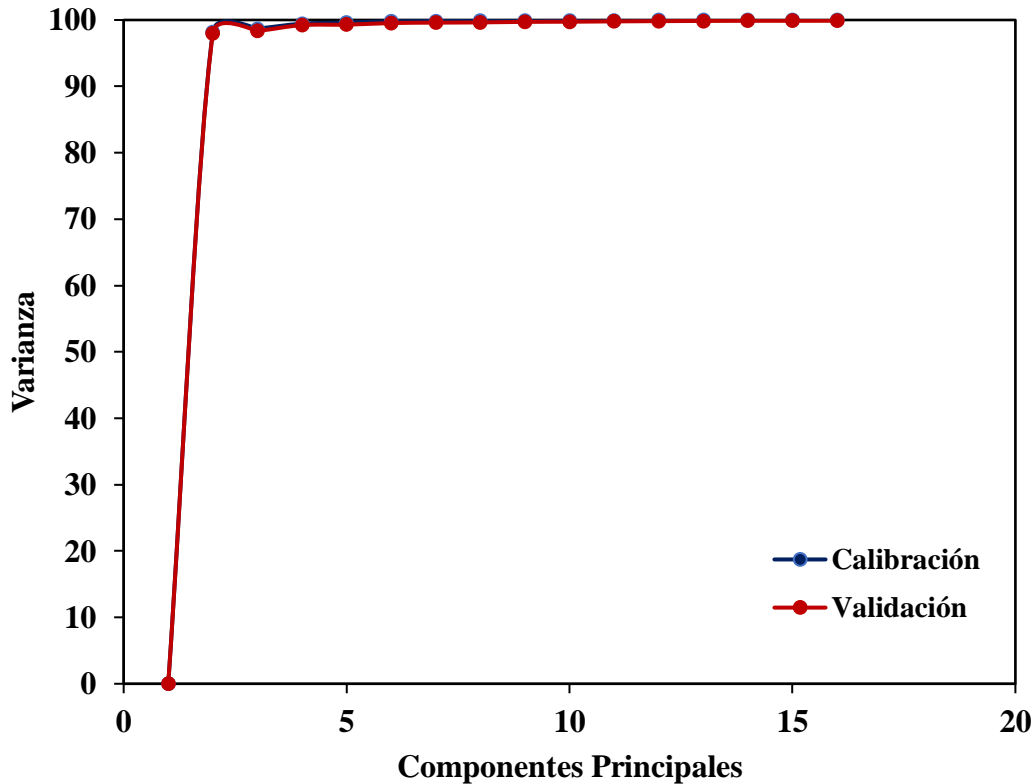


Nota: La figura 39, evidencia un ajuste lineal robusto para calibración y validación (azul y rojo) analizados respectivamente, ambos modelos presentan pendientes cercanas a la unidad de un 0.987 y 0.992 respectivamente, lo que indica una relación casi perfecta entre los valores predichos y reales, los coeficientes de determinación R^2 de 0.987 y 0.963 confirman un alto grado de precisión en la predicción, mientras que los valores bajos de RMSE de 0.495 y 0.301 reflejan errores mínimos en las estimaciones.

En conjunto, estos resultados indican que el modelo posee una mejor capacidad predictiva y estabilidad, siendo adecuado para su aplicación en la cuantificación de asfaltenos bajo las condiciones evaluadas.

Figura 40.

Cantidad Componentes Principales Modelo escogido, normalizado.



Nota: La figura 40 de varianza explicada, muestra que los dos primeros componentes principales PC1 y PC2 los cuales explican alrededor del 100% de la varianza total del conjunto de datos, lo cual es evidente por la rápida estabilización de la curva a partir del segundo componente, esto indica que la mayoría de la información contenida en las variables originales puede ser representada hasta con solo dos componentes principales, reduciendo significativamente la dimensionalidad sin pérdida relevante de información. En consecuencia, el modelo se beneficia de una mayor simplicidad, menor riesgo de sobreajuste y una interpretación más eficiente.

Apéndice B.

Modelos evaluados para asfaltenos < 5 (% peso)

Tabla 47.*Asfaltenos < 5%. base primera regresión.*

Parámetro estadístico	Asfaltenos base primera regresión.			
	Espectro 1	Espectro 2	Espectro 3	Espectro 4
Rcal	0.941	0.9434	0.927	0.939
Rval	0.923	0.921	0.904	0.918
RMSEcal	0.353	0.343	0.389	0.359
RMSEval	0.415	0.421	0.464	0.427
CP	3	3	3	3

Tabla 48.*Asfaltenos < 5%. base última regresión.*

Parámetro estadístico	Asfaltenos última regresión			
	Espectro 1	Espectro 2	Espectro 3	Espectro 4
Rcal	0.980	0.989	0.989 _ç	0.989
Rval	0.975	0.984	0.981 _ç	0.984
RMSEval	0.184	0.138	0.127 _ç	0.138
RMSEcal	0.217	0.174	0.176 _ç	0.174
CP	3	4	6	5

Nota: En la tabla 48 se puede observar que los coeficientes de correlación en calibración R_{cal} se ubicaron en 0.927 y 0.943, mientras que en validación R_{val} oscilaron entre 0.904 y 0.923. Posteriormente al hacer una depuración de 3 muestras equivalentes a menos del 10% de las muestras, se obtuvo una mejora significativa en la calidad de los modelos, alcanzando valores de R_{cal} entre 0.980 y 0.989 y R_{val} entre 0.975 y 0.984. de esta manera podemos confirmar que la eliminación de una cierta cantidad de componentes permitió incrementar de forma sustancial la robustez predictiva de los modelos, manteniendo un bajo número de componentes principales.

Pasando a los modelos calculados haciendo la respectiva normalización y por último línea base de los espectros tenemos las tablas

Tabla 49.

Asfaltenos normalizado primera regresión < 5%.

Asfaltenos normalizados primera regresión				
	Espectro 1	Espectro 2	Espectro 3	Espectro 4
Rcal	0.962	0.907	0.949	0.969
Rval	0.918	0.879	0.928	0.919
RMSEcal	0.282	0.440	0.203	0.252
RMSEval	0.429	0.519	0.253	0.426
CP	5	3	5	5

Tabla 50.*Asfaltenos normalizado última regresión < 5%.*

Asfaltenos normalizado última regresión				
	Espectro 1	Espectro 2	Espectro 3	Espectro 4
Rcal	0.985	0.994	0.979	0.990
Rval	0.985	0.981	0.969	0.982
RMSEcal	0.164	0.107	0.191	0.130
RMSEval	0.232	0.190	0.239	0.182
CP	5	6	5	5

Nota: La tabla 50 evidencia el desempeño de dichos modelos tras aplicar la normalización, mostrando coeficientes de correlaciones muy altos Rcal hasta 0.994 y Rval hasta 0.98) y unos errores bastante mínimos teniendo un RMSEcal de 0.10 - 0.19 y un RMSEval de 0.18 – 0.23 con un número reducido de componentes principales entre 5 a 6, confirmando que la selección adecuada de bandas espectrales mostrada en la tabla 27, que al ser combinada con un pretratamiento efectivo como la normalización vista en la tabla 2, permite maximizar la capacidad predictiva y robustez de modelos para muestras con bajo contenido de asfaltenos.

La figura 41 muestra la dispersión entre valores predichos y medidos tras la normalización de los espectros para asfaltenos < 5%. Se puede observar un alineamiento muy cercano a la recta de identidad, con baja dispersión de los puntos, lo que confirma un ajuste robusto del modelo PLS. Esto respalda la eficacia de la normalización para mejorar la precisión predictiva en concentraciones bajas de asfaltenos.

Figura 41.

Valores medidos vs predichos para el modelo de Asfaltenos < 5% en peso normalizado.

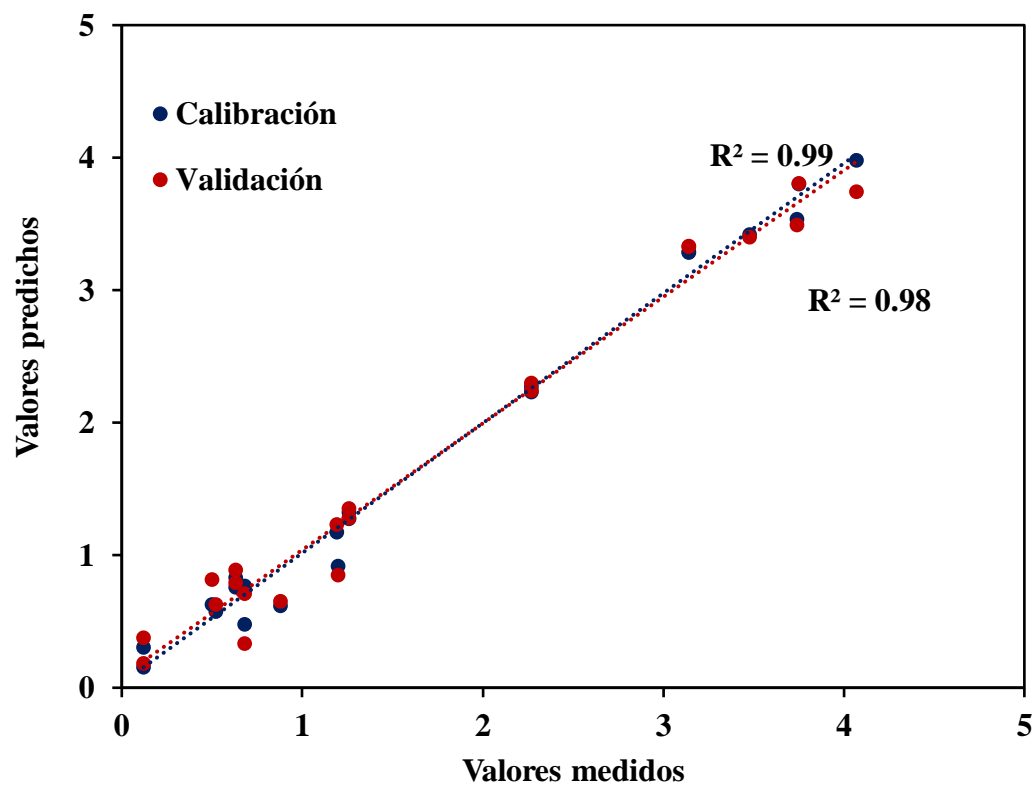
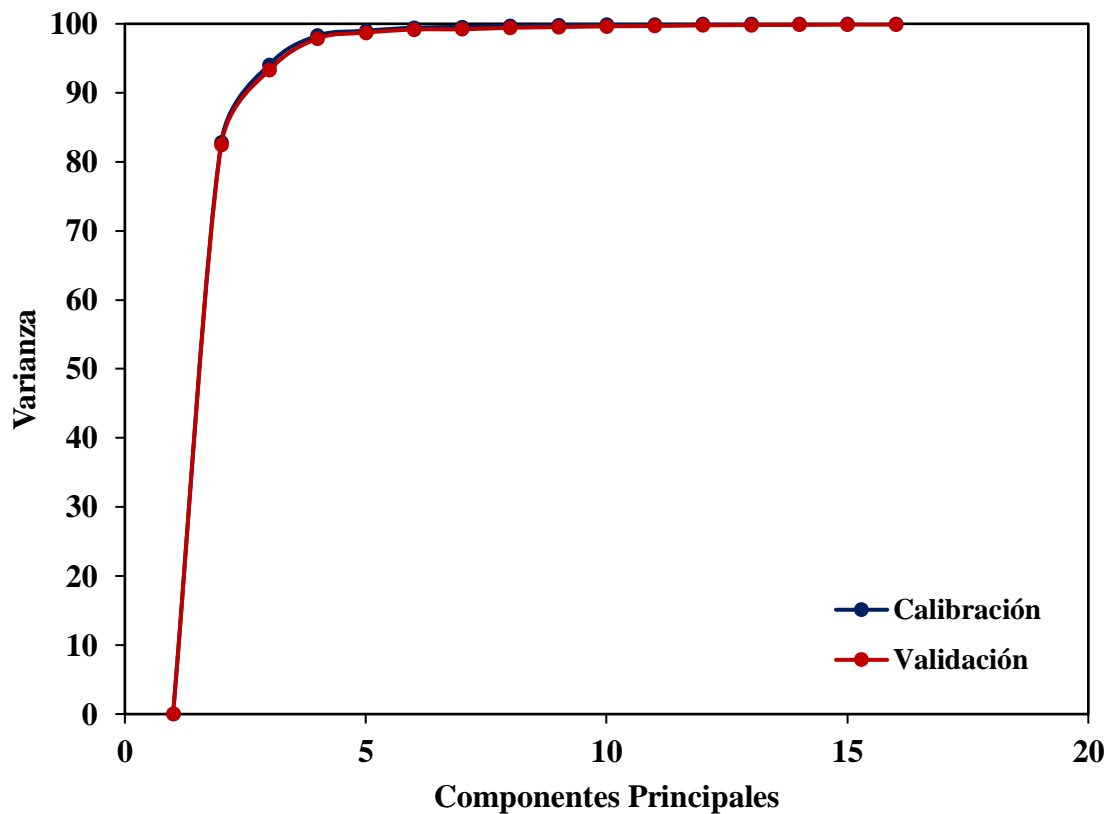


Figura 42.

Gráfico Componentes Principales Modelo escogido.



Nota: Se puede observar en la figura 42 la varianza explicada por los componentes principales del modelo. En donde se evidencia que los primeros factores concentran prácticamente toda la información relevante, estabilizando rápidamente la curva. Esto confirma que el modelo requiere un número reducido de componente para describir la variabilidad, garantizando simplicidad y alta eficiencia predictiva

Tabla 51.*Asfaltenos línea base primera regresión < 5%.*

Asfaltenos línea base primera regresión				
	Espectro 1	Espectro 2	Espectro 3	Espectro 4
Rcal	0.933	0.987	0.926	0.935
Rval	0.917	0.932	0.907	0.919
RMSEcal	0.373	0.168	0.393	0.367
RMSEval	0.430	0.389	0.455	0.423
CP	2	4	3	2

Nota: La Tabla 52 muestra resultados moderados, con coeficientes de validación entre 0.91 a 0.93 y errores en el rango de 0.39 a 0.46, lo que refleja un ajuste aceptable pero aún con dispersión atribuible a muestras atípicas.

Tabla 52.*Asfaltenos línea base última regresión < 5%.*

Asfaltenos última regresión				
	Espectro 1	Espectro 2	Espectro 3	Espectro 4
Rcal	0.976	0.994	0.983	0.982
Rval	0.971	0.958	0.974	0.978
RMSEcal	0.209	0.117	0.166	0.178
RMSEval	0.237	0.315	0.211	0.202
CP	2	5	4	2

Nota: Tras la depuración de datos, la tabla 53 evidencia una mejora sustancial, como se puede observar los coeficientes de calibración entre 0.98 y 0.99 con una validación de 0.96 a 0.97, estos resultados confirman que la corrección de línea base, combinada con la exclusión de un número reducido de muestras, fortalece la capacidad predictiva de los modelos para asfaltenos en bajas concentraciones, manteniendo además un número bajo de componentes principales.

Figura 43.

Valores medidos vs predichos para el modelo escogido, línea base asfaltenos < 5%.

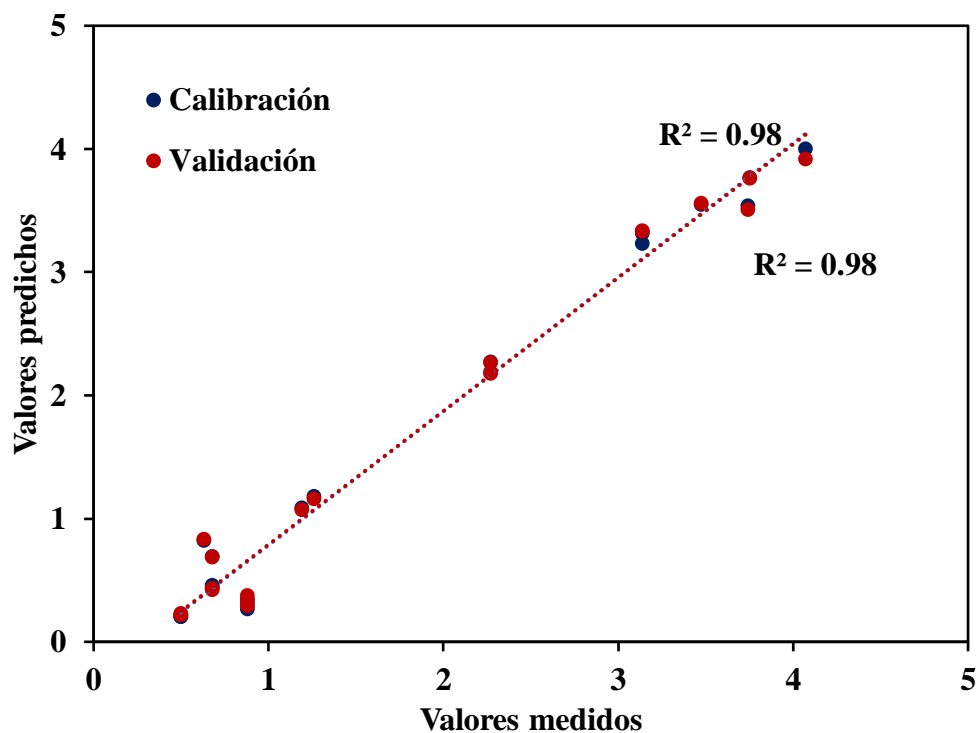
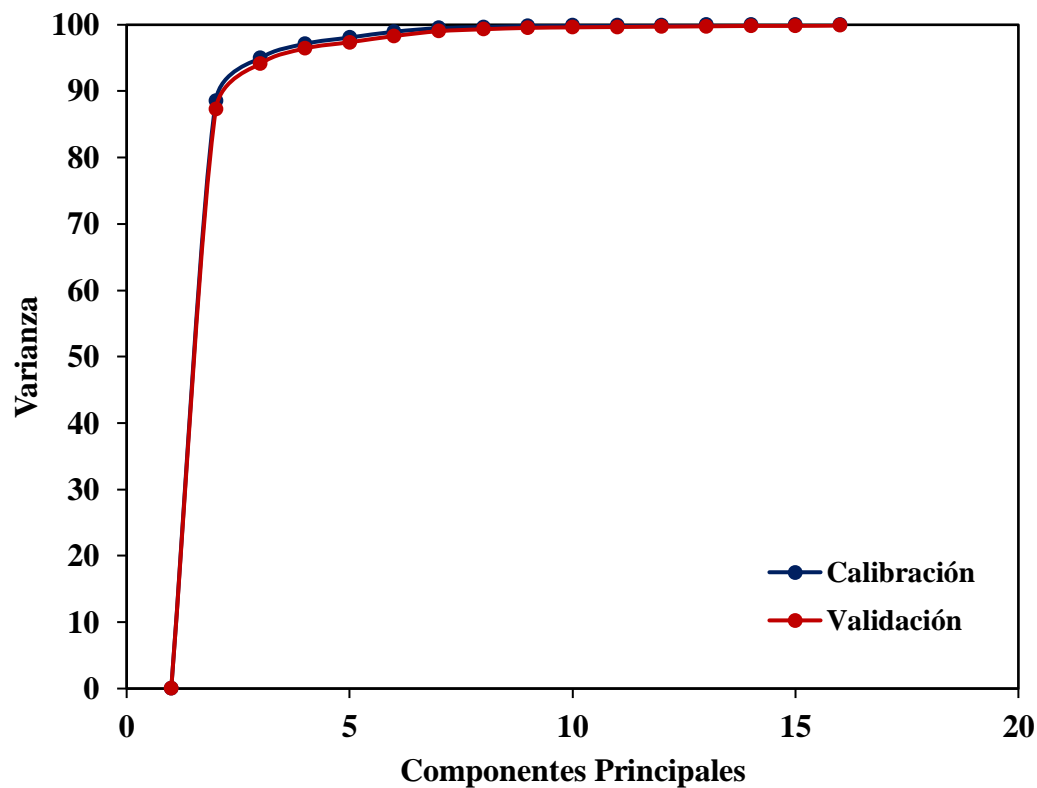


Figura 44.

Componentes Principales del modelo escogido.



Nota: La Figura 44 muestra la contribución de los componentes principales en el modelo corregido por línea base. Se observa que con un número reducido de factores se captura la mayor parte de la varianza explicativa del sistema, esto demuestra que el comportamiento de la estructura espectral de las muestras < 5% de asfaltenos es altamente correlacionada y que la variabilidad relevante se concentra en los primeros componentes. De este modo, el modelo combina alta precisión estadística con baja complejidad asociada a un $R^2 > 0.96$ en validación, lo cual optimiza tanto la robustez como la interpretabilidad de los resultados.

Apéndice C.

Modelos evaluados para densidades.

Tabla 53.

Modelo Densidad Primera regresión.

Densidad base primera regresión				
	Espectro 1	Espectro 2	Espectro 3	Espectro 4
Rcal	0.992	0.996	0.983	0.978
Rval	0.988	0.997	0.977	0.976
RMSEcal	0.006	0.005	0.009	0.010
RMSEval	0.008	0.006	0.011	0.011
CP	5	5	5	3

Nota: Los modelos de densidad evidencian una mejora significativa entre la primera y la última regresión hecha, en la tabla 54, el modelo inicial presentó un ajuste moderado con valores de correlación en calibración y validación aceptables pero acompañados de errores de predicción relativamente altos, lo que refleja cierta dispersión atribuida a la variabilidad de las muestras y condiciones experimentales.

Tabla 54.

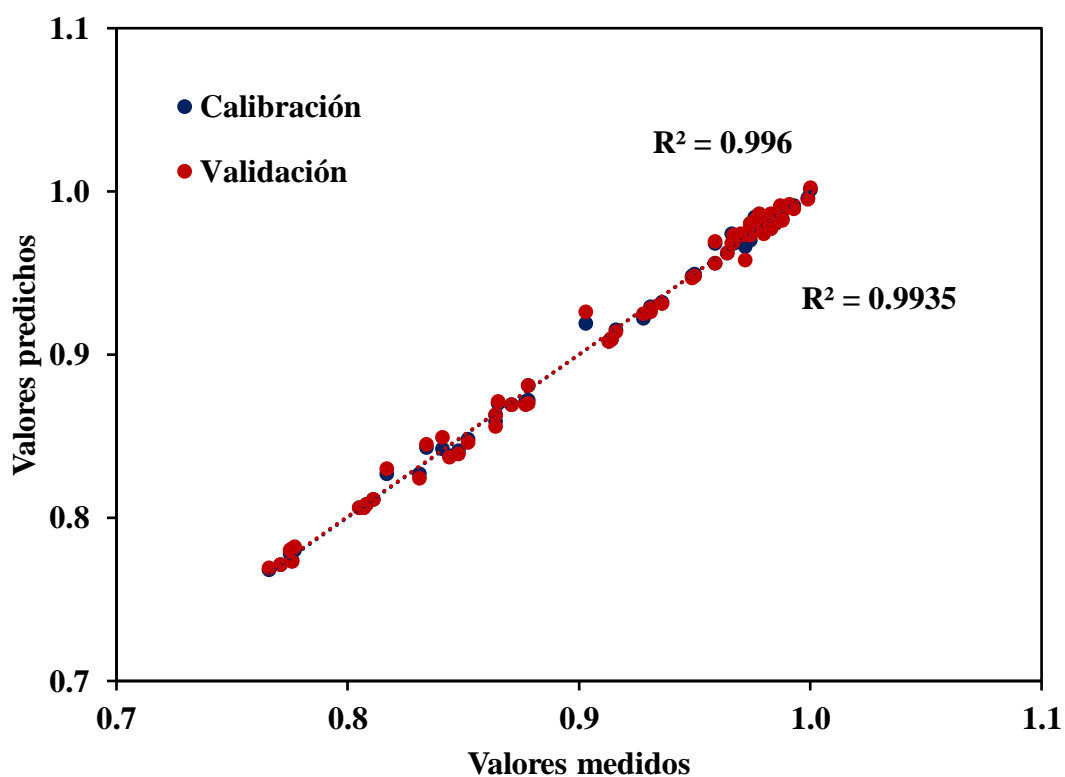
Modelo densidad última regresión.

Densidad base última regresión				
	Espectro 1	Espectro 2	Espectro 3	Espectro 4
Rcal	0.994	NO SE	0.986	0.985
Rval	0.989	DESARROLLA	0.983	0.980
RMSEcal	0.006	POR ALTA	0.008	0.009
RMSEval	0.007	CONVERGENCIA	0.009	0.010
CP	5	EN PRIMERA	5	3
		REGRESIÓN		

Nota: Por otro lado, la tabla 55 muestra un modelo mucho más robusto, con coeficientes R^2 superiores que varían entre 0.98 a 0.99, y una notable reducción en los errores de calibración y validación, lo que confirma una relación más consistente entre los espectros infrarrojos y la densidad.

Figura 45.

Valores medidos vs predichos para el modelo seleccionado para la densidad.

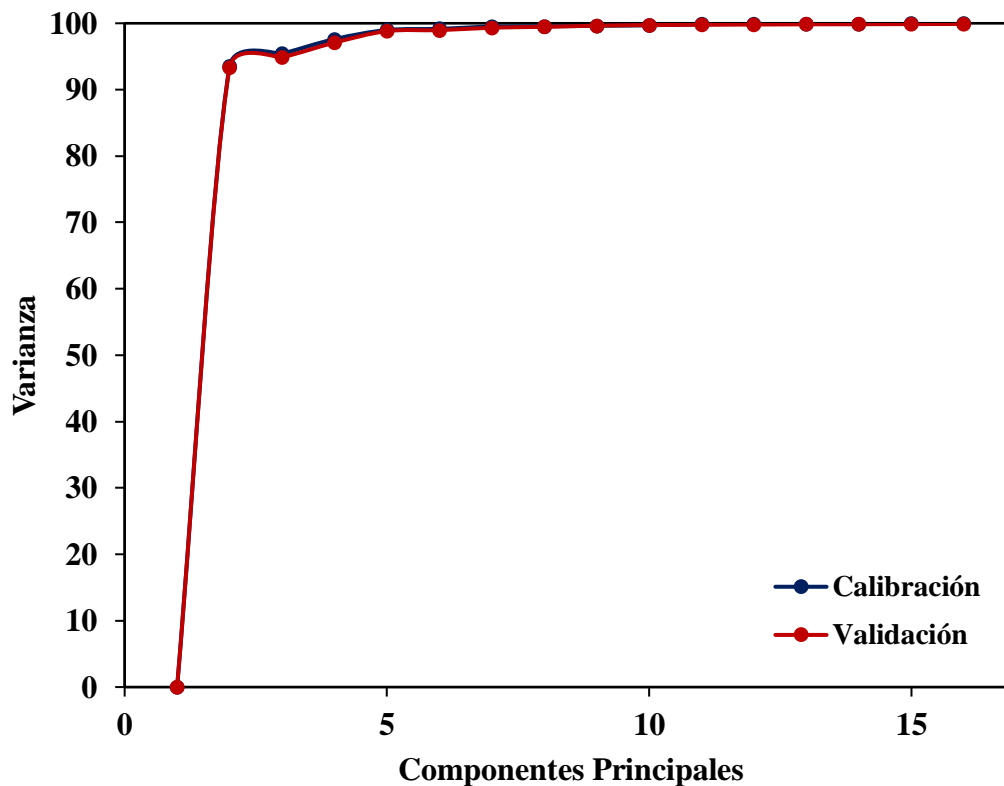


Nota: La Figura 45 muestra la dispersión entre los valores medidos y predichos para la densidad en la última regresión. Los puntos se alinean estrechamente con la recta de identidad, lo que refleja un ajuste robusto del modelo PLS. Este comportamiento está en concordancia con los resultados de la Tabla 42, donde se alcanzaron coeficientes de

correlación superiores a 0.98–0.99 y errores mínimos de calibración y validación, confirmando la confiabilidad del modelo en la predicción de densidad.

Figura 46.

Componentes principales modelo escogido.



Nota: La Figura 46 representa la varianza explicada por los componentes principales, en ella se observa que los primeros factores concentran casi la totalidad de la información relevante, en donde la curva se estabiliza rápidamente. Esto indica que el modelo requiere un número reducido de componentes para describir la variabilidad del sistema, garantizando simplicidad estructural y alta precisión estadística.

Luego de tener el primer modelo, se procede a desarrollar el modelo línea base iniciando con las primeras regresiones.

Tabla 55.

Modelo densidad línea base, primera regresión.

DENSIDAD LÍNEA BASE Última regresión				
	Espectro 1	Espectro 2	Espectro 3	Espectro 4
Rcal	0.991	0.998	0.974	0.983
Rval	0.987	0.998	0.968	0.980
RMSEcal	0.007	0.003	0.012	0.010
RMSEval	0.008	0.004	0.013	0.010
CP	4	5	3	3

Tabla 5756.

Modelo densidad línea base, última regresión.

Densidad línea base última regresión				
	Espectro 1	Espectro 2	Espectro 3	Espectro 4
Rcal	0.992	No se desarrolla	0.977	0.985
Rval	0.988	por alta	0.973	0.982
RMSEcal	0.006	convergencia en	0.011	0.009
RMSEval	0.008	primera regresión	0.012	0.010
CP	4		3	3

Tabla 5857.

Modelo densidad normalizado, primera regresión.

Densidad normalizada, última regresión				
	Espectro 1	Espectro 2	Espectro 3	Espectro 4
Rcal	0.977	0.953	0.948	0.952
Rval	0.966	0.939	0.942	0.945

RMSEcal	0.011	0.0158	0.017	0.015
RMSEval	0.013	0.018	0.018	0.017
CP	5	4	3	3

Nota: La Tabla 58, correspondiente a la primera regresión con normalización espectral, muestra un modelo de densidad con desempeño aceptable pero limitado, donde los coeficientes de correlación en validación se ubican entorno a R^2_{val} entre 0.92 a 0.94 y los errores de predicción son relativamente altos con un $RMSE_{val} > 0.35$, lo que refleja cierta dispersión y falta de robustez inicial.

Tabla 58.

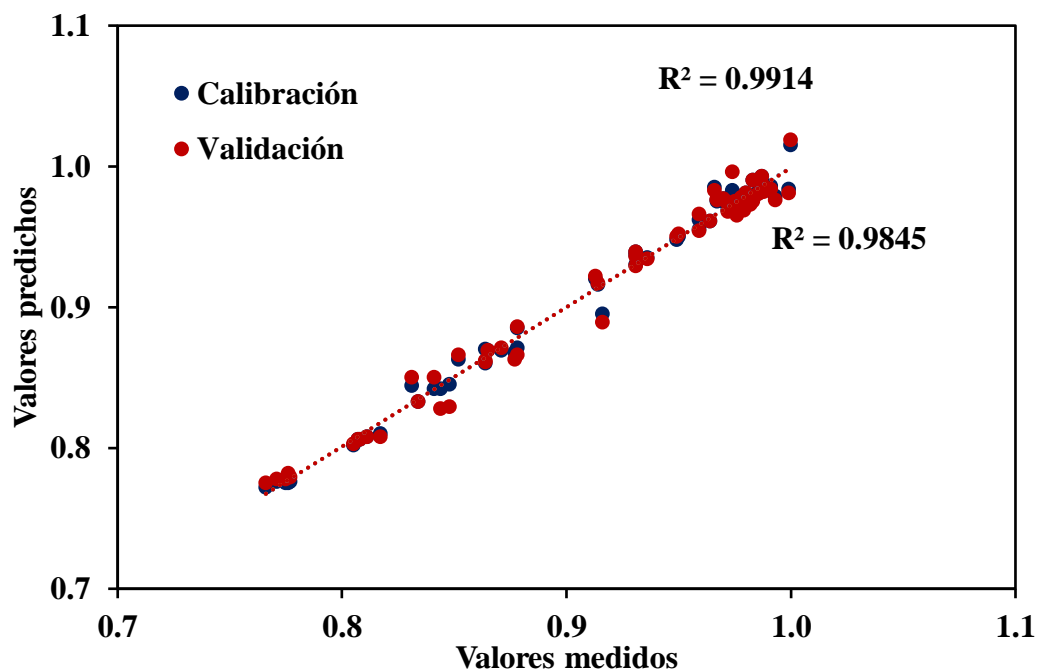
Modelo densidad normalizado, última regresión.

Densidad normalizada, última regresión				
	Espectro 1	Espectro 2	Espectro 3	Espectro 4
Rcal	0.991	0.976	0.979	0.991
Rval	0.985	0.964	0.974	0.985
RMSEcal	0.007	0.011	0.011	0.007
RMSEval	0.009	0.014	0.012	0.009
CP	5	5	5	5

Nota: En contraste, la Tabla 59, que recoge la última regresión de datos bajo este mismo pretratamiento, evidencia una mejora notable: los coeficientes de calibración y validación alcanzan valores superiores con un $R^2_{cal} \approx 0.99$ y un $R_{val} \approx 0.99$ con una reducción marcada en los errores $RMSE_{val} < 0.20$. Esto demuestra que la normalización de los espectros, junto con la depuración progresiva de muestras, permitió construir un modelo altamente preciso y estable para la predicción de la densidad, con un bajo número de componentes principales.

Figura 47.

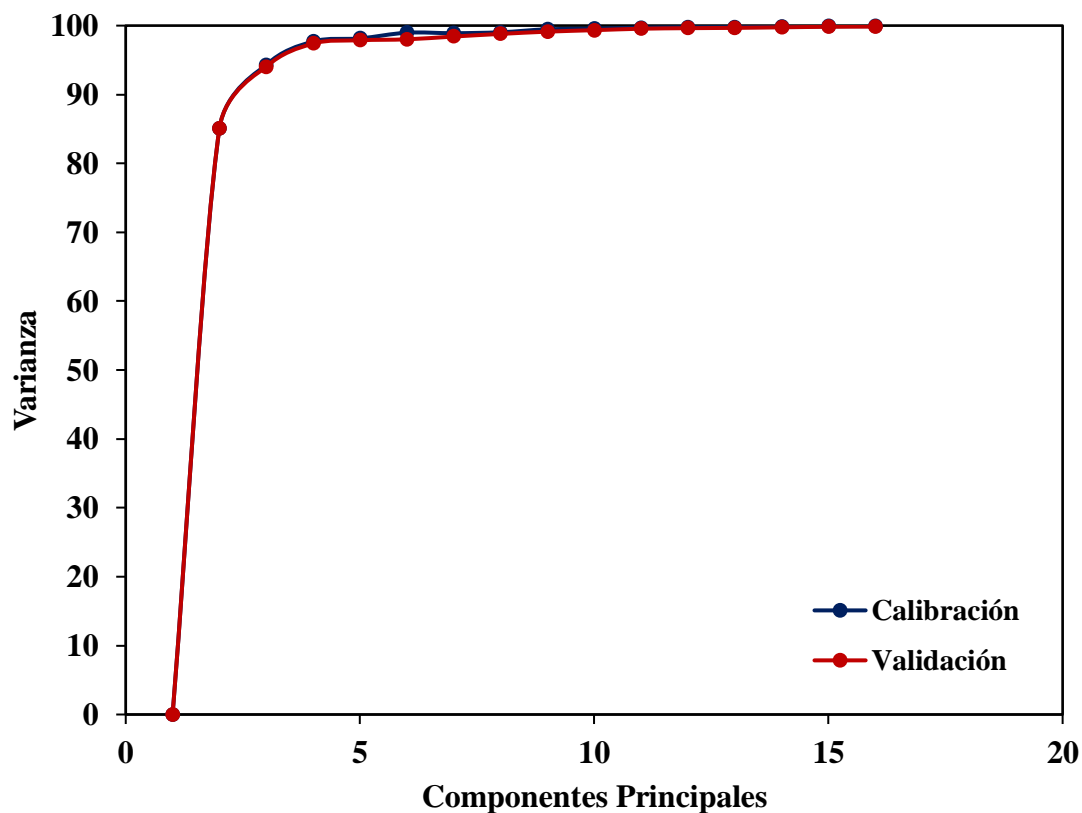
Valores medidos vs predichos para el modelo densidad normalizado.



Nota: La figura 47 representa la dispersión entre los valores medidos y predichos en el modelo de densidad con normalización. Los puntos se encuentran cercanos a la recta de calibración, mostrando un ajuste altamente confiable. Estos resultados concuerdan con la Tabla 59, donde se alcanzaron coeficientes de validación que van de 0.98 a 0.99 y errores mínimos RMSEval menores a 0.20, confirmando la eficacia de la normalización para mejorar la capacidad predictiva.

Figura 48.

Componentes principales modelo seleccionado.



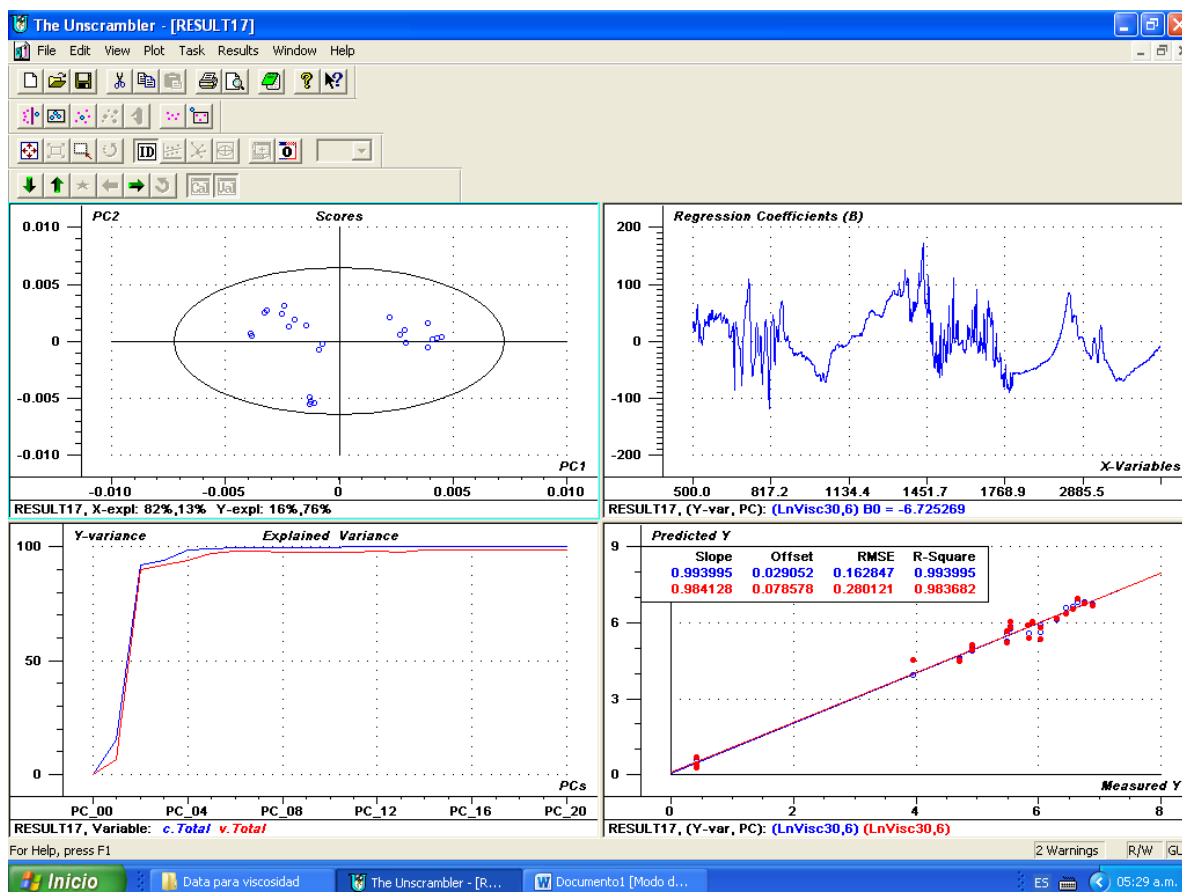
Nota: La Figura 48 muestra la varianza explicada por los componentes principales del modelo normalizado, donde la curva se estabiliza rápidamente, evidenciando que con un número reducido de factores de 3 a 5 CP se concentra prácticamente toda la información relevante. Esto demuestra que el modelo combina simplicidad estructural con alta precisión estadística, reduciendo el riesgo de sobreajuste y reforzando la robustez del enfoque de normalización

Apéndice D.

Vistas análisis gráfico de modelos

Figura 49.

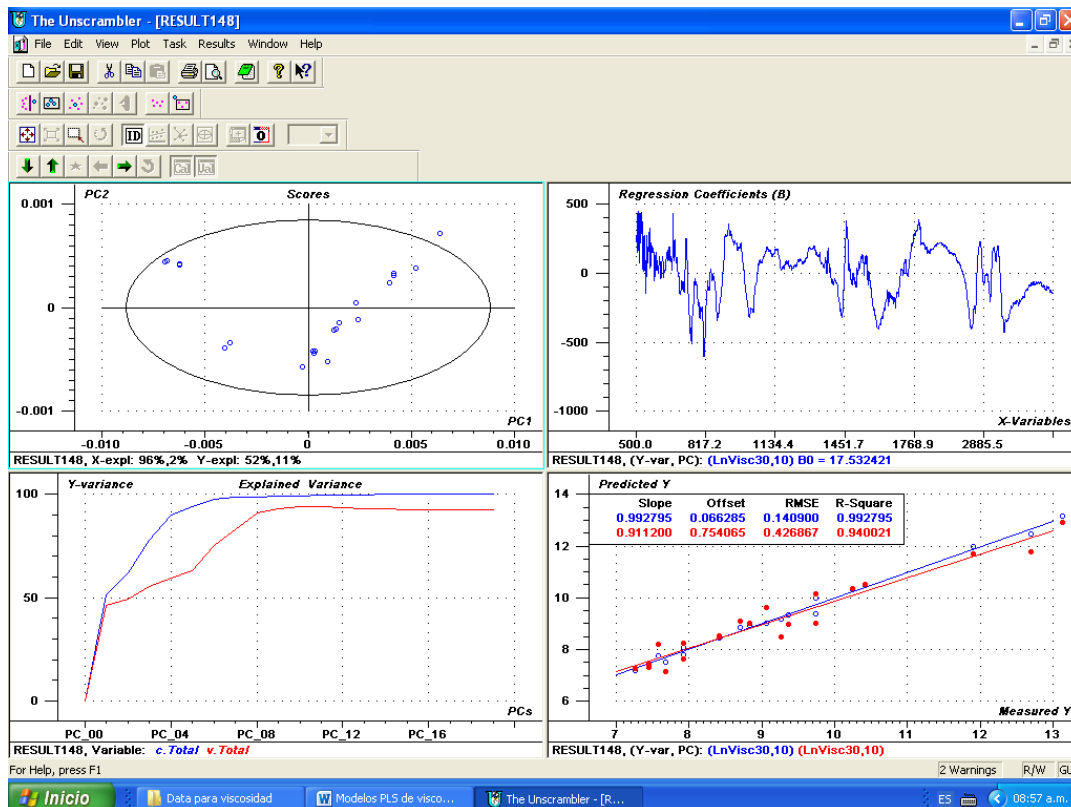
Vista análisis grafico general The Unscrambler Modelo Lnvisc @ 30°C < 1000 cP.



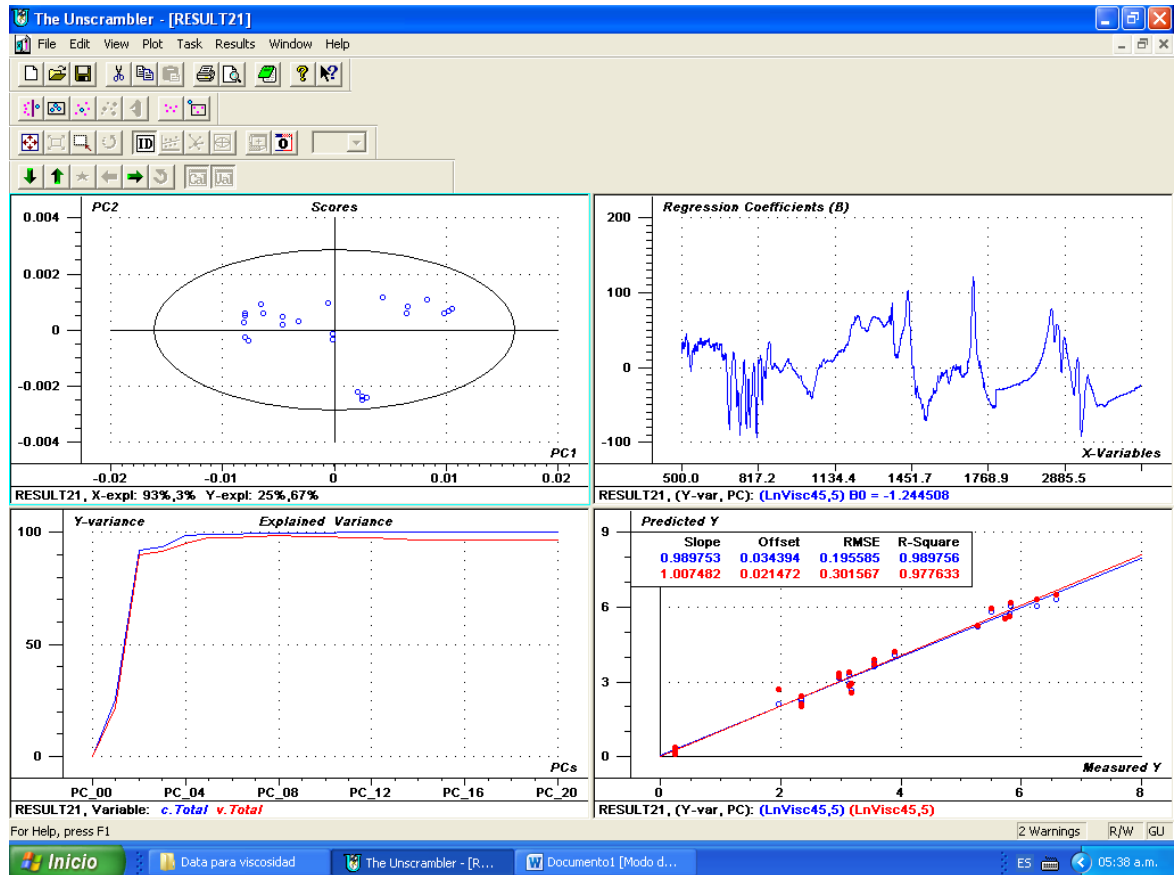
Nota: Imagen tomada del software The Unscrambler

Figura 50.

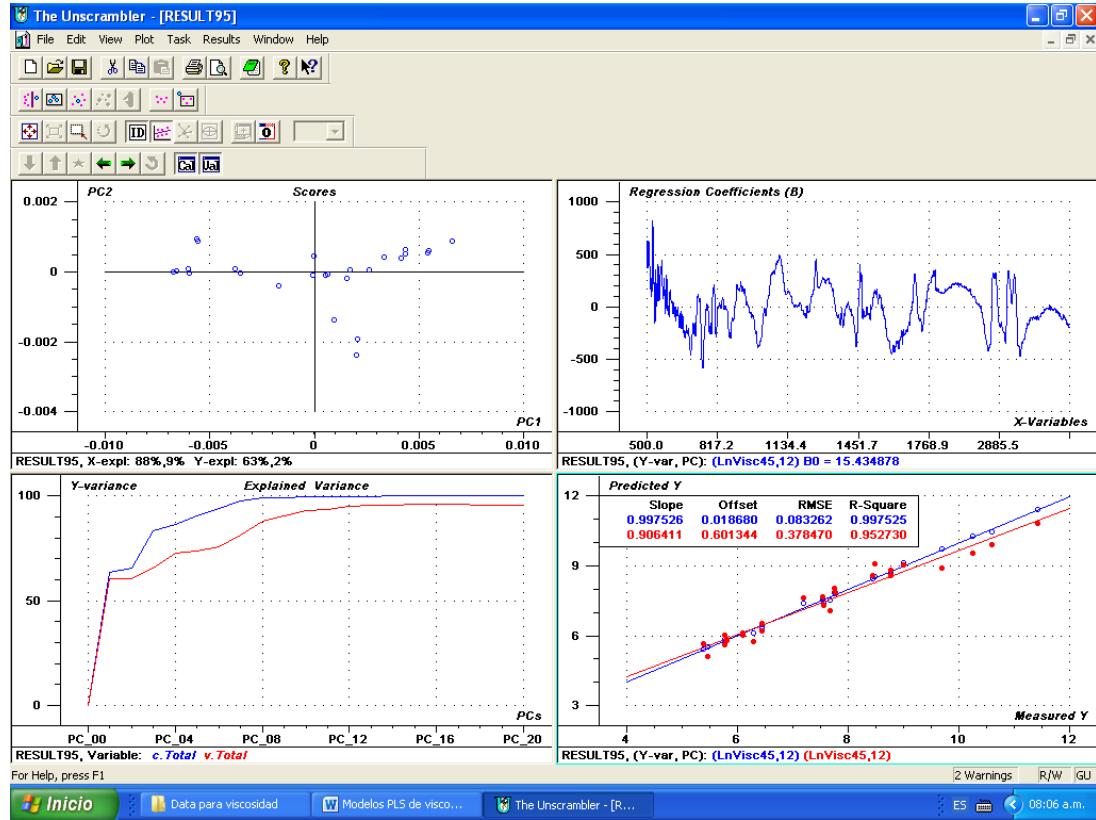
Vista análisis grafico general The Unscrambler para Modelo Lnvisc @ 30°C > 1000 cP.



Nota: Imagen tomada del software The Unscramble

Figura 51.*Vista análisis grafico general The Unscrambler*

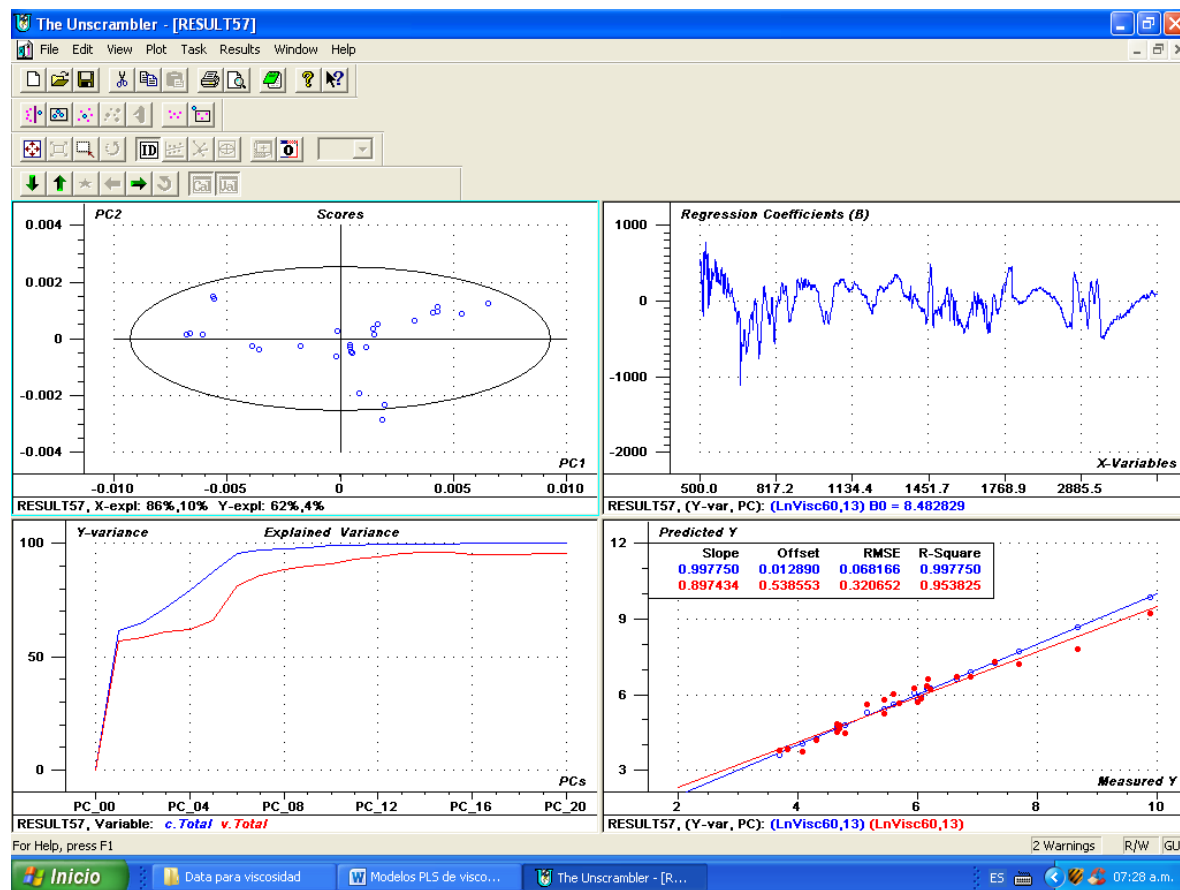
Nota: Imagen tomada del software The Unscramble

Figura 52.*Vista analisis grafico general The Unscrambler*

Nota: El gráfico scores mostrado en la Figura 52, evidencia que las observaciones se encuentran distribuidas de manera uniforme dentro de la elipse de Hotelling al 95%, sin presencia de valores atípicos significativos, lo cual confirma la estabilidad del modelo y la representatividad del conjunto de datos. Los coeficientes de regresión muestran picos definidos en distintas regiones espectrales, lo que indica la influencia de bandas específicas en la predicción de la viscosidad, asociadas a interacciones moleculares sensibles a la temperatura.

Figura 53.

Vista analisis grafico general The Unscrambler para Modelo Lnvisc @ 60°C mayor a 1000 cP



Nota: En el gráfico de scores visualizado en la Figura 53, se observa que las muestras se distribuyen homogéneamente dentro de la elipse de Hotelling al 95 %, sin evidencias claras de valores atípicos, lo que indica que el conjunto de datos es representativo y el modelo se construyó sobre una base estable.