

**PREDICION DE PROPIEDADES FISICOQUIMICAS DE DESTILADOS ATMOSFÉRICOS (199-371°C+) A
PARTIR DE SUS CRUDOS MADRE MEDIANTE FTIR-ATR Y ANALISIS MULTIVARIADO**

BEATRIZ MURCIA CELIS

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS BÁSICAS
ESCUELA DE QUÍMICA
MAESTRÍA EN QUÍMICA
BUCARAMANGA
2017**

PREDICCIÓN DE PROPIEDADES FÍSICOQUÍMICAS DE DESTILADOS ATMOSFÉRICOS (199-371°C+) A PARTIR DE SUS CRUDOS MADRE MEDIANTE FTIR-ATR Y ANÁLISIS MULTIVARIABLE

BEATRIZ MURCIA CELIS

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

Director

**ENRIQUE MEJÍA OSPINO, QUÍMICO, Ph.D
Director Grupo de Investigación Espectroscopia
Escuela de Química – UIS**

Co-Director

**YUSTINA RODRIGUEZ, QUÍMICO, Ph.D
Instituto Colombiano del Petróleo, Ecopetrol - ICP**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS BÁSICAS
ESCUELA DE QUÍMICA
MAESTRÍA EN QUÍMICA
BUCARAMANGA
2017**

DEDICATORIA

*A mi esposo, Javier, y a mi hijo Sebastián, por su
incondicional apoyo para recorrer este camino...*

AGRADECIMIENTOS

A Dios y a mi familia, por su incansable apoyo para llevar a feliz término este trabajo.

A los doctores Enrique Mejía Ospino y Yustina Rodríguez, por su orientación y aportes para la realización y culminación de este proyecto.

Al Instituto Colombiano del Petróleo (ICP) y la Universidad Industrial de Santander (UIS), por el apoyo brindado mediante el acuerdo de Cooperación No. 05 derivado del Convenio marco de Cooperación Tecnológica y Científica No. 5222395 UIS-ECOPETROL.

A todas aquellas personas que de una u otra forma colaboraron o participaron en la realización de este trabajo, extendiendo mi sincero agradecimiento.

TABLA DE CONTENIDO

| | |
|---|----|
| INTRODUCCION..... | 16 |
| 1. MARCO DE REFERENCIA..... | 18 |
| 1.1 PLANTEAMIENTO DEL PROBLEMA..... | 18 |
| 1.2 ESTADO DEL ARTE | 19 |
| 1.3 MARCO TEÓRICO | 25 |
| 1.3.1 EL Petróleo | 25 |
| 1.3.1.1 Fraccionamiento de los crudos | 26 |
| 1.3.1.2 Evaluación y caracterización de crudos | 28 |
| 1.3.2 Quimiometría | 33 |
| 1.3.3 Espectroscopia de Infrarrojo..... | 42 |
| 2. METODOLOGÍA EXPERIMENTAL | 50 |
| 2.1 PREPARACIÓN DE MUESTRAS..... | 50 |
| 2.1.1 Destilación Atmosférica ASTM D2892 | 50 |
| 2.1.2 Caracterización de fracciones atmosféricas | 50 |
| 2.2 CARACTERIZACIÓN POR ESPECTROSCOPIA FTIR-ATR | 54 |
| 2.2.1 Determinación de parámetros de adquisición de espectros..... | 54 |
| 2.2.2 Verificación desempeño del espectrómetro FTIR..... | 56 |
| 2.3 PROCESAMIENTO DE DATOS (PRETRATAMIENTO)..... | 57 |
| 2.4 TOMA DE ESPECTROS E IDENTIFICACIÓN DE SEÑALES IR..... | 59 |
| 3. RESULTADOS Y ANALISIS..... | 64 |
| 3.1 ANÁLISIS PRELIMINAR POR COMPONENTES PRINCIPALES (PCA) | 64 |
| 3.2 ANÁLISIS PRELIMINAR DE LA LINEALIDAD: MODELOS PLS INICIALES | 69 |
| 3.3 MODELOS PLS RENDIMIENTOS | 71 |
| 3.3.1 Modelo PLS Rendimientos queroseno..... | 72 |
| 3.3.2 Modelo PLS rendimientos acpm liviano..... | 81 |
| 3.3.3 Modelo PLS rendimientos acpm pesado | 85 |
| 3.3.4 Modelo PLS rendimientos crudo reducido | 89 |

| | |
|--|-----|
| 3.4 SELECCIÓN DE VARIABLES MEDIANTE ALGORITMOS GENETICOS (AG)..... | 93 |
| 3.4.1 Selección de variables para rendimientos queroseno..... | 99 |
| 3.4.2 Selección de variables para rendimientos acpm liviano..... | 102 |
| 3.4.3 Selección de variables para rendimientos acpm pesado..... | 105 |
| 3.4.5 Selección de variables para rendimientos crudo reducido..... | 107 |
| 3.4.6 Análisis regiones espectrales seleccionadas por AG..... | 109 |
| 3.5 MODELOS AG-PLS RENDIMIENTOS | 111 |
| 3.5.1 Modelo AG-PLS rendimientos queroseno | 112 |
| 3.5.2 Modelo AG-PLS rendimientos acpm liviano | 117 |
| 3.5.3 Modelo AG-PLS rendimientos acpm pesado | 118 |
| 3.5.4 Modelo AG-PLS rendimientos crudo reducido | 119 |
| 3.6 MODELOS PLS PROPIEDADES..... | 120 |
| 3.6.1 Modelo PLS API queroseno..... | 121 |
| 3.6.2 Modelo PLS azufre queroseno | 123 |
| 3.6.3 Modelo PLS Índice de cetano queroseno | 126 |
| 3.6.4 Modelos PLS API, azufre e índice de cetano acpm liviano..... | 129 |
| 3.6.5 Modelos PLS API, azufre e índice de cetano acpm pesado..... | 134 |
| 3.6.6 Modelos PLS API y azufre crudo reducido | 139 |
| 3.6.7 Reproducibilidad modelos PLS de propiedades..... | 142 |
| 4. CONCLUSIONES..... | 145 |
| 5. RECOMENDACIONES | 147 |
| BIBLIOGRAFIA | 148 |

LISTA DE TABLAS

| | |
|--|-----|
| <i>Tabla 1. Composición elemental del petróleo.....</i> | 25 |
| <i>Tabla 2. Clasificación del crudo según su densidad API.....</i> | 26 |
| <i>Tabla 3. Algunas fracciones de destilación del crudo.....</i> | 27 |
| <i>Tabla 4. Esquema de Fraccionamiento de Crudos (Assay Tipo II).....</i> | 29 |
| <i>Tabla 5. Esquema de Fraccionamiento de Crudos (Assay Tipo III).....</i> | 29 |
| <i>Tabla 6. Caracterización Fisicoquímica del Crudo y sus fracciones de destilación (Assay Tipo II). ..</i> | 30 |
| <i>Tabla 7. Caracterización Fisicoquímica del Crudo y sus fracciones de destilación (Assay Tipo III) ...</i> | 30 |
| <i>Tabla 8. Regiones espectrales de la radiación infrarroja.....</i> | 42 |
| <i>Tabla 9. Grados de libertad de moléculas Poliatómicas.....</i> | 44 |
| <i>Tabla 10. Características materiales para cristales ATR.....</i> | 47 |
| <i>Tabla 11. Asignación bandas IR para Crudos.....</i> | 48 |
| <i>Tabla 12. Principales propiedades set de crudos analizado.....</i> | 51 |
| <i>Tabla 13. Rendimientos fracciones atmosféricas.....</i> | 52 |
| <i>Tabla 14. Propiedades fisicoquímicas fracciones atmosféricas.....</i> | 53 |
| <i>Tabla 15. Valores de señal/ruido a resolución de 2.0 cm⁻¹.....</i> | 56 |
| <i>Tabla 16. Resultados Repetibilidad espectral Crudo No.1.....</i> | 57 |
| <i>Tabla 17. Valores de RSD obtenidos para tres números de onda característicos aplicando diferentes tipos de Normalización espectral.....</i> | 58 |
| <i>Tabla 18. Asignación de bandas de absorción para los espectros IR de crudos analizados.....</i> | 61 |
| <i>Tabla 19. Fracciones y propiedades a modelar.....</i> | 70 |
| <i>Tabla 20. Validación externa modelo PLS rendimientos (%V) fracción queroseno.....</i> | 81 |
| <i>Tabla 21. Validación externa del modelo PLS rendimientos (%V) fracción acpm liviano.....</i> | 85 |
| <i>Tabla 22. Validación externa del modelo rendimientos (%V) fracción acpm pesado.....</i> | 88 |
| <i>Tabla 23. Validación externa del modelo PLS rendimientos (%V) fracción crudo reducido.....</i> | 92 |
| <i>Tabla 24. Resultados modelos de calibración y validación PLS rendimientos (%V) destilados atmosféricos.....</i> | 93 |
| <i>Tabla 25. Optimización de parámetros del algoritmo genético (AG).....</i> | 96 |
| <i>Tabla 26. Repetibilidad del algoritmo genético rendimientos (%V) de queroseno.....</i> | 97 |
| <i>Tabla 27. Parámetros finales algoritmos genéticos para rendimientos destilados atmosféricos....</i> | 99 |
| <i>Tabla 28. Variables útiles y asignación de bandas vibracionales rendimientos (%V) fracción queroseno mediante AG.....</i> | 102 |
| <i>Tabla 29. Variables útiles y asignación de bandas vibracionales rendimientos acpm liviano.....</i> | 105 |
| <i>Tabla 30. Variables útiles y asignación de bandas vibracionales rendimientos acpm pesado (algoritmos genéticos).....</i> | 106 |
| <i>Tabla 31. Variables útiles y asignación de bandas vibracionales rendimientos crudo reducido</i> | 108 |

| | |
|---|------------|
| <i>Tabla 32. Comparación variables útiles y asignación de bandas vibracionales rendimientos fracciones atmosféricas</i> | <i>110</i> |
| <i>Tabla 33. Validación externa modelo AG-PLS rendimientos (%V) fracción queroseno.....</i> | <i>116</i> |
| <i>Tabla 34. Validación externa modelo AG-PLS rendimientos (%V) fracción acpm liviano</i> | <i>117</i> |
| <i>Tabla 35. Validación externa modelo AG-PLS rendimientos (%V) fracción acpm pesado</i> | <i>118</i> |
| <i>Tabla 36. Validación externa modelo AG-PLS rendimientos (%V) fracción crudo reducido.....</i> | <i>119</i> |
| <i>Tabla 37. Resultados modelos de calibración y validación AG-PLS rendimientos (%V) destilados atmosféricos.....</i> | <i>120</i> |
| <i>Tabla 38. Validación externa del modelo PLS API fracción queroseno.</i> | <i>123</i> |
| <i>Tabla 39. Validación externa del modelo PLS bajo azufre fracción queroseno</i> | <i>125</i> |
| <i>Tabla 40. Validación externa del modelo PLS alto azufre fracción queroseno</i> | <i>125</i> |
| <i>Tabla 41. Validación externa modelo PLS índice de cetano fracción queroseno</i> | <i>128</i> |
| <i>Tabla 42. Resultados modelos de calibración y validación PLS para fracción queroseno (199-249°C)</i> | <i>128</i> |
| <i>Tabla 43. Validación externa modelo PLS API fracción acpm liviano</i> | <i>130</i> |
| <i>Tabla 44. Validación externa modelo PLS índice de cetano fracción acpm liviano.....</i> | <i>131</i> |
| <i>Tabla 45. Validación externa del modelo PLS bajo azufre fracción acpm liviano</i> | <i>133</i> |
| <i>Tabla 46. Validación externa del modelo PLS alto azufre fracción acpm liviano.....</i> | <i>133</i> |
| <i>Tabla 47. Resultados modelos de calibración y validación PLS fracción acpm liviano (249-315°)</i> | <i>134</i> |
| <i>Tabla 48. Validación externa del modelo PLS API fracción acpm pesado.....</i> | <i>137</i> |
| <i>Tabla 49. Validación externa del modelo PLS azufre fracción acpm pesado</i> | <i>137</i> |
| <i>Tabla 50. Validación externa del modelo PLS índice de cetano fracción acpm pesado.....</i> | <i>138</i> |
| <i>Tabla 51. Resultados modelos de calibración y validación PLS fracción acpm pesado (315-371°C)</i> | <i>138</i> |
| <i>Tabla 52. Validación externa del modelo PLS API fracción crudo reducido</i> | <i>141</i> |
| <i>Tabla 53. Validación externa del modelo PLS azufre fracción crudo reducido</i> | <i>141</i> |
| <i>Tabla 54. Resultados modelos de calibración y validación PLS fracción crudo reducido (371°C+).</i> | <i>142</i> |
| <i>Tabla 55. Evaluación de la reproducibilidad modelos PLS fracción queroseno (199-249°C)</i> | <i>143</i> |
| <i>Tabla 56. Evaluación de la reproducibilidad modelos PLS fracción acpm liviano (249-315°).....</i> | <i>143</i> |
| <i>Tabla 57. Evaluación de la reproducibilidad modelos PLS fracción acpm pesado (315-371°C).....</i> | <i>143</i> |
| <i>Tabla 58. Evaluación de la reproducibilidad modelos PLS fracción crudo reducido (371°C+)</i> | <i>143</i> |

LISTA DE FIGURAS

| | |
|--|----|
| <i>Figura 1. Asignación bandas de absorción IR para crudos y espectros FTIR-ATR de crudos pesados.</i> | 22 |
| <i>Figura 2. Regiones seleccionadas para predicción de contenido de azufre por MIR con diferentes métodos de selección de variables.</i> | 24 |
| <i>Figura 3. Esquema general de destilación de crudos.</i> | 27 |
| <i>Figura 4. Representación geométrica de los componentes principales para dos variables.</i> | 36 |
| <i>Figura 5. Estructura del modelo PCA.</i> | 37 |
| <i>Figura 6. Representación geométrica del algoritmo PLS-R.</i> | 38 |
| <i>Figura 7. Esquema de optimización del algoritmo genético.</i> | 40 |
| <i>Figura 8. Esquema de la recombinación sencilla y doble.</i> | 41 |
| <i>Figura 9. Niveles de energía discreta cuantizados.</i> | 43 |
| <i>Figura 10. Esquema general de una celda de reflectancia total atenuada (ATR).</i> | 45 |
| <i>Figura 11. Espectrómetro infrarrojo Bruker Vertex 70V y accesorio ATR.</i> | 54 |
| <i>Figura 12. Espectros tomados a Resoluciones de 8, 4 y 2 cm⁻¹.</i> | 55 |
| <i>Figura 13. Espectros MIR-ATR de 54 Crudos sin pretratamiento.</i> | 58 |
| <i>Figura 14. Espectros MIR-ATR de 54 Crudos pretratados (Linear Baseline Correction + Normalización por Area + Mean center).</i> | 59 |
| <i>Figura 15. Espectros MIR-ATR de crudos extrapesado, pesado, medio y liviano.</i> | 60 |
| <i>Figura 16. Identificación de bandas de absorción en un espectro de crudo.</i> | 60 |
| <i>Figura 17. Regiones alifática 3050-2700 cm⁻¹ y huella dactilar 1100-650 cm⁻¹ de un crudo pesado y un crudo liviano.</i> | 62 |
| <i>Figura 18. Varianza explicada por los dos primeros componentes principales (PC).</i> | 64 |
| <i>Figura 19. Análisis PCA: Scores muestras de crudos.</i> | 65 |
| <i>Figura 20. Scores en los dos primeros componentes principales.</i> | 65 |
| <i>Figura 21. Varianzas residuales en X versus leverage.</i> | 66 |
| <i>Figura 22. Variables importantes en los espectros de muestras de crudos (PC-1).</i> | 67 |
| <i>Figura 23. Variables importantes en los espectros de muestras de crudos (PC-2).</i> | 69 |
| <i>Figura 24. Distribución de residuales para rendimientos (%V) de destilados atmosféricos.</i> | 71 |
| <i>Figura 25. Análisis Scores modelo PLS de rendimientos (%V) fracción queroseno.</i> | 72 |
| <i>Figura 26. Varianzas Residuales-X versus Leverage modelo PLS rendimientos queroseno.</i> | 73 |
| <i>Figura 27. Espectros FT-IR superpuestos de muestras de crudos y crudo No. 39.</i> | 73 |
| <i>Figura 28. Calibración modelo PLS rendimientos (%V) fracción queroseno.</i> | 74 |
| <i>Figura 29. Calibración/Validación modelo PLS rendimientos (%V) fracción queroseno.</i> | 75 |
| <i>Figura 30. Validación cruzada modelo PLD rendimientos (%V) fracción queroseno.</i> | 75 |
| <i>Figura 31. Residuales de rendimientos (%V) en muestras de calibración/validación modelo PLS fracción queroseno.</i> | 76 |
| <i>Figura 32. Varianza explicada modelo PLS rendimientos (%V) fracción queroseno.</i> | 76 |

| | |
|---|-----|
| <i>Figura 33. Error versus número de variables latentes para modelo PLS rendimientos (%V) fracción queroseno.</i> | 78 |
| <i>Figura 34. Variables importantes modelo PLS rendimientos (%V) fracción queroseno.</i> | 79 |
| <i>Figura 35. Validación externa modelo PLS rendimientos (%V) fracción queroseno.</i> | 80 |
| <i>Figura 36. Principales resultados modelo PLS rendimientos (%V) fracción acpm liviano.</i> | 82 |
| <i>Figura 37. Varianza explicada en X modelo PLS rendimientos (%V) fracción acpm liviano.</i> | 83 |
| <i>Figura 38. Variables importantes modelo PLS rendimientos (%V) fracción acpm liviano.</i> | 83 |
| <i>Figura 39. Validación externa modelo PLS rendimientos (%V) fracción acpm liviano.</i> | 84 |
| <i>Figura 40. Principales resultados modelo PLS para rendimientos (%V) fracción acpm pesado.</i> | 86 |
| <i>Figura 41. Varianza en X modelo PLS rendimientos (%V) fracción acpm pesado.</i> | 86 |
| <i>Figura 42. Variables importantes en el modelo de Rendimientos (%V) para acpm pesado.</i> | 87 |
| <i>Figura 43. Validación externa rendimientos (%V) fracción acpm pesado.</i> | 89 |
| <i>Figura 44. Principales resultados modelo PLS para rendimientos (%V) fracción crudo reducido.</i> | 90 |
| <i>Figura 45. Varianza en X modelo PLS rendimientos (%V) fracción crudo reducido.</i> | 90 |
| <i>Figura 46. Variables importantes en el modelo PLS de rendimientos (%V) fracción crudo reducido.</i> | 91 |
| <i>Figura 47. Validación externa modelo PLS rendimientos (%V) fracción crudo reducido.</i> | 92 |
| <i>Figura 48. RMSECV promedio y mejor RMSEC versus número de generaciones: (a) 25 generaciones. (b) 100 generaciones.</i> | 98 |
| <i>Figura 49. Ajuste RMSECV versus número de variables algoritmo genético rendimientos (%V) fracción queroseno.</i> | 100 |
| <i>Figura 50. Frecuencia de inclusión bloques de variables del algoritmo genético para rendimientos (%V) fracción queroseno.</i> | 101 |
| <i>Figura 51. Ajuste RMSECV versus número de variables algoritmo genético rendimientos fracción acpm liviano.</i> | 103 |
| <i>Figura 52. Frecuencia de inclusión bloques de variables del algoritmo genético para rendimientos fracción acpm liviano.</i> | 104 |
| <i>Figura 53. Ajuste RMSECV versus número de variables AG rendimientos acpm pesado.</i> | 105 |
| <i>Figura 54. Frecuencia de inclusión bloques de variables del algoritmo genético para rendimientos fracción acpm pesado.</i> | 106 |
| <i>Figura 55. Ajuste RMSECV versus número de variables algoritmo genético rendimientos fracción crudo reducido.</i> | 107 |
| <i>Figura 56. Frecuencia de inclusión bloques de variables del algoritmo genético para rendimientos (%V) fracción crudo reducido.</i> | 108 |
| <i>Figura 57. (a). Análisis scores y (b). Varianzas residuales, modelo AG-PLS rendimientos (%V) fracción queroseno.</i> | 112 |
| <i>Figura 58. Calibración/validación modelo AG-PLS rendimientos (%V) fracción queroseno.</i> | 113 |
| <i>Figura 59. Análisis Scores modelo AG-PLS rendimientos (%V) fracción queroseno.</i> | 113 |
| <i>Figura 60. Varianza explicada modelo AG-PLS de rendimientos (%V) fracción queroseno.</i> | 114 |

| | |
|---|------------|
| <i>Figura 61. Error versus número de variables latentes modelo AG-PLS rendimientos (%V) fracción queroseno.</i> | <i>115</i> |
| <i>Figura 62. Validación externa modelo AG-PLS rendimientos (%V) fracción queroseno.</i> | <i>115</i> |
| <i>Figura 63. Residuales de calibración y predicción modelo AG-PLS rendimientos (%V) fracción queroseno.</i> | <i>116</i> |
| <i>Figura 64. Parámetros estadísticos modelo AG-PLS rendimientos acpm liviano.....</i> | <i>117</i> |
| <i>Figura 65. Parámetros estadísticos modelo AG-PLS rendimientos acpm pesado.....</i> | <i>118</i> |
| <i>Figura 66. Parámetros estadísticos modelo AG-PLS rendimientos crudo reducido.</i> | <i>119</i> |
| <i>Figura 67. Principales gráficos del modelo PLS API fracción queroseno.</i> | <i>121</i> |
| <i>Figura 68. Validación externa modelo PLS API fracción queroseno.....</i> | <i>122</i> |
| <i>Figura 69. Principales gráficos modelos PLS fracción queroseno: (a) bajo azufre, (b) alto azufre.</i> | <i>124</i> |
| <i>Figura 70. Validación externa modelos PLS fracción queroseno: (a) bajo azufre, (b) alto azufre. .</i> | <i>125</i> |
| <i>Figura 71. Principales resultados modelo PLS índice de cetano fracción queroseno.</i> | <i>126</i> |
| <i>Figura 72. Validación externa modelo PLS índice de cetano fracción queroseno.....</i> | <i>127</i> |
| <i>Figura 73. Calibración y validación cruzada modelos PLS de API e índice de cetano fracción acpm liviano.....</i> | <i>129</i> |
| <i>Figura 74. Varianza explicada de calibración y validación modelos PLS acpm liviano: (a) API, (b) índice de cetano.</i> | <i>129</i> |
| <i>Figura 75. Validación externa modelos PLS de API e índice de cetano fracción acpm liviano.....</i> | <i>130</i> |
| <i>Figura 76. Principales gráficos modelos PLS fracción acpm liviano: (a) bajo azufre, (b) alto azufre.</i> | <i>132</i> |
| <i>Figura 77. Validación externa modelos PLS fracción acpm liviano: (a) bajo azufre, (b) alto azufre.</i> | <i>133</i> |
| <i>Figura 78. Calibración y validación cruzada modelos PLS de API, azufre e índice de cetano fracción acpm pesado.</i> | <i>135</i> |
| <i>Figura 79. Varianza explicada de calibración y validación modelos PLS de API, azufre e índice de cetano fracción acpm pesado.</i> | <i>135</i> |
| <i>Figura 80. Validación externa modelos PLS de API, azufre e índice de cetano fracción acpm pesado.</i> | <i>136</i> |
| <i>Figura 81. Principales resultados modelo PLS para API fracción crudo reducido.</i> | <i>139</i> |
| <i>Figura 82. Principales resultados modelo PLS para azufre fracción crudo reducido.</i> | <i>140</i> |
| <i>Figura 83. Validación externa modelos PLS de API y azufre fracción crudo reducido.</i> | <i>140</i> |

RESUMEN

TÍTULO:

PREDICCIÓN DE PROPIEDADES FÍSICOQUÍMICAS DE DESTILADOS ATMOSFÉRICOS (199-371°C+) A PARTIR DE SUS CRUDOS MADRE MEDIANTE FTIR-ATR Y ANÁLISIS MULTIVARIADO.*

AUTOR:

BEATRIZ MURCIA CELIS**

PALABRAS CLAVE:

FTIR-ATR, crudos, destilación atmosférica, calibración multivariada, PLS, algoritmos genéticos.

DESCRIPCIÓN:

Esta investigación describe el desarrollo de modelos predictivos para determinar los rendimientos de crudos y propiedades fisicoquímicas de destilados atmosféricos a partir de los espectros de sus crudos madre usando análisis multivariado. Un total de 54 muestras de crudos colombianos fueron usadas. Se realizó análisis preliminar por PCA (Análisis de Componentes Principales) y preprocesamiento espectral. Las regresiones PLS (Partial Least Squares) y AG-PLS (Algoritmos Genéticos – PLS) fueron desarrolladas usando los espectros de los crudos madre y los rendimientos de los destilados atmosféricos. Para los modelos PLS, fueron seleccionadas dos regiones del espectro FTIR-ATR: 3200-2500 cm^{-1} y 1800-650 cm^{-1} . Se realizaron análisis de regresión para rendimientos y para cada propiedad, con los mismos espectros FTIR-ATR de muestras de crudos. Los AG fueron usados para seleccionar las mejores variables para desarrollar los modelos de rendimientos AG-PLS. Las bandas vibracionales más importantes, responsables de la discriminación de las muestras para asociar las propiedades al espectro, fueron identificadas en los modelos PLS y AG-PLS.

Los coeficientes de correlación (R^2) obtenidos para los modelos de rendimientos fueron aceptables, entre 0.80 y 0.99. Los resultados predichos para los rendimientos son acordes con los valores de referencia, obteniendo variaciones menores al 20%. Para los modelos PLS (API, azufre e índice de cetano) los coeficientes de correlación (R^2) estuvieron entre 0.77 y 0.99, con errores relativos entre 5% y 25%, al comparar los valores predichos con los medidos. Los mejores parámetros estadísticos fueron obtenidos para el residuo atmosférico.

Se desarrolló una metodología FTIR-ATR rápida y confiable para predecir los rendimientos y propiedades fisicoquímicas de destilados atmosféricos a partir de los espectros de sus crudos madre. Esta metodología permite obtener información fundamental para evaluar los crudos y sus destilados, con errores relativos de predicción inferiores al 20% con respecto a los valores de referencia obtenidos mediante métodos estándar.

* Trabajo de Grado

**Facultad de Ciencias, Escuela de Química. Director: Enrique Mejía Ospino, Químico, Ph.D.

ABSTRACT

TITTLE:

PREDICTION OF PHYSICOCHEMICAL PROPERTIES OF ATMOSPHERIC DISTILLATES (199-371°C+) FROM ITS RAW CRUDE OILS BY FTIR-ATR AND MULTIVARIATE ANALYSIS. *

AUTHOR:

BEATRIZ MURCIA CELIS**

KEYWORDS:

FTIR-ATR, crude oil, atmospheric distillation, multivariate calibration, PLS, genetic algorithm.

DESCRIPTION:

This research describes the development of predictive models for determining the crude oil yields and the physicochemical properties of the atmospheric distillates from the analysis of the raw crude oil spectra using multivariate analysis. A total of 54 Colombian crude oil samples were used for this research. Preliminary PCA (Principal Component Analysis) was performed and spectral pre-processing. PLS (Partial Least Squares) and GA-PLS (Genetic Algorithm-PLS) regressions were performed using raw crude oil spectra and the yields of atmospheric distillate. For PLS models, two regions on the FTIR-ATR spectra were selected: 3200-2500cm⁻¹ and 1800-650cm⁻¹. Separate regression analysis was performed for yields and each property with the same FTIR-ATR spectra of the crude oils samples. GA was used to select the best variables to build the yield GA-PLS models. The most important vibrational bands, responsible for discrimination of samples to associate properties to spectra were identified in PLS and GA-PLS models.

The correlation coefficients (R^2) for the reference and predicted values for yield models were suitable, between 0.80 and 0.99. The results predicted for the yield are in agreement with the reference values, they have variations less than 20% when compared with the measured values. For the PLS models (API, sulfur and cetane index) the correlation coefficients (R^2) were between 0.77 y 0.99, with relative errors between 5% and 25% when compared with the measured values. The best statistical parameters were obtained for the atmospheric residuum.

A fast and reliable FTIR-ATR methodology was developed to predict the yields and physicochemical properties of atmospheric distillates from its raw crude oils spectra. This methodology allows getting fundamental information to evaluate the crude oil and its distillates, with prediction relative errors below 20% regarding to reference values obtained from standard methods.

* Thesis

**Faculty of Science, School of Chemistry. Director: Enrique Mejía Ospino, Chemist, Ph.D.

INTRODUCCION

El petróleo crudo es probablemente la más compleja mezcla orgánica, en términos del número de constituyentes químicos distintos. Para analizar estas mezclas se realizan múltiples separaciones químicas como extracción, precipitación, destilación y cromatografía, con el fin de reducir el número de componentes y simplificar la determinación de las diversas composiciones elementales moleculares y estructurales (Marshall & Rodgers 2004).

Las características fisicoquímicas de los crudos varían de acuerdo a su origen, condiciones de evolución y antigüedad. Los crudos tienen diferente composición en cuanto al tipo y cantidad de familias químicas, por lo tanto, sus posibilidades de aprovechamiento y refinación para obtener productos derivados serán diferentes, así como su valoración económica (Lluch Urpí 2004). Las refinerías se vuelven más complejas con el fin de procesar los crudos buscando una conversión total de los pesados y el cumplimiento de la normativa ambiental; por lo cual someten los crudos y sus mezclas a diferentes procesos de separación y transformación, con el fin de obtener productos más livianos y valiosos; siendo uno de los más importantes la destilación (Waquier 2004).

Durante este proceso es necesario conocer las características fisicoquímicas de los crudos y sus fracciones para determinar si el crudo recibido en refinería cumple con las expectativas, es decir, para que las valoraciones económicas proyectadas y las estrategias operacionales sean válidas (American Society for Testing and Materials 2003). Por lo anterior, los laboratorios de las refinerías de Ecopetrol S.A. y del ICP (Instituto Colombiano del Petróleo) realizan una serie de análisis fisicoquímicos para generar evaluaciones específicas para cada refinería según el tipo de crudo y como apoyo a los proyectos de investigación y servicios técnicos especializados.

En la caracterización fisicoquímica de crudos se usan dos esquemas de análisis: el preliminar y el detallado, basados en normas internacionales ASTM (American Society for Testing and Materials International), IP (Institute of Petroleum, actualmente Energy Institute) y API (American Petroleum Institute); los cuales requieren inversión de tiempo para su realización, altas cantidades de reactivos tóxicos y de muestra, recurso humano con experticia e infraestructura instrumental

especializada, lo que los hace costosos y dispendiosos. Adicionalmente, existe una gran cantidad de información embebida que no puede ser analizada fácilmente.

Como una alternativa rápida y viable a los métodos ASTM, IP y API, se han propuesto las técnicas espectroscópicas moleculares, debido a que teóricamente reflejan la composición molecular completa de una muestra. La mayoría de los estudios para caracterizar los crudos involucran la espectroscopia vibracional de infrarrojo medio (MIR) y de infrarrojo cercano (NIR). Las técnicas espectroscópicas son capaces de suministrar información sobre interacciones a escala molecular y contienen datos con información contenida en múltiples bandas de absorción en los espectros. Por esta razón, el potencial de varios métodos espectroscópicos para la caracterización de crudos y productos derivados se ha estudiado. La combinación de estas técnicas con el análisis de datos multivariados, puede ser muy útil para caracterizar y clasificar los crudos y fracciones de muchas maneras (de Peinder et al. 2009).

En esta investigación se usó análisis multivariado PLS (Partial Least Squares) para predecir los rendimientos y propiedades fisicoquímicas de fracciones de destilación atmosférica (199-371°C+) a partir de sus crudos madre, mediante espectroscopia de infrarrojo medio (FTIR-ATR), para obtener información relevante en corto tiempo y que sirve como insumo preliminar para el control de los procesos de refinación.

1. MARCO DE REFERENCIA

1.1 PLANTEAMIENTO DEL PROBLEMA

Las refinerías de Barrancabermeja y Cartagena procesan los crudos colombianos con el fin de obtener productos valiosos y suplir la demanda y consumo nacional de combustibles y petroquímicos (Jaime et al. 2013). El Instituto Colombiano del Petróleo (ICP), soporta a nivel de laboratorios y plantas piloto la destilación y caracterización fisicoquímica de los crudos y sus fracciones para apoyar la toma de decisiones en refinería. Estas evaluaciones son tareas lentas y costosas que requieren una infraestructura de análisis compleja, que involucra el uso de grandes cantidades de reactivos, solventes tóxicos y muestra; un seguimiento riguroso de los procedimientos y consumo de tiempo, factores que afectan la oportunidad en la entrega de resultados analíticos por parte de los laboratorios y que no permiten conocer características químicas más profundas del crudo.

La realización de las pruebas analíticas convencionales puede tardar entre una y varias semanas, incluso meses, tiempos de respuesta que limitan su aplicación como herramienta de control en el proceso, para la toma de decisiones en tiempo real. Es por esto que Ecopetrol requiere desarrollar metodologías que permitan obtener resultados de caracterización fisicoquímica de crudos y fracciones con una cantidad mínima de muestra, en un menor tiempo y con un alto grado de confiabilidad. En este sentido, la espectroscopia de infrarrojo es una excelente alternativa para la caracterización de crudos y fracciones por su velocidad de respuesta, instrumentación relativamente sencilla y posibilidad de acoplarse en línea a los procesos industriales.

De acuerdo con la revisión bibliográfica realizada, hasta el momento no se han construido modelos quimiométricos para predicción de propiedades de las fracciones de destilación atmosférica a partir de los crudos madre, que permitan no sólo predecir las propiedades fisicoquímicas más importantes de las fracciones, sino también los rendimientos de éstas. Por lo anterior, surgen las *preguntas ¿es posible correlacionar las respuestas espectrales de los crudos con la información de las propiedades fisicoquímicas de sus destilados atmosféricos y desarrollar*

modelos quimiométricos para predecir estas propiedades? Y, adicionalmente, es posible proponer un esquema de cálculo para la predicción de los rendimientos de las fracciones de la destilación atmosférica de los crudos?. Como respuesta a estas preguntas se plantea la hipótesis: Es posible correlacionar las respuestas espectrales de los crudos madre con la información de las propiedades fisicoquímicas de las fracciones de destilación atmosférica (199-371°C) para desarrollar modelos quimiométricos y predecir las propiedades fisicoquímicas y rendimientos de las fracciones.

En este contexto, en la presente investigación se desarrollaron modelos de predicción para determinar los rendimientos y algunas propiedades fisicoquímicas de las fracciones de destilación atmosférica (199-371°C+) a partir del análisis espectral de los crudos madre, mediante el uso de la espectroscopia de infrarrojo medio, empleando análisis multivariado de datos (PLS) y algoritmos genéticos (AG-PLS). Este desarrollo genera información que permitirá evaluar la calidad de las fracciones en términos de sus propiedades físico-químicas y proyección de rendimientos a productos valiosos.

1.2 ESTADO DEL ARTE

El uso de técnicas sofisticadas y métodos de investigación de hidrocarburos ha generado avances en la comprensión de la química del petróleo. Entre estos avances, se encuentra la espectroscopia vibracional, la cual ha sido empleada con éxito para la determinación de múltiples propiedades del crudo y sus fracciones desde hace más de 20 años. Los principales campos de investigación relacionados con la aplicación de la espectroscopia de infrarrojo y métodos quimiométricos en el análisis de productos del petróleo han sido la aplicación de NIR y MIR para caracterización y predicción de propiedades fisicoquímicas, control de calidad, análisis estructural, inspección de adulteración de combustibles, control de calidad en línea y en estudios ambientales relacionados con el petróleo (Khanmohammadi et al. 2012). De acuerdo con la mayoría de trabajos publicados, las aplicaciones más exitosas han sido el uso de la espectroscopia NIR y MIR combinada con PLS como método de regresión y, posteriormente, el uso de métodos de selección de variables o regiones espectrales para obtener modelos menos complejos de interpretar y más estables. En

destilados livianos y medios, se han generado modelos MLR en un NIR portátil para la determinación del contenido de aromáticos totales y poliaromáticos, azufre, ALC, y temperatura de destilación (Baldrich F. & Novoa M. 2006), modelos para predecir el análisis tipo hidrocarburo de destilados medios y gasóleos de vacío empleando espectroscopía UV-Vis (Baldrich F. & Novoa M. 2007), modelos de propiedades fisicoquímicas como densidad, CCR, contenido de S y N, metales y análisis SAR para gasóleos de vacío, empleando espectroscopia UV-Vis y NIR (Ferrer 2008) (Baldrich et al. 2010). Metodologías análogas a las mencionadas se han desarrollado para predecir propiedades de biodiesel (Liliana de Fátima Bezerra de Lira et al 2010), kerosene (M.P. Gómez-Carracedo, J.M. Andrade, M. Calviño, E. Fernández, D. Prada 2003), gasolina (Andreas A. Kardamakis 2010) y diesel (Mohammad Al-Ghouti, Yahya Al-Degs 2010), y modelos quimiométricos MIR y NIR para clasificar los residuos de destilación de gasolinas con y sin aditivos dispersantes (Silva et al. 2014).

En crudos y fracciones pesadas, se han desarrollado modelos para estimar el análisis SARA, la densidad, la viscosidad, el análisis elemental y el carbón micro a partir de los espectros MIR-ATR de residuos de fondos atmosféricos, de vacío y residuos hidrogenados (Hongfu Y. et al. 2006); modelos quimiométricos para estimar el análisis elemental, el análisis SARA, la densidad y el peso molecular de fracciones residuales de crudos a partir de NIR (Satya et al. 2007), modelos para predecir la gravedad API y la curva de destilación (TBP) de crudos brasileños e internacionales por NIR (Pasquini & Bueno 2007), desarrollo de correlaciones PLS para predicción de rendimientos de productos de la destilación del crudo y predicción de propiedades fisicoquímicas y destilación simulada de los mismos, a partir del espectro RMN de ^1H y PLS (Molina et al. 2007). También se han desarrollado modelos para predicción de gravedad API, viscosidad cinemática y contenido de agua en crudos medios y pesados, mediante FTIR-ATR y métodos de calibración multivariada no lineal (Máquina de Vectores, SVR) y PLS (Filgueiras et al. 2014), así como modelos para el análisis SARA de crudos usando espectroscopias NIR y MIR, empleando análisis PCA y PLS (Aske et al. 2001).

El análisis de la información contenida en múltiples bandas de absorción en los espectros de infrarrojo ha permitido el desarrollo y cálculo de índices estructurales y funcionales a partir de las relaciones de bandas de absorción, desarrollados para caracterizar la composición química de los

crudos y fracciones. En diversas publicaciones, se han identificado las regiones del espectro infrarrojo de mayor importancia para desarrollar los modelos y se ha realizado la asignación de las bandas del espectro IR con el fin de establecer las relaciones entre la estructura y funcionalidad de diferentes tipos de muestras.

En publicaciones de los últimos diez años, se han identificado las bandas más importantes para caracterización de la estructura química de las diferentes fracciones de crudos e hidrocarburos pesados. Abbas O. y otros autores, obtuvieron parámetros para predecir la calidad de crudos mediante FTIR-ATR y fluorescencia ultravioleta asíncrona (SUUVF) usando análisis multivariado (PLS). La gravedad API y los índices funcionales y estructurales que caracterizan las estructuras alifáticas y aromáticas presentes en los crudos se estimaron mediante el cálculo del índice de aromaticidad, índice de alifaticidad e índice alifáticos/aromáticos, con base en la identificación de las regiones espectrales que describen las tensiones del anillo aromático a 1600 cm^{-1} y las flexiones de los grupos $-\text{CH}_2-$ y $-\text{CH}_3$ a 1460 cm^{-1} y 1376 cm^{-1} , respectivamente (Abbas et al. 2012). Meléndez y otros autores, desarrollaron modelos predictivos para el análisis SARA de crudos colombianos, mediante espectroscopia FTIR-ATR y PCA, PLS. Identificaron las regiones de mayor importancia en el espectro infrarrojo para construir los modelos en los intervalos de $690\text{-}1600\text{ cm}^{-1}$ y $2800\text{-}3000\text{ cm}^{-1}$ (Meléndez et al. 2012). Orrego-Ruiz y otros autores, desarrollaron una metodología basada en espectroscopia MIR-ATR y PLS-R para predicción del contenido de asfaltenos en residuos de vacío de crudos colombianos. Hallaron las variables X más importantes para correlacionar estructura y funcionalidad de los asfaltenos y obtuvieron parámetros moleculares a partir del espectro FTIR. Concluyeron que la estructura química de los asfaltenos está muy relacionada con los grupos funcionales específicos que están contenidos en tres regiones del espectro IR: $3000\text{ - }2800\text{ cm}^{-1}$, $1600\text{ - }1250\text{ cm}^{-1}$ y $870\text{ - }715\text{ cm}^{-1}$ (Orrego-Ruiz et al. 2011).

Se han identificado y asignado los diferentes tipos de vibraciones C-H en varios intervalos (figura 1) para crudos y fracciones. Las bandas del primer intervalo se caracterizan por fuertes absorciones, provenientes de las vibraciones de tensión y flexión C-H de los grupos metilo ($-\text{CH}_3$) y metileno ($-\text{CH}_2-$) en el intervalo de $2950\text{ a }2850$ y $1446\text{-}1377\text{ cm}^{-1}$. En el segundo intervalo se encuentra una mezcla de bandas asociadas a vibraciones de grupos alifáticos y aromáticos, destacándose la banda asociada a vibraciones de tensión del C=C aromático, y las bandas entre 900 y 700 cm^{-1}

asignadas a vibraciones del C-H aromático de flexión dentro y fuera del plano (Castro & Vazquez 2009), (Li et al. 2013), (Gaweł et al. 2014).

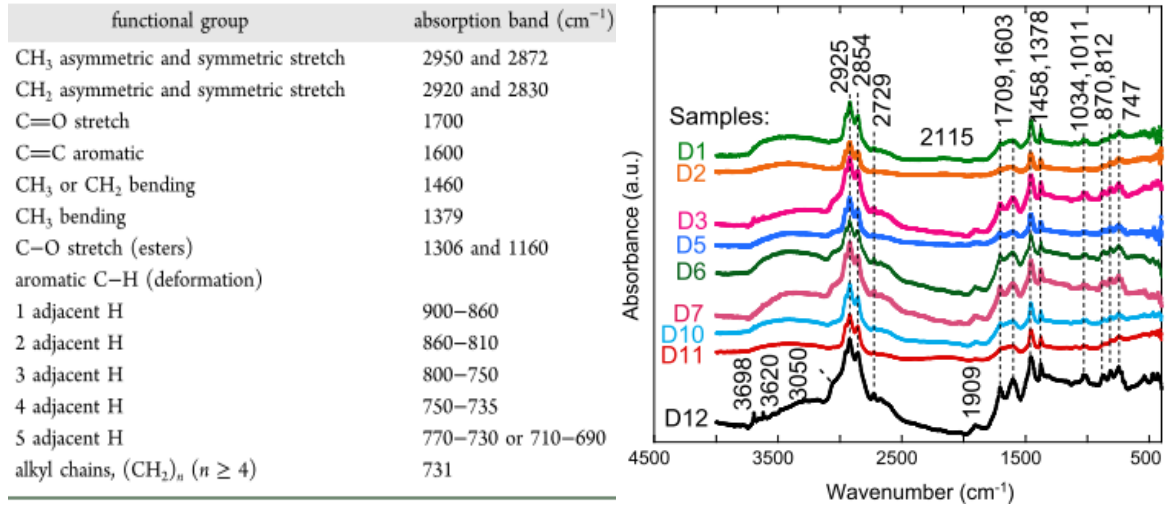


Figura 1. Asignación bandas de absorción IR para crudos y espectros FTIR-ATR de crudos pesados (Gaweł et al. 2014), (Li et al. 2013).

De los trabajos publicados hasta el presente, pocos se acercan a lo propuesto en esta investigación. Chu y otros autores, propusieron una estrategia para la identificación rápida de crudos y predicción del assay básico de crudos basada en el método del coeficiente de correlación de ventana móvil y librería espectral NIR combinada con la base de datos de assays de crudos. Usaron 256 crudos para construir una librería espectral NIR y la base de datos de assay con las propiedades fisicoquímicas (API, número de ácido, residuo de carbón, azufre, nitrógeno, contenido de ceras, asfaltenos y curva TBP). El método permite distinguir diferencias muy pequeñas entre dos espectros. Los resultados mostraron que no existen crudos idénticos en la librería NIR. La metodología permite obtener la información analítica de un crudo desconocido si la base de datos de assay contiene el crudo, si no lo contiene, predice los datos del crudo más similar a la muestra (Chu et al. 2011).

Algoritmos genéticos. Los algoritmos genéticos (AG) han sido usados con el fin de seleccionar las mejores variables para calibración multivariada y mejorar así la habilidad predictiva de los modelos al reducir su complejidad. Son herramientas de optimización que llevan a cabo una

búsqueda global al azar, dentro de un espacio de alta dimensión. Se han aplicado a muchos datos espectrales y se ha demostrado que dan mejores resultados que los modelos obtenidos usando todo el espectro. Su uso se ha extendido a aplicaciones relacionadas con datos espectroscópicos como aditivos en polímeros, en la determinación espectrofotométrica de multicomponentes, en la selección de píxeles en imágenes, en la determinación de la composición cuantitativa de textiles por NIR, entre otros (Durand et al. 2007).

Leardi y diversos autores han usado algoritmos genéticos (AG) para seleccionar los mejores conjuntos de variables y combinarlos con PLS para mejorar la capacidad de predicción de los modelos obtenidos por PLS (Leardi & Lupiáñez González 1998), (Leardi et al. 2002), (Ghasemi et al. 2003), (Sarkhosh et al. 2014). Se presentan los riesgos de modelar el ruido al realizar muchas evaluaciones ó de obtener muy pocas variables si se realizan pocas evaluaciones. Si una pequeña parte del dominio es explorado, entonces el resultado final puede depender de la generación al azar realizada sobre la población original, por lo cual, los resultados de diferentes corridas pueden ser sustancialmente diferentes. Debido a lo anterior, se aconseja realizar un alto número de diferentes corridas para extraer la información más relevante. Esta información se puede obtener de la frecuencia con la cual una variable es seleccionada en cada corrida. Se concluye que además de mejorar la habilidad predictiva de los modelos, los AG *también ayudan en la interpretación espectral, a entender cuáles son las regiones espectrales que correlacionan con una característica específica de la muestra*. Una de las principales desventajas de los AG es el riesgo de sobreajuste.

En otra publicación (Leardi et al. 2002), los autores realizaron selección de variables para calibración multivariada usando algoritmos genéticos para concentraciones de aditivos en películas poliméricas mediante FT-IR. Confirmaron que mediante AG se pueden seleccionar variables que dan buenas soluciones en términos de la habilidad predictiva e interpretativa. La metodología permite al usuario remover variables y llevar a cabo el AG con tamaños de ventana más pequeños, compatibles con las características espectrales de interés y sin perder características finas del espectro que son importantes para el modelo de calibración. Encontraron que el desempeño del algoritmo disminuye cuando se usan más de 200 variables, aumentando el riesgo de sobreajuste. Las características espectrales cuyas variables han sido seleccionadas en

todos los modelos pueden considerarse que son las que aportan información relevante y deben ser retenidas en el modelo final.

La eficiencia de la selección de variables también ha sido evaluada usando otras técnicas como el Intervalo de mínimos cuadrados parciales (iPLS), por eliminación de variables no informativas por mínimos cuadrados parciales (UVE), (Rocha et al. 2016). Los autores obtuvieron mejores resultados para los modelos MIR (en términos de los errores de predicción, RMSEP) al comparar estas técnicas con los modelos PLS globales (usando todas las variables), con algoritmos genéticos (GA) y Sinergia de intervalo de mínimos cuadrados parciales (siPLS), en fracciones de destilados de crudo en el intervalo de 15 a 500°C, para la propiedad azufre (figura 2).

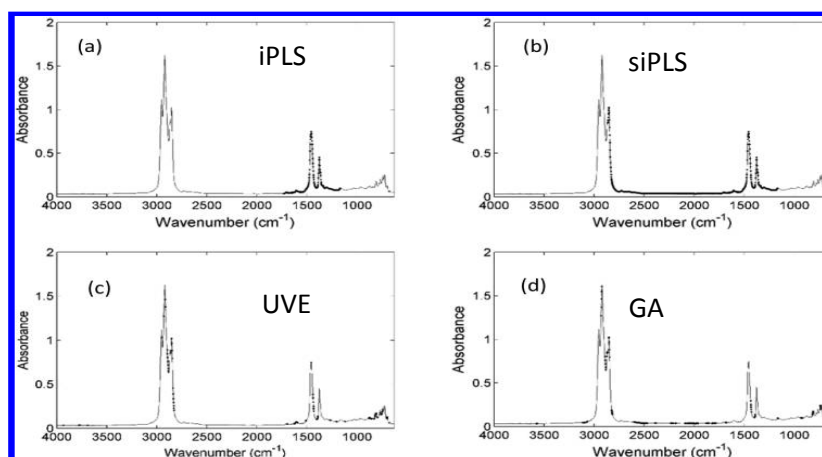


Figura 2. Regiones seleccionadas para predicción de contenido de azufre por MIR con diferentes métodos de selección de variables (Rocha et al. 2016).

De acuerdo con la revisión bibliográfica realizada, existe una relación entre las propiedades fisicoquímicas de una muestra de crudo o fracción con su estructura molecular. Dado que los crudos contienen las fracciones atmosféricas, es lógico suponer que las propiedades de éstas pueden correlacionarse con el espectro de los crudos mediante el desarrollo de modelos quimiométricos. Se ha demostrado que la selección de ciertas regiones del espectro infrarrojo permite obtener modelos de mayor precisión que usando todo el espectro, lo cual sustenta la decisión de seleccionar las variables más informativas para la construcción de los modelos mediante algoritmos genéticos. Adicionalmente, hasta el momento no se han desarrollado

modelos quimiométricos para predicción de propiedades de destilados a partir de sus crudos madre, aspecto que lo convierte en un tema de investigación a explorar.

1.3 MARCO TEÓRICO

1.3.1 EL Petróleo

El petróleo es una mezcla compleja de hidrocarburos líquidos y gaseosos, de origen fósil, formados por carbono e hidrógeno, con trazas de compuestos que contienen azufre, nitrógeno y oxígeno y metales (níquel y vanadio y sales de calcio, sodio y magnesio). Están presentes los grupos de parafinas, naftenos y aromáticos (Riazi 2005), además de los compuestos orgánicos heteroatómicos (azufre, oxigenados y nitrogenados) y organometálicos de níquel y vanadio. En general, la composición elemental del petróleo varía dentro de los intervalos presentados en la tabla 1.

Tabla 1. Composición elemental del petróleo

| <i>Elemento</i> | <i>Concentración (% en</i> |
|---------------------|----------------------------|
| Carbono (C) | 83.0-87.0 |
| Hidrógeno (H) | 10.0-14.0 |
| Nitrógeno (N) | 0.1-2.0 |
| Oxígeno (O) | 0.05-1.5 |
| Azufre (S) | 0.05-6.0 |
| Metales (Ni, V, Cu) | <1000 ppm (0.1%) |

(Bro & Smilde 2014)

Los crudos pueden clasificarse según varios criterios, dos de los más relevantes son la densidad API (ver tabla 2) y la composición química. La densidad definida como Gravedad API por el American Petroleum Institute, se basa en la comparación de la densidad del crudo a una temperatura dada con la densidad del agua a la misma temperatura. A mayor gravedad API el petróleo será más liviano.

Tabla 2. Clasificación del crudo según su densidad API

| <i>Aceite crudo</i> | <i>Densidad (g/cm³)</i> | <i>Gravedad API</i> |
|----------------------------|---|----------------------------|
| Extra pesado | >1.0 | 10 |
| Pesado | 1.0 – 0.92 | 10.0 – 22.3 |
| Mediano | 0.92 – 0.83 | 22.3 – 31.1 |
| Ligero | 0.87 – 0.83 | 31.1 – 39 |
| Superligero | < 0.83 | >39 |

(Pemex n.d.)

Según su composición química los crudos se clasifican en Parafínicos (livianos a medios), Nafténicos (crudos pesados, viscosos, con alto contenido de azufre) y Aromáticos (crudos pesados y extrapesados).

1.3.1.1 Fraccionamiento de los crudos

Para conocer los rendimientos y la calidad de las fracciones resultantes del proceso de refinación es necesario estudiar y evaluar cada crudo en unidades de destilación a nivel de laboratorio (ver figura 3). Durante la destilación, el crudo se convierte en una serie de fracciones donde cada una es una mezcla de hidrocarburos con un rango específico de puntos de ebullición. El residuo atmosférico tiene compuestos con número de carbonos mayores a 25 y el residuo de vacío, mayores a 50 (Riazi 2005). El número de cortes obtenidos depende de las necesidades puntuales de la refinería, y puede variar entre 4 y 24 cortes.

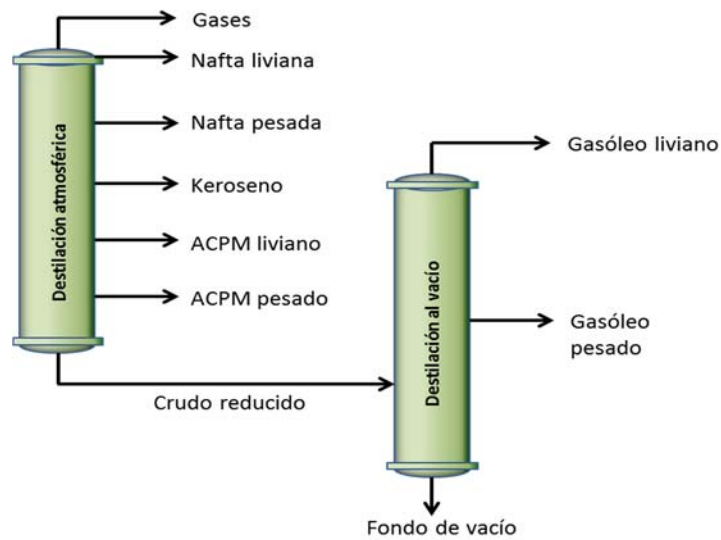


Figura 3. Esquema general de destilación de crudos. (Ecopetrol S.A. - ICP).

En la tabla 3 se presentan algunas de las fracciones obtenidas de las columnas de destilación atmosférica y al vacío.

Tabla 3. Algunas fracciones de destilación del crudo

| <i>Fracción</i> | <i>Intervalo aproximado de carbonos</i> | <i>Intervalo de Ebullición (°C)</i> |
|-----------------------------|---|-------------------------------------|
| Gases livianos | C2-C4 | -90 a 1 |
| Gasolina (liviana y pesada) | C4 - C10 | -1 a 200 |
| Naftas (liviana y pesada) | C4 - C11 | -1 a 205 |
| Jet | C9 – C14 | 150 – 255 |
| Keroseno | C11 - C14 | 205 – 255 |
| Diesel | C11 - C16 | 205 – 290 |
| Gasóleo liviano | C14 - C18 | 255 – 315 |
| Gasóleo pesado | C18 - C28 | 315 – 425 |
| Ceras | C18 - C36 | 315 – 500 |
| Bases Lubricantes | > C25 | > 400 |
| Gasóleo de vacío | C28 - C55 | 425 – 600 |
| Residuo de vacío | > C55 | > 600 |

(Riazi 2005)

A nivel de laboratorio, el fraccionamiento se realiza mediante la destilación del punto de ebullición real ó TBP (True Boiling Point), bajo la norma ASTM D2892 (American Society for Testing and Materials, 2013). Este método puede usarse para todas las fracciones hasta un punto de corte máximo de presión reducida alrededor de 350°C, equivalente a la temperatura atmosférica (AET, Atmospheric Equivalent Temperature) utilizada para minimizar el craqueo de la muestra. Más allá de la AET, es necesario continuar la destilación a presión reducida aplicando la norma ASTM D1160 (American Society for Testing and Materials, 2013a) o la ASTM D 5236 (American Society for Testing and Materials 2013c) para destilaciones al vacío. De esta manera, se obtienen los cortes para posterior caracterización fisicoquímica.

1.3.1.2 Evaluación y caracterización de crudos

La caracterización de los crudos se realiza para definir su calidad, seleccionar esquemas de procesamiento en las refinerías, ingeniería de diseño de plantas, evaluar la viabilidad económica de nuevos descubrimientos, y para el diseño de nuevos equipos y desarrollo de correlaciones (American Society for Testing and Materials 2003).

Inspección preliminar. Este tipo de evaluación, también denominada Assay Tipo I, se realiza con el objetivo de identificar cambios en la calidad del crudo y realizar una evaluación económica preliminar. Permite obtener las propiedades fisicoquímicas básicas del crudo (sin fraccionar), como contenido de agua y sedimento, gravedad API, entre otras.

Evaluación detallada. Comprende los Assays Tipo II y III, en los que se realizan diferentes cortes. El crudo y las fracciones son caracterizados fisicoquímicamente con base en normas internacionales, para obtener información de su calidad y rendimientos. En las tablas 4 y 5 se presentan las fracciones de las destilaciones atmosférica y al vacío.

Tabla 4. Esquema de Fraccionamiento de Crudos (Assay Tipo II)

| <i>Fracción</i> | <i>Intervalo de Ebullición (°C)</i> |
|------------------|-------------------------------------|
| Gases | C1 – C4 |
| Nafta Liviana | C5 – 71.1 |
| Nafta pesada | 71.1 – 176.6 |
| Destilado medio | 176.6 – 343.3 |
| Gasóleo de vacío | 343.3 - 528.5 |
| Fondo de vacío | 528.5+ |

Ecopetrol S.A., Instituto Colombiano del Petróleo ICP

Tabla 5. Esquema de Fraccionamiento de Crudos (Assay Tipo III)

| <i>Fracción</i> | <i>Intervalo de Ebullición (°C)</i> |
|--------------------------------|-------------------------------------|
| <i>Destilación Atmosférica</i> | |
| Gases | C1 – C4 |
| Nafta 1 | 15 – 60 |
| Nafta 2 | 60 – 107 |
| Nafta 3 | 107 – 152 |
| Nafta 4 | 152 – 199 |
| Keroseno | 199 – 249 |
| Destilado liviano | 249 – 315 |
| Destilado pesado | 315 – 371 |
| Crudo reducido | 371+ |
| <i>Destilación al Vacío</i> | |
| Gasóleo liviano | 371 – 427 |
| Gasóleo medio | 427 – 482 |
| Gasóleo pesado | 482 – 521 |
| Fondo de vacío | 521+ |

Ecopetrol S.A., Instituto Colombiano del Petróleo ICP

Adicional a los análisis realizados al crudo madre, las fracciones obtenidas son caracterizadas fisicoquímicamente de acuerdo con los parámetros presentados en las Tablas 6 y 7.

Las propiedades de las fracciones atmosféricas objeto de estudio para construir los modelos quimiométricos en esta investigación incluyen los rendimientos de las fracciones (%v), la gravedad API, el contenido de azufre y el índice de cetano, debido a la relevancia de éstas para los procesos de refinación.

Tabla 6. Caracterización Físicoquímica del Crudo y sus fracciones de destilación (Assay Tipo II).

| Análisis | Nafta liviana | Nafta pesada | Destilado medio | Gasóleo de vacío | Fondo de vacío |
|------------------|---------------|--------------|-----------------|------------------|----------------|
| | C5-71.1°C | 71.1-176.6°C | 176.6-343.3°C | 343.3-528.5°C | 528.5+°C |
| Rendimientos | X | X | X | X | X |
| API | X | X | X | X | X |
| Densidad 15°C | X | X | X | X | X |
| % Azufre | X | X | X | X | X |
| Viscosidad 140°C | | | | | X |
| Viscosidad 160°C | | | | | X |
| Factor K | X | X | X | X | |
| Destilación | | D - 86 | | D - 1160 | Simulada |

Ecopetrol S.A, Instituto Colombiano del Petróleo ICP

Tabla 7. Caracterización Físicoquímica del Crudo y sus fracciones de destilación (Assay Tipo III)

| ANÁLISIS | Nafta | | | | | Acpm | | Gasóleo | Gasóleo | Gasóleo | Crudo | Fondo de |
|--|-------|--------|---------|-------------|---------|---------|---------|---------------|---------|---------|---------------|----------|
| | 1 | 2 | 3 | 4 | Quero | liviano | pesado | liviano | medio | pesado | Red. | vacío |
| °C | 15-60 | 60-107 | 107-152 | 152-199 | 199-249 | 249-315 | 315-371 | 371-427 | 427-482 | 482-521 | 371+ | 521+ |
| API | x | x | x | x | x | x | x | x | x | x | x | x |
| Densidad | x | x | x | x | x | x | x | x | x | x | x | x |
| Azufre | x | x | x | x | x | x | x | x | x | x | x | x |
| Carbón Conradson | | | | | | | | x | x | x | x | x |
| Acidez | | x | x | x | x | x | x | | | | x | |
| RVP | x | x | x | x | | | | | | | | |
| Pto. Fluidez | | | | | x | x | x | x | x | x | x | x |
| Ins n-C7 | | | | | | | | x | x | x | x | x |
| Factor K (UOP) | x | x | x | x | x | x | x | x | x | x | | |
| Viscosidad y temperatura | | | | | x | x | x | x | x | x | x | x |
| Cenizas | | | | | | | | | | | x | |
| Vanadio, níquel | | | | | | | | x | x | x | x | x |
| Sodio, hierro, cobre, aluminio, magnesio, calcio | | | | | | | | | | | x | x |
| Ceras | | | | | | | | | | | x | x |
| Chispa | | | | x | x | x | x | | | | | |
| Destilación | | | | D - 86 (°C) | | | | D - 1160 (°C) | | | Simulada (°C) | |
| Rendimientos | x | x | x | x | x | x | x | x | x | x | x | x |
| Número de Octano MON | x | x | x | x | | | | | | | | |
| Número de Octano RON | x | x | x | x | | | | | | | | |
| Nitrógeno Básico | | | | | | x | x | x | x | x | x | x |
| Nitrógeno Total | | | | | x | x | x | x | x | x | x | x |
| Índice de Refracción | | | | | | | x | | | | | |
| Pto. Anilina | | | | | | x | | | | | | |
| Análisis PIANO | x | x | x | x | | | | | | | | |
| Naftalenos | | | x | x | x | | | | | | | |
| Pto. Congelación | | | x | x | x | | | | | | | |
| Pto. Humo | | | x | x | x | | | | | | | |
| Ind. cetano | | | | x | x | x | x | | | | | |
| Punto de Nube | | | | x | x | x | x | | | | | |
| Pto. Ablandamiento | | | | | | | | | | | | x |
| Penetración | | | | | | | | | | | | x |
| Azufre Mercaptano | x | x | x | x | x | x | | | | | | |
| Tipo de Carbono Aromático | | | x | x | x | x | x | x | x | x | x | x |
| Análisis SARA | | | | | | | | | | x | x | x |
| Estabilidad Térmica | | | | | x | x | x | | | | | |

Ecopetrol S.A, Instituto Colombiano del Petróleo ICP.

Rendimientos de las fracciones (%V). Las muestras de crudo previamente deshidratadas, son destiladas en una unidad de 50 Litros, a presión atmosférica. Se obtiene la masa y la densidad de cada corte o fracción. Los rendimientos en masa de la destilación, son calculados con base en la masa de todas las fracciones obtenidas, incluyendo el corte del gas licuado y el residuo atmosférico. Los rendimientos (en volumen) de la destilación son calculados a partir de todas las fracciones y el residuo a 15°C, son calculados a partir de la masa y la densidad.

Gravedad API. La densidad definida como Gravedad API se basa en la comparación de la densidad del crudo a una temperatura dada con la densidad del agua a la misma temperatura. Esta propiedad física, parámetro fundamental en la definición de la calidad de un crudo o producto, se usa para convertir volúmenes medidos a la temperatura estándar de 15.56°C ó 60°F y medidas de masa de crudo en unidades de volumen y se representa por la siguiente ecuación:

$$\text{Gravedad API, grados} = \left(\frac{141.5}{\text{Gravedad específica}_{60/60^{\circ}\text{F}}} \right) - 131.5 \quad (\text{Ecuación 1})$$

Donde la gravedad específica es la relación de la masa de un volumen dado de crudo o producto líquido a una temperatura específica y la masa de un volumen igual de agua pura a la misma o a diferente temperatura. La densidad y densidad relativa fueron determinadas bajo la norma ASTM D5002 (Standard Test Method for Density and Relative Density of crude Oils by Digital Density Analyzer) y la gravedad API fue calculada bajo la norma ASTM D1250 (Standard Guide for Use of the Petroleum Measurement Tables).

Contenido de Azufre. Es uno de los más importantes atributos de calidad del crudo y puede variar desde menos del 0.1% hasta 5% en masa. Los compuestos de azufre contribuyen a la corrosión de los equipos de refinería y al envenenamiento de los catalizadores, causan corrosividad en los productos terminados y contribuyen a la polución del medio ambiente como resultado de las emisiones de óxidos de azufre producto de la combustión. Esta propiedad fue determinada mediante Espectroscopia de Fluorescencia de Rayos X bajo el método estándar ASTM D4294 (Standard Test Method for Sulfur in Petroleum and Petroleum Products by Energy Dispersive X-ray Fluorescence Spectrometry).

Índice de Cetano. Es un parámetro que permite realizar un cálculo aproximado del número de cetano. El número de cetano es un indicativo de la capacidad de ignición del combustible diésel. El desempeño de un motor diésel es una función de la relación de compresión, del tiempo de inyección, y de la manera en la cual el combustible y el aire se mezclan y el tiempo de ignición resultante desde el inicio de la inyección hasta el comienzo de la combustión. La naturaleza del combustible es un factor importante en la reducción del tiempo de ignición. Las características físicas como la viscosidad, gravedad y el punto de ebullición intermedio influyen. La composición de los hidrocarburos también es importante porque afecta las características físicas y de combustión del combustible. Los combustibles parafínicos de cadena lineal logran ignición rápidamente bajo compresión, mientras que los parafínicos ramificados y los aromáticos reaccionan más lentamente. La escala del número de cetano usa dos combustibles de referencia primarios. Al n-hexadecano (el normal cetano), le fue asignado arbitrariamente el número de cetano de 100. El segundo combustible, el α -metilnaftaleno, tiene pobres cualidades de ignición y le fue asignado un número de cetano de cero. Posteriormente el α -metilnaftaleno fue reemplazado con heptametilnonano, el cual fue calibrado contra el combustible original y asignado un número de cetano de 15. El número de cetano es definido por la siguiente ecuación:

$$\text{Número de cetano} = \% n - \text{cetano} + 0.15 (\% \text{ heptametilnonano}) \quad (\text{Ecuación 2})$$

Los combustibles de referencia primarios son usados para calibrar dos combustibles de referencia secundarios de mezclas de diésel, y son designados como T y U. El combustible T tiene un número de cetano aproximado de 75 y el combustible U tiene un número de 20. Cada set de estos combustibles es usado para calibrar los motores que definen los números de cetano de mezclas volumétricas de estos dos combustibles de referencia secundarios. A mayor número de cetano del combustible, el motor mejora la eficiencia y se reducen las emisiones. Para asegurar un desempeño aceptable, muchos motores diésel requieren un número de cetano mínimo de 40 y algunos fabricantes lo han elevado a 50 para cumplir con los requerimientos de emisiones (American Society for Testing and Materials 2003).

1.3.2 Quimiometría

La Quimiometría se define como una disciplina química que usa métodos matemáticos y estadísticos que emplean lógica formal para diseñar o seleccionar procedimientos experimentales y de medida óptimos con el fin de obtener la máxima información química relevante para analizar datos químicos (Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y., Kaufman 2003), (Martens & Naes 1992).

Pre-procesamiento espectral. El pre-procesamiento de datos consiste en realizar uno o varios procedimientos a la data original antes de aplicar cualquier algoritmo de modelamiento. El objetivo del pre-procesamiento es remover fuentes de variación que no son de interés (ruido, desviación de línea base, background, no linealidad de los datos), con el fin de enfocar el análisis en la variación de interés. Por ejemplo, en los sistemas espectroscópicos que generan dispersión (scattering), es necesario separar efectos físicos de efectos químicos. Una aproximación general para realizar el preprocesamiento espectral consiste en aplicar métodos de filtrado para remover background, ruido y corrimientos de línea base, luego normalizar para lograr que todos los datos estén a la misma escala o lograr una distribución más uniforme de las varianzas y los valores promedio y, finalmente, centrar o escalar los datos para asegurar que éstos se encuentren en niveles o escala comparables antes de construir los modelos. Algunos de los métodos de pre-procesamiento usados se describen a continuación.

Corrección de Línea Base. Se usa para ajustar el cero espectral y eliminar efectos no químicos del levantamiento del espectro, ya sea ajustando los datos al punto mínimo de la data espectral (offset) o haciendo una corrección lineal basada en dos variables definidas como cero, transformando una línea base inclinada en una línea base horizontal. La fórmula para la corrección de línea base offset es:

$$f(x) = x - \min X \quad (\text{Ecuación 3})$$

Donde x es una variable y X denota todas las variables seleccionadas para esa muestra.

Normalización. Es un método usado para lograr que todos los datos espectrales estén aproximadamente a la misma escala. Existen varios tipos de normalización:

Normalización por área: calcula el área bajo la curva del espectro. Las áreas se hacen igual a 1.

Normalización por vector unitario: normaliza el espectro respecto al vector unitario, haciendo la longitud del vector igual a 1. Usada en aplicaciones de reconocimiento de patrones.

Normalización media: normaliza el espectro dividiendo por el valor medio de cada espectro. Se usa en cromatografía.

Normalización máxima: es una normalización que polariza el espectro. A los espectros con picos de valores positivos se les asigna +1, los que presentan valores de ambos signos se les asigna -1.

Normalización por rango: escala las muestras a un intervalo común, por ejemplo, entre 0 y +1.

Normalización por pico: normaliza una muestra como la relación entre cada valor del pico y el valor de una variable seleccionada (por ejemplo, longitud de onda).

Centrado o escalado. Consiste en centrar los datos espectrales respecto a algún punto de referencia, como la media. El centrado usa el valor medio para asegurar que los datos resultantes puedan ser interpretados en términos de la variación alrededor de la media. Calcula la media de cada columna y sustrae este valor de la columna.

Calibración Multivariada. La calibración multivariada es la combinación de quimiometría con datos espectrales y consiste en modelar o correlacionar una propiedad química, biológica o física de una muestra y su respuesta medida (Lavine & Workman 2013). El álgebra matricial permite expresar la estructura de los datos como vectores y matrices, y representar conjuntos de datos en un espacio multivariado para desarrollar métodos de calibración multivariada, permitiendo la reducción del ruido y la incorporación de señales interferentes (Bro & Smilde 2014). En el caso de la espectroscopia FTIR-ATR, donde se obtienen respuestas de intensidad a varios cientos de números de onda, se requiere establecer la relación entre las intensidades de los espectros de los crudos y las propiedades de las fracciones, lo cual se logra usando técnicas de calibración multivariada.

Un espectro puede representarse como una matriz de datos espectrales X de dimensiones $[i \times j]$, donde i espectros de calibración son medidos a j discretas frecuencias. En esta matriz el número de filas corresponde al número de muestras y el número de columnas, al número de frecuencias. El conjunto de datos que contiene los valores de referencia de las muestras de calibración

(propiedad fisicoquímica) también puede representarse como una matriz **Y** con dimensiones [**i** x **k**]:

$$\begin{array}{cccccc}
 X_{11} & X_{12} & X_{13} & \dots & X_{1j} & & Y_{11} & Y_{12} & Y_{13} & \dots & Y_{1k} \\
 X_{21} & X_{22} & X_{23} & \dots & X_{2j} & & Y_{21} & Y_{22} & Y_{23} & \dots & Y_{2k} \\
 X_{31} & X_{32} & X_{33} & \dots & X_{3j} & & Y_{31} & Y_{32} & Y_{33} & \dots & Y_{3k} \\
 \dots & \dots & \dots & \dots & \dots & & \dots & \dots & \dots & \dots & \dots \\
 X_{i1} & X_{i2} & X_{i3} & \dots & X_{ij} & & Y_{i1} & Y_{i2} & Y_{i3} & \dots & Y_{ik}
 \end{array}$$

Matriz **X**
Matriz **Y**

Si **i** espectros de calibración son medidos a **j** discretas longitudes de onda, entonces **X**, la matriz de datos espectral, se define como una matriz [**i** x **j**] que contiene los espectros como filas. De manera similar, **y** es un vector de dimensión [**i** x 1] que contiene los valores de referencia para las muestras de calibración. El objetivo es determinar la propiedad **Y** de un sistema a partir de un observable experimental **X**, donde **X** y **Y** están correlacionados por un vector **p** de predicción, de dimensión [**j** x 1] que resuelve la siguiente ecuación (American Society for Testing and Materials 2012):

$$y = X^t p + e \quad \text{(Ecuación 4)}$$

Donde **X^t** es la transpuesta de la matriz **X**, **e** es el vector de dimensión [**i** x 1] llamado vector de error. El vector **e** se calcula como la diferencia entre los valores de referencia **y** y los valores estimados por el modelo. El vector **p** de predicción se calcula minimizando la suma de los cuadrados de los errores mediante la ecuación 2:

$$e^T e = \|e^2\| = (y - X^t p)^t (y - X^t p) \quad \text{(Ecuación 5)}$$

Debido a que **X** generalmente no es una matriz cuadrada, la ecuación 3, no puede ser resuelta directamente. Una alternativa para solucionarlo es determinar la matriz pseudo-inversa de **X**, **X⁺** y calcular el vector de predicción **p** mediante la siguiente ecuación:

$$X^+ y = (X X^t)^{-1} X y = p \quad \text{(Ecuación 6)}$$

La dimensionalidad de la matriz \mathbf{X} generada por los extensos rangos de frecuencias en los que se miden los espectros hace que el número de valores de absorbancia por espectro exceda el número de espectros de calibración lo que complica la predicción del vector \mathbf{p} por los métodos de mínimos cuadrados ordinarios. Por lo anterior, es necesario utilizar técnicas como la regresión por mínimos cuadrados parciales (PLS) (Vandeginste, B.G.M., Massart, D.L., Buydens, L.M.C., De Jong, S., Lewi, P.J., Smeyers-Verbeke 2003) . La solución consiste en descomponer la matriz \mathbf{X} por análisis de componentes principales (PCA) y realizar la regresión entre los vectores score que resultan y la respuesta.

Análisis de Componentes Principales (PCA). El propósito de PCA es extraer la información esencial contenida en la matriz de datos espectrales, expresando la información principal de las variables \mathbf{X} en un número menor de variables llamadas componentes principales de \mathbf{X} , reduciendo la dimensionalidad de \mathbf{j} a una combinación lineal de las variables originales de menor dimensión. El análisis PCA se realiza sobre la matriz de datos espectrales \mathbf{X} , produciendo un set de vectores ortogonales que son totalmente no correlacionados y seleccionando un subconjunto de los vectores para ayudar a mejorar la relación señal/ruido (Martens & Naes 1992). En la figura 4 se representa gráficamente la descomposición en componentes principales, de un conjunto de muestras definidas por dos únicas variables, donde la estructura original de los datos es mejor preservada en el PC1 (I_1):

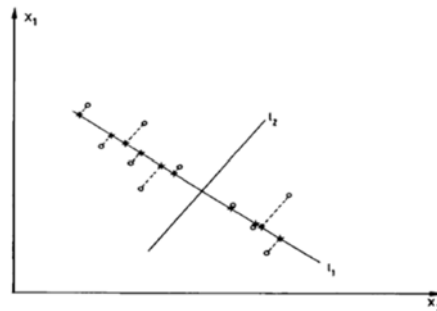


Figura 4. Representación geométrica de los componentes principales para dos variables. (Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y., Kaufman 2003).

Esta reducción es representada por la siguiente ecuación:

$$X = t_1 p'_1 + t_2 p'_2 + \dots + t_A p'_A + E \quad (\text{Ecuación 7})$$

máxima correlación con la información suministrada por la variable propiedad. PLS intenta encontrar los factores que capturan la varianza y logran la correlación, es decir, intenta maximizar la covarianza (figura 6).

PLS involucra la descomposición de la matriz de datos espectrales X , en el producto de matrices. El vector y se utiliza para obtener la descomposición de la matriz X (American Society for Testing and Materials 2012). PLS puede usarse para relacionar múltiples variables X a múltiples variables Y (PLS2) o a una única variable Y (PLS1).

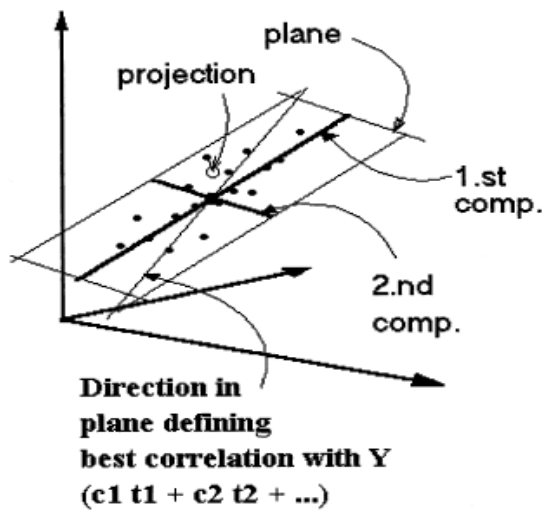


Figura 6. Representación geométrica del algoritmo PLS-R (Wold et al. 2001)

PLS se realiza mediante una serie de pasos, los cuales se repiten en un ciclo. Cada vez que los pasos se repiten, un vector de pesos w_i (de dimensión $j \times 1$), un vector de scores s_i (de dimensión $i \times 1$), un coeficiente de regresión b_i (un escalar) y un vector de loadings l_i (de dimensión $j \times 1$) son calculados. El subíndice i indica el número de veces que el ciclo se ha ejecutado, inicialmente es 1. Matemáticamente, el algoritmo PLS se ejecuta de la siguiente forma (American Society for Testing and Materials 2012):

1. Se calcula el vector de pesos de dimensión $j \times 1$:

$$X^t = y w_i^t + Z \quad (\text{Ecuación 8})$$

$$\hat{w}_i = X y \quad (\text{Ecuación 9})$$

2. Se escala el vector de pesos $\hat{\mathbf{w}}_i$ y se calcula el vector de scores normalizado \hat{s}_i de dimensión $i \times 1$:

$$X^t = s_i \hat{\mathbf{w}}_i^t + Z \quad (\text{Ecuación 10})$$

$$\hat{s} = X^t \hat{\mathbf{w}}_i^t \quad (\text{Ecuación 11})$$

$$\hat{\mathbf{w}}_i = \hat{\mathbf{w}}_i (\hat{s}^t \hat{s})^{1/2} \quad (\text{Ecuación 12})$$

$$\hat{s}_i = \frac{\hat{s}_i}{(\hat{s}^t \hat{s})^{1/2}} \quad (\text{Ecuación 13})$$

3. Se correlaciona el vector scores contra los valores de referencia para obtener un coeficiente de regresión, \mathbf{b}_i :

$$y = \hat{s}_i b_i + e \quad (\text{Ecuación 14})$$

$$b_i = \hat{s}_i^t y \quad (\text{Ecuación 15})$$

4. Se calcula el vector loading \mathbf{l}_i de dimensión $j \times 1$:

$$X = l_i \hat{s}_i^t + Z \quad (\text{Ecuación 16})$$

$$l_i = X \hat{s}_i \quad (\text{Ecuación 17})$$

5. Se calculan los residuales:

$$Z_i = X - l_i \hat{s}_i^t \quad (\text{Ecuación 18})$$

$$e_i = y - b_i \hat{s}_i \quad (\text{Ecuación 19})$$

6. Se continúa a través del ciclo, la matriz \mathbf{X} se reemplaza con los residuales de la matriz \mathbf{Z}_{i-1} del ciclo previo, y el vector \mathbf{y} se reemplaza con los residuales del vector \mathbf{e}_{i-1} . El ciclo se repite k veces para obtener los k pesos, los vectores scores y loadings, y k coeficientes de regresión.

Algoritmos Genéticos. Los algoritmos genéticos fueron desarrollados por Holland y se basan en la teoría de la evolución biológica y la teoría de Darwin de la selección natural. Es una técnica de optimización que ha sido muy empleada en varios campos y ha sido exitosamente aplicada a problemas de selección de frecuencias. Generalmente se usa para implementar un procedimiento automatizado de selección de variables para usarlo en la construcción de modelos de calibración multivariada (Xiaobo et al. 2010), (Roman M. Balabin 2011), (Laxalde et al. 2011), (Roman M. Balabin 2011), (Mehmood et al. 2012). Es un método apropiado para seleccionar longitudes de onda para PLS, MLR, Redes neuronales, entre otras. Se realiza en dos etapas: la inicial y la de

evolución (figura 7). La primera consiste en la creación de la primera generación de individuos (soluciones). En la primera etapa de su ejecución, se crea una población inicial con un determinado número de individuos (población), seleccionados al azar. Debido a que se requiere tratar matemáticamente un cromosoma, se requiere un código. Esto se resuelve representando cada variable/ventana (gen) con un código binario en un vector (cromosoma), con una celda para cada variable/ventana. El cromosoma es entonces perturbado al azar para construir la población inicial. La siguiente etapa (evolución) consiste en cinco pasos: la evaluación, la selección, la recombinación, la mutación y la reinsertión. Estos pasos son repetidos hasta que el criterio de optimización se cumple.

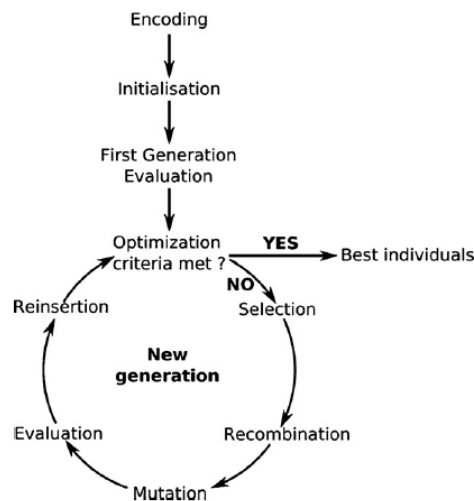


Figura 7. Esquema de optimización del algoritmo genético (Laxalde et al. 2011).

La **evaluación** consiste en la valoración de los individuos mediante una función de aptitud (fitness) que describe la calidad del modelo. Se realiza haciendo un modelo PLS para cada cromosoma, el objetivo es minimizar el error de predicción mediante validación cruzada, evaluando el error de validación (RMSECV). En el siguiente paso, la mitad de la población es **seleccionada** según su valor fitness, de modo que todas las soluciones obtenidas de la generación puedan tomar parte en la creación de una nueva generación. Los individuos con mayor fitness, tienen más probabilidad de ser seleccionados y ésta población (padres) serán usados para la creación de la población hija, por **recombinación** de los cromosomas originales. Esta recombinación es realizada mediante uno de dos mecanismos de cruce (figura 8): cruce sencillo, el cual está basado en dos cromosomas padres que son cortados en dos piezas, cada uno en un punto escogido al azar. La primera parte del gen

del individuo A es intercambiado con la primera parte del gen del individuo B, y los dos genes híbridos forman 2 nuevos individuos (C y D). El cruce doble es similar al anterior, excepto que los puntos al azar en el gen son seleccionados y las posiciones medias de los dos genes son intercambiadas. En este punto se cruzan y se colocan juntos nuevamente para formar dos cromosomas hijos, que reemplazarán a los cromosomas padres en la nueva generación (Barry M. Wise, Neal B. Gallager, Rasmus Bro, Jeremy M. Shaver, Willem Windig 2006).

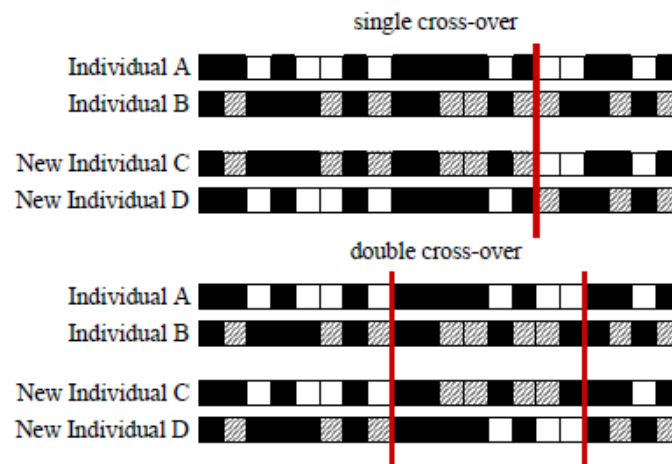


Figura 8. Esquema de la recombinación sencilla y doble
(Barry M. Wise, Neal B. Gallager, Rasmus Bro, Jeremy M. Shaver, Willem Windig 2006).

Estos cromosomas hijos son sometidos a **mutaciones**, que consisten en una inversión al azar de los valores del bit (inversión de un gen en un cromosoma). La mutación soluciona el inconveniente de que si hay alguna variable que no fue seleccionada de ninguno de los cromosomas originales, nunca podría ser seleccionada en las generaciones siguientes si las mutaciones no estuvieran presentes. Adicionalmente, reducen el riesgo de sobre-ajuste o sub-ajuste. El usuario define una tasa de mutación determinada. Finalmente, la mitad de la nueva generación está compuesta por los individuos de la generación anterior con los mejores valores de fitness (**reinserción**) y la otra mitad por los hijos generados. El algoritmo es repetido hasta que la condición de terminación se cumpla. Esta condición está basada en un criterio de convergencia, donde el algoritmo es terminado cuando un cierto porcentaje de los cromosomas son idénticos o después de un número finito de iteraciones.

El algoritmo genético combinado con PLS (AG-PLS) es uno de los métodos más usados en espectroscopia de infrarrojo. El AG encuentra valores óptimos para varias variables asociadas con el modelo de calibración, y el procedimiento PLS integra la función objetivo direccionando la optimización. El AG-PLS muestra superioridad sobre los otros métodos multivariados debido a la selección de longitudes de onda en la calibración PLS, usando un algoritmo genético sin pérdida de la capacidad predictiva y además, proporcionando información útil sobre el sistema químico. Sin embargo, los algoritmos genéticos tienen dos inconvenientes: son muy lentos en tiempo de ejecución comparado con los métodos escalonados; y significan un desafío para el usuario debido a los numerosos parámetros que se deben ajustar y que afectan el resultado del AG.

1.3.3 Espectroscopia de Infrarrojo

Introducción. La espectroscopia de infrarrojo se desarrolló a partir de 1881, con la toma del primer espectro vibracional medido por Abney y Festing. En 1905, Coblentz publicó los espectros de absorción de 120 compuestos orgánicos. Entre 1930 y 1950 se implementó el uso de la espectroscopia de infrarrojo para aplicaciones industriales (Bougeard et al. 1995), se introdujeron las rejillas de difracción y, posteriormente, el interferómetro de Michelson y la transformada de Fourier (FTIR) (Stuart 2004).

Fundamentos de absorción en la región de infrarrojo. En el espectro electromagnético, el infrarrojo corresponde a la región de números de onda comprendida entre 12800 y 10 cm^{-1} y se divide en varias regiones, como se muestra en la tabla 8.

El proceso de absorción de radiación infrarroja es un proceso cuantizado, como se muestra en la figura 9.

Tabla 8. Regiones espectrales de la radiación infrarroja

| <i>Región</i> | <i>Longitud de onda λ (μm)</i> | <i>Número de onda ν (cm^{-1})</i> |
|---------------|---|--|
| Cercano (NIR) | 0.78 – 2.5 | 12800 – 4000 |
| Medio (MIR) | 2.5 – 25 | 4000 – 400 |
| Lejano (FIR) | 25 – 1000 | 400 – 10 |

(Bougeard et al. 1995)(Donald L. Pavia, Gary M. Lampman 2001)

En un acople de moléculas, habrá una distribución de todas las moléculas entre varios niveles de energía. Siempre que una molécula interactúa con la radiación electromagnética, un cuanto de energía o fotón será emitido o absorbido. Esta energía corresponde a la diferencia neta de energía entre dos niveles vibracionales: $E_1 - E_0$ o $E_2 - E_1$, etc. Adicionalmente, para que una molécula muestre absorciones en el infrarrojo debe producirse un cambio en el momento dipolar eléctrico durante la vibración (Donald L. Pavia, Gary M. Lampman 2001).

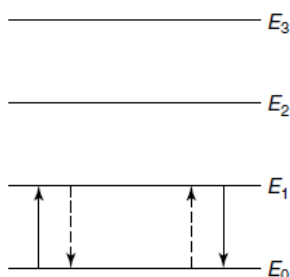


Figura 9. Niveles de energía discreta cuantizados (Stuart 2004).

En espectroscopia de infrarrojo las moléculas diatómicas vibrarán como un oscilador armónico, y la energía, E_{vib} , está dada por la ecuación 17:

$$E_{vib} = \left(\nu + \frac{1}{2} \right) h \nu_m \quad (\text{Ecuación 20})$$

Donde ν es el número cuántico vibracional, h es la constante de Planck, ν_m es la frecuencia natural del oscilador, que corresponde a la frecuencia de vibración fundamental del enlace que produce una banda de absorción en la región del infrarrojo medio.

Este modelo permite únicamente las transiciones entre niveles energéticos adyacentes ($\Delta \nu = \pm 1$) y asume que los niveles energéticos se encuentran igualmente espaciados. Sin embargo, las vibraciones en moléculas poliatómicas implican el movimiento complejo de sus átomos constituyentes y tienden a ser anarmónicas (vibraciones asimétricas sobre la posición de equilibrio). Esta anarmonicidad permite las transiciones entre estados de energía vibracional no consecutivos, ($\Delta \nu = \pm 2, \pm 3, \dots$), generando las bandas de absorción llamadas sobretonos, las cuales

son múltiplos de las frecuencias fundamentales de vibración, aunque su intensidad es mucho menor. Los niveles de energía para la vibración anarmónica se pueden escribir como:

$$E_{vib} = \left(\nu + \frac{1}{2} \right) hc \nu_0 - \left(\nu + \frac{1}{2} \right)^2 hc \nu_0 X_e \quad (\text{Ecuación 21})$$

Donde c es la velocidad de la luz en el vacío y X_e es la constante de anarmonicidad.

Modos Normales de Vibración. Las interacciones de la radiación infrarroja con la materia deben entenderse en términos de los cambios en los dipolos moleculares asociados con las vibraciones y las rotaciones. Las moléculas diatómicas tienen tres grados de libertad traslacional, dos grados de libertad rotacional y un grado de libertad vibracional. Las moléculas poliatómicas contienen muchos átomos (N) que tendrán $3N$ grados de libertad, según se observa en la tabla 9:

Tabla 9. Grados de libertad de moléculas Poliatómicas.

| <i>Tipo de grados</i> | <i>Lineal</i> | <i>No Lineal</i> |
|-----------------------|---------------|------------------|
| Traslacional | 3 | 3 |
| Rotacional | 2 | 3 |
| Vibracional | $3N - 5$ | $3N - 6$ |
| Total | $3N$ | $3N$ |

(Stuart 2004)

Una molécula únicamente puede absorber radiación cuando la radiación de infrarrojo es de la misma frecuencia que uno de los modos fundamentales de vibración de la molécula. Existen dos tipos de vibraciones: tensiones (cambio en la longitud del enlace) y flexiones (cambio en el ángulo de enlace), reflejados en las principales bandas de absorción para espectroscopia MIR.

Las propiedades de transformación de los modos vibracionales pueden determinarse mediante un análisis de coordenadas normal, por lo cual los espectros vibracionales de sistemas moleculares pueden predecirse a partir de las propiedades de simetría de los mismos. Cuando una vibración fundamental se acopla con un sobretono o bandas de combinación cercanas y de la misma simetría, la vibración acoplada se denomina Resonancia Fermi. Aunque las frecuencias rotacionales de la molécula no son activas en infrarrojo, ellas se acoplan con las vibraciones de

tensión y flexión en la molécula. Una de las razones por las que una banda es más ancha que otra en el espectro infrarrojo es debido al acoplamiento rotacional (Donald L. Pavia, Gary M. Lampman 2001), (Wörner & Merkt 2011), (Lindon 2000).

Es así como la espectroscopia IR se ha convertido en una herramienta fundamental en el análisis de crudos y productos petroquímicos, debido a que los espectros vibracionales reflejan la composición molecular cualitativa y cuantitativa de una muestra, al contener datos ricos en información embebida en múltiples bandas de absorción. Sin embargo, es necesaria la conversión de la respuesta instrumental de interés mediante el uso de técnicas de calibración multivariada. Estos métodos son necesarios para extraer información analítica, teniendo en cuenta que estas muestras son altamente complejas en términos de su composición, y las bandas espectrales resultantes están altamente solapadas. Varios estudios han determinado relaciones derivadas de bandas específicas para predecir propiedades, relaciones de absorción de aromáticos/alifáticos, índices estructurales y funcionales calculados de la relación de diversas áreas de bandas de absorción en el infrarrojo (Abbas et al. 2012), (Orrego-Ruiz et al. 2011), (Kuptsov & Arbuzova 2011).

Espectroscopia Infrarroja con Reflectancia Total Atenuada (FTIR-ATR). La espectroscopia FTIR unida a la reflectancia total atenuada (ATR) es una técnica ampliamente usada para muestras que son difíciles de analizar mediante métodos de transmitancia convencionales. ATR permite el análisis de muestras con poca o ninguna preparación, lo cual agiliza el análisis. Usa los fenómenos de reflexión interna y refracción (figura 10). La ley de refracción de Snell, se describe mediante la siguiente ecuación:

$$n_1 \sin \theta_i = n_2 \sin \theta_t \quad (\text{Ecuación 22})$$

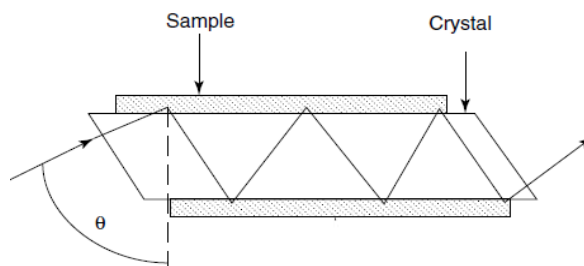


Figura 10. Esquema general de una celda de reflectancia total atenuada (ATR)b (Stuart 2004).

Donde θ_i es el ángulo del haz incidente, θ_t es el ángulo del haz refractado, n_1 y n_2 son los índices de refracción de los medios 1 y 2, respectivamente. La reflexión interna sólo puede ocurrir cuando θ_t es mayor que θ_i . Lo anterior significa que el índice de refracción n_2 debe ser menor que n_1 . La región de reflexión total comienza cuando θ_t llega a ser 90° , es decir, en el ángulo crítico, θ_c .

En ATR, un haz de radiación se direcciona a un elemento de reflexión interna IRE (Internal Reflection Element) ópticamente denso, con un alto índice de refracción. El haz entra con un determinado ángulo. El haz de IR se refleja en la superficie interna del IRE, y crea una onda evanescente en la interfase IRE/muestra, la cual se proyecta más allá de la superficie del IRE hasta la muestra que está en contacto con la superficie del mismo. La onda evanescente penetra una pocas micras ($0.5 - 5 \mu$), dependiendo de los índices de refracción del IRE y del medio, así como del ángulo de incidencia. El haz penetra una fracción de una longitud de onda más allá del cristal hasta la muestra, perdiendo energía a la longitud de onda a la que la muestra la absorbe. Parte de la energía de la onda evanescente es absorbida por la muestra y la radiación reflejada (atenuada) es captada por el detector (Stuart 2004), (Lindon 2000).

Varios aspectos afectan el espectro final obtenido. Uno de ellos es la profundidad de penetración del haz incidente, el cual es función de la longitud de onda λ , del índice de refracción del IRE (n_1), del ángulo de la radiación incidente θ con respecto a la superficie del IRE y del índice de refracción de la muestra (n_2). La profundidad de penetración (dp) está definida por la siguiente ecuación:

$$dp = \frac{\lambda/n_1}{2\pi\sqrt{\sin^2\theta - (n_2/n_1)^2}} \quad (\text{Ecuación 23})$$

Otros factores afectan la profundidad de penetración del haz incidente: el ángulo crítico, el número de reflexiones internas, el íntimo contacto de la muestra con el IRE y las características de éste. El ángulo crítico se cumple, si el ángulo de incidencia θ o el índice de refracción del IRE (n_1) son muy bajos o si el índice de refracción de la muestra (n_2) es muy alto; o una combinación de estos tres factores. En este punto se observará un espectro puro de ATR. Si el ángulo crítico no se cumple, se observará un resultado combinado de ATR y reflectancia externa.

El aumento del número de reflexiones internas permite observar componentes minoritarios en la muestra. Al aumentar la longitud efectiva del paso óptico, se incrementa el número de reflexiones

internas del cristal ATR y, por consiguiente, se aumenta la posibilidad de observar bandas menores en la muestra al obtenerse espectros más intensos. Típicamente se usan 10, 25 o más reflexiones internas. Existen diversos materiales que son usados como IRE, los cuales varían dependiendo del tipo de muestras a analizar, de las propiedades de las mismas, del intervalo espectral deseado, entre otros. Para que suceda la reflectancia interna, el índice de refracción del IRE debe ser mucho mayor que el de la muestra (para muestras orgánicas n es aproximadamente 1.5). En general, los cristales usados en ATR tienen valores de índice de refracción entre 2.38 y 4.01 (tabla 10), y son muy poco solubles en agua.

Tabla 10. Características materiales para cristales ATR

| Material | Rango espectral (cm ⁻¹) | Índice de Refracción (n_1) | Dureza (Kg/mm) | Propiedades |
|-------------------------|-------------------------------------|--------------------------------|----------------|---|
| KRS-5 (yoduro de talio) | 17000-250 | 2.4 | 40 | Soluble en bases, ligeramente soluble en agua, insoluble en ácidos. Muy tóxico. |
| ZnSe | 20000-500 | 2.4 | 130 | Insoluble en agua, solventes orgánicos, ácidos y bases diluídas (uso a pH de 4 a 8). Tóxico. |
| Ge | 5000-550 | 4.0 | 550 | Insoluble en agua, muy frágil. Puede sufrir quiebres o rompimiento del cristal. |
| Diamante | 50000 – 50 | 2.4 | 9000 | Robusto debido a su alta dureza. Ampliamente usado en diversas aplicaciones. |
| Diamante/ZnSe | 20000 – 500 | 2.4 | 5700 | El ZnSe es usado como elemento que restringe el rango espectral. Muy blando, puede sufrir arañazos, quiebres o rompimiento del cristal. |

(Stuart 2004), (Bruker OptiK International 2012).

Asignación de bandas espectrales IR. De acuerdo con la literatura revisada, en la tabla 11 se resumen las principales bandas espectrales identificadas y asignadas a muestras de crudos, las cuales reflejan su composición molecular.

Tabla 11. Asignación bandas IR para Crudos

| <i>Grupo Funcional</i> | <i>Modo vibracional</i> | <i>Bandas de absorción (cm⁻¹)</i> | <i>Referencia</i> |
|---|-------------------------|--|--|
| O-H (libre) | Tensión | 3650-3500, 3620 | |
| O-H ó N-H (enlace de H intermolecular) | Tensión | 3500-3200, 3698-3479 | Castro y Vásquez, 2009, K. Li et al., 2013 |
| C-H (C=C y aromáticos) | Tensión | 3050-3000, 3050 | |
| R-CH ₃ | Tensión asimétrica | 2950, 2951 | Castro y Vásquez, 2009 y Gawel B. et al., 2014, K. Li et al., 2013 |
| R ₂ -CH ₂ | Tensión asimétrica | 2920, 2925 | |
| R-CH ₃ | Tensión simétrica | 2872 | Castro y Vásquez, 2009 y Gawel B. et al., 2014 |
| R ₂ -CH ₂ | Tensión simétrica | 2830, 2854 | Castro y Vásquez, 2009 y Gawel B. et al., 2014, K. Li et al 2013 |
| CH ₃ (de anillo fenílico metil sustituido) | Tensión simétrica | 2729, 2659 | K. Li et al., 2013 |
| R-COO-R, Ar-COO-Ar (ésteres) | Tensión | 1750-1730 | |
| C=O (aldehídos) | Tensión | 1740-1730 | |
| C=O (cetona) | Tensión | 1735-1705 | |
| C=O (ácidos, R-COOH ó Ar-COOH) | Tensión | 1760 ó 1720 | Castro y Vásquez, 2009 |
| C=O (amida libre) | Tensión | 1690-1650 | |
| C=O (amida asociada) | Tensión | 1650-1640 | |
| C=O | Tensión | 1709, 1700 | K. Li et al., 2013, Gawel B. et al., 2014 |
| C = C conjugado | Tensión | 1650 y 1600 | Castro y Vásquez, 2009 |
| C = C aromático | Tensión | 1600, 1603 | Castro y Vásquez, 2009 y Gawel B. et al., 2014, K. Li et al., 2013 |
| R-CH ₃ ó R ₂ -CH ₂ | Flexión | 1460, 1458 | Castro y Vásquez, 2009 y Gawel B. et al., 2014, K. Li et al., 2013 |
| R-CH ₃ | Flexión | 1379, 1378 | |
| C-O (éster) | Tensión | 1300 y 1050, 1306 y 1160 | Castro y Vásquez, 2009, Gawel B. et al., 2014 |

| | | | |
|--|-------------------------------|------------------------------|---|
| Éster alifático | | 1150-1070 | Castro y Vásquez, 2009 |
| Éster aromático | | 1275-1200 | |
| Sulfóxido | Tensión | 1033, 1034 | Castro y Vásquez, 2009, K. Li et al., 2013 |
| Anillo fenílico | Distorsión | 1312 | K. Li et al., 2013 |
| Anillo fenílico | Tensión, balanceo en el plano | 1155, 1112 | |
| Deformación C-H aromático | | | |
| 1 H adyacente | Deformación | 900-860, 870 | Castro y Vásquez, 2009 y Gawel B. et al., 2014, K. Li et al., 2013 |
| 2 H adyacente | Deformación | 860-810, 812 | |
| 3 H adyacente | Deformación | 800-750, 747 | |
| 4 H adyacente | Deformación | 750-735 | Castro y Vásquez 2009 y Gawel B. et al., 2014 Castro y Vásquez, 2009 y Gawel B. et al 2014, K. Li et al 2013 |
| 5 H adyacente | Deformación | 770-730 ó 710-690, 721 ó 702 | |
| (CH ₂) _n ; n >= 4, cadenas alquílicas | Balanceo | 731 | Castro y Vásquez, 2009 y Gawel B. et al., 2014 |

(Castro & Vazquez 2009), (Li et al. 2013), (Gawel et al. 2014).

2. METODOLOGÍA EXPERIMENTAL

2.1 PREPARACIÓN DE MUESTRAS

Un set de 54 muestras de crudos procedentes de diferentes regiones del país, fueron suministradas por Ecopetrol S.A., ICP. Los crudos fueron destilados y analizados durante un período de 3 años en los diferentes laboratorios del ICP, según el esquema de Assay Tipo III. Esta caracterización permite determinar las propiedades del crudo y sus fracciones para conocer cuál será el desempeño del crudo en los diferentes procesos en los que se vaya a utilizar, como refinación, almacenamiento y transporte, entre otros. Se aplicó la metodología estándar ASTM D2892 para la destilación y obtención de las fracciones atmosféricas. Las principales propiedades del set de crudos analizado se presentan en la tabla 12.

2.1.1 Destilación Atmosférica ASTM D2892

La destilación atmosférica de los crudos se realizó en una unidad de 50 litros que tiene una eficiencia a reflujo total de 14 a 18 platos teóricos, donde se obtienen las fracciones a diferentes alturas de la torre. Este método permitió la separación de las fracciones de queroseno (199-249°C), acpm liviano (249-315°C), acpm pesado (315-371°C) y crudo reducido (371°C+). En la tabla 13 se muestran los resultados de los rendimientos obtenidos para cada una de las fracciones de destilación atmosférica.

2.1.2 Caracterización de fracciones atmosféricas

Las fracciones de queroseno, acpm liviano, acpm pesado y crudo reducido fueron caracterizadas bajo las siguientes metodologías estándar: densidad y API (ASTM D5002, ASTM D4052), azufre (ASTM D4294) e índice de cetano (ASTM D4737). En la tabla 14 se presenta la caracterización fisicoquímica en términos del API, Azufre e índice de cetano.

Tabla 12 . Principales propiedades set de crudos analizado

| Crudo No. | API Crudo | Azufre (%w) | Número de ácido |
|------------------|------------------|--------------------|------------------------|
| 33 | 7.3 | 3.27 | 0.300 |
| 23 | 7.9 | 2.95 | 1.090 |
| 25 | 8.0 | 3.27 | 0.173 |
| 11 | 8.2 | 3.14 | ND |
| 40 | 8.2 | 3.23 | 0.154 |
| 13 | 8.2 | 3.15 | 0.144 |
| 37 | 8.5 | 3.32 | 0.403 |
| 27 | 9.8 | 2.83 | 0.119 |
| 50 | 11.2 | 1.64 | 7.689 |
| 49 | 11.9 | 1.60 | 7.428 |
| 52 | 11.9 | 2.80 | 0.668 |
| 24 | 11.9 | 2.39 | 0.381 |
| 43 | 12.2 | 2.43 | <0.100 |
| 14 | 12.5 | 2.26 | 0.267 |
| 34 | 12.8 | 1.44 | 5.265 |
| 9 | 13.3 | 1.25 | 0.495 |
| 5 | 13.4 | 1.85 | 0.872 |
| 4 | 13.6 | 1.27 | 0.505 |
| 31 | 13.7 | 1.20 | 0.265 |
| 28 | 13.7 | 1.25 | 0.295 |
| 30 | 13.9 | 1.20 | 0.093 |
| 7 | 14.1 | 1.23 | 0.297 |
| 38 | 14.4 | 1.73 | 0.116 |
| 22 | 15.4 | 1.87 | 0.142 |
| 47 | 15.5 | 1.26 | 4.774 |
| 53 | 18.1 | 1.54 | 1.328 |
| 44 | 19.7 | 1.52 | 0.337 |
| 29 | 20.0 | 2.09 | 0.208 |
| 46 | 20.0 | 2.00 | 0.617 |
| 18 | 21.1 | 0.91 | 6.690 |
| 21 | 21.3 | 1.90 | 1.010 |
| 51 | 21.9 | 1.98 | 0.300 |
| 19 | 22.0 | 1.33 | 4.991 |
| 20 | 23.0 | 1.80 | 0.121 |
| 10 | 23.1 | 0.97 | 0.373 |
| 54 | 23.3 | 1.19 | 1.736 |
| 3 | 23.6 | 1.02 | 1.669 |
| 8 | 24.3 | 1.92 | 0.218 |
| 2 | 24.5 | 0.92 | 0.400 |
| 36 | 24.5 | 1.90 | 0.375 |
| 15 | 25.3 | 1.12 | 2.526 |
| 45 | 25.4 | 0.88 | 1.626 |
| 32 | 26.2 | 1.17 | 0.299 |
| 25 | 26.6 | 0.41 | 0.681 |
| 6 | 27.2 | 0.68 | 0.180 |
| 35 | 28.5 | 0.75 | <0.100 |
| 42 | 28.8 | 0.45 | 0.314 |
| 16 | 29.7 | 0.63 | 1.454 |
| 1 | 30.0 | 0.44 | 0.132 |
| 41 | 31.2 | 0.69 | <0.100 |
| 26 | 31.8 | 0.38 | 0.183 |
| 12 | 32.9 | 1.47 | <0.100 |
| 17 | 34.1 | 0.43 | 0.726 |
| 39 | 42.0 | 0.16 | <0.100 |
| Mínimo | 7.3 | 0.16 | 0.093 |
| Máximo | 42.0 | 3.32 | 7.689 |

Tabla 13. Rendimientos fracciones atmosféricas

| Crudo No. | QUERO Rendimiento (%V) | ACPM LIVIANO Rendimiento (%V) | ACPM PESADO Rendimiento (%V) | CRUDO REDUCIDO Rendimiento (%V) |
|---------------|---------------------------|----------------------------------|---------------------------------|------------------------------------|
| 1 | 10.54 | 15.38 | 12.19 | 40.73 |
| 2 | 7.44 | 14.20 | 11.78 | 48.55 |
| 3 | 7.47 | 12.47 | 11.19 | 53.78 |
| 4 | 4.39 | 13.87 | 15.80 | 64.80 |
| 5 | 4.34 | 12.15 | 14.14 | 66.72 |
| 6 | 9.02 | 14.00 | 9.02 | 46.11 |
| 7 | 3.45 | 13.88 | 13.30 | 68.06 |
| 8 | 6.14 | 10.61 | 9.42 | 58.45 |
| 9 | 3.62 | 12.63 | 14.18 | 69.23 |
| 10 | 6.93 | 12.99 | 11.30 | 51.81 |
| 11 | 2.22 | 7.78 | 8.57 | 78.08 |
| 12 | 9.19 | 13.22 | 10.10 | 39.44 |
| 13 | 2.51 | 7.80 | 9.32 | 78.02 |
| 14 | 3.90 | 10.23 | 9.91 | 67.71 |
| 15 | 7.92 | 11.01 | 8.77 | 51.05 |
| 16 | 8.53 | 12.22 | 9.26 | 45.49 |
| 17 | 9.16 | 12.80 | 10.12 | 41.06 |
| 18 | 8.78 | 12.76 | 10.92 | 54.38 |
| 19 | 6.59 | 10.39 | 9.71 | 55.75 |
| 20 | 4.93 | 8.81 | 8.02 | 65.69 |
| 21 | 4.99 | 8.43 | 8.34 | 66.91 |
| 22 | 6.07 | 13.18 | 11.14 | 61.01 |
| 23 | 2.88 | 6.70 | 8.24 | 78.70 |
| 24 | 3.74 | 9.57 | 10.08 | 69.54 |
| 25 | 9.13 | 10.28 | 10.91 | 48.26 |
| 26 | 7.86 | 15.17 | 10.33 | 35.96 |
| 27 | 3.05 | 8.89 | 9.34 | 73.31 |
| 28 | 3.51 | 13.49 | 14.05 | 67.74 |
| 29 | 5.87 | 10.53 | 8.04 | 61.54 |
| 30 | 4.28 | 13.88 | 12.82 | 68.20 |
| 31 | 4.22 | 13.94 | 15.48 | 65.10 |
| 32 | 7.41 | 9.93 | 8.66 | 52.92 |
| 33 | 2.47 | 7.29 | 8.26 | 79.91 |
| 34 | 5.14 | 8.32 | 11.21 | 73.39 |
| 35 | 8.96 | 12.49 | 10.15 | 46.06 |
| 36 | 6.40 | 10.18 | 8.68 | 59.31 |
| 37 | 2.66 | 7.08 | 8.43 | 79.67 |
| 38 | 4.84 | 10.90 | 10.85 | 68.08 |
| 39 | 9.58 | 12.66 | 9.23 | 24.69 |
| 40 | 2.59 | 7.34 | 8.57 | 79.72 |
| 41 | 9.43 | 15.16 | 11.00 | 39.12 |
| 42 | 8.73 | 16.08 | 10.35 | 43.74 |
| 43 | 3.92 | 9.37 | 8.06 | 72.16 |
| 44 | 4.00 | 9.83 | 10.96 | 57.53 |
| 45 | 7.14 | 12.07 | 9.98 | 53.97 |
| 46 | 4.02 | 9.34 | 6.00 | 70.54 |
| 47 | 5.48 | 12.69 | 12.79 | 67.17 |
| 48 | 2.35 | 6.39 | 9.25 | 80.09 |
| 49 | 3.44 | 7.87 | 12.28 | 75.54 |
| 50 | 3.45 | 6.49 | 13.68 | 75.68 |
| 51 | 6.99 | 6.66 | 8.07 | 65.12 |
| 52 | 1.55 | 10.61 | 10.77 | 70.43 |
| 53 | 5.23 | 12.00 | 10.30 | 59.87 |
| 54 | 8.42 | 11.93 | 9.72 | 55.51 |
| Mínimo | 1.55 | 6.39 | 6.00 | 24.69 |
| Máximo | 10.54 | 16.08 | 15.80 | 80.09 |
| Media | 5.68 | 11.04 | 10.43 | 60.95 |

Tabla 14. Propiedades fisicoquímicas fracciones atmosféricas

| Crudo No. | QUERO | | | ACPM LIVIANO | | | ACPM PESADO | | | CRUDO REDUCIDO | |
|---------------|-------------------|--------------|---------------|-------------------|--------------|---------------|-------------------|--------------|---------------|-------------------|--------------|
| | API 15.6°C (60°F) | Azufre (%w) | Indice Cetano | API 15.6°C (60°F) | Azufre (%w) | Indice Cetano | API 15.6°C (60°F) | Azufre (%w) | Indice Cetano | API 15.6°C (60°F) | Azufre (%w) |
| 1 | 42.5 | 0.033 | 50.1 | 36.0 | 0.164 | 56.1 | 30.5 | 0.394 | 61.2 | 15.1 | 0.848 |
| 2 | 38.1 | 0.088 | 40.4 | 31.3 | 0.305 | 45.6 | 26.5 | 0.621 | 51.4 | 11.7 | 1.540 |
| 3 | 37.9 | 0.113 | 40.3 | 31.4 | 0.389 | 46.1 | 25.9 | 0.778 | 50.0 | 12.3 | 1.520 |
| 4 | 31.5 | 0.054 | 29.3 | 27.3 | 0.300 | 38.5 | 22.6 | 0.763 | 43.4 | 9.0 | 1.620 |
| 5 | 32.4 | 0.142 | 29.7 | 26.6 | 0.371 | 35.9 | 23.0 | 0.733 | 44.6 | 7.7 | 2.380 |
| 6 | 40.2 | 0.080 | 42.2 | 34.2 | 0.302 | 51.5 | 29.4 | 0.586 | 58.5 | 13.9 | 1.100 |
| 7 | 31.7 | 0.058 | 23.1 | 27.5 | 0.259 | 38.6 | 23.4 | 0.644 | 44.5 | 9.6 | 1.540 |
| 8 | 41.3 | 0.345 | 46.9 | 34.3 | 1.120 | 52.2 | 28.6 | 1.710 | 55.2 | 14.7 | 2.540 |
| 9 | 31.0 | 0.060 | 28.5 | 27.2 | 0.236 | 36.2 | 23.2 | 0.581 | 42.9 | 8.4 | 1.580 |
| 10 | 37.4 | 0.091 | 38.5 | 30.5 | 0.327 | 43.3 | 25.8 | 0.614 | 50.2 | 10.9 | 1.520 |
| 11 | 32.8 | 0.519 | 30.4 | 25.2 | 1.370 | 34.3 | 19.7 | 2.020 | 37.8 | 3.7 | 3.520 |
| 12 | 40.0 | 0.585 | 43.6 | 32.2 | 1.540 | 46.9 | 27.2 | 1.960 | 53.2 | 17.7 | 2.310 |
| 13 | 33.5 | 0.539 | 30.5 | 25.6 | 1.320 | 34.3 | 19.9 | 1.980 | 37.9 | 4.1 | 3.370 |
| 14 | 35.0 | 0.207 | 34.7 | 26.5 | 0.699 | 35.9 | 21.3 | 1.260 | 40.2 | 5.5 | 2.790 |
| 15 | 38.6 | 0.132 | 39.9 | 30.4 | 0.532 | 42.6 | 24.6 | 1.150 | 47.3 | 12.9 | 1.720 |
| 16 | 41.0 | 0.040 | 44.8 | 34.2 | 0.259 | 51.3 | 29.1 | 0.689 | 57.7 | 14.7 | 1.080 |
| 17 | 42.0 | 0.035 | 47.9 | 36.1 | 0.174 | 55.1 | 31.3 | 0.527 | 62.6 | 19.4 | 0.796 |
| 18 | 32.3 | 0.131 | 30.2 | 25.3 | 0.423 | 34.0 | 20.3 | 1.010 | 37.2 | 13.2 | 1.310 |
| 19 | 35.5 | 0.189 | 33.4 | 26.8 | 0.638 | 35.2 | 21.2 | 1.260 | 38.3 | 11.7 | 1.880 |
| 20 | 40.1 | 0.376 | 44.4 | 33.9 | 1.140 | 51.1 | 29.9 | 1.530 | 59.2 | 14.6 | 2.400 |
| 21 | 39.7 | 0.382 | 42.7 | 32.6 | 1.150 | 47.7 | 28.0 | 1.620 | 54.2 | 12.7 | 2.370 |
| 22 | 36.8 | 0.196 | 38.8 | 28.4 | 0.682 | 40.3 | 22.7 | 1.210 | 43.6 | 6.8 | 2.510 |
| 23 | 33.8 | 0.575 | 32.6 | 25.6 | 1.410 | 34.9 | 20.2 | 2.000 | 37.4 | 3.3 | 3.260 |
| 24 | 34.3 | 0.261 | 33.2 | 26.0 | 0.765 | 35.4 | 20.9 | 1.340 | 39.7 | 4.8 | 2.900 |
| 25 | 35.4 | 0.046 | 41.3 | 32.6 | 0.127 | 51.0 | 29.9 | 0.263 | 58.6 | 15.2 | 0.770 |
| 26 | 40.0 | 0.039 | 42.1 | 32.8 | 0.150 | 47.2 | 29.2 | 0.351 | 57.7 | 14.2 | 0.723 |
| 27 | 33.6 | 0.336 | 31.9 | 25.6 | 0.989 | 34.5 | 20.4 | 1.670 | 38.2 | 4.5 | 3.300 |
| 28 | 31.7 | 0.062 | 29.7 | 27.6 | 0.269 | 38.1 | 23.2 | 0.666 | 42.4 | 8.8 | 1.580 |
| 29 | 39.2 | 0.442 | 41.5 | 32.3 | 1.190 | 46.4 | 26.9 | 1.720 | 51.3 | 9.7 | 2.800 |
| 30 | 31.5 | 0.062 | 28.9 | 27.5 | 0.270 | 38.2 | 23.1 | 0.680 | 42.7 | 9.0 | 1.600 |
| 31 | 31.1 | 0.063 | 29.9 | 27.3 | 0.283 | 38.4 | 22.7 | 0.737 | 43.8 | 9.2 | 1.580 |
| 32 | 40.1 | 0.085 | 44.9 | 34.4 | 0.472 | 52.2 | 29.2 | 0.975 | 57.5 | 12.7 | 1.860 |
| 33 | 31.6 | 0.676 | 28.5 | 24.9 | 1.550 | 33.9 | 19.6 | 2.180 | 36.9 | 3.8 | 3.640 |
| 34 | 31.8 | 0.262 | 29.8 | 25.0 | 0.658 | 32.6 | 19.5 | 1.200 | 33.5 | 9.3 | 1.650 |
| 35 | 40.6 | 0.048 | 45.9 | 34.9 | 0.283 | 53.1 | 29.5 | 0.751 | 58.4 | 15.2 | 1.390 |
| 36 | 41.4 | 0.291 | 47.3 | 34.6 | 0.994 | 52.8 | 29.1 | 1.600 | 56.9 | 13.8 | 2.510 |
| 37 | 31.4 | 0.726 | 28.2 | 24.9 | 1.550 | 34.1 | 19.7 | 2.160 | 36.8 | 4.3 | 3.650 |
| 38 | 36.7 | 0.166 | 36.7 | 28.1 | 0.729 | 37.9 | 22.6 | 1.150 | 42.6 | 13.8 | 2.150 |
| 39 | 39.0 | 0.015 | 42.9 | 32.7 | 0.082 | 48.4 | 28.8 | 0.222 | 56.9 | 18.7 | 0.519 |
| 40 | 32.1 | 0.640 | 29.0 | 25.3 | 1.520 | 34.8 | 19.7 | 2.190 | 36.7 | 4.1 | 3.560 |
| 41 | 40.0 | 0.070 | 42.7 | 34.1 | 0.311 | 51.6 | 29.5 | 0.599 | 58.7 | 16.0 | 1.120 |
| 42 | 40.4 | 0.044 | 42.6 | 33.4 | 0.238 | 48.0 | 28.1 | 0.458 | 55.2 | 16.7 | 0.782 |
| 43 | 32.3 | 0.435 | 35.6 | 25.8 | 0.945 | 38.4 | 21.1 | 1.550 | 40.7 | 5.1 | 2.960 |
| 44 | 33.4 | 0.152 | 32.4 | 27.2 | 0.450 | 38.5 | 22.3 | 0.905 | 42.5 | 7.2 | 2.250 |
| 45 | 37.7 | 0.088 | 40.2 | 31.4 | 0.374 | 46.0 | 26.2 | 0.830 | 50.8 | 14.8 | 1.330 |
| 46 | 40.1 | 0.337 | 42.5 | 32.7 | 1.150 | 49.3 | 27.8 | 1.640 | 56.2 | 12.9 | 2.420 |
| 47 | 34.3 | 0.184 | 31.8 | 26.8 | 0.544 | 36.2 | 20.9 | 1.100 | 38.5 | 11.4 | 1.540 |
| 48 | 32.0 | 0.646 | 29.0 | 25.8 | 1.540 | 34.4 | 20.3 | 2.170 | 37.1 | 3.8 | 3.880 |
| 49 | 31.1 | 0.249 | 29.2 | 25.3 | 0.605 | 34.0 | 19.4 | 1.170 | 34.6 | 9.4 | 1.820 |
| 50 | 31.3 | 0.280 | 29.0 | 25.2 | 0.641 | 33.3 | 19.2 | 1.220 | 34.5 | 8.8 | 1.840 |
| 51 | 39.2 | 0.456 | 41.5 | 31.8 | 1.360 | 45.2 | 27.3 | 1.740 | 51.9 | 14.8 | 2.380 |
| 52 | 36.0 | 0.254 | 32.9 | 27.7 | 0.662 | 34.5 | 21.0 | 1.410 | 37.2 | 6.1 | 3.320 |
| 53 | 36.9 | 0.142 | 38.9 | 28.7 | 0.500 | 42.7 | 24.1 | 0.882 | 42.9 | 8.7 | 2.130 |
| 54 | 38.7 | 0.143 | 41.1 | 31.4 | 0.530 | 45.5 | 26.3 | 1.060 | 51.3 | 14.5 | 1.740 |
| Mínimo | 31.0 | 0.015 | 23.1 | 24.9 | 0.082 | 32.6 | 19.2 | 0.222 | 33.5 | 3.3 | 0.519 |
| Máximo | 42.5 | 0.726 | 50.1 | 36.1 | 1.550 | 56.1 | 31.3 | 2.190 | 62.6 | 19.4 | 3.880 |
| Media | 36.1 | 0.235 | 36.7 | 29.5 | 0.682 | 42.1 | 24.5 | 1.152 | 46.9 | 10.6 | 2.059 |

2.2 CARACTERIZACIÓN POR ESPECTROSCOPIA FTIR-ATR

El instrumento empleado fue un espectrómetro infrarrojo con transformada de Fourier VERTEX 70V de Bruker (figura 11). Este equipo cuenta con fuentes de radiación que cubren las regiones espectrales del infrarrojo lejano, medio y cercano en el rango de 50 a 12500 cm^{-1} , con un sistema óptico al vacío, no higroscópico que permite la operación en ambientes sin humedad. La adquisición de espectros se realizó en la región del infrarrojo medio en el modo de reflectancia total atenuada sobre una celda ATR de diamante con ángulo fijo de reflexión (45°), en el rango de 4000-650 cm^{-1} , usando un detector DTGS (sulfato de triglicerina deuterado), a una temperatura alrededor de $20^\circ\text{C} \pm 2^\circ\text{C}$ (temperatura del laboratorio), y humedad relativa en el intervalo de 35% a 60%.

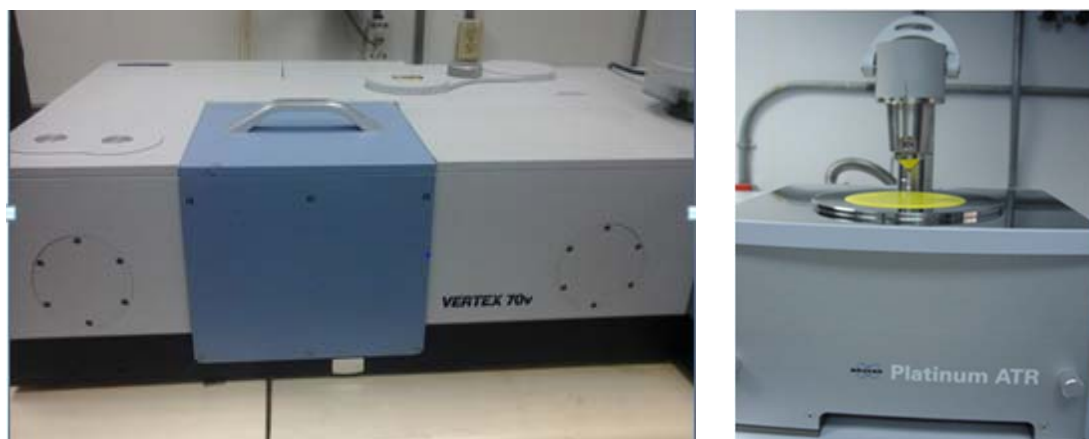


Figura 11. Espectrómetro infrarrojo Bruker Vertex 70V y accesorio ATR.

2.2.1 Determinación de parámetros de adquisición de espectros

Los espectros de los crudos fueron adquiridos usando el software OPUS Versión 6.5. Los parámetros de adquisición (resolución, velocidad del interferómetro y número de scans) fueron evaluados mediante la relación señal/ruido en la región comprendida entre 2000 cm^{-1} y 1800 cm^{-1} , en la cual sólo se observa el fondo espectral, para un nivel de confianza del 99% en el cálculo de la desviación estándar (Douglas A. Skoog, F. James Holler 2001).

$$\frac{S}{R} = \frac{\text{media de las medidas de la señal}}{\text{desviación estándar de la media}} = \frac{5\mu}{S_{\text{máxima}} - S_{\text{mínima}}} \quad (\text{Ecuación 24})$$

La desviación estándar puede estimarse para un nivel de confianza del 99% dividiendo por cinco la diferencia entre la señal máxima y la mínima, suponiendo que las variaciones respecto de la media son aleatorias. El 99% de los datos bajo una curva normal de error caen dentro de $\pm 2.5\sigma$ de la media. De esta forma se puede decir con un 99% de certeza que la diferencia entre el máximo y el mínimo abarca 5σ .

Inicialmente se evaluó la resolución a 2.0, 4.0 y 8.0 cm^{-1} . La resolución de 2.0 permite resolver mejor las bandas de absorción en la zona de alifáticos y aromáticos con respecto a las resoluciones de 4.0 y 8.0 cm^{-1} , según se observa en la figura 12, aunque se incrementa ligeramente el ruido. Se tomó como resolución de trabajo 2.0 cm^{-1} , teniendo en cuenta el interés en la obtención de la máxima resolución de bandas para asociarlas con determinados grupos funcionales.

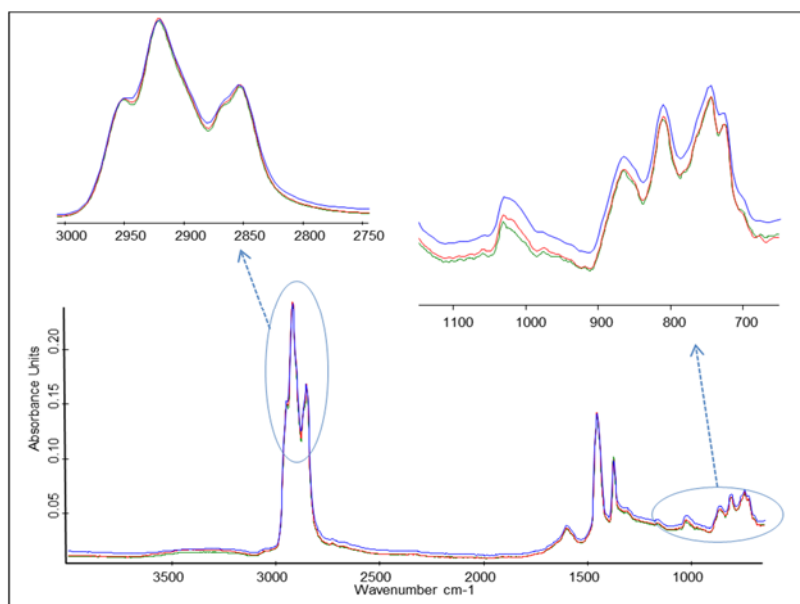


Figura 12. Espectros tomados a Resoluciones de 8, 4 y 2 cm^{-1} .

Se determinó la relación señal/ruido del fondo espectral en la región 2000 a 1800 cm^{-1} . Como se observa en la tabla 15, al aumentar la velocidad del interferómetro, baja la relación señal/ruido.

Esto se debe a que a una alta velocidad del interferómetro, el tiempo de interacción de la radiación con la muestra es corto, con lo cual disminuye la intensidad de las señales.

Tabla 15. Valores de señal/ruido a resolución de 2.0 cm⁻¹

| Velocidad Interferómetro (KHz) | No. Scans | | |
|--------------------------------|-----------|--------|--------|
| | 16 | 32 | 64 |
| 5 | 0.9720 | 0.9333 | 1.1337 |
| 10 | 0.5723 | 0.7417 | 1.0980 |
| 20 | 0.2753 | 0.3907 | 0.5910 |

Al aumentar el número de scans, aumenta la relación señal/ruido. La mejor relación señal/ruido se obtiene con 64 scans a 5 KHz. Sin embargo, debido a que la relación señal/ruido a las velocidades de 5 y 10 KHz no varía sustancialmente y que el tiempo de adquisición a 5KHz es el doble que a 10KHz, se definió como velocidad de trabajo 10KHz. Finalmente, se definieron como condiciones de toma espectral las siguientes: resolución de 2.0 cm⁻¹, velocidad del interferómetro 10 KHz y número de scans 64.

2.2.2 Verificación desempeño del espectrómetro FTIR

La repetibilidad espectral del instrumento fue evaluada tomando 10 espectros de la muestra de Crudo No. 34, de acuerdo con los parámetros de adquisición determinados previamente. Se calculó la desviación estándar relativa (RSD) de tres señales de interés presentes en el espectro (tabla 16).

Las desviaciones calculadas para cada pico de absorbancia fueron inferiores a 0.000262, con desviación estándar relativa máxima de 0.68%, asegurando de esta manera la repetibilidad de las mediciones en el espectrómetro IR. Una vez verificados los RSD, los espectros de los crudos fueron tomados por triplicado.

Tabla 16. Resultados Repetibilidad espectral Crudo No.1

| Espectro No. | 2922 (cm⁻¹) | 1601 (cm⁻¹) | 744 (cm⁻¹) |
|-------------------------|-----------------------------------|-----------------------------------|----------------------------------|
| 1 | 0.229592 | 0.027499 | 0.050522 |
| 2 | 0.229382 | 0.027317 | 0.050507 |
| 3 | 0.229381 | 0.027411 | 0.050740 |
| 4 | 0.229258 | 0.027188 | 0.050331 |
| 5 | 0.229850 | 0.027569 | 0.051181 |
| 6 | 0.229279 | 0.027008 | 0.050395 |
| 7 | 0.229478 | 0.027223 | 0.050690 |
| 8 | 0.229267 | 0.027403 | 0.050443 |
| 9 | 0.229801 | 0.027499 | 0.050312 |
| 10 | 0.229329 | 0.027597 | 0.050398 |
| Desv | 0.000218 | 0.000187 | 0.000262 |
| Media | 0.229462 | 0.027371 | 0.050552 |
| RSD | 0.09 | 0.68 | 0.52 |

2.3 PROCESAMIENTO DE DATOS (PRETRATAMIENTO)

Los espectros MIR-ATR obtenidos se analizaron en el software Unscrambler X Versión 10.2 licenciado por Ecopetrol - ICP y en el software PLS Toolbox Versión 8.0.2, licenciado por la Universidad Industrial de Santander, UIS.

Inicialmente se realizó una corrección lineal de línea base para corregir la dispersión en la señal espectral, que consiste en la transformación de la línea base con pendiente en línea base horizontal. Para determinar el pretratamiento óptimo para remover la varianza no deseada de muestras individuales, los espectros fueron evaluados mediante el cálculo de la desviación estándar relativa (RSD) a tres números de onda determinados después de aplicar varios tipos de normalización, según se presenta en la tabla 17.

En los números de onda 811.92cm⁻¹ y 1604.56 cm⁻¹ los valores más bajos de RSD se obtuvieron para la normalización por área, sin embargo, para el número de onda 2953.58 cm⁻¹ el menor RSD se obtuvo para la normalización máxima. Por lo anterior, se evaluaron estos dos tipos de

normalización, previa corrección de línea base, obteniéndose los menores RSD para el procesamiento corrección de línea base + normalización por área.

Tabla 17. Valores de RSD obtenidos para tres números de onda característicos aplicando diferentes tipos de Normalización espectral

| Pre-tratamiento | 811.92 (cm⁻¹) | 1604.56 (cm⁻¹) | 2953.58 (cm⁻¹) |
|------------------------------|-------------------------------------|--------------------------------------|--------------------------------------|
| Espectros originales | 31.99 | 39.99 | 4.89 |
| Normalización Area | 6.75 | 12.86 | 21.94 |
| Normalización Vector | 20.38 | 28.21 | 7.55 |
| Normalización Media | 6.75 | 12.86 | 21.94 |
| Normalización Máxima | 31.15 | 39.14 | 6.07 |
| Normalización Rango | 32.34 | 40.37 | 6.72 |
| Normalización Picos | 37.68 | 29.98 | 59.25 |
| Corr Línea Base + Norm. Area | 10.13 | 13.06 | 13.93 |
| Corr Línea Base + Norm. Máx | 25.76 | 29.94 | 5.89 |

En las figuras 13 y 14 se observan los espectros de los crudos sin preprocesamiento y después de aplicar el procesamiento seleccionado (Corrección Línea Base + Normalización por área). El procedimiento Mean center fue aplicado al realizar el PCA y el PLS.

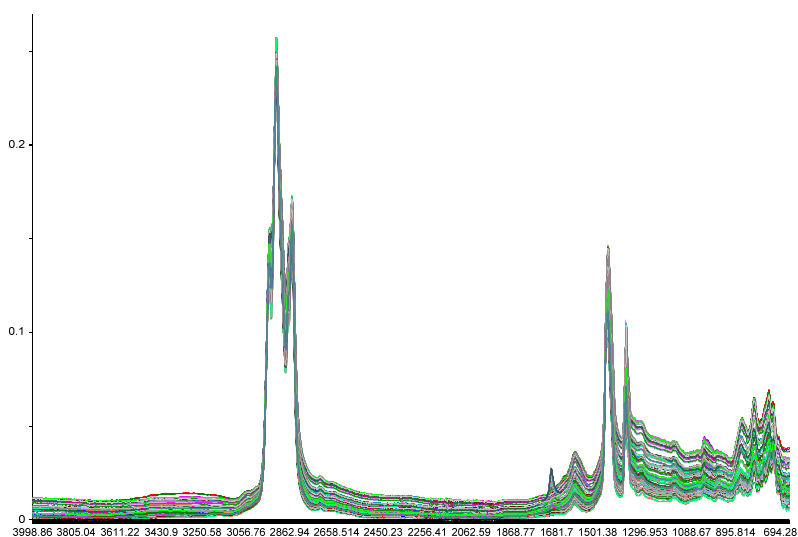


Figura 13. Espectros MIR-ATR de 54 Crudos sin pretratamiento.

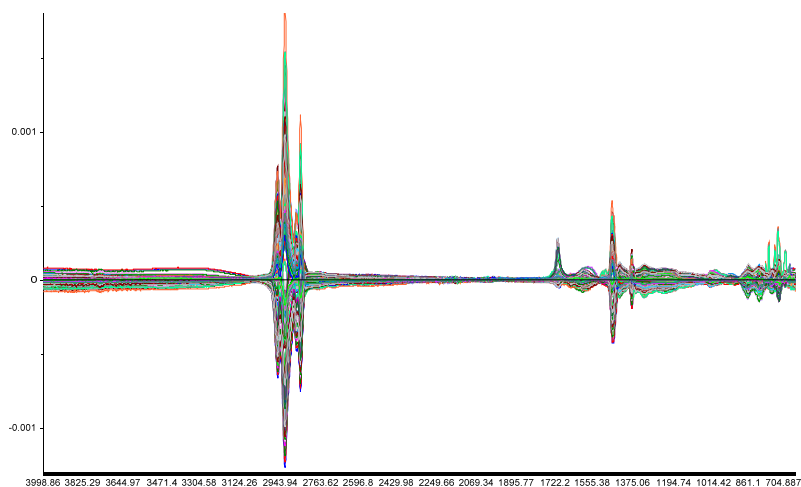


Figura 14. Espectros MIR-ATR de 54 Crudos pretratados (Linear Baseline Correction + Normalización por Area + Mean center)

La normalización por área corrige el espectro de las muestras para lograr que los datos estén a la misma escala (por ejemplo para pasos ópticos indeterminados). En este caso, el paso óptico varía dependiendo de la densidad de la muestra, por lo cual este tipo de normalización permite un buen ajuste de los espectros. El centrado de la media (Mean Center) asegura que los datos resultantes puedan ser interpretados en términos de la variación alrededor de la media. Una vez pretratados los espectros, se procedió a realizar el análisis PCA y la construcción de los modelos PLS y AG-PLS (algoritmos genéticos - PLS).

2.4 TOMA DE ESPECTROS E IDENTIFICACIÓN DE SEÑALES IR

Los espectros FTIR-ATR fueron tomados bajo las condiciones anteriormente seleccionadas: resolución de 2.0 cm^{-1} , velocidad del interferómetro 10 KHz y número de scans 64. Los espectros presentan bandas características, correspondientes a dos regiones bien diferenciadas: región de los grupos funcionales ($3200\text{-}2700 \text{ cm}^{-1}$ y $1800\text{-}1160 \text{ cm}^{-1}$) y región de huella dactilar ($1100\text{-}650 \text{ cm}^{-1}$). En la figura 15 se muestran superpuestos los espectros típicos de una muestra de crudo extrapesado, uno pesado, uno medio y uno liviano.

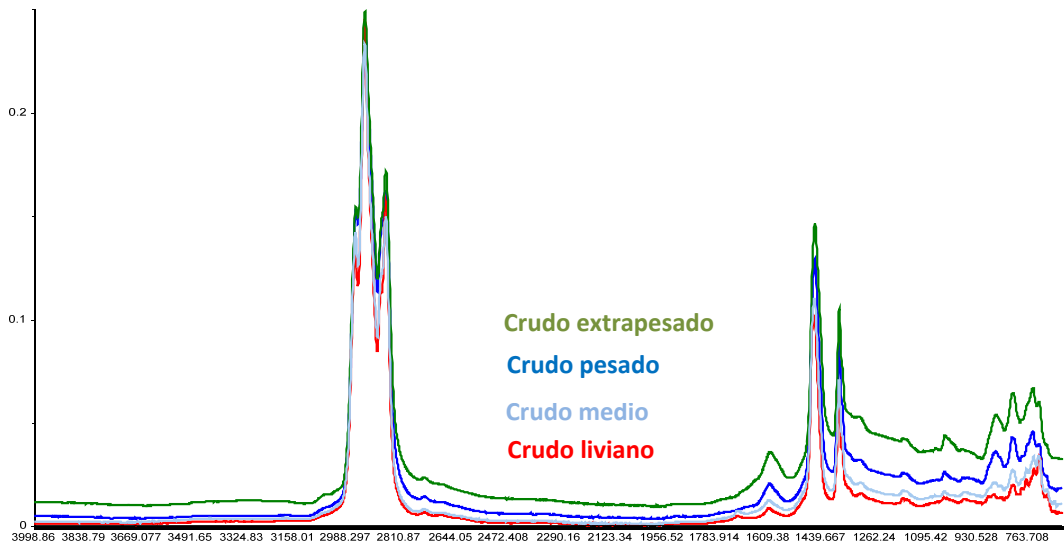


Figura 15. Espectros MIR-ATR de crudos extrapesado, pesado, medio y liviano.

En la figura 16 se identifican las principales bandas de absorción de los espectros FTIR-ATR de la totalidad de las muestras estudiadas. En la tabla 18 se muestra la asignación de estas bandas, de acuerdo con la literatura publicada (Castro & Vazquez 2009), (Li et al. 2013), (Gawet et al. 2014).

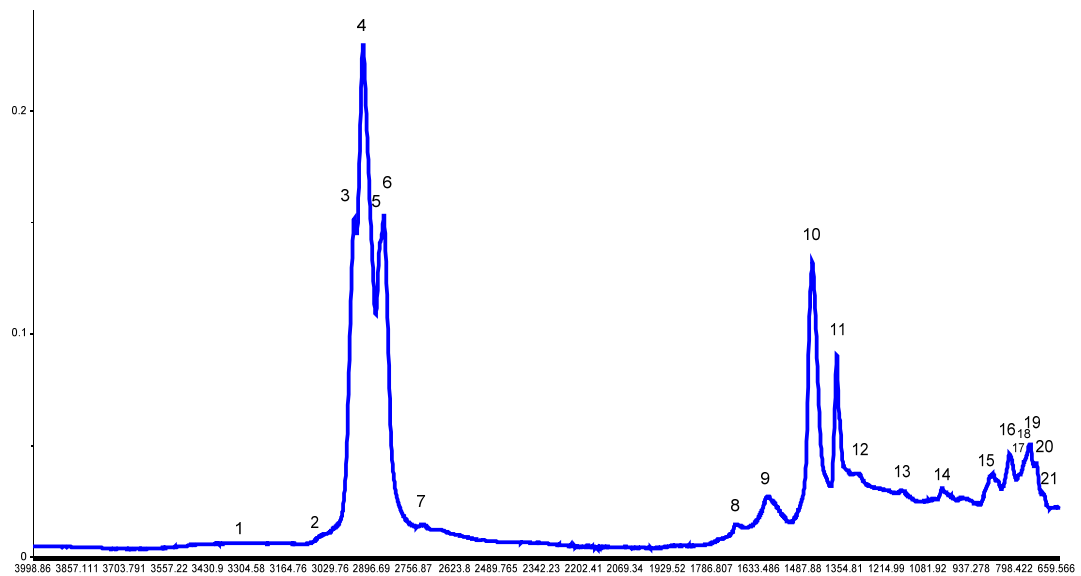


Figura 16. Identificación de bandas de absorción en un espectro de crudo.

Tabla 18. Asignación de bandas de absorción para los espectros IR de crudos analizados

| <i>Banda No.</i> | <i>Bandas de absorción (cm⁻¹)</i> | <i>Grupo Funcional</i> | <i>Modo vibracional</i> |
|------------------|--|---|-------------------------------|
| 1 | 3100-3500 | -OH, NH- (enlace H intermolecular) | Tensión |
| 2 | 3090 – 3025 | C-H (C=C y aromáticos) | Tensión |
| 3 | 2952 | R-CH ₃ | Tensión asimétrica |
| 4 | 2920 | R ₂ -CH ₂ | Tensión asimétrica |
| 5 | 2869 | R-CH ₃ | Tensión simétrica |
| 6 | 2852 | R ₂ -CH ₂ | Tensión simétrica |
| 7 | 2727 | CH ₃ (de anillo fenílico metil sustituido) | Tensión simétrica |
| 8 | 1705 | C=O | Tensión |
| 9 | 1601 | C = C aromático | Tensión |
| 10 | 1455 | R-CH ₃ ó R ₂ -CH ₂ | Flexión |
| 11 | 1376-1365 | R-CH ₃ | Flexión |
| 12 | 1305 | Anillo fenílico | Distorsión |
| 13 | 1160 | Anillo fenílico | Tensión, balanceo en el plano |
| 14 | 1032 | Sulfóxido | Tensión |
| 15 | 865 | 1 H adyacente | Deformación C-H aromático |
| 16 | 810 | 2 H adyacente | |
| 17 | 782 | 3 H adyacente | |
| 18 | 766 | 4 H adyacente | |
| 19 | 742 | 5 H adyacente | |
| 20 | 725 | (CH ₂) _n ; n >= 4 | Balanceo |
| 21 | 699 ó 674 | Anillo aromático ó bencenos ó xilenos | Flexión |

(Castro & Vazquez 2009), (Li et al. 2013), (Gaweł et al. 2014)

Las bandas del primer intervalo se caracterizan por fuertes absorciones, provenientes de las vibraciones de tensión y flexión C-H de los grupos metilo (-CH₃) y metileno (-CH₂-). En el segundo intervalo se encuentra una mezcla de bandas asociadas a vibraciones de grupos alifáticos y aromáticos. En la figura 17 se presentan los espectros de un crudo pesado y uno liviano, se puede observar que hay diferencias significativas en la resolución e intensidades de las bandas según el tipo de crudo.

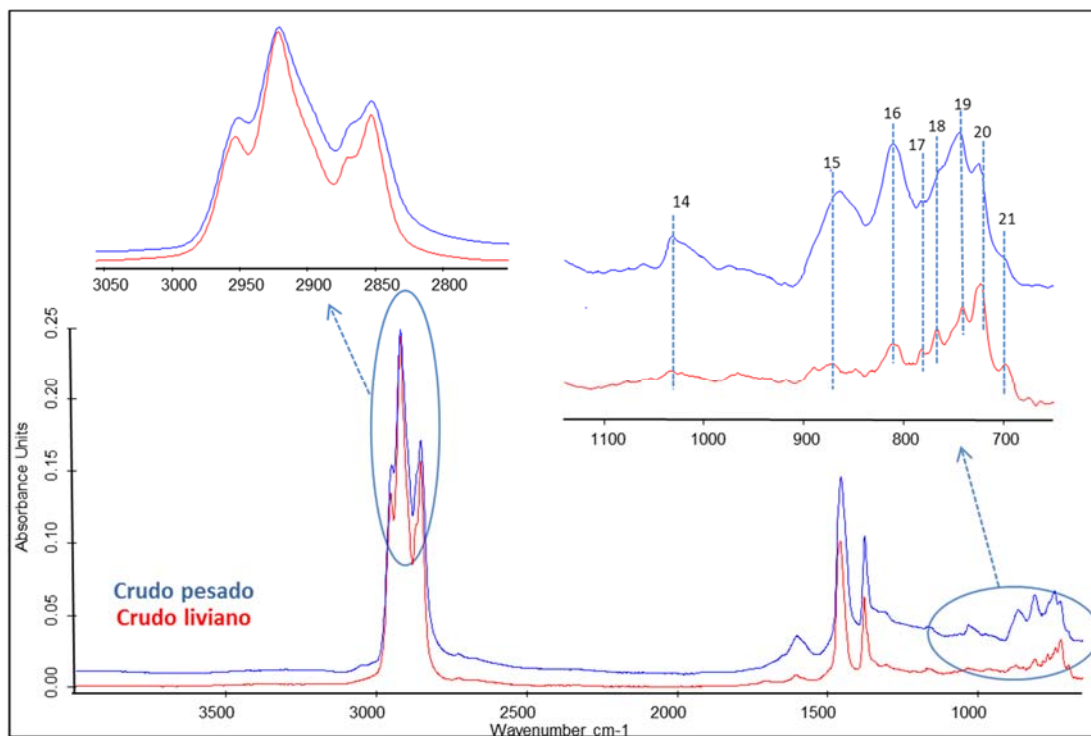


Figura 17. Regiones alifática $3050\text{-}2700\text{ cm}^{-1}$ y huella dactilar $1100\text{-}650\text{ cm}^{-1}$ de un crudo pesado y un crudo liviano

Los espectros infrarrojo de los crudos son muy similares a los espectros de crudos publicados previamente en la literatura (Castro & Vazquez 2009), (Li et al. 2013), (Gawel et al. 2014). Los espectros de los crudos analizados presentan bandas fuertes de tensión y flexión de los grupos metilo y metileno alifáticos alrededor de $2952\text{-}2852\text{ cm}^{-1}$ y 1455 cm^{-1} , respectivamente. Se observa una banda doble de intensidad media a $1376\text{-}1365\text{ cm}^{-1}$, correspondiente a las vibraciones de flexión simétrica del grupo metilo alifático. También se observa la banda $(\text{CH}_2)_n$, $n > 4$ a 725 cm^{-1} , el cual indica la presencia de cadenas alifáticas más largas, observada en todos los espectros. La banda a 1032 cm^{-1} muestra la presencia de sulfóxidos $\text{S}=\text{O}$.

Se puede observar que las bandas de tensión de los grupos metilo a 2950 cm^{-1} y 2872 cm^{-1} son más intensas y se resuelven mejor en los espectros de crudos livianos. Algo similar sucede con la banda de flexión de este mismo grupo a $1376\text{-}1365\text{ cm}^{-1}$. A 1601 cm^{-1} se observa una banda débil debida a la tensión de múltiples enlaces $\text{C}=\text{C}$ asociados a aromáticos.

En la región de huella dactilar, se observa que los crudos pesados presentan bandas más anchas en la región de deformación del C-H aromático a $900-700\text{ cm}^{-1}$, mientras que los livianos presentan bandas más agudas y, adicionalmente, dos bandas alrededor de 698 y 674 cm^{-1} que podrían ser atribuidas a la flexión del anillo aromático o al encogimiento de pequeñas moléculas de aromáticos como bencenos o xilenos (de Peinder et al. 2009).

Las bandas a 865 y 810 cm^{-1} asociadas a vibraciones del C-H aromático con 1 y 2 hidrógenos adyacentes son más intensas en los espectros de crudos pesados que en los livianos, lo cual es lógico si se tiene en cuenta que los crudos pesados contienen mayores cantidades de asfaltenos (compuestos de moléculas policíclicas aromáticas de varios anillos aromáticos fusionados) que los crudos livianos, es decir, tienen compuestos aromáticos con anillos tetra y pentasustituídos. Las bandas a 782 y 767 cm^{-1} asociadas a vibraciones del C-H aromático con 3 y 4 hidrógenos adyacentes son más intensas en los espectros de crudos livianos debido a que hay menos sustituciones en los anillos aromáticos, por presencia de compuestos aromáticos con anillos di y trisustituídos.

Otra característica que puede observarse en los espectros de crudos pesados es su línea base elevada, principalmente en la región comprendida entre 1750 y 650 cm^{-1} . Esta característica se debe a la mayor dispersión de la radiación por la presencia de cúmulos de estructuras complejas como los asfaltenos. También se presenta la banda del grupo carbonilo, C=O, a 1705 cm^{-1} , asociada a compuestos oxidados. Esta banda se observa en los crudos ácidos (con número de ácido entre 1.0 y 7.7), independiente del API del crudo.

Las diferencias anteriores, originadas por las vibraciones de los diversos grupos funcionales presentes en los crudos, permiten asociar las propiedades de los destilados atmosféricos (variables Y) a los espectros (variables X) mediante los modelos quimiométricos.

3. RESULTADOS Y ANALISIS

3.1 ANÁLISIS PRELIMINAR POR COMPONENTES PRINCIPALES (PCA)

Se realizó una inspección preliminar de los datos mediante el análisis de componentes principales (PCA) para detectar el nivel de agrupación de las muestras e identificar muestras discrepantes. Este análisis se realizó con la totalidad de las muestras de crudo disponibles (54 muestras).

En el gráfico de la varianza total explicada (figura 18), se observa que el primer componente principal explica el 96% de la varianza de los datos y el segundo PC explica el 2%. Estos dos componentes resumen la mayor variación existente en los datos que cualquier otro par de componentes.

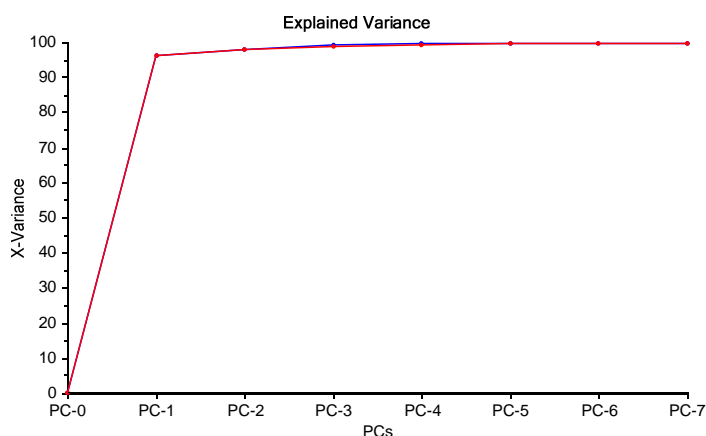


Figura 18. Varianza explicada por los dos primeros componentes principales (PC).

El gráfico de Scores (puntajes) de la figura 19, describe las relaciones entre las muestras. Se evidenciaron grupos de muestras de crudos según su densidad API, lo cual está asociado a la naturaleza química de cada muestra de crudo.

En el cuadrante izquierdo se ubican las muestras de crudos superlivianos (SL), livianos (L) y medios (M) y parte de los pesados (P); en el cuadrante derecho se ubican la mayoría de muestras de

crudos pesados (P) y extrapesados (EP). Todas las muestras están ubicadas en espacios bidimensionales cuyos ejes son los componentes principales que mejor explican la variabilidad en los espectros.

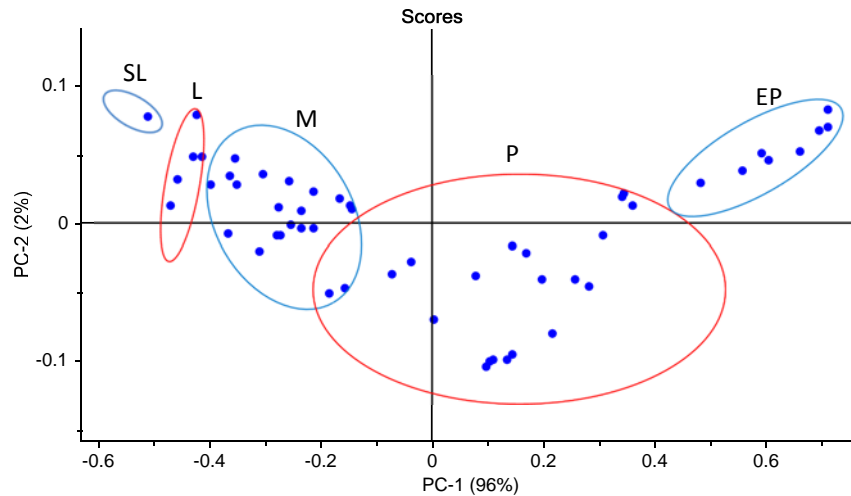


Figura 19. Análisis PCA: Scores muestras de crudos.

Para detectar posibles muestras discrepantes en el set de muestras de crudos se aplicó el test de *outliers* sobre la Ellipse T^2 de Hotelling, usando como límite crítico un nivel de confianza del 95%. Las muestras fuera de la ellipse son consideradas posibles outliers. Como se observa en la figura 20, la muestra de Crudo No. 11 se encuentra en el límite de la ellipse.

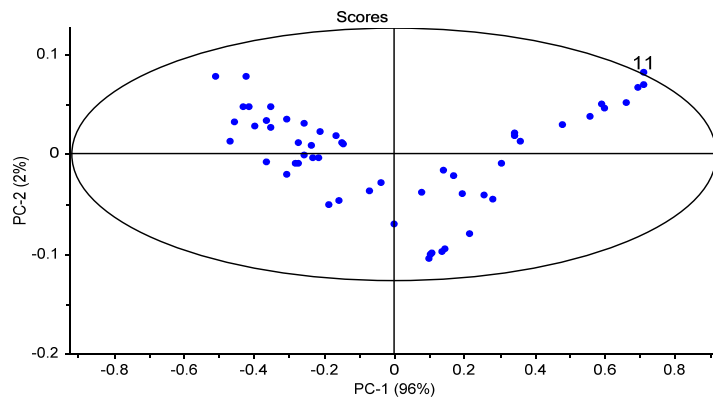


Figura 20. Scores en los dos primeros componentes principales.

El test de Hotelling T^2 sugiere que la muestra No. 11 puede ser un outlier. Sin embargo, para confirmar lo anterior, se evaluó el gráfico de influencia de los residuales espectrales (definen qué tan bien las variables se ajustan al modelo) frente al leverage de las muestras (distancia de la muestra proyectada al centro del modelo), el cual se presenta en la figura 21.

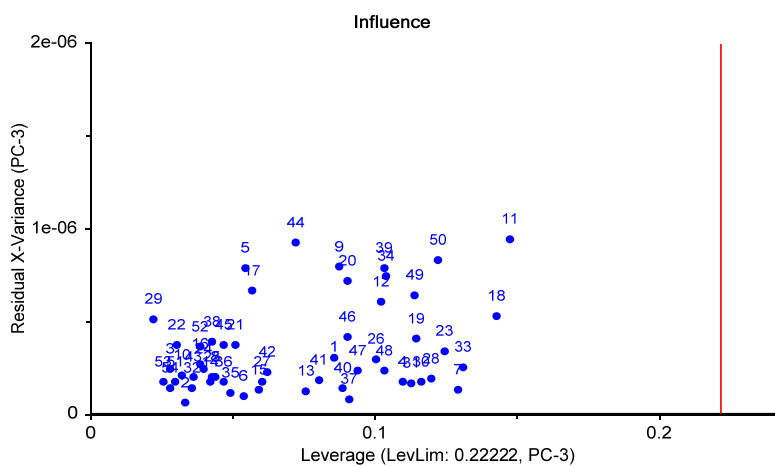


Figura 21. Varianzas residuales en X versus leverage.

Se observa que la muestra de crudo No.11 aparece con residual similar a un gran número de muestras del grupo, por lo que se descarta como posible outlier. Las muestras No. 11 y 18 aparecen con un leverage alto pero con residual similar al grupo de muestras No. 5, 9, 12, 17, 20, 29, 34, 39, 44, 49 y 50, por lo que se consideran muestras influyentes en el modelo pero no necesariamente outliers.

Los resultados de la química analítica de este grupo de muestras indican que existe una verdadera diferencia entre este grupo de muestras y el resto de crudos, debido a determinadas características químicas y no a diferencias al azar o a la toma de los espectros. Éstas pertenecen a un grupo de muestras diferente que adiciona variedad al set de muestras, y no necesariamente se consideran outliers. Estos crudos se caracterizan por su alto contenido de azufre, por ser crudos ácidos o por ser mezclas de crudos preparadas en el laboratorio. Por lo tanto, las anteriores características químicas de los crudos se reflejan en cada uno de los destilados atmosféricos que

son generados a partir de los crudos base. Teniendo en cuenta lo anterior, no se descartó ninguna muestra del grupo de 54 muestras de crudos como outlier.

El gráfico de Loadings (cargas) describe las relaciones entre las variables y las contribuciones de éstas según el componente especificado. En las figuras 22 y 23 se presentan los gráficos de Loadings para el primer y segundo componente, los cuales muestran las variables importantes y cómo se relacionan en los espectros de las muestras de crudos.

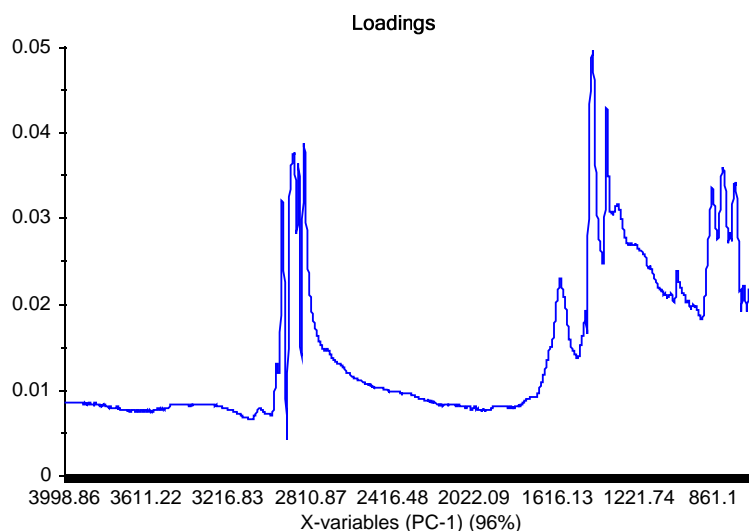


Figura 22. Variables importantes en los espectros de muestras de crudos (PC-1).

En la figura 22 se observa que las señales asociadas a compuestos alifáticos que aumentan o aparecen son las que corresponden a vibraciones del grupo metilo a 2952 cm⁻¹ (tensión asimétrica de CH₃), 2869 cm⁻¹ (tensión simétrica de CH₃), 1455 cm⁻¹ (flexión CH₃ ó CH₂) y 1376-1365 cm⁻¹ (flexión CH₃), mientras que las variables que disminuyen corresponden a vibraciones del grupo metileno a 2920 cm⁻¹ (tensión asimétrica del CH₂), 2852 cm⁻¹ (tensión simétrica del CH₂) y 725 cm⁻¹ (balanceo de alifáticos (CH₂)_n, n>4).

En general, hay una correlación inversa entre las variables relacionadas con vibraciones de grupos CH₃ y CH₂. Mientras que la intensidad de las vibraciones de los grupos metílicos aumenta, la de los grupos metileno disminuye. En los crudos livianos y medios se espera mayor presencia de

compuestos parafínicos con cadenas alifáticas más largas que en los crudos pesados y extrapesados, esto se refleja en el aumento de los grupos metileno y disminución de los grupos metílicos al alargarse las cadenas alifáticas.

En la región de huella dactilar, también se encontró una relación inversa entre las bandas que aparecen a 865 cm^{-1} y 810 cm^{-1} , y las bandas a 782 cm^{-1} , 767 cm^{-1} , y 699 cm^{-1} . Las bandas a 865 cm^{-1} (C-H aromático con 1 hidrógeno adyacente) y 810 cm^{-1} (C-H aromático con 2 hidrógenos adyacentes) aumentan su intensidad en la medida que las bandas a 782 cm^{-1} (C-H aromático con 3 hidrógenos adyacentes), 766 cm^{-1} (C-H aromático con 4 hidrógenos adyacentes) y 699 cm^{-1} (bencenos o xilenos) la disminuyen. Las dos primeras bandas están asociadas a vibraciones del C-H aromático con 1 y 2 hidrógenos adyacentes, es decir, a crudos con compuestos pesados (aromáticos con anillos tetra y pentasustituídos) que corresponderían a crudos pesados, mientras que las bandas a 782 cm^{-1} , 767 cm^{-1} y 699 cm^{-1} están asociadas a compuestos aromáticos con anillos di y trisustituídos, los cuales corresponderían a crudos más livianos o medios.

Las bandas vibracionales a 1601 cm^{-1} (tensión del C=C aromático) y 1032 cm^{-1} (banda de sulfóxidos, de intensidad media) aumentan en la medida que también lo hacen las bandas de los grupos metílicos y del C-H aromático con 1 y 2 hidrógenos adyacentes, asociado a la presencia de compuestos pesados en las muestras de crudos.

En la figura 23 se presenta el gráfico de Loadings del segundo componente, en el cual se observa que las variables importantes están asociadas a las vibraciones de los grupos CH_3 y CH_2 . Este componente muestra las variables importantes en las regiones alifáticas. De forma similar a lo observado en el PC-1, los grupos CH_3 presentan una correlación inversa con los grupos CH_2 . Mientras que las intensidades de las vibraciones de los grupos CH_2 aumentan, las de los grupos CH_3 disminuyen.

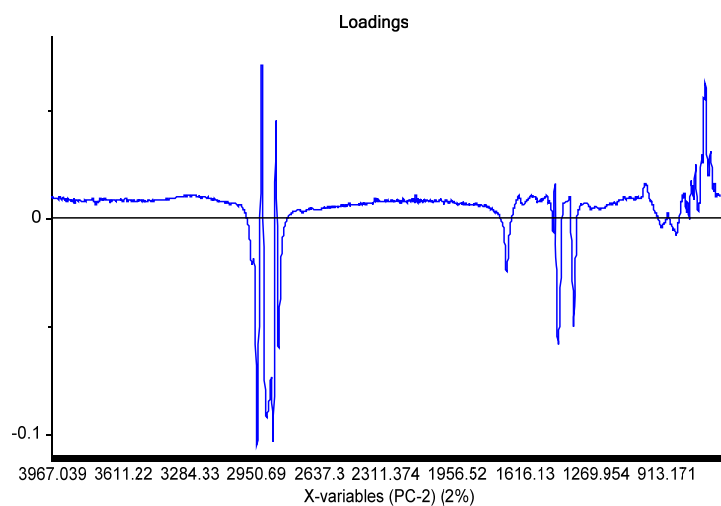


Figura 23. Variables importantes en los espectros de muestras de crudos (PC-2).

Adicionalmente, se observa la aparición de la banda correspondiente a la vibración de tensión del grupo carbonilo (C=O) a 1705 cm^{-1} , con tendencia positiva similar a los grupos CH_2 , lo que podría indicar que los compuestos oxigenados asociados a este grupo estarían presentes en compuestos parafínicos con cadenas alifáticas largas presentes en los crudos livianos y medios.

Las variables anteriores son las responsables de las diferencias espectrales entre las muestras de crudos, lo cual permite que las variables sean espectralmente distinguibles a la matemática multivariada.

3.2 ANÁLISIS PRELIMINAR DE LA LINEALIDAD: MODELOS PLS INICIALES

Para explorar las relaciones entre la data instrumental química y la data de propiedades se elaboraron modelos de regresión PLS iniciales. Teniendo en cuenta que el objetivo de la investigación es realizar la predicción de las propiedades fisicoquímicas de los destilados a partir de los espectros de sus crudos madre, se realizó un análisis preliminar de la *linealidad* entre las variables espectrales y las propiedades fisicoquímicas de los destilados, mediante la construcción de modelos PLS iniciales.

El número de variables se optimizó seleccionando las regiones del espectro comprendidas entre 3200-2500 cm^{-1} y 1800-650 cm^{-1} , con pretratamiento definido en el numeral 2.3, reduciendo el número de variables de 3474 a 1921. Las regiones del espectro a modelar fueron seleccionadas según inspección visual, removiendo las zonas en las que no se observaron bandas de absorción. Se eliminaron variables con el fin de evitar que las predicciones del modelo sean inestables, es decir, que pequeños cambios en el espectro en el orden del ruido espectral den cambios estadísticamente significativos en las predicciones. Las propiedades y fracciones modeladas se presentan en la siguiente tabla:

Tabla 19. Fracciones y propiedades a modelar

| Destilado | Quero | Acpm liviano | Acpm pesado | Crudo reducido |
|-------------------|--------------|---------------------|--------------------|-----------------------|
| Propiedad | (199-249°C) | (249-315°C) | (315-371°C) | (371°C+) |
| Rendimientos (%V) | X | X | x | X |
| API | X | X | x | X |
| Azufre (%w) | X | X | x | X |
| Índice de Cetano | X | X | x | NA |

NA: No aplica para esta fracción.

Se desarrollaron los modelos PLS iniciales y se analizaron los gráficos de residuales para detectar no linealidades o carencia de ajuste. Los modelos iniciales mostraron una distribución aleatoria de los residuales, con algunos outliers influyentes que fueron descartados o evaluados en términos de errores en los datos analíticos y/o eliminados durante la etapa de afinación de los modelos. En la figura 24 se muestra la distribución de residuales para los modelos iniciales de rendimientos (%volumen) de queroseno, acpm liviano, acpm pesado y crudo reducido.

Los gráficos de residuales para las propiedades fisicoquímicas de los destilados (API, azufre, índice de cetano) mostraron tendencias similares a los de rendimientos. Después de verificar la linealidad en los modelos iniciales, se desarrollaron los modelos definitivos.

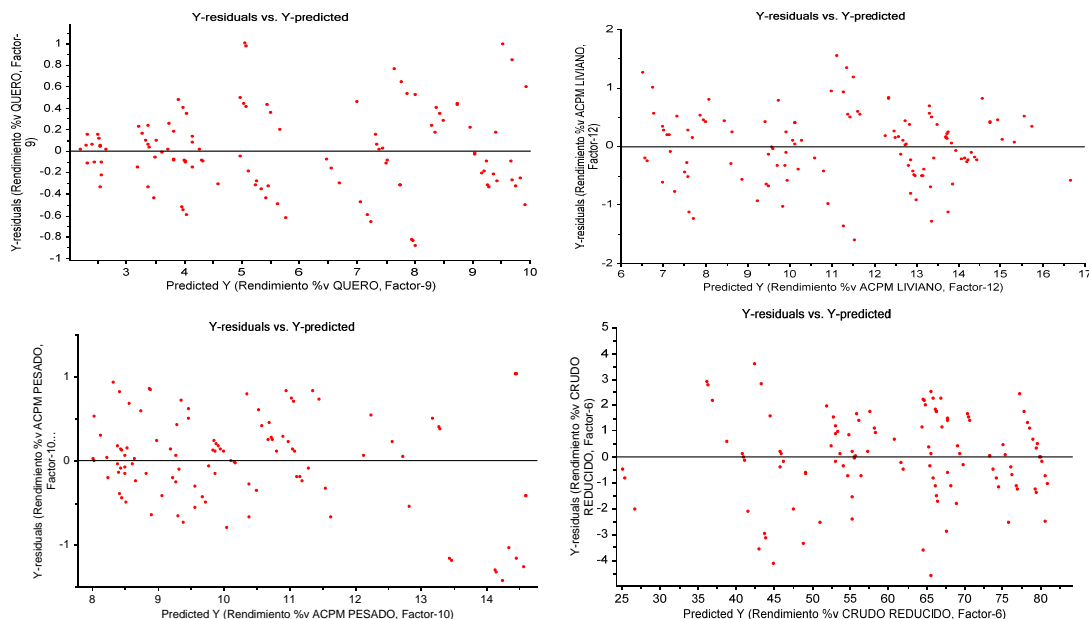


Figura 24. Distribución de residuales para rendimientos (%V) de destilados atmosféricos.

3.3 MODELOS PLS RENDIMIENTOS

El grupo total de espectros de crudos (54) se dividió en aproximadamente 70% para calibración (38 muestras) y 30% para predicción (16 muestras), se organizaron con base en su API, teniendo en cuenta que este parámetro está muy relacionado con la composición química de los crudos y sus destilados. Para evitar obtener una distribución normal de valores de referencia en el set de calibración se realizó una distribución lo más uniforme posible, dividiendo el rango total de valores de referencia en varios intervalos (crudos livianos, medios, pesados y extrapesados). En cada intervalo, se seleccionaron al azar algunas muestras, las cuales en su conjunto dan un set de calibración uniformemente distribuido, garantizando así la variabilidad de las muestras de calibración y validación.

Los modelos de regresión PLS fueron desarrollados usando los espectros de los crudos y los valores de las propiedades de sus respectivos destilados atmosféricos. El análisis de regresión se llevó a cabo para cada propiedad usando aproximadamente el mismo set de espectros de muestras de crudos.

A continuación se muestran los resultados obtenidos para los modelos de rendimientos para las fracciones de destilados atmosféricos queroseno, acpm liviano, acpm pesado y crudo reducido. Se detalla el desarrollo de los modelos de rendimientos para las cuatro fracciones mencionadas y el análisis de las variables importantes (regiones espectrales responsables de la discriminación de las muestras según la propiedad a modelar). Teniendo en cuenta que se aplicó la misma metodología para la obtención de los modelos de API, azufre e índice de cetano, se muestra el resumen de los resultados obtenidos para estas propiedades y los aspectos más relevantes observados en su desarrollo, al final del capítulo, después de los modelos AG-PLS de rendimientos.

3.3.1 Modelo PLS Rendimientos queroseno

Se desarrolló el modelo de calibración para rendimientos (% en volumen) de la fracción queroseno y se revisaron posibles muestras discrepantes, aplicando el mismo procedimiento usado para la detección de outliers usado en el PCA (numeral 3.1). En la figura 25 se observa el gráfico de Scores, el cual no propone muestra alguna como outlier.

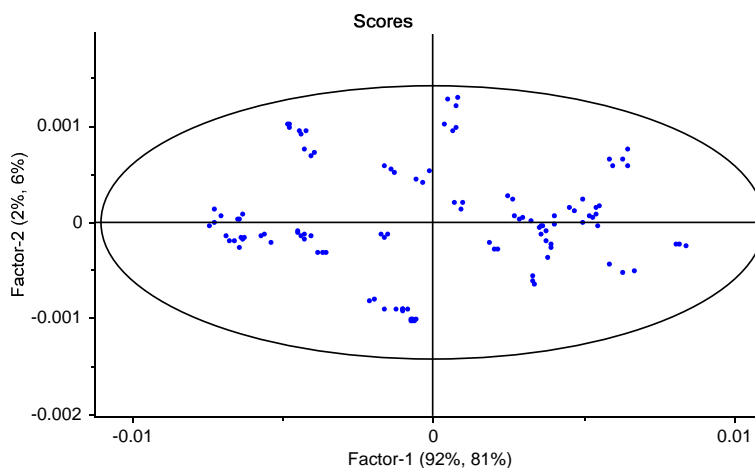


Figura 25. Análisis Scores modelo PLS de rendimientos (%V) fracción queroseno.

Al evaluar el gráfico de influencia de los residuales espectrales frente al leverage de las muestras (figura 26), se observa que la muestra de crudo No.39 aparece con alto residual, confirmado en el gráfico de Residuales Q versus Hotelling T^2 de la misma figura al quedar fuera del límite del valor crítico del Q residual. Por lo tanto, se considera esta muestra como outlier y se descartó del grupo

de calibración, reprocesando el PLS. Para los modelos de API, azufre e índice de cetano, también se descartó esta muestra del grupo de calibración.

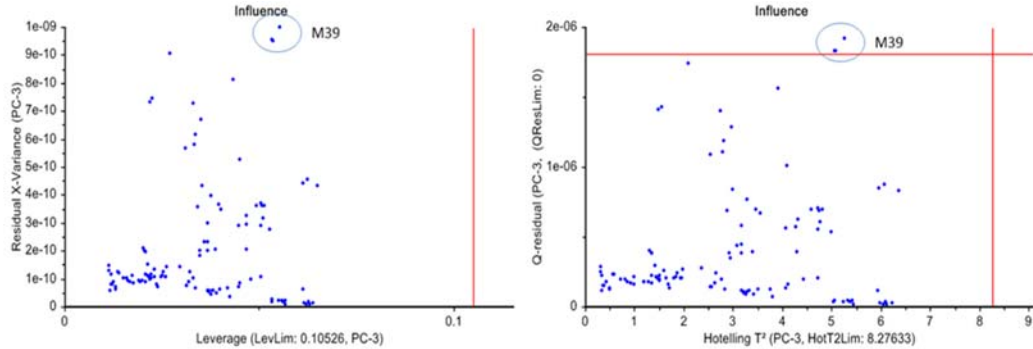


Figura 26. Varianzas Residuales-X versus Leverage modelo PLS rendimientos queroseno.

La muestra descartada como outlier corresponde a una mezcla de tres crudos en diferentes proporciones, no incluidos en el set de calibración, preparada en el laboratorio. Observando los espectros, esta es la única muestra que presenta diferencias espectrales visibles a simple vista en la región de huella dactilar (región aromática), según se presenta en la figura 27:

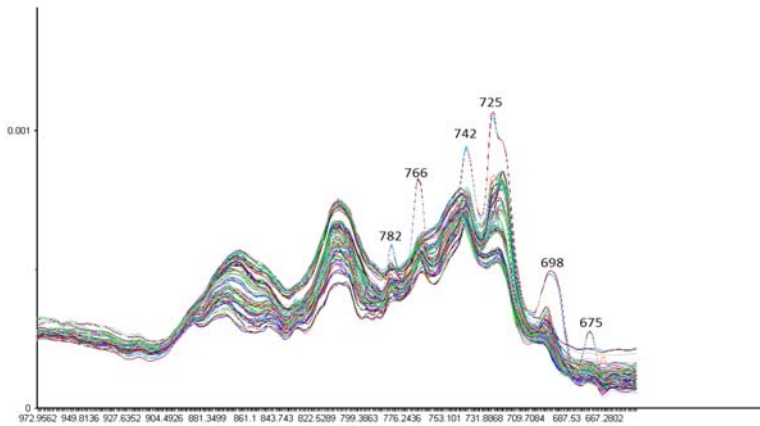


Figura 27. Espectros FT-IR superpuestos de muestras de crudos y crudo No. 39

En la muestra de Crudo No. 39 (con API de 42) las bandas a 782, 766 y 742 cm^{-1} , asociadas a vibraciones del C-H aromático con 3,4 y 5 hidrógenos adyacentes (que caracterizan a los crudos

livianos), son más intensas que en el resto de las muestras de crudos. También se aprecian las bandas a 698 y 674 cm^{-1} atribuibles a la flexión del anillo aromático o encogimiento de bencenos o xilenos.

La calidad del modelo de regresión se refleja en el gráfico de valores predichos versus valores medidos o de referencia. En la figura 28 se presentan los resultados de la regresión obtenida para la calibración del modelo, con coeficiente de correlación de 0.983 y error de calibración de 0.324, en un intervalo entre 1 y 11 % de rendimientos para esta fracción.

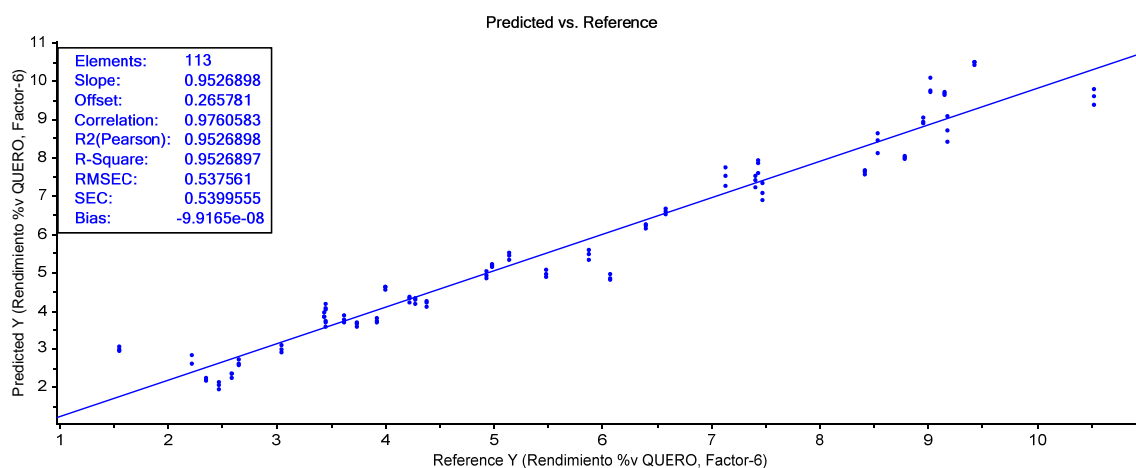


Figura 28. Calibración modelo PLS rendimientos (%V) fracción queroseno.

Validación cruzada. La validación del modelo se realiza para determinar qué tan bien el modelo podrá predecir nuevos datos del mismo tipo que los usados para desarrollar el modelo, estimando la incertidumbre para futuras predicciones.

Se realizó validación cruzada para estimar la habilidad predictiva del modelo. Un grupo de muestras es removido de la matriz de datos y el modelo se construye sobre las muestras remanentes. Una vez realizada la calibración, se prueba el modelo con los datos del grupo de muestras excluido. Se repite el proceso de manera sucesiva dejando cada vez un grupo de muestras fuera del proceso de cálculo. La incertidumbre de futuras predicciones se evalúa mediante el RMSEC.

En la figura 29 se presenta el gráfico de Scores, mostrando las relaciones entre las muestras de calibración y de validación. Las muestras en azul fueron empleadas para calibración y las muestras en rojo fueron usadas en la validación cruzada. Se observa que ninguna muestra fue dejada fuera de la elipse (outlier).

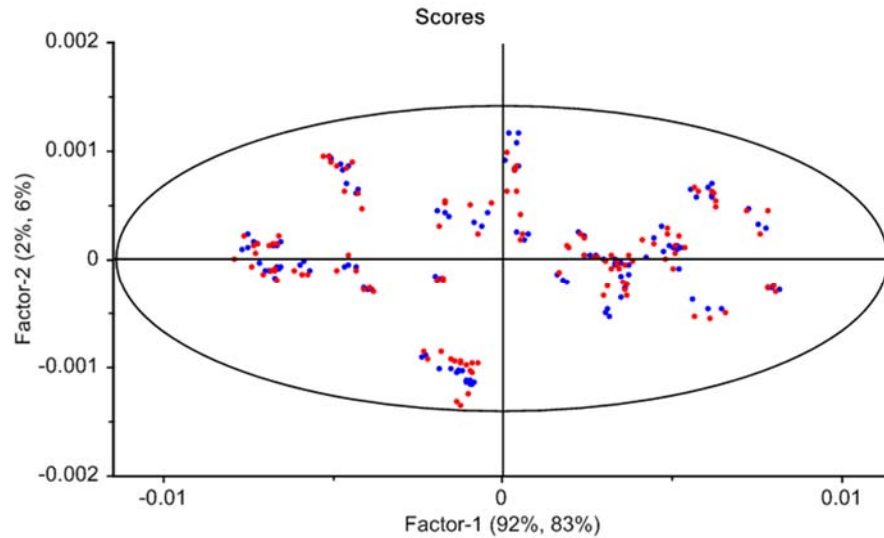


Figura 29. Calibración/Validación modelo PLS rendimientos (%V) fracción queroseno.

La calidad del modelo de regresión para la validación cruzada se evaluó mediante el gráfico de valores predichos versus valores medidos. En la figura 30 se presentan los resultados de la regresión obtenida para la validación del modelo, con coeficiente de correlación 0.971 y error de validación de 0.430, en un intervalo entre 1 y 11 % de rendimientos.

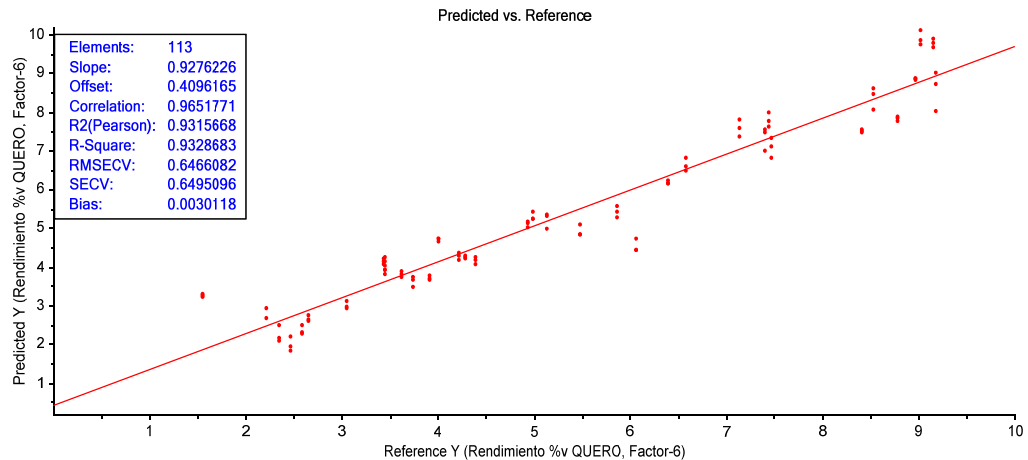


Figura 30. Validación cruzada modelo PLD rendimientos (%V) fracción queroseno.

Los errores RMSEC y RMSECV, no varían significativamente entre sí, resultados que demuestran el buen desempeño del modelo según validación cruzada.

El gráfico de la figura 31 muestra la distribución de los residuales, los cuales se distribuyen aleatoriamente, lo que implica que estas variaciones son debidas sólo al ruido y no a las muestras. La varianza residual máxima es de 0.8%.

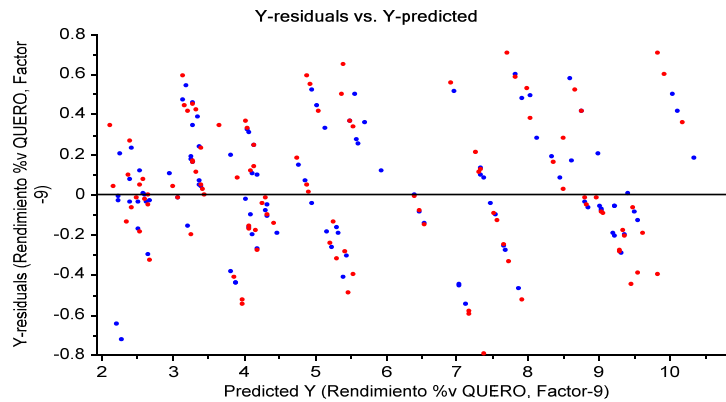


Figura 31. Residuales de rendimientos (%V) en muestras de calibración/validación modelo PLS fracción queroseno.

Varianza explicada. En la figura 32 se observan las varianzas explicadas de calibración y validación del modelo para los datos X y Y. Estos gráficos muestran cuánto de la variación en las respuestas es descrito por cada componente diferente.

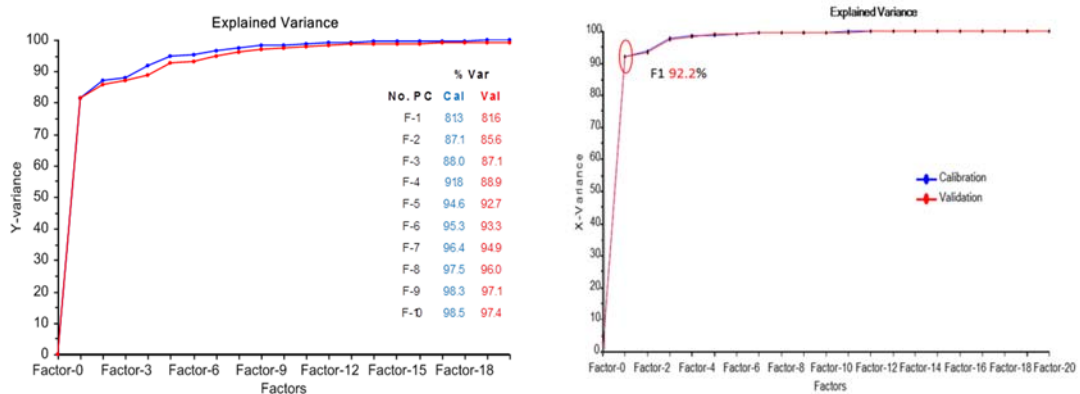


Figura 32. Varianza explicada modelo PLS rendimientos (%V) fracción queroseno.

Se observa que con cuatro factores o componentes se explica más del 91% de la varianza de calibración y 88% de la varianza de validación, en los datos de rendimientos en volumen (datos Y). Con un factor se explica más del 92% de la varianza de calibración y validación en los datos de los espectros de los crudos (datos X). En cada caso, las dos varianzas no difieren sustancialmente, lo que significa que los datos de calibración y de validación cruzada son representativos, de modo que el modelo tiene la capacidad para describir correctamente nuevos datos.

Selección del número de componentes principales. La elección del número de componentes principales o variables latentes (VL) es un punto clave en la utilización de cualquier técnica de calibración que realice reducción de variables. Para seleccionar el número de componentes óptimo en cada modelo PLS, se analizó el error estándar de calibración (RMSEC) y el error estándar de validación cruzada (RMSECV) frente al número de variables latentes, seleccionando como óptimo el mínimo de la curva o, en su defecto, el componente a partir del cual no se observa una disminución significativa del error. Sin embargo, teniendo en cuenta que si se usan pocos componentes el modelo resultante será menos preciso, o por el contrario, si se usan muchos componentes el modelo puede ser inestable (debido a que pequeños cambios en el espectro del orden del ruido espectral pueden producir cambios estadísticamente significativos en los estimados), no se aplicó en todos los casos el concepto del mínimo propuesto según el RMSEC/RMSECV.

El error estándar de calibración se define como la desviación estándar para la diferencia entre los valores del análisis de referencia y los valores estimados para las muestras incluidas en el set de calibración. El error estándar de validación se calcula de forma similar pero para las muestras del proceso iterativo de validación cruzada.

El máximo número de variables latentes que debería usarse en un modelo está relacionado con el número de componentes espectralmente distinguibles (funcionalidades) que están presentes en el set de calibración. Los componentes (o funcionalidades) son espectralmente distinguibles si dan absorciones que no están linealmente correlacionadas entre las muestras de calibración, y si el cambio en las absorciones entre los espectros de calibración es mucho mayor que el ruido espectral (American Society for Testing and Materials 2012).

La figura 33 muestra el comportamiento del error de calibración (RMSEC) para este modelo, el cual disminuye paralelamente al error de validación (RMSECV). Para determinar el número óptimo de variables latentes, se desarrollaron diferentes modelos usando desde cuatro hasta nueve variables latentes. Para cada modelo se evaluó el RMSECV, RMSEP y el coeficiente de correlación R^2 . Se observó que para el modelo desarrollado con cuatro variables latentes, el error de predicción (RMSEP) fue más alto y el R^2 desmejoró, al compararlos con los modelos de cinco hasta siete variables latentes. El RMSEP y R^2 de los modelos con cinco, seis y siete variables latentes mostraron mejor desempeño con respecto a los modelos desarrollados usando ocho y nueve variables latentes. Entre los modelos de cinco, seis y siete variables latentes, el modelo de seis variables latentes mostró el mayor R^2 y mejor RMSEP. Por esta razón, se seleccionaron seis variables latentes. Una aproximación similar fue usada para seleccionar el número de variables latentes óptimo en la mayoría de los modelos desarrollados en esta investigación.

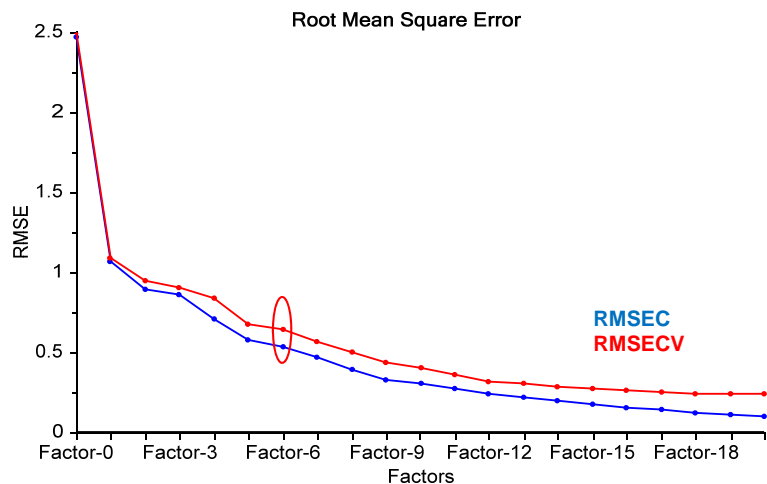


Figura 33. Error versus número de variables latentes para modelo PLS rendimientos (%V) fracción queroseno.

Variables importantes. Con el fin de detectar las señales responsables de la discriminación de las muestras en los primeros componentes, se analizaron los gráficos de Loadings. Estos gráficos tienen un perfil similar a los datos originales y muestran las regiones espectrales de importancia.

Si la suma de las varianzas explicadas para los dos o tres primeros componentes está entre 70-80%, estos componentes muestran una gran cantidad de información en los datos, de modo que las relaciones pueden ser interpretadas con un alto grado de certidumbre. Si el total de la varianza explicada es menor que 70-80% es necesario tomar más componentes o podría haber poca información útil en los datos bajo estudio. En la figura 34, se presenta el gráfico de Loadings que muestra las variables importantes para los primeros componentes del modelo PLS de rendimientos en la fracción queroseno.

Se observa que la suma de las varianzas explicadas para los dos primeros componentes es del 94% en X y del 87% en Y, lo cual sugiere que estos dos componentes muestran las principales relaciones entre los datos.

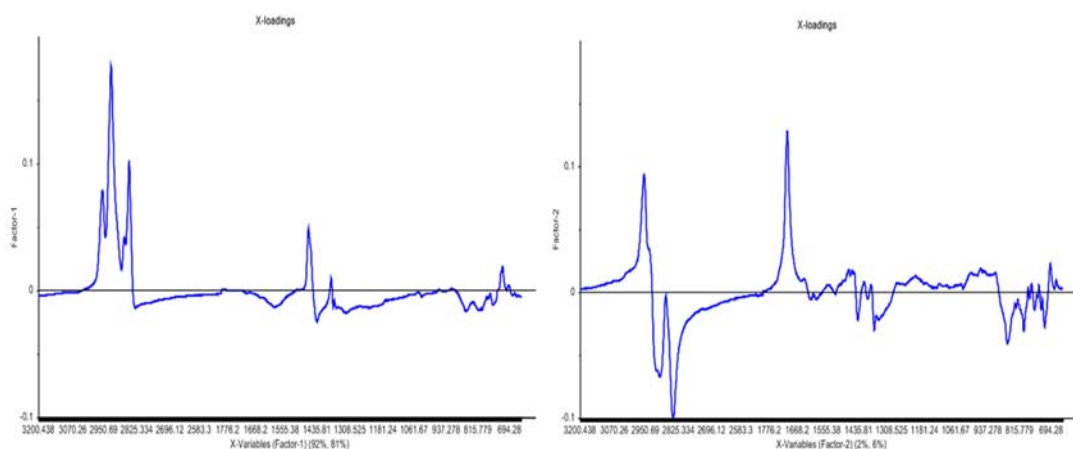


Figura 34. Variables importantes modelo PLS rendimientos (%V) fracción queroseno.

Las características espectrales relacionadas con los rendimientos de la fracción queroseno están definidas por las bandas de tensión y flexión de los grupos metilo y metileno alifáticos. Las bandas de los grupos metileno alifáticos a 2920 y 2852 cm^{-1} son las que presentan mayor contribución a este componente, seguidas de las bandas de los grupos metílicos a 2952, 2869 y 1455 cm^{-1} , respectivamente. También se observa la banda a 1376 cm^{-1} , correspondiente a las vibraciones de flexión simétrica del grupo metilo alifático y la banda $(\text{CH}_2)_n$, $n > 4$ a 725 cm^{-1} , la cual indica la presencia de cadenas alifáticas más largas. Estas bandas son las identificadas en el primer componente, el cual explica el 92% de la varianza en X y el 81% de la varianza en Y.

En el segundo componente, el cual explica el 2% de la varianza en X y el 6% de la varianza en Y, se identifican principalmente las bandas de los grupos metilo alifáticos a 2952 cm^{-1} y 2869 cm^{-1} con crecimiento positivo y la banda del grupo metileno a 2852 cm^{-1} , las bandas de flexión del grupo metilo/metileno a 1455 cm^{-1} y 1376 cm^{-1} con efecto contrario, similar al de la banda a 725 cm^{-1} correspondiente al grupo metileno (balanceo de alifáticos $(\text{CH}_2)_n$, $n > 4$). Otra banda con un aporte importante en este componente es la banda de tensión del grupo carbonilo a 1705 cm^{-1} . Las bandas con crecimiento positivo (grupos metilo y carbonilo) en contraposición a las bandas con crecimiento negativo (grupos metileno) sugieren que esta propiedad está definida por los grupos que generan vibraciones en la región alifática y compuestos oxidados (grupo carbonilo), asociados al tipo de crudo (pesado o liviano).

Validación externa del modelo. La validación externa del modelo PLS desarrollado se realizó con un set de 12 muestras no usadas para la calibración, mediante la predicción o estimación de los valores de respuesta Y desconocidos. Las muestras de crudo No. 26 y 42 fueron eliminadas del set de calibración debido a que deterioraban los parámetros estadísticos del modelo. Las muestras de predicción reportaron valores de referencia de rendimientos (%V) de queroseno en el intervalo de 2.51% a 9.13%. En la figura 35 se presentan los resultados de la regresión obtenida para la validación externa del modelo, con coeficiente de correlación de 0.943 y error de predicción de 0.465.

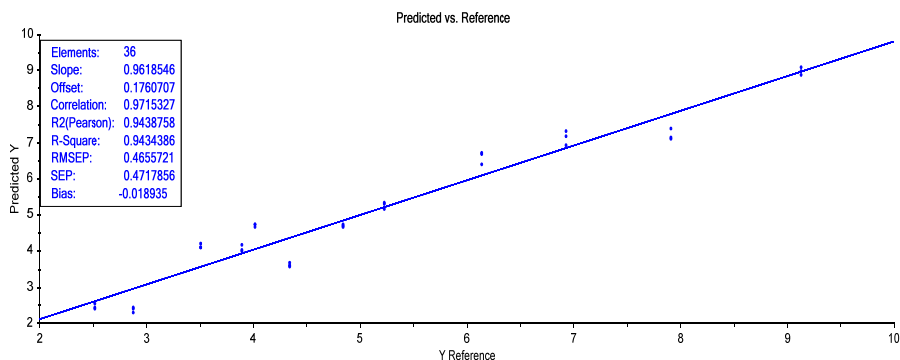


Figura 35. Validación externa modelo PLS rendimientos (%V) fracción queroseno.

En la tabla 20 se muestran los resultados de la validación externa para el modelo desarrollado:

Tabla 20. Validación externa modelo PLS rendimientos (%V) fracción queroseno

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|-----------|------------------|----------------|------------|--------------------|
| 5 | 4.34 | 3.60 | 0.75 | 17.2 |
| 8 | 6.14 | 6.58 | 0.45 | 7.29 |
| 10 | 6.93 | 7.14 | 0.21 | 3.05 |
| 13 | 2.51 | 2.44 | 0.07 | 2.82 |
| 14 | 3.90 | 4.05 | 0.15 | 3.93 |
| 15 | 7.92 | 7.20 | 0.72 | 9.03 |
| 23 | 2.88 | 2.36 | 0.52 | 18.0 |
| 25 | 9.13 | 8.97 | 0.16 | 1.72 |
| 28 | 3.51 | 4.13 | 0.62 | 17.6 |
| 38 | 4.84 | 4.68 | 0.16 | 3.29 |
| 46 | 4.02 | 4.70 | 0.68 | 17.0 |
| 53 | 5.23 | 5.26 | 0.03 | 0.47 |

Los valores de predicción de rendimientos son comparados con los valores de referencia obtenidos mediante métodos analíticos estándar. Se observa que los errores relativos obtenidos son iguales o inferiores al 18%, lo que refleja un desempeño aceptable del modelo en la predicción de nuevas muestras.

3.3.2 Modelo PLS rendimientos acpm liviano

Los parámetros más importantes para el modelo de rendimientos (%V) de la fracción acpm liviano se presentan en la figura 36. Los modelos de calibración y de validación fueron desarrollados bajo la misma metodología empleada para los modelos de la fracción queroseno.

En la figura 36 (a) se presenta el gráfico de Scores, el cual muestra las relaciones entre las muestras de calibración y de validación. No se proponen muestras adicionales como outliers (previa eliminación de la muestra No.39).

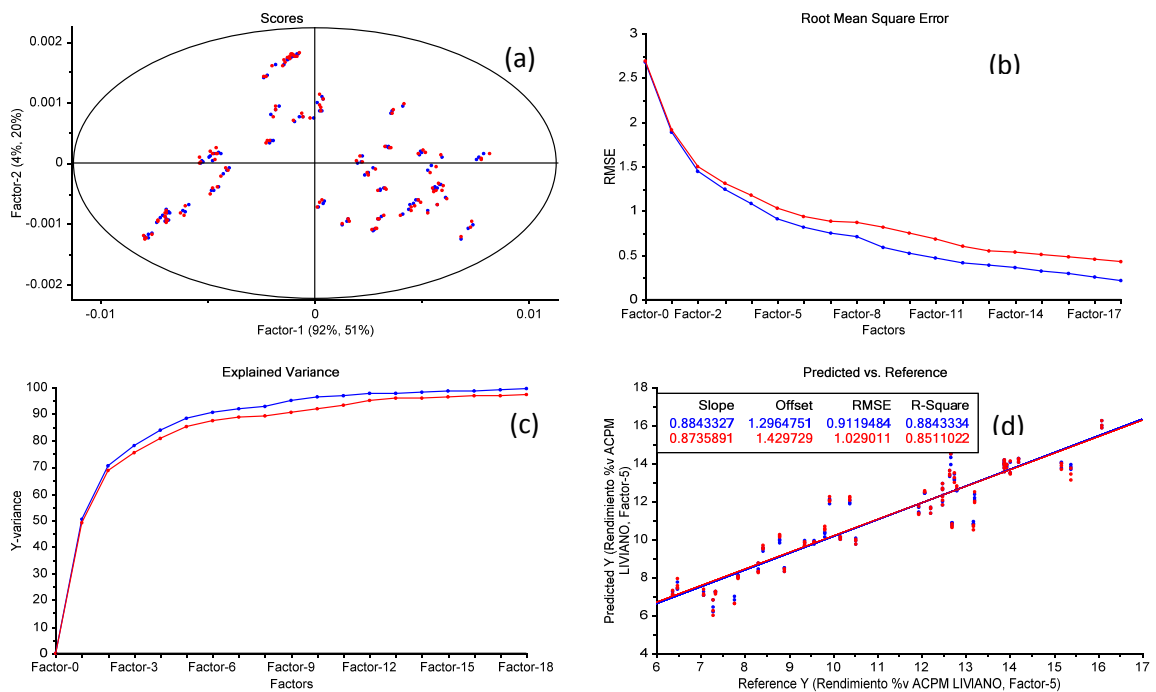


Figura 36. Principales resultados modelo PLS rendimientos (%V) fracción acpm liviano.

En la figura 36 (c) se observa la varianza explicada de calibración y validación del modelo PLS de rendimientos (%V) en la fracción acpm liviano. Con cinco factores o componentes se explica más del 94% de la varianza de calibración y 92% de la varianza de validación, en los datos de rendimientos en volumen. Teniendo en cuenta que las varianzas son similares, se infiere que el modelo puede describir correctamente nuevos datos.

Los resultados de las regresiones obtenidas para la calibración y validación del modelo son mostrados en la figura 36 (d). Para el modelo de calibración, se obtuvo un coeficiente de correlación de 0.976 y error de calibración de 0.41, en un intervalo de rendimientos entre 6 y 17%. Para la validación, se obtuvo un coeficiente de correlación del 0.950 y error de validación de 0.597, para el mismo intervalo de rendimientos. Fueron seleccionadas doce variables latentes como el número de componentes óptimo. El error RMSEC fue del 0.32% y 0.43% para RMSECV, en un intervalo de valores de rendimientos de 6 a 17%.

En el siguiente gráfico se presentan las varianzas en X, con un factor se explica el 92.2% y con dos factores el 93.9% de la varianza.

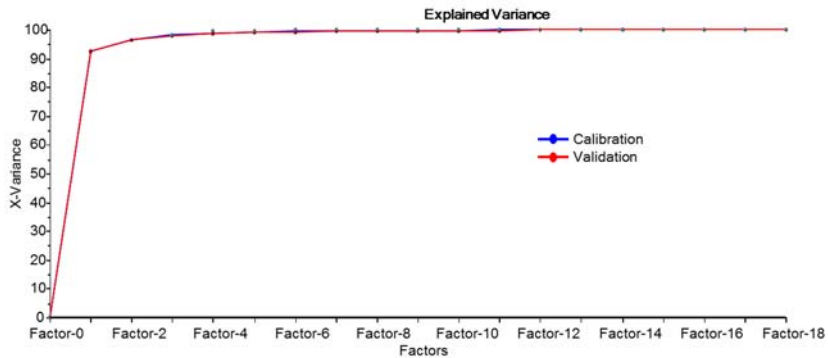


Figura 37. Varianza explicada en X modelo PLS rendimientos (%V) fracción acpm liviano.

Variables importantes. Las regiones espectrales de importancia para los dos primeros componentes del modelo PLS de rendimientos en la fracción acpm liviano se muestran en la figura 38.

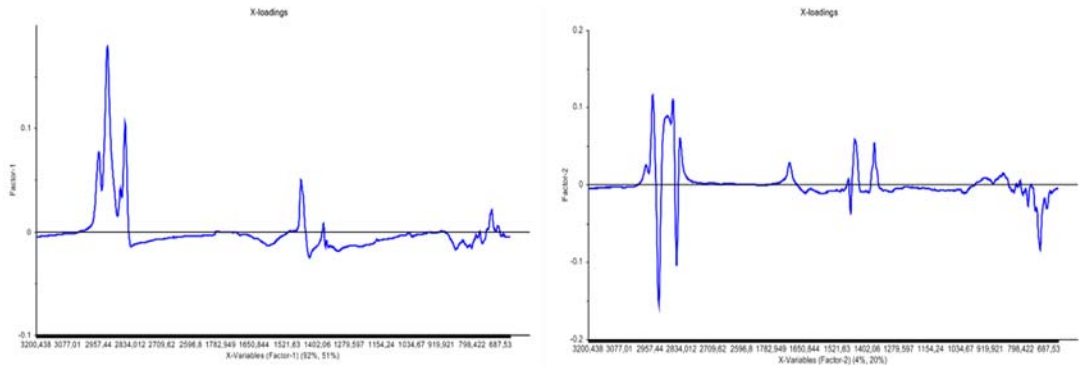


Figura 38. Variables importantes modelo PLS rendimientos (%V) fracción acpm liviano.

La suma de las varianzas explicadas para los dos primeros componentes es del 96% en X y 71% en Y. Para el primer componente, el cual explica el 92% de la varianza en X y el 51% de la varianza en Y, las características espectrales relacionadas con los rendimientos de la fracción acpm liviano están definidas por las mismas bandas de tensión y flexión de los grupos metilo y metileno alifáticos observadas previamente para los rendimientos de la fracción queroseno, con el mismo grado de aporte (valor de intensidad en el gráfico).

Sin embargo, en el segundo componente, el cual explica el 4% de la varianza en X y el 20% de la varianza en Y, el perfil es diferente con respecto al de rendimientos en la fracción queroseno. Se identifican principalmente las bandas de los grupos metilo alifáticos a 2952 cm^{-1} y 2869 cm^{-1} , las bandas de flexión del grupo metilo/metileno a 1455 y 1376 cm^{-1} con crecimiento positivo (contrario al comportamiento mostrado en los rendimientos de la fracción queroseno) y las bandas de los grupos metileno a 2920 cm^{-1} (no observada en los rendimientos del queroseno para el mismo componente) y 2852 cm^{-1} con efecto contrario, similar al de la banda a 725 cm^{-1} correspondiente al grupo metileno (balanceo de alifáticos $(\text{CH}_2)_n$, $n>4$). La banda del grupo carbonilo a 1705 cm^{-1} aparece con menor intensidad respecto a los rendimientos del queroseno.

En este componente, además del crecimiento positivo o negativo de las bandas, éstas también se diferencian por su aporte (intensidad en el gráfico).

Validación externa del Modelo. La validación externa del modelo PLS desarrollado se realizó con un set de 13 muestras no usadas para calibración. En la figura 39 se presentan los resultados de la regresión obtenida para la validación externa del modelo, con coeficiente de correlación de 0.811 y error de predicción de 0.940.

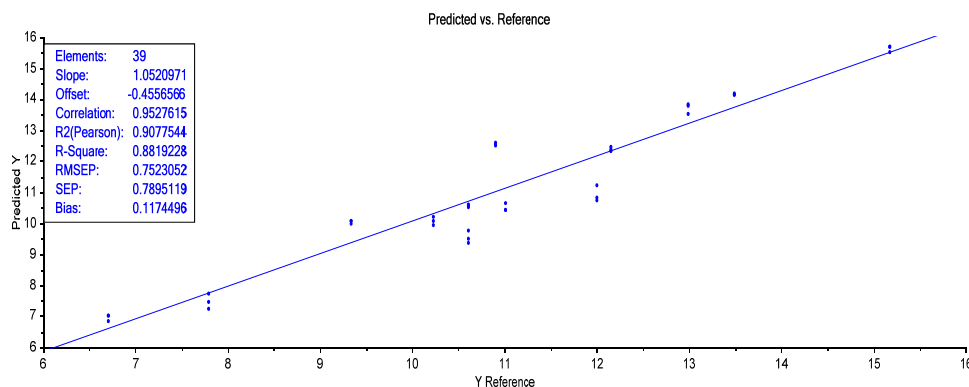


Figura 39. Validación externa modelo PLS rendimientos (%V) fracción acpm liviano.

En la predicción de muestras externas se obtuvieron errores relativos máximos de 14.6%, según se muestra en la tabla 21.

Tabla 21. Validación externa del modelo PLS rendimientos (%V) fracción acpm liviano

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|-----------|------------------|----------------|------------|--------------------|
| 5 | 12.15 | 12.39 | 0.24 | 1.94 |
| 8 | 10.61 | 10.57 | 0.04 | 0.41 |
| 10 | 12.99 | 13.71 | 0.72 | 5.57 |
| 13 | 7.798 | 7.485 | 0.31 | 4.01 |
| 14 | 10.23 | 10.06 | 0.17 | 1.69 |
| 15 | 11.01 | 10.50 | 0.51 | 4.62 |
| 23 | 6.703 | 6.970 | 0.27 | 3.99 |
| 26 | 15.17 | 15.71 | 0.53 | 3.50 |
| 28 | 13.49 | 14.14 | 0.65 | 4.83 |
| 38 | 10.90 | 12.49 | 1.59 | 14.6 |
| 46 | 9.335 | 10.05 | 0.72 | 7.68 |
| 52 | 10.61 | 9.552 | 1.06 | 10.0 |
| 53 | 12.00 | 10.92 | 1.08 | 8.99 |

3.3.3 Modelo PLS rendimientos acpm pesado

Los parámetros más importantes para el modelo de rendimientos (%V) de la fracción acpm pesado se presentan en la figura 40. En este modelo la suma de las varianzas explicadas (figura 40 a), para los dos primeros componentes, es del 97% en X y del 58% en Y.

En la figura 40 (c), se observa que con seis factores se explica más del 89% de la varianza de calibración y 81% de la varianza de validación, en los datos de rendimientos en volumen (datos Y). Las dos varianzas no difieren sustancialmente, lo que significa que los datos de calibración y de validación cruzada son representativos.

Para el modelo de calibración (figura 40 d), se obtuvo un coeficiente de correlación de 0.914 y error de calibración de 0.634, en un intervalo de rendimientos entre 6 y 16%. Para la validación, se obtuvo un coeficiente de correlación del 0.890 y error de validación de 0.724, para el mismo intervalo de rendimientos.

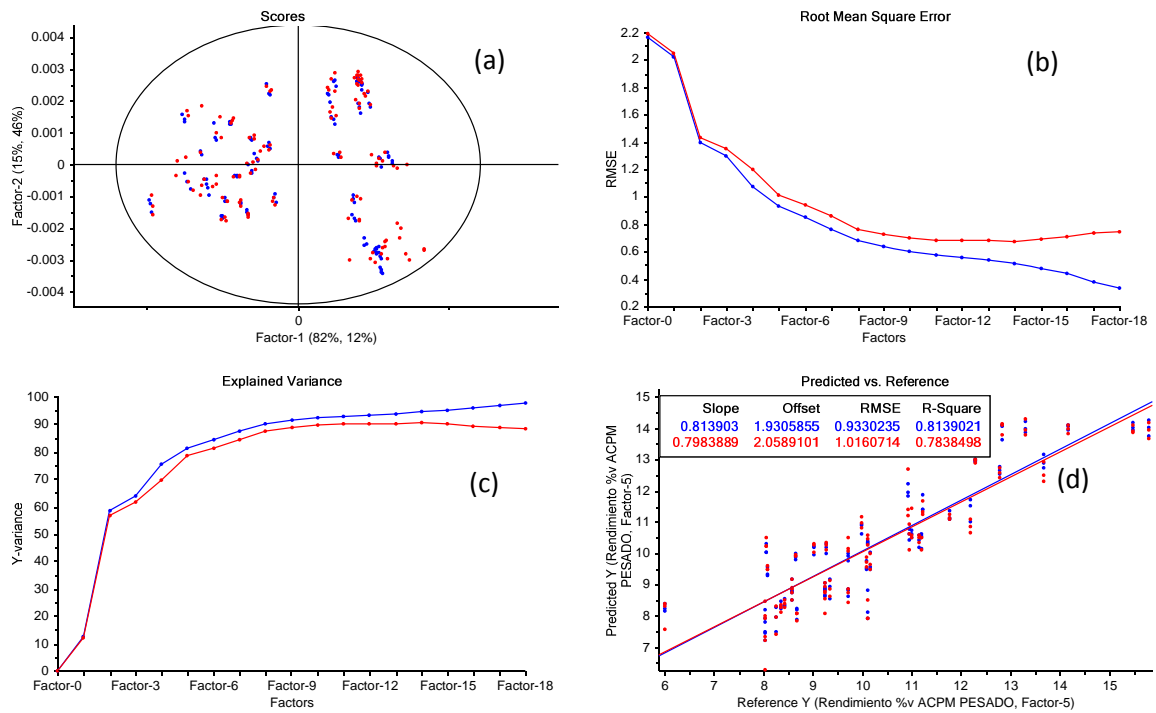


Figura 40. Principales resultados modelo PLS para rendimientos (%V) fracción acpm pesado.

Se seleccionó como cinco el número de variables latentes óptimo. El error RMSE fue del 0.634% y 0.724% para RMSECV, en un intervalo de valores de rendimientos de 6 a 16%. En el siguiente gráfico se presenta la varianza explicada en X, con un factor se explica el 82% y con dos factores, el 96.4% de la varianza de calibración y validación en los datos de los espectros de los crudos (datos X).

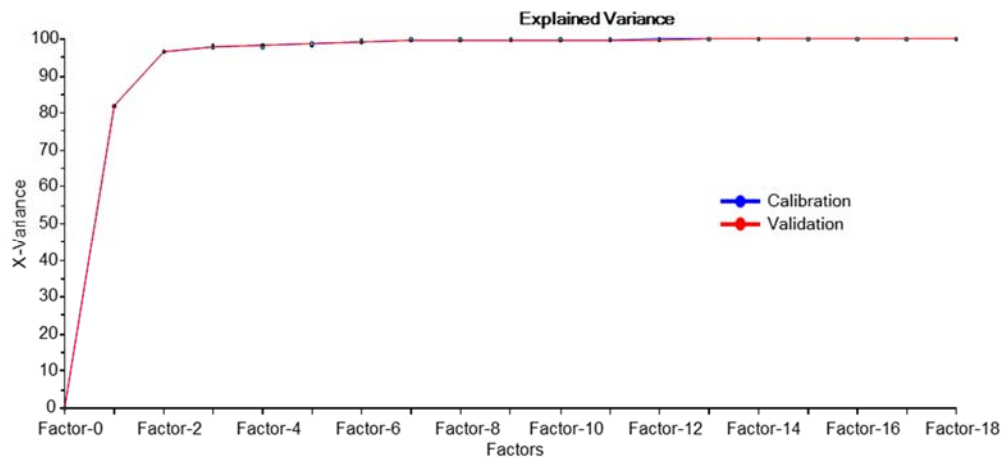


Figura 41. Varianza en X modelo PLS rendimientos (%V) fracción acpm pesado.

VARIABLES IMPORTANTES. Debido a que los dos primeros componentes principales no explican mínimo el 70% de la varianza explicada en X, se tomaron cuatro componentes. La suma de las varianzas explicadas para X fue de 99% y 75% para Y. Los cuatro componentes principales del modelo PLS de rendimientos en la fracción acpm pesado se muestran en la figura 42.

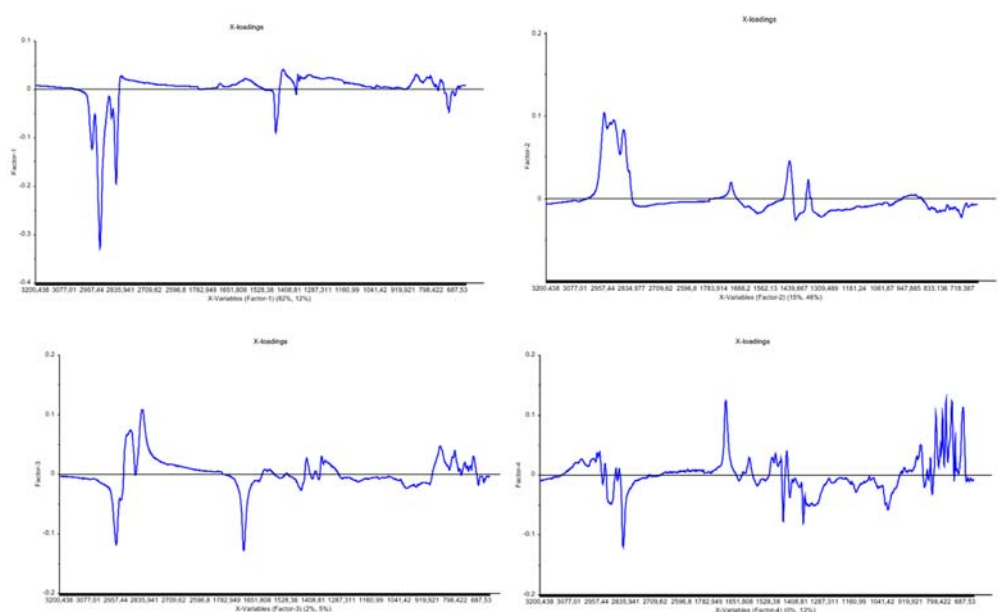


Figura 42. Variables importantes en el modelo de Rendimientos (%V) para acpm pesado.

En este modelo más del 90% de la varianza en X explica el 58% en Y para los dos primeros componentes (97%, 58%). Lo anterior sugiere que son necesarios más componentes para explicar mínimo el 70% de la varianza en Y, lo que podría indicar que la información de los datos bajo estudio no es altamente significativa.

El perfil del gráfico de Loadings para el primer componente (con varianzas de 82% para X y 12% para Y) identifica como variables importantes las mismas bandas identificadas para el queroseno y acpm liviano pero en sentido contrario (crecimiento negativo) y con intensidades mayores. En el segundo componente, el cual explica el 15% de la varianza en X y el 46% de la varianza en Y, las variables importantes crecen en sentido positivo. Se identifican principalmente las bandas de los grupos metilo alifáticos a 2952 cm^{-1} y 2869 cm^{-1} , así como las bandas de flexión del grupo metilo/metileno a 1455 cm^{-1} y 1376 cm^{-1} con crecimiento positivo, no se presentan las bandas del

grupo metileno a 2920 cm^{-1} y 2852 cm^{-1} ni la banda a 725 cm^{-1} correspondiente al grupo metileno (balanceo de alifáticos $(\text{CH}_2)_n$, $n>4$). La banda de tensión del grupo carbonilo a 1705 cm^{-1} aparece con una baja intensidad, no siendo importante en este componente.

En el tercer componente se presenta la banda del grupo carbonilo a 1705 cm^{-1} como banda principal y las bandas de los grupos metilo alifáticos a 2952 cm^{-1} y 2869 cm^{-1} , con sentido negativo. También se observan las bandas del grupo metileno a 2852 cm^{-1} y 725 cm^{-1} en sentido positivo. En este componente también se presentan las bandas a 865 cm^{-1} y 810 cm^{-1} asociadas a vibraciones del C-H aromático con 1 y 2 hidrógenos adyacentes, características de los crudos pesados. Lo anterior sugiere que para discriminar esta propiedad en el acpm pesado, el tercer componente tiene en cuenta los grupos funcionales en la región aromática para asociarlos al tipo de crudo (pesado o liviano).

Validación externa del Modelo. La validación externa del modelo PLS desarrollado se realizó con un set de 13 muestras no usadas para la calibración. En la tabla 22 se presenta el set de muestras de crudo usado para validación externa. El máximo error relativo de predicción es del 10.4%, lo cual demuestra el buen desempeño del modelo y su capacidad para predecir nuevas muestras.

Tabla 22. Validación externa del modelo rendimientos (%V) fracción acpm pesado

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|-----------|------------------|----------------|------------|--------------------|
| 8 | 9.42 | 8.44 | 0.98 | 10.4 |
| 10 | 11.3 | 10.8 | 0.53 | 4.73 |
| 13 | 9.32 | 8.80 | 0.52 | 5.58 |
| 14 | 9.91 | 9.82 | 0.09 | 0.95 |
| 15 | 8.77 | 9.66 | 0.90 | 10.2 |
| 23 | 8.24 | 8.44 | 0.20 | 2.42 |
| 25 | 10.9 | 10.5 | 0.39 | 3.61 |
| 26 | 10.3 | 10.3 | 0.06 | 0.59 |
| 28 | 14.0 | 14.4 | 0.35 | 2.51 |
| 38 | 10.9 | 11.4 | 0.57 | 5.24 |
| 42 | 10.3 | 11.0 | 0.65 | 6.25 |
| 53 | 10.3 | 10.9 | 0.62 | 5.98 |

La regresión obtenida (figura 43) para la validación externa del modelo arrojó un coeficiente de correlación de 0.800 y error de predicción de 0.762.

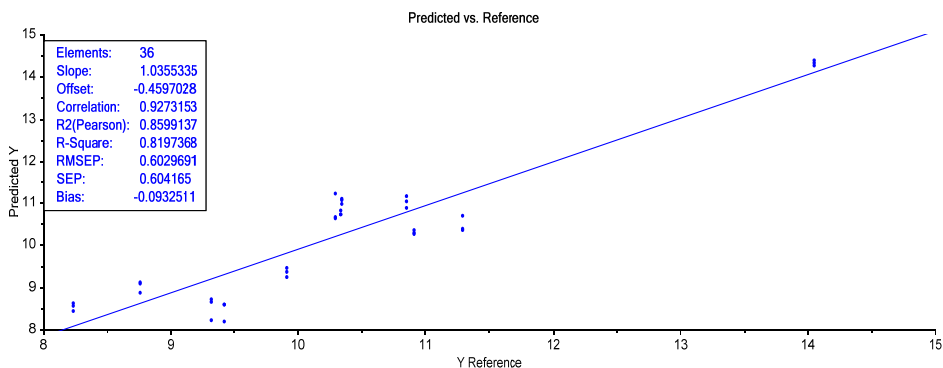


Figura 43. Validación externa rendimientos (%V) fracción acpm pesado.

3.3.4 Modelo PLS rendimientos crudo reducido

El modelo de rendimientos para la fracción crudo reducido arrojó mejores resultados en la estadística de desempeño que los modelos de rendimientos para las fracciones queroseno, acpm liviano y acpm pesado, según se observa en la figura 44.

La suma de las varianzas explicadas para los dos primeros componentes es del 93% en X y 94% en Y (figura 44 a), lo cual sugiere que estos dos componentes muestran las principales relaciones entre los datos y una gran cantidad de información.

En la figura 44 (c) se observa que con dos factores se explica más del 90% de las varianzas de calibración y de validación, en los datos de rendimientos en volumen (datos Y), adicionalmente, las dos varianzas son muy similares.

De acuerdo con el gráfico de valores predichos versus valores medidos de la figura 44 (d), en el modelo de calibración se obtuvo un coeficiente de correlación de 0.988 y error de calibración de 1.488, en un intervalo de rendimientos entre 20 y 80%. Para la validación, se obtuvo un

coeficiente de correlación del 0.985 y error de validación de 1.696, para el mismo intervalo de rendimientos.

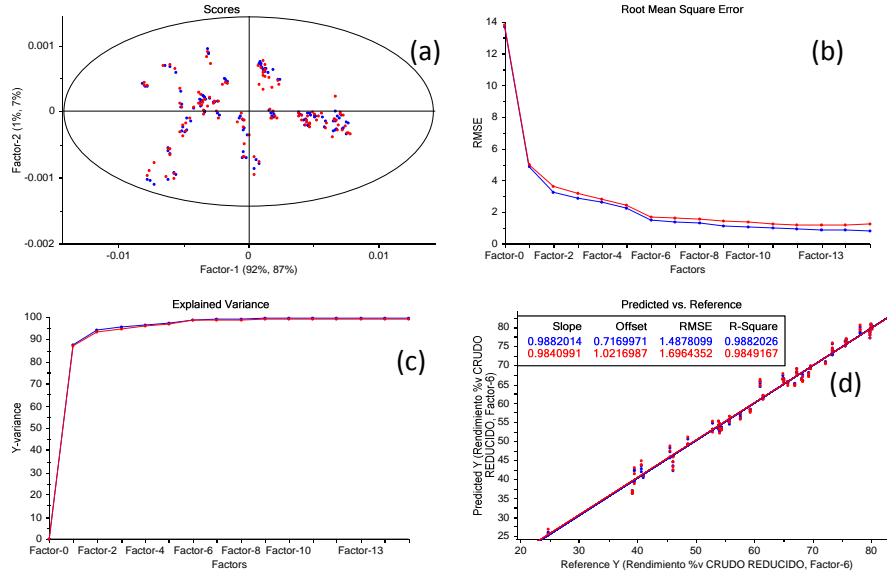


Figura 44. Principales resultados modelo PLS para rendimientos (%V) fracción crudo reducido.

Se seleccionaron seis factores como el número de variables latentes óptimo. El error RMSEC fue del 1.488% y 1.696% para RMSECV, en un intervalo de valores de rendimientos de 20% a 80%. En la figura 45 se presenta la varianza explicada en X, con un factor se explica el 92.2%, y con dos factores se explica el 93.5% de la varianza de calibración y validación en los datos espectrales de los crudos (datos X).

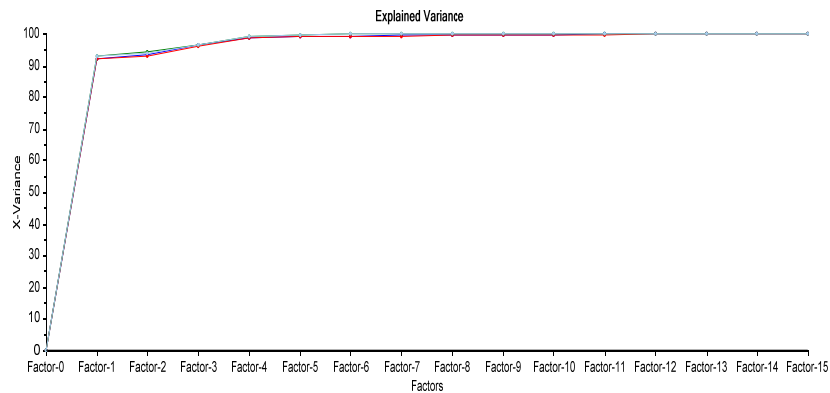


Figura 45. Varianza en X modelo PLS rendimientos (%V) fracción crudo reducido.

Variables importantes. Las regiones espectrales de importancia para los dos primeros componentes del modelo PLS de rendimientos en la fracción crudo reducido se muestran en la figura 46.

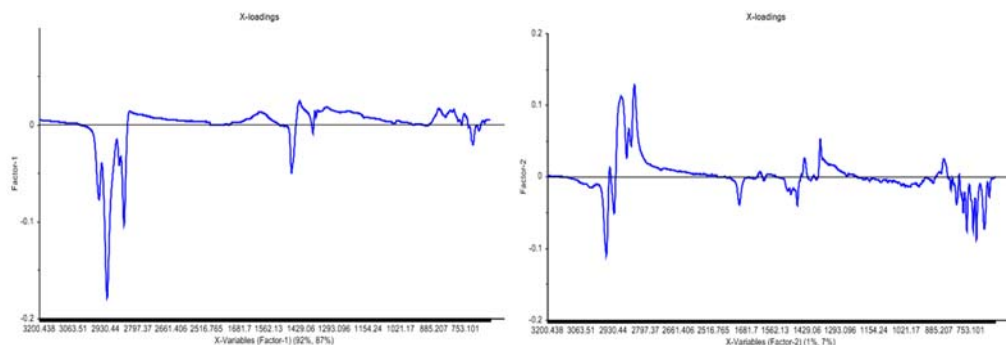


Figura 46. Variables importantes en el modelo PLS de rendimientos (%V) fracción crudo reducido.

El perfil del gráfico de Loadings para el primer componente identifica como variables importantes las mismas bandas identificadas para las fracciones queroseno y acpm liviano pero en sentido contrario (crecimiento negativo) y con intensidades similares. El perfil es similar al del acpm pesado pero las intensidades no lo son, lo cual significa que el grado de aporte de las variables X es diferente entre el acpm pesado y el crudo reducido, aunque su perfil sea muy similar (bandas vibracionales correspondientes a los mismos grupos funcionales).

En el segundo componente, además de tener en cuenta las bandas de alifáticos explicadas para el primer componente, se observan las bandas asociadas a vibraciones del grupo carbonilo a 1705 cm^{-1} y compuestos aromáticos. La banda a 865 cm^{-1} , correspondiente a vibraciones del C-H aromático con 1 hidrógeno adyacente, crece en sentido positivo mientras que las bandas a 810 cm^{-1} , 782 cm^{-1} , 766 cm^{-1} , 742 cm^{-1} y 699 cm^{-1} correspondientes a vibraciones del C-H aromático con 2 a 5 hidrógenos adyacentes y anillo aromático, respectivamente, lo hacen en sentido negativo. Lo anterior sugiere que este componente discrimina según el tipo de crudo (pesado o liviano).

Validación externa del Modelo. La validación externa del modelo PLS desarrollado se realizó con un set de 14 muestras. La regresión obtenida para la validación externa del modelo (figura 47) arrojó un coeficiente de correlación de 0.975 y error de predicción de 1.644.

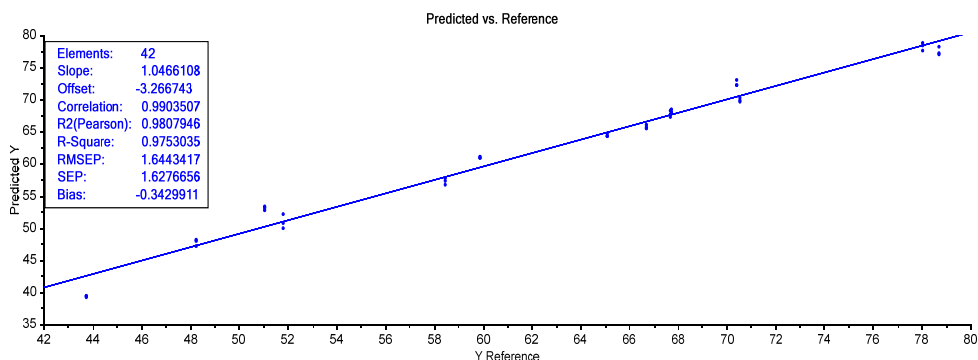


Figura 47. Validación externa modelo PLS rendimientos (%V) fracción crudo reducido.

En la tabla 23 se presentan los resultados de las predicciones realizadas para el set de muestras de validación externa.

Tabla 23. Validación externa del modelo PLS rendimientos (%V) fracción crudo reducido

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|-----------|------------------|----------------|------------|--------------------|
| 5 | 66.72 | 66.01 | 0.71 | 1.07 |
| 8 | 58.45 | 57.21 | 1.24 | 2.12 |
| 10 | 51.81 | 50.90 | 0.91 | 1.76 |
| 13 | 78.02 | 78.32 | 0.30 | 0.38 |
| 14 | 67.71 | 67.66 | 0.05 | 0.07 |
| 15 | 51.05 | 53.11 | 2.05 | 4.02 |
| 23 | 78.70 | 77.46 | 1.24 | 1.58 |
| 25 | 48.26 | 47.69 | 0.57 | 1.18 |
| 26 | 35.96 | 31.11 | 4.85 | 13.5 |
| 28 | 67.74 | 68.13 | 0.39 | 0.57 |
| 42 | 43.74 | 39.35 | 4.39 | 10.0 |
| 46 | 70.54 | 69.89 | 0.65 | 0.92 |
| 51 | 65.12 | 64.46 | 0.66 | 1.01 |
| 52 | 70.43 | 72.51 | 2.08 | 2.95 |
| 53 | 59.87 | 60.98 | 1.10 | 1.85 |

Se obtuvieron resultados con errores relativos máximos del 10%, lo que permite concluir que el modelo tiene un buen desempeño en la predicción de nuevas muestras.

El resumen de los parámetros estadísticos de los modelos PLS para rendimientos de las fracciones queroseno, acpm liviano, acpm pesado y crudo reducido se presentan en la tabla 24.

Tabla 24. Resultados modelos de calibración y validación PLS rendimientos (%V) destilados atmosféricos.

| Modelo | Media Calibración | R ² Calibración | RMSEC | R ² Validación | RMSECV | VL | Media Val. externa | R ² Predicción | RMSEP |
|------------|-------------------|----------------------------|-------|---------------------------|--------|----|--------------------|---------------------------|-------|
| Queroseno | 5.69 | 0.953 | 0.538 | 0.933 | 0.647 | 6 | 5.66 | 0.943 | 0.466 |
| Acpm liv. | 11.18 | 0.884 | 0.912 | 0.851 | 1.029 | 5 | 10.66 | 0.882 | 0.752 |
| Acpm pes. | 10.36 | 0.814 | 0.933 | 0.784 | 1.016 | 5 | 10.62 | 0.820 | 0.603 |
| Crudo red. | 60.77 | 0.988 | 1.488 | 0.985 | 1.696 | 6 | 61.39 | 0.975 | 1.644 |

El modelo PLS de rendimientos para la fracción crudo reducido arrojó los mejores parámetros estadísticos de calibración, validación y predicción externa. El modelo de rendimientos de acpm pesado arrojó el menor coeficiente de correlación en la predicción (0.820) con respecto a los demás modelos, sin embargo, los errores relativos de predicción fueron aceptables (inferiores al 15%).

3.4 SELECCIÓN DE VARIABLES MEDIANTE ALGORITMOS GENÉTICOS (AG)

Teniendo en cuenta que el objetivo principal de esta investigación fue correlacionar los rendimientos de los destilados atmosféricos con los espectros de los crudos de los cuales provienen estas fracciones, se realizó selección de variables mediante la técnica de algoritmos genéticos (AG) con el fin de optimizar los modelos. Se empleó el software PLS Toolbox Versión 8.0.2, licenciado por la Universidad Industrial de Santander, UIS para la selección de las variables y construcción de los modelos PLS.

Esta técnica fue empleada con el fin de escoger las variables que mejor correlacionan con los datos de rendimientos de los destilados al eliminar variables que no tienen relación con la propiedad de

interés. La técnica de algoritmos genéticos ayuda a identificar un subconjunto de las variables medidas (regiones espectrales) que son las más útiles para un modelo de regresión. El algoritmo busca obtener el subconjunto de variables que obtienen el más bajo error de validación cruzada (RMSECV). Teniendo un bloque de datos predictores X y un bloque de valores a predecir Y, se puede escoger un subconjunto al azar de variables de X y, mediante validación cruzada y cualquier método de regresión (en este caso PLSR), determinar el error de validación (RMSECV) obtenido usando sólo ese subconjunto de variables en un modelo de regresión. El algoritmo genético usa esta aproximación iterativamente para localizar los subconjuntos de variables que dan el RMSECV más bajo, que es la respuesta a maximizar (Barry M. Wise, Neal B. Gallager, Rasmus Bro, Jeremy M. Shaver, Willem Windig 2006).

La ejecución del algoritmo involucra una serie de pasos: genera subconjuntos de variables al azar, evalúa cada subconjunto según su aptitud para predecir Y, descarta la peor mitad de individuos, engendra individuos remanentes, permite la mutación y repite los pasos anteriores hasta que el criterio final se cumpla. Estos pasos son realizados por el algoritmo mediante los parámetros evaluados en la tabla 25 y que son descritos a continuación.

Tamaño de la población: corresponde al pool de todos los individuos probados. El tamaño de la población es un parámetro cuya selección requiere indagación. Considerar unos pocos individuos puede llevar a la exploración de una pequeña parte de la población, mientras que lo contrario penaliza el algoritmo desde el punto de vista computacional. Si el tamaño es 64 (subconjuntos de variables) y el número de variables es 3474, se tendrán 54 variables en cada subconjunto (3474/64).

Ancho de ventana: indica cuántas variables adyacentes o sucesivas deberían agruparse a la vez. Cuando variables adyacentes contienen información correlacionada y se busca incluir o excluir variables adyacentes como un bloque (por ejemplo, medidas de baja resolución), las variables originales pueden agruparse y ser incluidas o excluidas en bloques.

Porcentaje de términos iniciales: especifica el número aproximado de variables incluidas en los subconjuntos iniciales de variables. Seleccionar 20 indica que se incluye inicialmente el 20% del número total de variables.

Máximas Generaciones: es el número máximo de iteraciones para generar una población inicial, evaluar la adaptación de los individuos en la población, y crear una nueva población mediante la selección, cruce y mutación.

Porcentaje de convergencia: valor al cual se espera que los individuos de la población hayan convergido, de modo que al final del proceso la población sea muy similar y se tenga una representación de la mayor parte de la población. El criterio es un cierto porcentaje de la población que es idéntico y generalmente se maneja un 50%.

Tasa de mutación: es la probabilidad con la que mutan los dos descendientes de dos individuos para ser insertados en la nueva población. Generalmente se usa una probabilidad de mutación del 0.5% (valor de 0.005) o 1% (valor de 0.01).

Cruzamiento: permite seleccionar uno de dos métodos para realizar el engendramiento de individuos (single o double) para obtener los descendientes.

Método de validación cruzada: permite realizar una estimación del desempeño de un modelo cuando se aplica a datos desconocidos. El método genera subconjuntos de objetos (número de splits), remueve un subconjunto de objetos y construye un modelo usando los objetos remanentes y luego aplica el modelo resultante a los objetos removidos. Se puede seleccionar el tipo de validación cruzada (al azar o contigua).

El método de regresión usado fue mínimos cuadrados parciales (PLS) y el número máximo de variables latentes permitidas fue de 10, el cual corresponde al número máximo de variables latentes empleado en la validación cruzada de los modelos PLS construidos previamente.

Con el fin de obtener las mejores variables se realizó optimización de los parámetros presentados en la tabla 25 para los espectros y datos de rendimientos de la fracción queroseno. El método de engendramiento seleccionado fue el de cruzamiento doble. Para iniciar se usaron los parámetros por default del algoritmo (AG1) contra el cual se compararon los resultados en términos del valor promedio del RMSECV obtenido para cada algoritmo genético.

Tabla 25. Optimización de parámetros del algoritmo genético (AG)

| Parámetro | AG 1 | AG 2 | AG 3 | AG 4 | AG 5 | AG 6 | AG 7 | AG 8 | AG 9 | AG 10 | AG 11 | AG 12 | Final |
|----------------------|--------|------------|-----------|-----------|-------------|-----------|-------------|-------------|---------------|-----------|----------|----------|-------------|
| Tamaño población | 64 | 128 | 64 | 64 | 64 | 64 | 64 | 64 | 64 | 64 | 64 | 64 | 64 |
| Ancho de ventana | 25 | 25 | 40 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 40 |
| % Términos iniciales | 20 | 20 | 20 | 30 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 30 |
| Max. generaciones | 10 | 10 | 10 | 10 | 25 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 25 |
| % Convergencia | 25 | 25 | 25 | 25 | 25 | 50 | 25 | 25 | 25 | 25 | 25 | 25 | 50 |
| Tasa de mutación | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.01 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 |
| Número VL | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 10 | 5 | 5 | 5 | 5 | 10 |
| Método validación | Contig | Contig | Contig | Contig | Contig | Contig | Contig | Contig | Random | Contig | Contig | Contig | Random |
| Número splits | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 10 | 5 | 5 | 5 |
| Número iteraciones | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 5 | 1 | 1 |
| Réplicas de corridas | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 5 | 5 |
| RESULTADOS | | | | | | | | | | | | | |
| Mejor RMSECV | 0.98 | 0.82 | 0.90 | 0.97 | 0.76 | 0.75 | 0.98 | 0.55 | 0.61 | 0.89 | 0.75 | 0.84 | 0.30 |
| Promedio RMSECV | 1.02 | 0.97 | 1.02 | 0.99 | 0.84 | 0.91 | 1.01 | 0.65 | 0.71 | 0.99 | 0.94 | 0.97 | 0.33 |

Los resultados mostraron que los tres parámetros de mayor influencia en el algoritmo son el número máximo de generaciones, el número de variables latentes y el método de validación cruzada, dado que se obtuvieron los menores valores de RMSECV al incrementar los valores de los dos primeros y seleccionar como método de validación cruzada el método al azar (random). La variación de los demás parámetros no mostró una mejora sustancial del RMSECV. En la última columna se muestra la combinación de valores seleccionados y aplicados al algoritmo genético final. Se obtuvieron valores de RMSECV de 0.30 como mejor RMSECV y 0.33 como valor promedio de RMSECV.

Los parámetros finales involucraron un tamaño de la población de 64 individuos seleccionados al azar, es decir, 64 subconjuntos de variables con un ancho de ventana espectral de 25 (bloques de 25 variables). Lo anterior permitió reducir el número de bloques de variables o ventanas

espectrales a 139 (3474/25), distribuidas en todo el dominio espectral para todas las fracciones analizadas (queroseno, acpm liviano, acpm pesado y crudo reducido).

Finalmente, el algoritmo genético fue corrido de manera independiente tres veces, aplicando los parámetros escogidos anteriormente. Las variables X seleccionadas fueron almacenadas como tres bloques de datos X independientes, los cuales fueron usados para construir los modelos PLS-AG preliminares con el fin de evaluar la repetibilidad en la selección de variables (tabla 26).

El promedio de RMSECV obtenido estuvo en el intervalo de 0.73 a 0.82 y el número de variables seleccionadas para el bloque de datos X, entre 775 a 899. Los modelos PLS-AG construidos con los tres bloques de datos X mostraron variaciones significativas en los errores medios de calibración y validación, especialmente en la réplica PLS-AG R2, con valores altos de RMSEC y RMSECV.

Tabla 26. Repetibilidad del algoritmo genético rendimientos (%V) de queroseno

| Resultados algoritmos genéticos (AG) | | | |
|---|------------------|------------------|------------------|
| | AG R1 | AG R2 | AG R3 |
| Mejor RMSECV | 0.72 | 0.73 | 0.68 |
| Promedio RMSECV | 0.80 | 0.82 | 0.73 |
| No. Variables selecc. | 899 | 800 | 775 |
| Resultados modelos PLS-GA | | | |
| | PLS-AG R1 | PLS-AG R2 | PLS-AG R3 |
| RMSEC | 0.52 | 0.89 | 0.36 |
| R ² Calibración | 0.97 | 0.88 | 0.98 |
| RMSECV | 0.66 | 0.90 | 0.44 |
| R ² Validación | 0.93 | 0.87 | 0.97 |
| RMSEP | 0.98 | 0.92 | 1.00 |
| R ² Predicción | 0.85 | 0.85 | 0.80 |

Los resultados del modelo PLS-AG varían de una réplica a otra. Esto podría deberse a que el método de validación cruzada al azar (random), usado para seleccionar las mejores variables durante la corrida del algoritmo genético, tiende a mostrar variabilidad en los resultados, debido únicamente a la selección de las variables al azar. De acuerdo con lo descrito en literatura (Leardi & Lupiáñez González 1998), una corrida sencilla del algoritmo genético (AG) no es suficiente debido a que sólo unas pocas variables estarían presentes y la ventaja de usar PLS no sería

totalmente explotada. Por lo tanto, se incrementaron las réplicas de corridas de 5 a 100, para reducir la probabilidad de seleccionar variables no informativas, dado que cuando el algoritmo genético es ejecutado consecutivamente varias veces, las variables importantes tienden a ser repetidamente seleccionadas, mientras que las menos importantes son seleccionadas al azar.

Adicionalmente, se incrementó el número de generaciones de 25 a 100 para que la condición de terminación del algoritmo fuera 100 generaciones. Lo anterior debido a que se observó que la línea del RMSECV promedio (línea azul) y el mejor RMSECV (línea verde) no convergían, lo que significa que la población no es similar, necesitando un mayor número de generaciones para permitir esta convergencia. Lo anterior se observa en la figura 48. La línea punteada roja corresponde al RMSECV usando todas las variables.

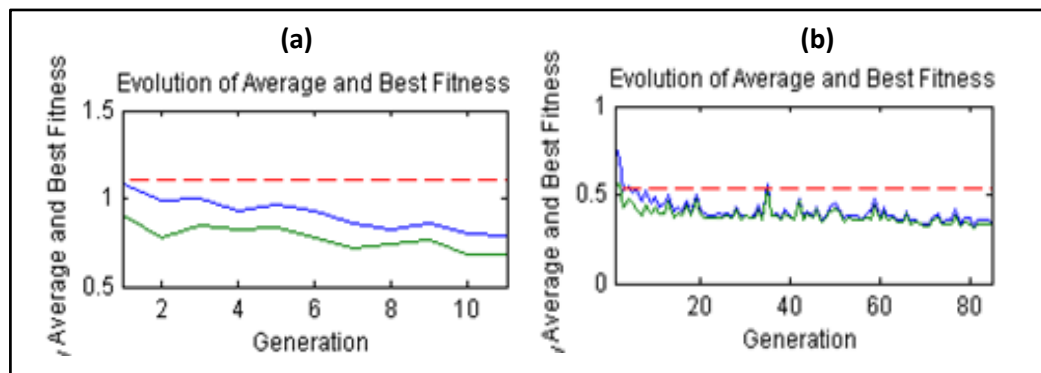


Figura 48. RMSECV promedio y mejor RMSEC versus número de generaciones: (a) 25 generaciones. (b) 100 generaciones.

Teniendo en cuenta que el algoritmo genético es repetido hasta que la condición de terminación se cumpla, ya sea cuando alcance el número de generaciones (en este caso 100) o cuando se cumpla el criterio de convergencia (50% de la población es idéntica), el aumento del número de generaciones permitió que el algoritmo terminara según el criterio de convergencia, dado que el número de generaciones alcanzó valores inferiores a 100 y en los gráficos del RMSECV versus el número de generaciones se observa que los RMSECV promedio y el mejor RMSECV convergen antes de las 100 generaciones, según se observa en la Figura 48 (b).

Sin embargo, los cambios anteriores significaron un costo en tiempo de procesamiento, aumentando en promedio de 1 a 16 horas la corrida para cada algoritmo genético. En la tabla 27 se presentan los resultados de los parámetros definitivos obtenidos en la selección de variables mediante algoritmos genéticos para los rendimientos de los destilados atmosféricos.

Tabla 27. Parámetros finales algoritmos genéticos para rendimientos destilados atmosféricos

| Parámetro | Valor |
|---------------------------|--------------|
| Tamaño de la población | 64 |
| Ancho de ventana | 25 |
| % Términos iniciales | 30 |
| Max. generaciones | 100 |
| % Convergencia | 50 |
| Tasa de mutación | 0.005 |
| Método de regresión | PLS |
| Número variables latentes | 10 |
| Método validación | Random |
| Número de splits | 5 |
| Número de iteraciones | 1 |
| Réplicas de corridas | 100 |

3.4.1 Selección de variables para rendimientos queroseno

Los gráficos de ajuste del RMSECV y frecuencia de uso de las ventanas en los modelos incluyen los resultados compilados de todas las réplicas de las corridas realizadas. El gráfico del error medio de validación cruzada (RMSECV) versus el número de variables permite buscar tendencias en el uso de variables. Las variables que aparecen en la parte inferior del gráfico son las que mejoran el RMSECV, las que aparecen en la parte superior desmejoran el modelo. Las variables que aparecen en todos los modelos son variables útiles y las que no aparecen en absoluto, no lo son.

La figura 49 muestra los resultados de 2012 modelos después de correr 100 generaciones del algoritmo genético para los rendimientos de la fracción queroseno. Se observa que las regiones seleccionadas caen sobre o cerca de los picos espectrales del espectro medio, superpuesto como referencia en el gráfico.

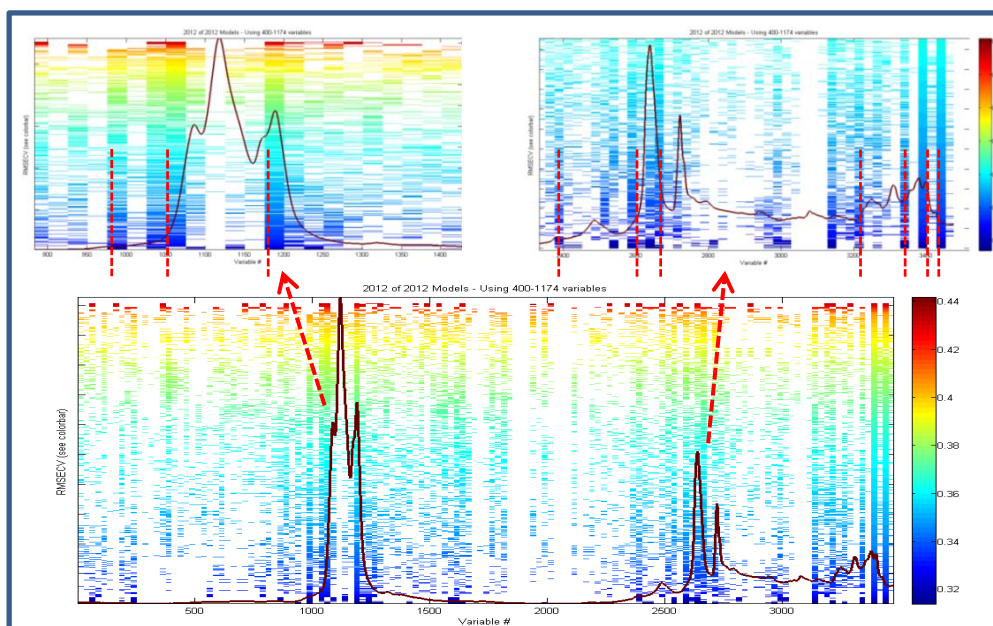


Figura 49. Ajuste RMSECV versus número de variables algoritmo genético rendimientos (%V) fracción queroseno.

Las ventanas de variables útiles (demarcadas con líneas rojas) corresponden a las identificadas en los intervalos de ventanas y bandas vibracionales mostradas en la tabla 28, regiones que contienen total o parcialmente las bandas espectrales respectivas.

El algoritmo genético consideró algunas variables como no útiles o que van en detrimento del error medio de validación (RMSECV). Las bandas fuertes de tensión asimétrica del grupo metileno (R_2-CH_2) a 2920 cm^{-1} y tensión simétrica del grupo metilo ($R-CH_3$) a 2869 cm^{-1} fueron usadas por muy pocos modelos.

La figura 50 muestra la frecuencia de uso de los bloques de variables seleccionadas en los modelos. Los bloques con mayor frecuencia de inclusión sugieren que éstas son las variables de mayor utilidad para construir los modelos AG-PLS. Estas variables, originadas por las vibraciones de los diferentes grupos funcionales presentes en los espectros de los crudos, pueden ser asociadas a las propiedades (variables Y) mediante los modelos quimiométricos.

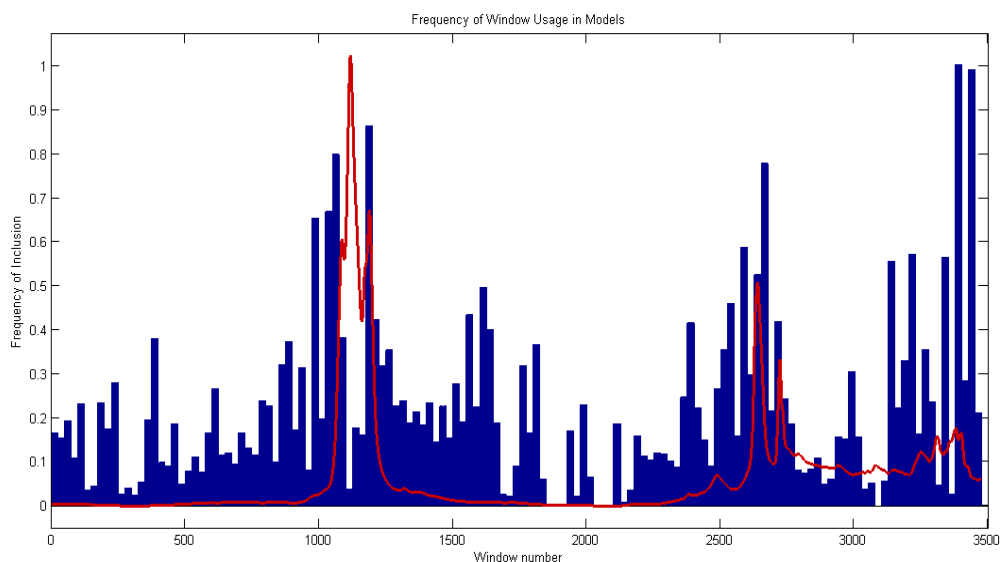


Figura 50. Frecuencia de inclusión bloques de variables del algoritmo genético para rendimientos (%V) fracción queroseno.

De los bloques o ventanas de variables seleccionados previamente por el algoritmo genético, las más usadas son las que aparecen con mayor frecuencia. Estas ventanas contienen las vibraciones de tensión C-H asimétrica del grupo metílico ($R-CH_3$) y simétrica del grupo metileno (R_2-CH_2) alrededor de 2952 cm^{-1} y 2852 cm^{-1} , la tensión del C-H aromático en el intervalo $3100-3500\text{ cm}^{-1}$ y flexión del grupo metílico/metileno ($R-CH_3$ ó R_2-CH_2) a 1455 cm^{-1} .

También se observan las vibraciones de deformación del C-H aromático para 1, 3 y 5 hidrógenos adyacentes, así como la banda $(CH_2)_n$, $n > 4$ a 725 cm^{-1} . Las bandas alrededor de 699 cm^{-1} y 674 cm^{-1} , son atribuidas a la flexión del anillo aromático o encogimiento de bencenos o xilenos. La banda del grupo carbonilo ($C=O$) a 1705 cm^{-1} se presenta con relativa baja frecuencia (tabla 28).

Adicionalmente, el algoritmo genético seleccionó bloques de variables que no contienen señales que correspondan a vibraciones de grupos funcionales definidos sino, al parecer, a regiones con ruido. Lo mismo sucedió en la selección de variables de los espectros FTIR-ATR para las fracciones de acpm liviano, acpm pesado y crudo reducido.

Tabla 28. Variables útiles y asignación de bandas vibracionales rendimientos (%V) fracción queroseno mediante AG

| <i>Ventana No.</i> | <i>Bandas de absorción (cm⁻¹)</i> | <i>Grupo Funcional</i> | <i>Modo vibracional</i> |
|--------------------|--|---|-------------------------------------|
| 975-1000 | 3090 – 3025 | C-H (C=C y aromáticos) | Tensión |
| 1025-1075 | 2952 | R-CH ₃ | Tensión asimétrica |
| 1175-1200 | 2852 | R ₂ -CH ₂ | Tensión simétrica |
| 2375-2400 | 1705 | C=O | Tensión |
| 2575-2675 | 1455 | R-CH ₃ ó R ₂ -CH ₂ | Flexión |
| 3200-3225 | 865 | 1 H adyacente | Deformación C-H |
| 3325-3350 | 782 | 3 H adyacente | aromático |
| 3375-3400 | 742, 725 | 5 H adyacente, (CH ₂) _n ; n >= 4 | Deformación C-H aromático, Balanceo |
| 3425-3450 | 699, 674 | Anillo aromático ó bencenos ó xilenos | Flexión |

Lo anterior es lógico si se tiene en cuenta que los algoritmos genéticos seleccionan variables distribuidas a través de todo el espectro, no en rangos espectrales específicos. El set entero de variables debe considerarse como un todo en lugar de considerar cada variable como útil por sí misma. Esto se debe a que una variable puede ser útil para predicción sólo cuando es usada en conjunto con otras variables incluidas en un subconjunto de variables individuales (Barry M. Wise, Neal B. Gallager, Rasmus Bro, Jeremy M. Shaver, Willem Windig 2006).

El interés al usar esta técnica es eliminar variables que no tienen relación con la propiedad bajo estudio, sin embargo, el algoritmo genético no sólo selecciona variables importantes, sino que también selecciona variables aparentemente no informativas (Rocha et al. 2016).

3.4.2 Selección de variables para rendimientos acpm liviano

En la figura 51 se presentan los resultados de 2037 modelos después de correr 100 generaciones del algoritmo genético para los rendimientos de la fracción acpm liviano.

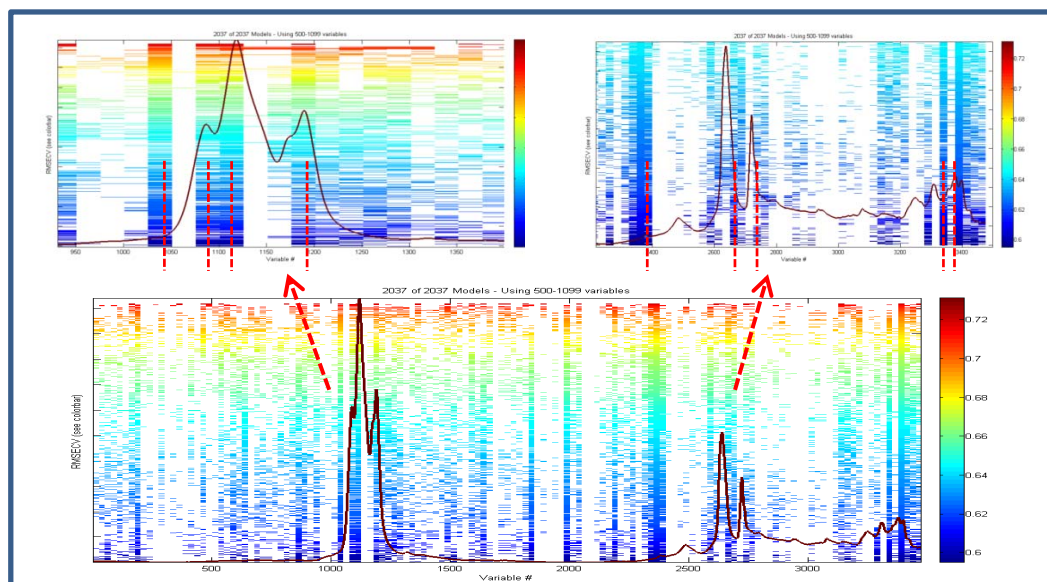


Figura 51. Ajuste RMSECV versus número de variables algoritmo genético rendimientos fracción acpm liviano.

Las ventanas de variables más útiles (demarcadas con líneas rojas) corresponden a las identificadas en los intervalos de ventanas y bandas vibracionales mostradas en la tabla 29. El algoritmo genético consideró algunas variables como no útiles o que van en detrimento del error medio de validación (RMSECV). Tal es el caso de la banda fuerte de tensión simétrica del grupo metilo ($R-CH_3$) a 2869 cm^{-1} , usada por muy pocos modelos.

La figura 52 muestra la frecuencia de uso de los bloques de variables seleccionadas en los modelos. De las ventanas de variables seleccionadas previamente por el algoritmo genético, las más usadas son las que aparecen con mayor frecuencia. Estas ventanas contienen las vibraciones de tensión C-H asimétrica del grupo metílico ($R-CH_3$) alrededor de 2952 cm^{-1} y tensiones asimétrica y simétrica del grupo metileno (R_2-CH_2) alrededor de 2920 cm^{-1} y 2852 cm^{-1} ; la tensión del $-OH$, $NH-$ en el intervalo $3100-3500\text{ cm}^{-1}$, la tensión del C-H aromático a $3090-3025\text{ cm}^{-1}$ y la tensión del grupo carbonilo ($C=O$) a 1705 cm^{-1} .

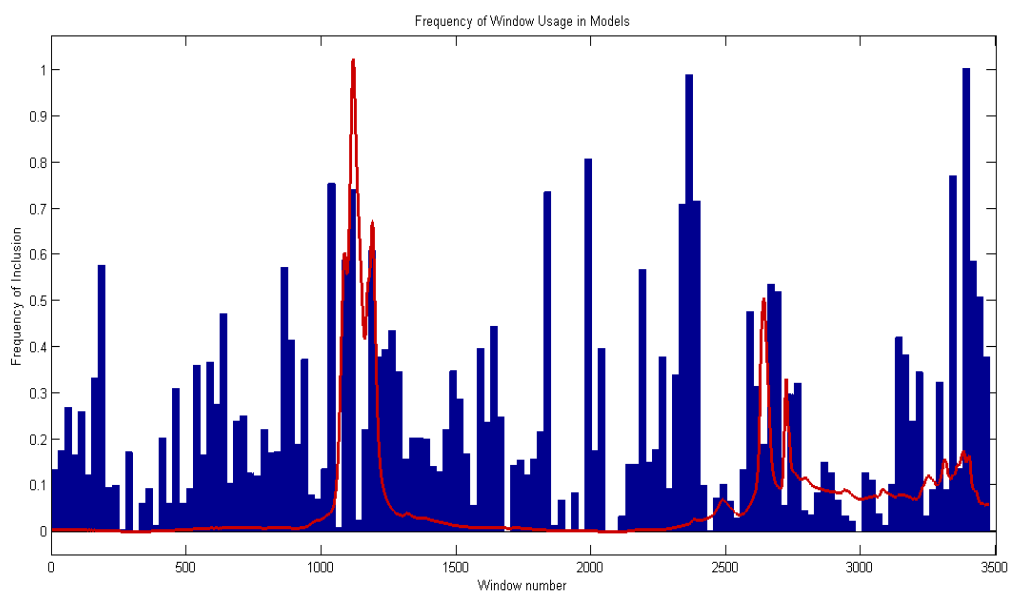


Figura 52. Frecuencia de inclusión bloques de variables del algoritmo genético para rendimientos fracción acpm liviano.

También se observan las vibraciones de deformación del C-H aromático para 3 hidrógenos adyacentes, así como la banda $(CH_2)_n$, $n > 4$ a 725 cm^{-1} . La banda de flexión simétrica del grupo metílico/metileno ($R-CH_3$ ó R_2-CH_2) a 1455 cm^{-1} se presenta con relativa baja frecuencia en esta fracción, así como la banda de flexión asimétrica del grupo metílico ($R-CH_3$) alrededor de $1376-1365 \text{ cm}^{-1}$. Estas bandas presentan baja frecuencia o sólo una parte de la ventana es tenida en cuenta en los modelos.

Las bandas seleccionadas por el algoritmo genético se presentan en la tabla 29.

Tabla 29. Variables útiles y asignación de bandas vibracionales rendimientos acpm liviano

| <i>Ventana No.</i> | <i>Bandas de absorción (cm⁻¹)</i> | <i>Grupo Funcional</i> | <i>Modo vibracional</i> |
|--------------------|--|---|------------------------------|
| 1025-1050 | 3090 – 3025 | C-H (C=C y aromáticos) | Tensión |
| 1075-1100 | 2952 | R-CH ₃ | Tensión asimétrica |
| 1100-1125 | 2920 | R ₂ -CH ₂ | Tensión asimétrica |
| 1175-1200 | 2852 | R ₂ -CH ₂ | Tensión simétrica |
| 2325-2400 | 1705 | C=O | Tensión |
| 2650-2700 | 1455 | R-CH ₃ ó R ₂ -CH ₂ | Flexión |
| 2725-2750 | 1376-1365 | R-CH ₃ | Flexión |
| 3325-3350 | 782 | 3 H adyacente | Deformación C-H aromático |
| 3375-3400 | 742, 725 | 5 H adyacente, (CH ₂) _n ; n >= 4 | Def. C-H aromático, Balanceo |

3.4.3 Selección de variables para rendimientos acpm pesado

El algoritmo genético para esta fracción arrojó 2064 modelos (figura 53). Como característica particular, no seleccionó regiones cerca de los picos espectrales correspondientes a las bandas vibracionales de los grupos metílico y metileno entre 2952 cm⁻¹ y 2852 cm⁻¹, fueron consideradas por el algoritmo como variables no útiles.

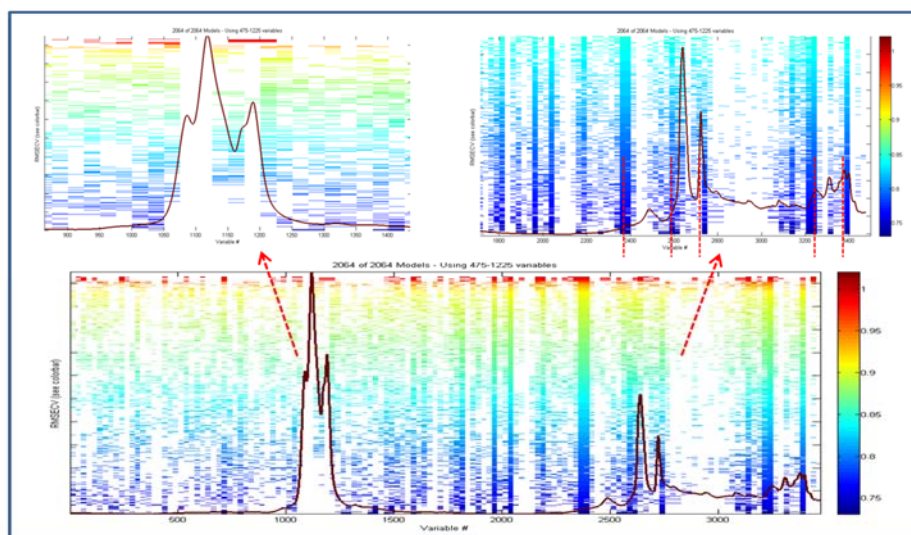


Figura 53. Ajuste RMSECV versus número de variables AG rendimientos acpm pesado.

Las ventanas de variables más útiles (demarcadas con líneas rojas) corresponden a las identificadas en los intervalos de ventanas y bandas vibracionales mostradas en la tabla 30. La figura 54 muestra la frecuencia de uso de los bloques de estas variables.

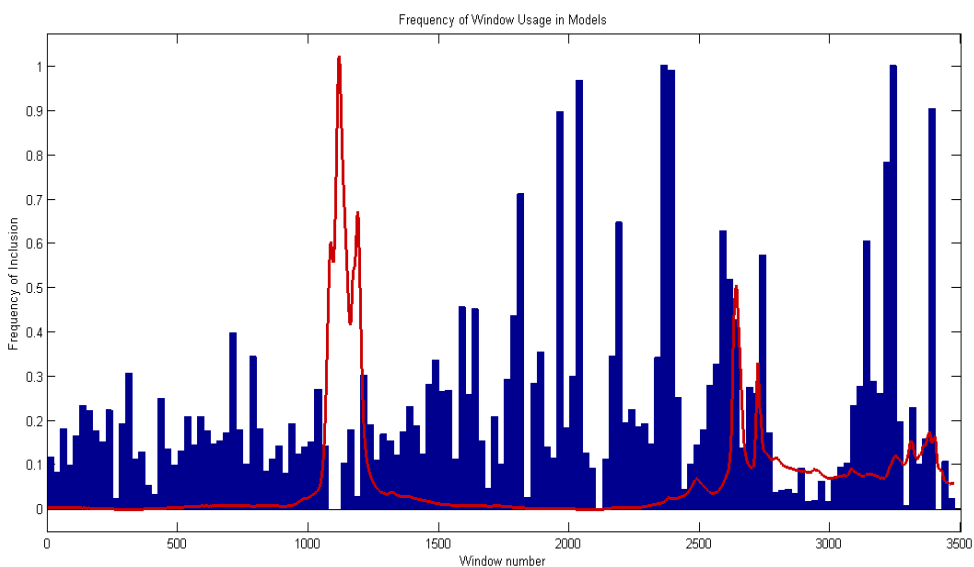


Figura 54. Frecuencia de inclusión bloques de variables del algoritmo genético para rendimientos fracción acpm pesado.

Tabla 30. Variables útiles y asignación de bandas vibracionales rendimientos acpm pesado (algoritmos genéticos)

| <i>Ventana No.</i> | <i>Bandas de absorción (cm⁻¹)</i> | <i>Grupo Funcional</i> | <i>Modo vibracional</i> |
|--------------------|--|---|------------------------------|
| 2350-2400 | 1705 | C=O | Tensión |
| 2575-2625 | 1455 | R-CH ₃ ó R ₂ -CH ₂ | Flexión |
| 2725-2750 | 1376-1365 | R-CH ₃ | Flexión |
| 3200-3250 | 865 | 1 H adyacente | Deformación C-H aromático |
| 3375-3400 | 742, 725 | 5 H adyacente, (CH ₂) _n ; n >= 4 | Def. C-H aromático, Balanceo |

Las ventanas de variables que aparecen con mayor frecuencia contienen las vibraciones de tensión del grupo carbonilo (C=O) a 1705 cm^{-1} , la banda de flexión simétrica del grupo metílico/metileno (R-CH_3 ó $\text{R}_2\text{-CH}_2$) a 1455 cm^{-1} , así como la banda de flexión asimétrica del grupo metílico (R-CH_3) alrededor de $1376\text{-}1365\text{ cm}^{-1}$. También se observan las vibraciones de deformación del C-H aromático para 1 y 5 hidrógenos adyacentes alrededor de 865 cm^{-1} y 742 cm^{-1} , así como la banda $(\text{CH}_2)_n$, $n>4$ a 725 cm^{-1} .

3.4.5 Selección de variables para rendimientos crudo reducido

El algoritmo genético para esta fracción obtuvo 2022 modelos. Las variables seleccionadas son mostradas en la figura 55.

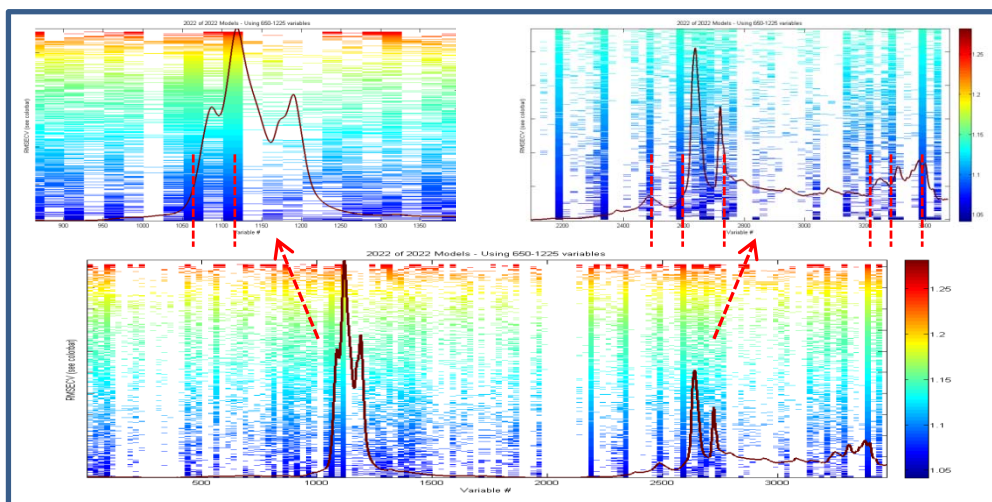


Figura 55. Ajuste RMSECV versus número de variables algoritmo genético rendimientos fracción crudo reducido.

Los bloques de variables anteriormente seleccionadas y su frecuencia de inclusión son mostrados en la figura 56.

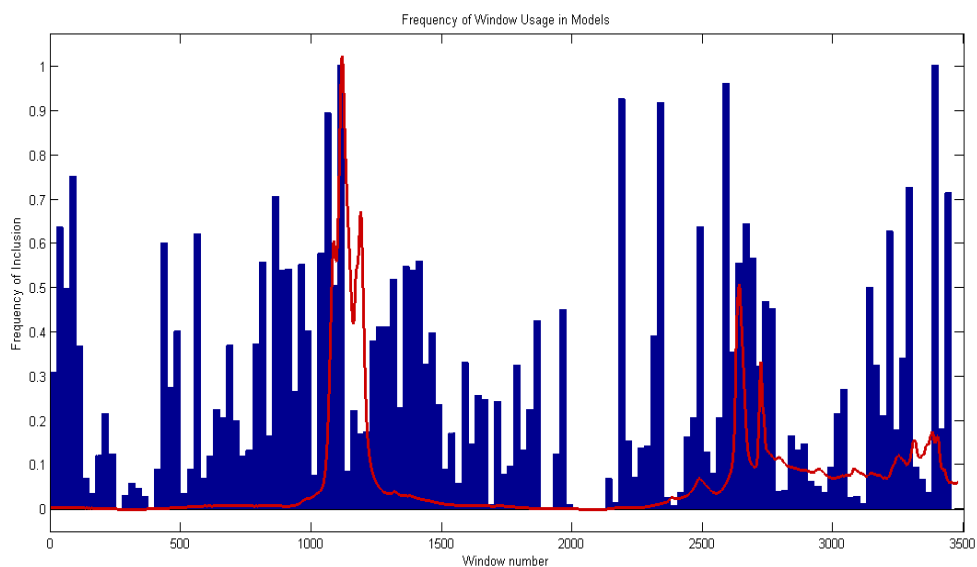


Figura 56. Frecuencia de inclusión bloques de variables del algoritmo genético para rendimientos (%V) fracción crudo reducido.

Las variables de mayor frecuencia de inclusión seleccionadas por el algoritmo genético para esta fracción se presentan en la tabla 31.

Tabla 31. Variables útiles y asignación de bandas vibracionales rendimientos crudo reducido

| <i>Ventana No.</i> | <i>Bandas de absorción (cm⁻¹)</i> | <i>Grupo Funcional</i> | <i>Modo vibracional</i> |
|-------------------------|--|--|------------------------------------|
| 1050-1075 | 2952 | R-CH ₃ | Tensión asimétrica |
| 1100-1125 | 2920 | R ₂ -CH ₂ | Tensión asimétrica |
| 2475-2500 | 1601 | C = C aromático | Tensión |
| 2575-2600, 2625-2700 | 1455 | R-CH ₃ ó R ₂ -CH ₂ | Flexión |
| 2725-2750 | 1376-1365 | R-CH ₃ | Flexión |
| 3200-3225 | 865 | 1 H adyacente | Deformación C-H |
| 3275-3300 | 810 | 2 H adyacente | aromático |
| 3375-3400 | 742, 725 | 5 H adyacente, (CH ₂) _n ; n >= 4 | Def. C-H aromático, Balanceo |

Se destaca la inclusión de la banda de tensión C=C aromático a 1601 cm^{-1} , y las bandas correspondientes a las vibraciones de deformación del C-H aromático para 1 y 2 hidrógenos adyacentes alrededor de 865 cm^{-1} y 810 cm^{-1} , las cuales son más intensas en los espectros de crudos pesados que en los livianos, de acuerdo a lo explicado en la sección 2.4.

3.4.6 Análisis regiones espectrales seleccionadas por AG

En la tabla 32 se relacionan las bandas vibracionales que podrían corresponder a las ventanas de las variables más útiles, seleccionadas por los algoritmos genéticos para las fracciones de destilados atmosféricos. Arbitrariamente se seleccionaron las ventanas con frecuencia de uso superior al 50%, con el fin de analizar las bandas vibracionales más relevantes. Sin embargo, los modelos tienen en cuenta otras variables, como regiones con ruido, según lo explicado en la sección 3.4.1.

Las bandas del intervalo de alifáticos correspondientes a las tensiones de los grupos metilo ($-\text{CH}_3$) y metileno ($-\text{CH}_2-$) a 2952 cm^{-1} , 2920 cm^{-1} , 2869 cm^{-1} y 2852 cm^{-1} , permitieron al algoritmo genético discriminar los rendimientos para las fracciones atmosféricas, mediante diversos valores de frecuencia de inclusión. Lo anterior es lógico si se tiene en cuenta que bandas como la del grupo metilo a 2952 cm^{-1} son más intensas y se resuelven mejor en los espectros de crudos livianos que en los de crudos pesados, por consiguiente, es lógico suponer que estas bandas permitan al algoritmo genético discriminar los rendimientos de los destilados que provienen de estos crudos.

La banda del grupo carbonilo ($\text{C}=\text{O}$) a 1705 cm^{-1} , observada principalmente en los espectros de crudos ácidos, presentó frecuencias de inclusión que aumentaron en la medida que la fracción se tornó más pesada, desde la fracción queroseno hasta el acpm pesado. Sin embargo, esta banda no fue tomada en cuenta en los modelos de algoritmos genéticos para crudo reducido (frecuencia de inclusión muy baja). Lo anterior está de acuerdo con la tendencia observada en los datos del número de ácido de los destilados atmosféricos: la acidez se distribuye en las fracciones de forma ascendente hasta el acpm pesado. En el crudo reducido los valores de esta propiedad disminuyen notoriamente.

Tabla 32. Comparación variables útiles y asignación de bandas vibracionales rendimientos fracciones atmosféricas

| <i>Bandas de absorción (cm⁻¹)</i> | <i>Grupo Funcional</i> | <i>Quero</i> | <i>Acpm liviano</i> | <i>Acpm pesado</i> | <i>Crudo red.</i> |
|--|---|--------------|---------------------|--------------------|-------------------|
| 3100-3500 | -OH, NH- (enlace H intermolecular) | | | | |
| 3090 – 3025 | C-H (C=C y aromáticos) | X | X | | |
| 2952 | R-CH ₃ | X | X | | X |
| 2920 | R ₂ -CH ₂ | | X | | X |
| 2869 | R-CH ₃ | | | | |
| 2852 | R ₂ -CH ₂ | X | X | | |
| 2727 | CH ₃ (de anillo fenílico metil sustituido) | | | | |
| 1705 | C=O | X | X | X | |
| 1601 | C = C aromático | | | | X |
| 1455 | R-CH ₃ ó R ₂ -CH ₂ | X | X | X | X |
| 1376-1365 | R-CH ₃ | | X | X | X |
| 1305 | Anillo fenílico | | | | |
| 1160 | Anillo fenílico | | | | |
| 1032 | Sulfóxido | | | | |
| 865 | 1 H adyacente | X | | X | X |
| 810 | 2 H adyacente | | | | X |
| 782 | 3 H adyacente | X | X | | |
| 766 | 4 H adyacente | | | | |
| 742 | 5 H adyacente | X | X | X | X |
| 725 | (CH ₂) _n ; n >= 4 | X | X | X | X |
| 699, 674 | Anillo aromático ó bencenos ó xilenos | X | | | |

La banda débil de tensión de enlaces C=C asociados a aromáticos en 1601 cm⁻¹, tiene una alta frecuencia de inclusión en el modelo de crudo reducido, lo cual demuestra que esta banda permitió discriminar la fracción de destilado más pesada, dado que el crudo reducido es la fracción con mayor cantidad de moléculas con varios anillos aromáticos, si se compara con los demás destilados atmosféricos.

La banda a 865 cm^{-1} (vibraciones del C-H aromático con 1 hidrógeno adyacente) contribuye a los modelos de acpm pesado y crudo reducido, con una alta frecuencia de inclusión. Adicionalmente, la banda a 810 cm^{-1} (vibraciones del C-H aromático con 2 hidrógenos adyacentes) contribuye al modelo de crudo reducido, estas dos bandas son más intensas en los espectros de crudos pesados que en los livianos, de modo que es lógico que aparezcan con una alta frecuencia de inclusión para esta fracción y baja frecuencia de inclusión para las fracciones más livianas.

Las bandas a 782 cm^{-1} y 742 cm^{-1} , la primera con alta frecuencia de inclusión en las fracciones para queroseno y acpm liviano, y la segunda con una frecuencia de inclusión similar para las cuatro fracciones, denotan en su conjunto una asociación con las fracciones livianas, dado que estas bandas están asociadas a vibraciones del C-H aromático con 3 y 5 hidrógenos adyacentes y son más intensas en los espectros de crudos livianos que en los pesados, debido a que hay menos sustituciones en los anillos aromáticos, por presencia de compuestos aromáticos con anillos tri y mono sustituidos.

Las dos bandas a 699 cm^{-1} y 674 cm^{-1} , asociadas a la flexión del anillo aromático o bencenos o xilenos, aparecen con alta frecuencia de inclusión sólo para la fracción queroseno. Lo anterior es lógico pues estas bandas se asocian a hidrocarburos aromáticos livianos.

Bandas como la del grupo sulfóxido a 1032 cm^{-1} aparecen con una baja frecuencia de inclusión. Es probable que estas bandas aumenten su frecuencia de inclusión en la selección de variables si se corren los algoritmos genéticos para propiedades asociadas con compuestos azufrados.

3.5 MODELOS AG-PLS RENDIMIENTOS

Se siguió la misma metodología empleada en la construcción de los modelos PLS en Unscrambler. Con base en los resultados anteriores, se desarrollaron cuatro modelos para predicción de rendimientos de las fracciones queroseno, acpm liviano, acpm pesado y crudo reducido. Los bloques de datos X correspondientes a las variables seleccionadas se usaron para construir los modelos AG-PLS usando el software Toolbox.

3.5.1 Modelo AG-PLS rendimientos queroseno

En la figura 57 se observa el gráfico de Scores, el cual propone como outliers las muestras de crudo No. 39 y 11, correspondientes a un crudo superligero y uno extrapesado.

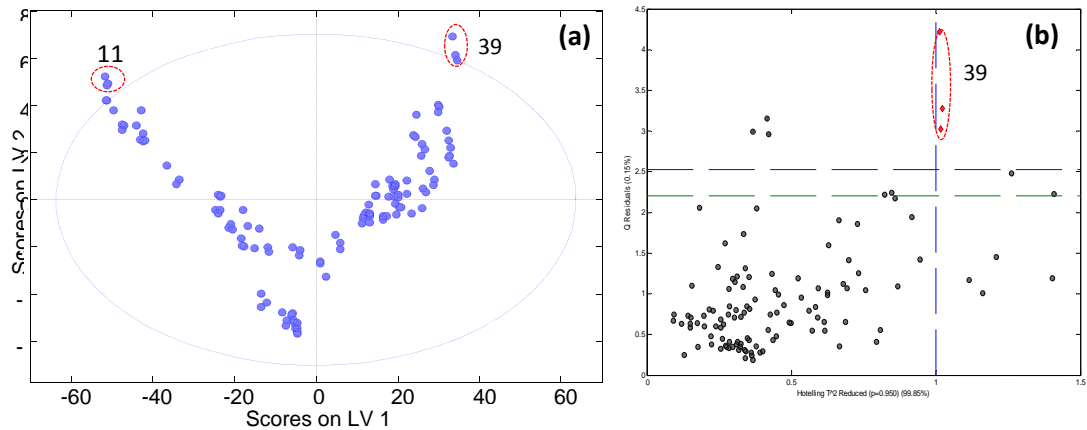


Figura 57. (a). Análisis scores y (b). Varianzas residuales, modelo AG-PLS rendimientos (%V) fracción queroseno.

Al evaluar el gráfico de influencia de los residuales espectrales frente al leverage de las muestras (figura 57 b), se observa que la muestra de crudo No.39 aparece con alto residual y alto leverage, mientras la muestra No. 11 presenta alto leverage. Por lo anterior, la muestra No.39 es considerada como outlier y se descartó del grupo de calibración, reprocesando el modelo AG-PLS.

Las características espectrales de la muestra de crudo No. 39 fueron explicadas en el numeral 3.3.1 de los modelos PLS desarrollados en Unscrambler, en los que también se consideró como outlier. La muestra No. 11 corresponde a un crudo extrapesado, con API de 8.2, el cual no reporta acidez como característica particular, dado que los demás crudos extrapesados y pesados del set de muestras reportan valores de acidez superiores a 0.093 mg KOH/g. Esta muestra no fue descartada del grupo de calibración porque se considera una muestra influyente pero no outlier. En la calibración del modelo, se obtuvo un coeficiente de correlación de 0.990 y error de calibración de 0.249, en un intervalo entre 1 y 11% de rendimientos para esta fracción (figura 58).

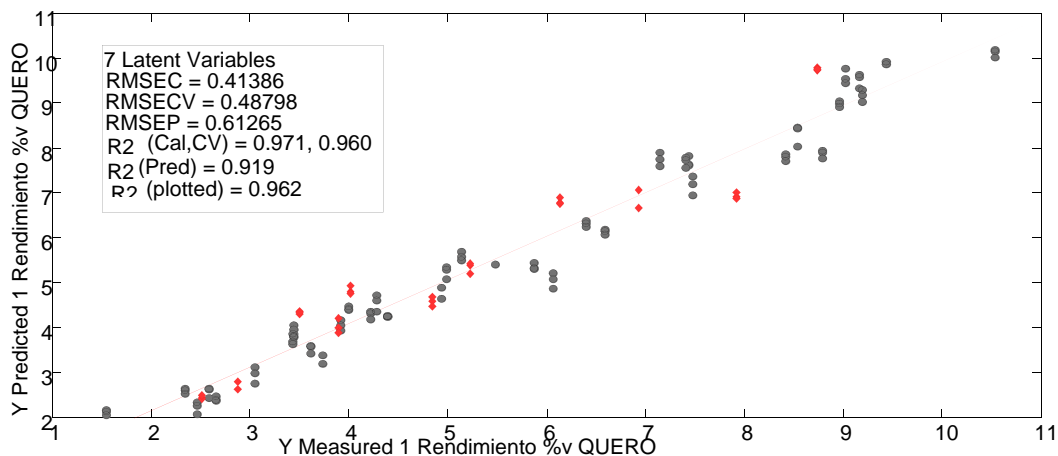


Figura 58. Calibración/validación modelo AG-PLS rendimientos (%V) fracción queroseno.

El gráfico de Scores de la figura 59, muestra las relaciones entre las muestras de calibración y de validación cruzada. En la validación del modelo, se obtuvo un coeficiente de correlación de 0.985 y error de validación de 0.301, en un intervalo entre 1 y 11% de rendimientos.

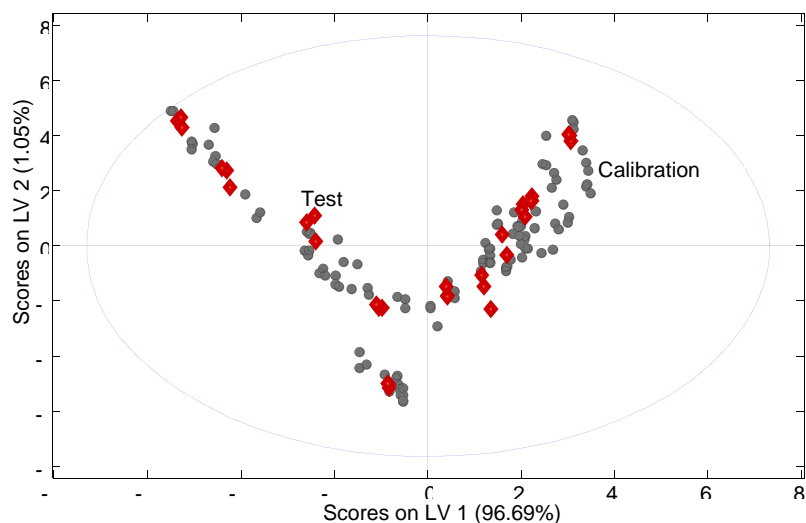


Figura 59. Análisis Scores modelo AG-PLS rendimientos (%V) fracción queroseno.

Como se observa en la figura 60, con cinco variables latentes se explica más del 90% de la varianza de calibración y validación, en los datos de rendimientos en volumen (datos Y) para la fracción queroseno, demostrando que los datos son representativos.

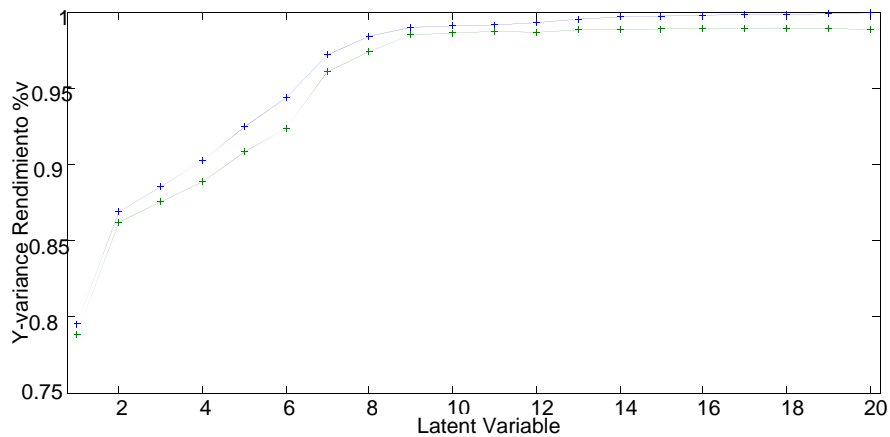


Figura 60. Varianza explicada modelo AG-PLS de rendimientos (%V) fracción queroseno.

Para seleccionar el número de componentes óptimo en cada modelo AG-PLS, se analizó el error estándar de calibración (RMSEC) y el error estándar de validación cruzada (RMSECV) frente al número de variables latentes, de manera similar al análisis realizado para los modelos PLS. La figura 61 muestra el comportamiento del error de calibración (RMSEC), el cual disminuye paralelamente al error de validación (RMSECV) hasta la variable latente once.

Se seleccionó como el número de variables latentes óptimo la variable siete, con valores de RMSEC de 0.414% y 0.488% para RMSECV, en un intervalo de valores de rendimientos de 1 a 11%.

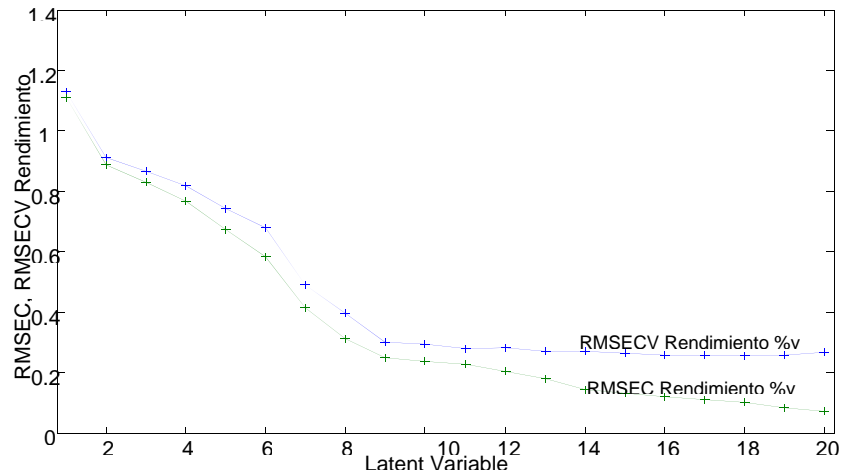


Figura 61. Error versus número de variables latentes modelo AG-PLS rendimientos (%V) fracción queroseno.

Validación externa del modelo. En la figura 62 se presentan los resultados de la regresión obtenida para la validación externa del modelo, realizada con 11 muestras. Se obtuvo un coeficiente de correlación de 0.919 y error de predicción de 0.613.

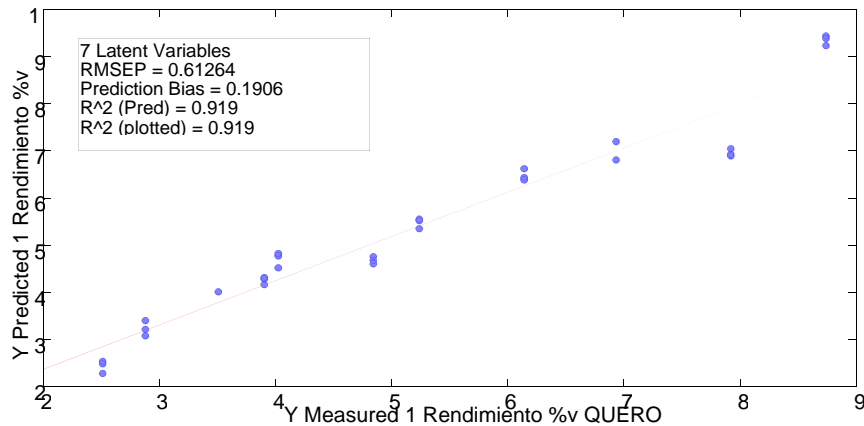


Figura 62. Validación externa modelo AG-PLS rendimientos (%V) fracción queroseno.

En la tabla 33 se muestran los resultados de la validación externa para el modelo desarrollado. Los errores relativos obtenidos son iguales o inferiores al 17%, lo que refleja un desempeño aceptable del modelo en la predicción de nuevas muestras.

Tabla 33. Validación externa modelo AG-PLS rendimientos (%V) fracción queroseno.

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|-----------|------------------|----------------|------------|--------------------|
| 8 | 6.14 | 6.48 | 0.34 | 5.55 |
| 10 | 6.93 | 7.00 | 0.07 | 1.04 |
| 13 | 2.51 | 2.44 | 0.08 | 3.13 |
| 14 | 3.90 | 4.25 | 0.35 | 9.10 |
| 15 | 7.92 | 6.95 | 0.97 | 12.2 |
| 23 | 2.88 | 3.23 | 0.35 | 12.2 |
| 28 | 3.51 | 4.01 | 0.50 | 14.2 |
| 38 | 4.84 | 4.68 | 0.17 | 3.42 |
| 42 | 8.73 | 9.35 | 0.62 | 7.08 |
| 46 | 4.02 | 4.71 | 0.69 | 17.1 |
| 53 | 5.23 | 5.47 | 0.24 | 4.53 |

La figura 63 muestra la distribución de residuales en las muestras de calibración y predicción. Las muestras usadas en la calibración se presentan en negro y en rojo las de predicción externa.

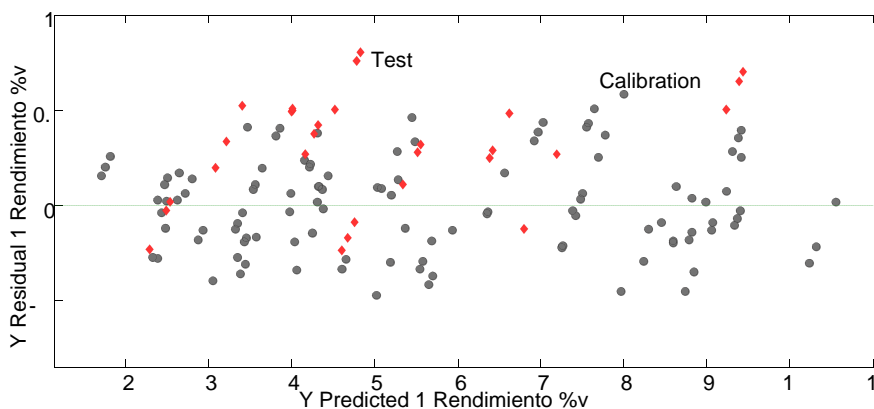


Figura 63. Residuales de calibración y predicción modelo AG-PLS rendimientos (%V) fracción queroseno.

El modelo presenta buen desempeño al mostrar un comportamiento aleatorio para todas las muestras y un valor de Y-residual máximo de 0.75%. El gráfico de residuales demuestra que el modelo predice adecuadamente las variaciones en los rendimientos, dado que los residuales se distribuyen aleatoriamente.

3.5.2 Modelo AG-PLS rendimientos acpm liviano

Los coeficientes de correlación del modelo (figura 64) para calibración, validación y predicción fueron 0.909, 0.888 y 0.909, respectivamente. Con seis variables latentes se explica más del 90% de la varianza de calibración y validación en los datos de rendimientos para esta fracción.

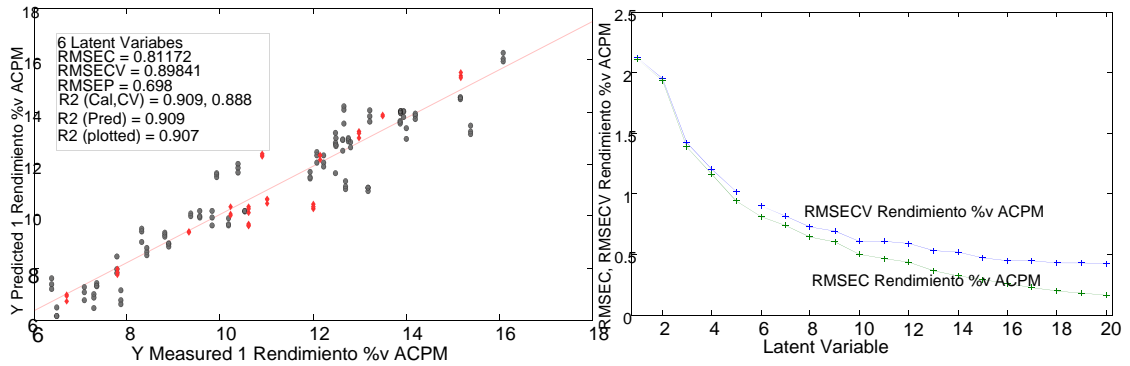


Figura 64. Parámetros estadísticos modelo AG-PLS rendimientos acpm liviano.

La validación externa del modelo AG-PLS muestra errores relativos máximos del 13.3% (tabla 34).

Tabla 34. Validación externa modelo AG-PLS rendimientos (%V) fracción acpm liviano

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|-----------|------------------|----------------|------------|--------------------|
| 5 | 12.1 | 12.3 | 0.14 | 1.12 |
| 8 | 10.6 | 10.2 | 0.36 | 3.41 |
| 10 | 13.0 | 13.1 | 0.15 | 1.17 |
| 13 | 7.80 | 7.85 | 0.05 | 0.69 |
| 14 | 10.2 | 10.2 | 0.08 | 0.78 |
| 15 | 11.0 | 10.60 | 0.41 | 3.75 |
| 23 | 6.70 | 6.88 | 0.18 | 2.63 |
| 26 | 15.2 | 15.4 | 0.23 | 1.51 |
| 28 | 13.5 | 13.8 | 0.35 | 2.58 |
| 38 | 10.9 | 12.4 | 1.45 | 13.3 |
| 46 | 9.34 | 9.41 | 0.08 | 0.83 |
| 52 | 10.6 | 9.6 | 0.99 | 9.28 |
| 53 | 12.0 | 10.5 | 1.54 | 12.8 |

3.5.3 Modelo AG-PLS rendimientos acpm pesado

Los coeficientes de correlación del modelo (figura 65) para calibración, validación y predicción fueron 0.902, 0.877 y 0.870, respectivamente. Con siete variables latentes se explica más del 90% de la varianza de calibración y 87% de la varianza de validación. Se seleccionaron siete variables latentes para este modelo.

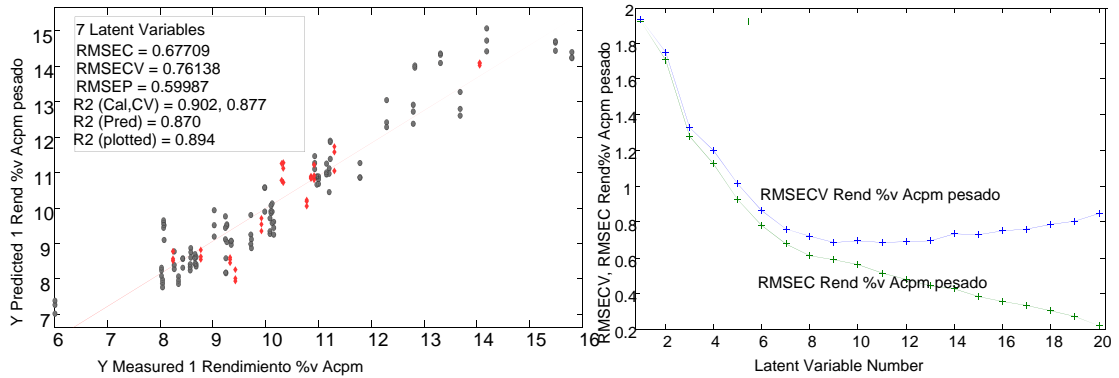


Figura 65. Parámetros estadísticos modelo AG-PLS rendimientos acpm pesado.

La validación externa arrojó errores relativos máximos del 14.2% (tabla 35). Lo anterior muestra que el modelo tiene un desempeño aceptable.

Tabla 35. Validación externa modelo AG-PLS rendimientos (%V) fracción acpm pesado

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|-----------|------------------|----------------|------------|--------------------|
| 5 | 14.1 | 12.3 | 1.84 | 13.0 |
| 8 | 9.42 | 8.08 | 1.34 | 14.2 |
| 10 | 11.3 | 11.5 | 0.17 | 1.51 |
| 13 | 9.32 | 8.54 | 0.78 | 8.36 |
| 14 | 9.91 | 9.54 | 0.37 | 3.75 |
| 15 | 8.77 | 8.67 | 0.10 | 1.14 |
| 23 | 8.24 | 8.62 | 0.38 | 4.65 |
| 25 | 10.9 | 11.0 | 0.06 | 0.58 |
| 26 | 10.3 | 11.0 | 0.71 | 6.84 |
| 28 | 14.0 | 14.1 | 0.02 | 0.13 |
| 38 | 10.9 | 10.9 | 0.01 | 0.12 |
| 52 | 10.8 | 10.2 | 0.62 | 5.73 |
| 53 | 10.3 | 10.9 | 0.65 | 6.27 |

3.5.4 Modelo AG-PLS rendimientos crudo reducido

Para este modelo se obtuvieron coeficientes de correlación (figura 66) para calibración, validación y predicción de 0.995, 0.980 y 0.964, respectivamente. Con siete variables latentes se explica más del 95% de las varianzas de calibración y de validación.

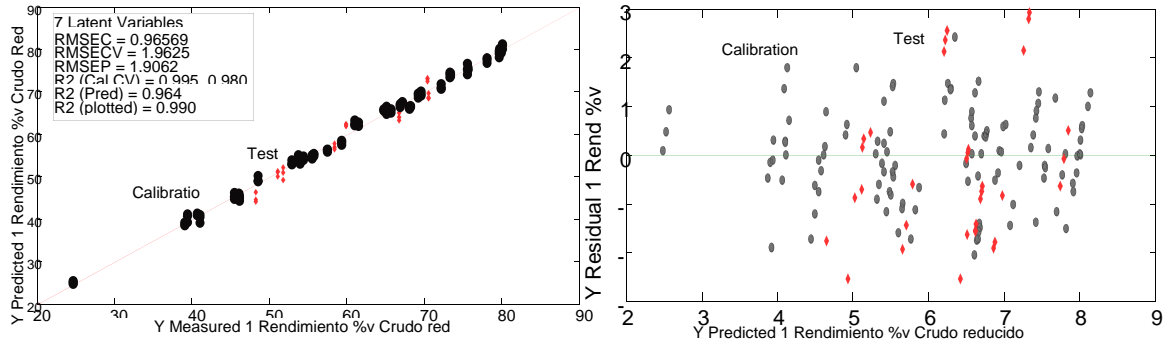


Figura 66. Parámetros estadísticos modelo AG-PLS rendimientos crudo reducido.

La validación externa del modelo (tabla 35) muestra errores relativos máximos del 11.4%.

Tabla 36. Validación externa modelo AG-PLS rendimientos (%V) fracción crudo reducido

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|-----------|------------------|----------------|------------|--------------------|
| 5 | 66.72 | 63.85 | 2.88 | 4.31 |
| 8 | 58.45 | 57.31 | 1.15 | 1.96 |
| 10 | 51.81 | 52.02 | 0.21 | 0.40 |
| 13 | 78.02 | 77.76 | 0.26 | 0.33 |
| 14 | 67.71 | 66.34 | 1.37 | 2.03 |
| 15 | 51.05 | 50.79 | 0.27 | 0.52 |
| 23 | 78.70 | 75.19 | 3.51 | 4.46 |
| 25 | 48.26 | 44.85 | 3.41 | 7.07 |
| 28 | 67.74 | 66.79 | 0.96 | 1.41 |
| 38 | 68.08 | 62.04 | 6.04 | 8.9 |
| 42 | 43.74 | 38.75 | 4.99 | 11.4 |
| 46 | 70.54 | 68.66 | 1.88 | 2.67 |
| 51 | 65.12 | 65.25 | 0.13 | 0.20 |
| 52 | 70.43 | 72.68 | 2.25 | 3.19 |
| 53 | 59.87 | 62.35 | 2.48 | 4.13 |

El resumen de los parámetros estadísticos de los modelos AG-PLS para rendimientos de las fracciones queroseno, acpm liviano, acpm pesado y crudo reducido se presenta en la tabla 37.

Tabla 37. Resultados modelos de calibración y validación AG-PLS rendimientos (%V) destilados atmosféricos.

| Modelo | Media Calibración | R ² Calibración | RMSEC | R ² Validación | RMSECV | VL | Media Val. externa | R ² Predicción | RMSEP |
|------------|-------------------|----------------------------|-------|---------------------------|--------|----|--------------------|---------------------------|-------|
| Queroseno | 5.69 | 0.971 | 0.414 | 0.960 | 0.488 | 7 | 5.66 | 0.919 | 0.613 |
| Acpm liv. | 11.18 | 0.909 | 0.812 | 0.888 | 0.898 | 6 | 10.66 | 0.909 | 0.698 |
| Acpm pes. | 10.36 | 0.902 | 0.677 | 0.877 | 0.761 | 7 | 10.62 | 0.870 | 0.600 |
| Crudo red. | 60.77 | 0.995 | 0.966 | 0.980 | 1.962 | 7 | 61.39 | 0.964 | 1.906 |

El modelo AG-PLS de rendimientos para la fracción crudo reducido arrojó los mejores parámetros estadísticos de calibración, validación y predicción externa, similar a lo ocurrido en los modelos PLS, en los cuales, este modelo también obtuvo los mejores resultados.

Como característica particular, los modelos PLS y AG-PLS de rendimientos para la fracción acpm pesado, fueron los de más bajo desempeño. Las variables importantes identificadas, tanto en los modelos de rendimientos PLS como en los de algoritmos genéticos, son muy similares. En los dos casos, los modelos de rendimientos de acpm pesado arrojaron un desempeño aceptable. Sin embargo, en PLS se necesitaron cuatro componentes para explicar mínimo el 70% de la varianza en Y. Adicionalmente, no fueron seleccionadas como variables importantes las bandas del grupo metileno a 2920 cm⁻¹ ni a 2852 cm⁻¹. Lo anterior concuerda con la selección de variables en AG, pues este fue el único modelo en el que el algoritmo no seleccionó regiones cerca de los picos espectrales correspondientes a las bandas vibracionales de los grupos metílico y metileno alifáticos entre 2952cm⁻¹ y 2852 cm⁻¹, consideradas como variables no útiles.

3.6 MODELOS PLS PROPIEDADES

A continuación se presentan los resultados obtenidos para los modelos de API, azufre e índice de cetano para las fracciones de destilados atmosféricos queroseno, acpm liviano, acpm pesado y crudo reducido.

3.6.1 Modelo PLS API queroseno

En la figura 67 se presentan los principales gráficos para el modelo PLS de API desarrollado para la fracción queroseno. El gráfico de Scores (figura 67 a) muestra las relaciones entre las muestras de calibración y de validación. Al igual que para el modelo PLS de rendimientos en la fracción queroseno, se descartó la muestra de crudo No. 39 por tratarse de un outlier. La figura 67 (c) explica cuánto de la variación en las respuestas es descrito por cada componente diferente. Con dos factores se explica más del 94% de la varianza de calibración y 93% de la varianza de validación en los datos de API (datos Y) para la fracción queroseno. Lo anterior muestra que el modelo de API tiene la capacidad para describir correctamente nuevos datos.

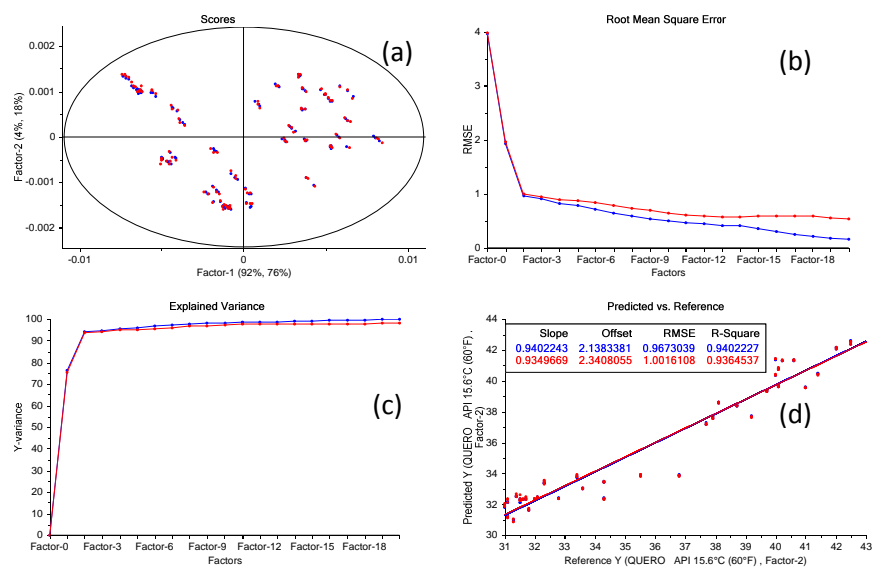


Figura 67. Principales gráficos del modelo PLS API fracción queroseno.

El comportamiento del error de calibración (RMSEC) y del error de validación (RMSECV) es mostrado en la figura 67 (b). Se observa que a partir del segundo componente no existe una disminución significativa del error. Se seleccionaron los dos primeros componentes como el número óptimo de variables latentes para este modelo, con errores de 0.97 para RMSEC y 1.00 para el RMSECV, en un intervalo de valores de API entre 31 y 43.

En la figura 67 (d) se presentan los resultados de las regresiones obtenidas para la calibración y validación del modelo. Para el modelo de calibración, se obtuvo como coeficiente de correlación 0.940 y error de calibración de 0.967, en un intervalo de API entre 31 y 41. Para la validación, se obtuvo un coeficiente de correlación del 0.936 y error de validación de 1.00, para el mismo intervalo de API.

La validación externa del modelo API se realizó con un set de 16 muestras no usadas para la calibración. En la figura 68 se presentan los resultados de la validación externa. La regresión obtenida muestra un coeficiente de correlación inferior a 0.8, y error de predicción de 1.40.

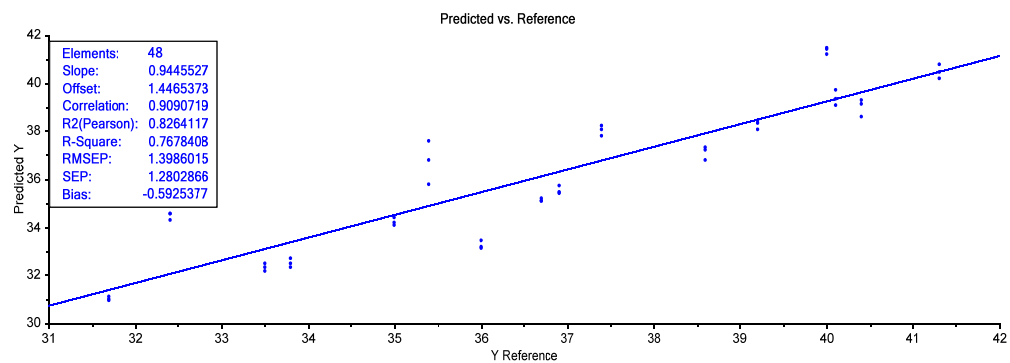


Figura 68. Validación externa modelo PLS API fracción queroseno.

Los resultados de las predicciones son mostrados en la tabla 38. Los errores relativos obtenidos son inferiores al 8%, reflejando un desempeño aceptable del modelo en la predicción de nuevas muestras.

Tabla 38. Validación externa del modelo PLS API fracción queroseno.

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|-----------|------------------|----------------|------------|--------------------|
| 5 | 32.4 | 34.5 | 2.1 | 6.39 |
| 8 | 41.3 | 40.5 | 0.8 | 1.95 |
| 10 | 37.4 | 38.0 | 0.6 | 1.70 |
| 13 | 33.5 | 32.3 | 1.2 | 3.48 |
| 14 | 35.0 | 34.2 | 0.8 | 2.22 |
| 15 | 38.6 | 37.1 | 1.5 | 3.82 |
| 23 | 33.8 | 32.5 | 1.3 | 3.81 |
| 25 | 35.4 | 36.7 | 1.3 | 3.73 |
| 26 | 40.0 | 41.4 | 1.4 | 3.43 |
| 28 | 31.7 | 31.0 | 0.7 | 2.16 |
| 38 | 36.7 | 35.1 | 1.6 | 4.25 |
| 42 | 40.4 | 39.0 | 1.4 | 3.46 |
| 46 | 40.1 | 39.4 | 0.7 | 1.79 |
| 51 | 39.2 | 38.3 | 0.9 | 2.33 |
| 52 | 36.0 | 33.3 | 2.7 | 7.58 |
| 53 | 36.9 | 35.5 | 1.4 | 3.72 |

3.6.2 Modelo PLS azufre queroseno

El modelo inicial desarrollado para esta propiedad no arrojó resultados satisfactorios. El 50% de los errores relativos obtenidos para las muestras de predicción externa fueron superiores al 20%. Los errores relativos de predicción fueron mayores en las concentraciones de azufre iguales o inferiores a 0.25%, para los cuales se obtuvieron errores en el intervalo del 18% al 61%. Debido a que los valores de referencia en la concentración de azufre variaron en un orden de magnitud (desde 0.015% a 0.726%), se desarrollaron modelos adicionales, agrupando las muestras en dos niveles de concentración: bajo azufre (0.015% – 0.207%) y alto azufre (0.207 – 0.726%).

Los gráficos de la figura 69 presentan los principales resultados obtenidos para los dos modelos PLS desarrollados. Las varianzas de calibración y de validación para cada modelo no difieren

sustancialmente una de la otra. Con cinco componentes se explica más del 90% de la varianza de calibración y de validación para el modelo de bajo azufre, y con siete componentes más del 90% de las varianzas para el modelo de alto azufre.

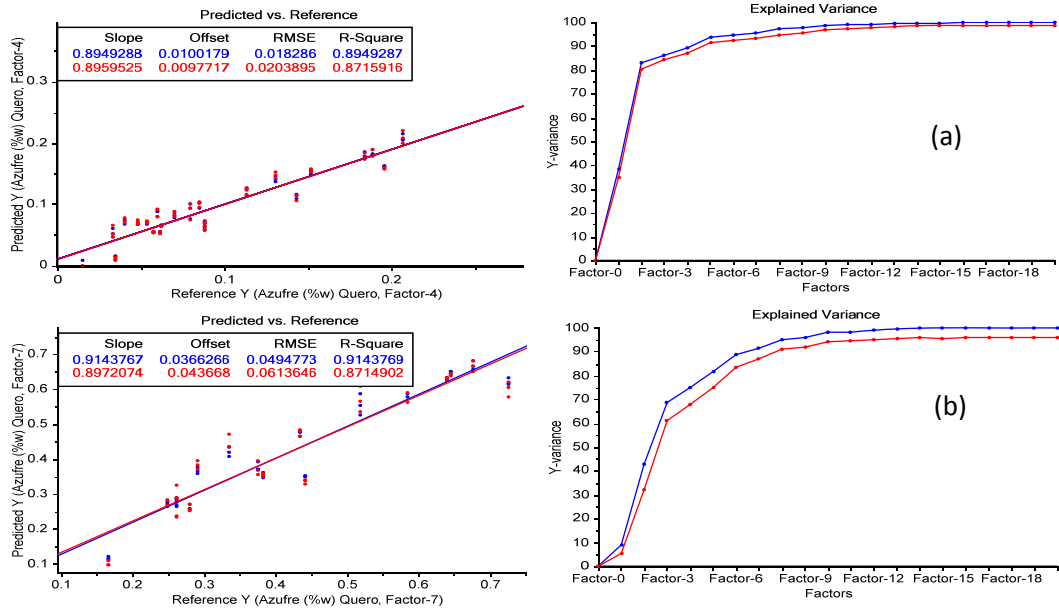


Figura 69. Principales gráficos modelos PLS fracción queroseno: (a) bajo azufre, (b) alto azufre.

Para los dos modelos fueron seleccionadas cuatro y siete variables latentes respectivamente, como el número óptimo de componentes principales. Las regresiones de calibración y validación de los dos modelos (figura 69, superior izquierda e inferior izquierda) muestran coeficientes entre 0.871 y 0.914.

La validación de los modelos PLS se realizó con dos sets de 6 y 7 muestras no usadas en la calibración. Fue eliminada la muestra No. 26 debido a que desmejoraba el modelo. Los errores de predicción RMSEP (figura 70) para bajo y alto azufre fueron de 0.010% y 0.031%, en los intervalos de azufre de 0.04 al 0.15% y 0.2 al 0.6%, respectivamente. En las tablas 39 y 40 se presentan los resultados de las predicciones realizadas en los dos sets de muestras externas.

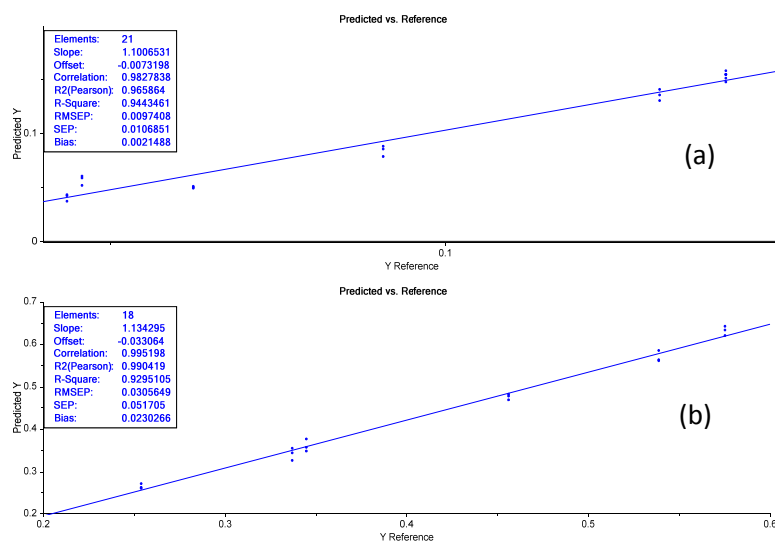


Figura 70. Validación externa modelos PLS fracción queroseno: (a) bajo azufre, (b) alto azufre.

Tabla 39. Validación externa del modelo PLS bajo azufre fracción queroseno

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|-----------|------------------|----------------|------------|--------------------|
| 5 | 0.1420 | 0.1556 | 0.0136 | 9.55 |
| 10 | 0.0908 | 0.0841 | 0.0067 | 7.42 |
| 15 | 0.1320 | 0.1353 | 0.0033 | 2.53 |
| 25 | 0.0458 | 0.0570 | 0.0112 | 24.5 |
| 28 | 0.0624 | 0.0501 | 0.0123 | 19.8 |
| 42 | 0.0435 | 0.0408 | 0.0027 | 6.28 |
| 53 | 0.1420 | 0.1507 | 0.0087 | 6.15 |

Tabla 40. Validación externa del modelo PLS alto azufre fracción queroseno

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|-----------|------------------|----------------|------------|--------------------|
| 8 | 0.345 | 0.360 | 0.015 | 4.35 |
| 13 | 0.539 | 0.570 | 0.031 | 5.83 |
| 23 | 0.575 | 0.633 | 0.058 | 10.1 |
| 46 | 0.337 | 0.341 | 0.004 | 1.18 |
| 51 | 0.456 | 0.476 | 0.020 | 4.42 |
| 52 | 0.254 | 0.264 | 0.010 | 3.85 |

Los errores relativos máximos obtenidos fueron 25.5% para bajo azufre y 10.1% para alto azufre, lo cual demuestra un desempeño aceptable para estos dos modelos.

3.6.3 Modelo PLS Índice de cetano queroseno

En este modelo (figura 71 a) se observa que los dos primeros componentes explican el 96% de la varianza en X y el 91% en Y. El gráfico de Scores no propone muestra alguna como outlier, dado que ya se había descartado la muestra No.39.

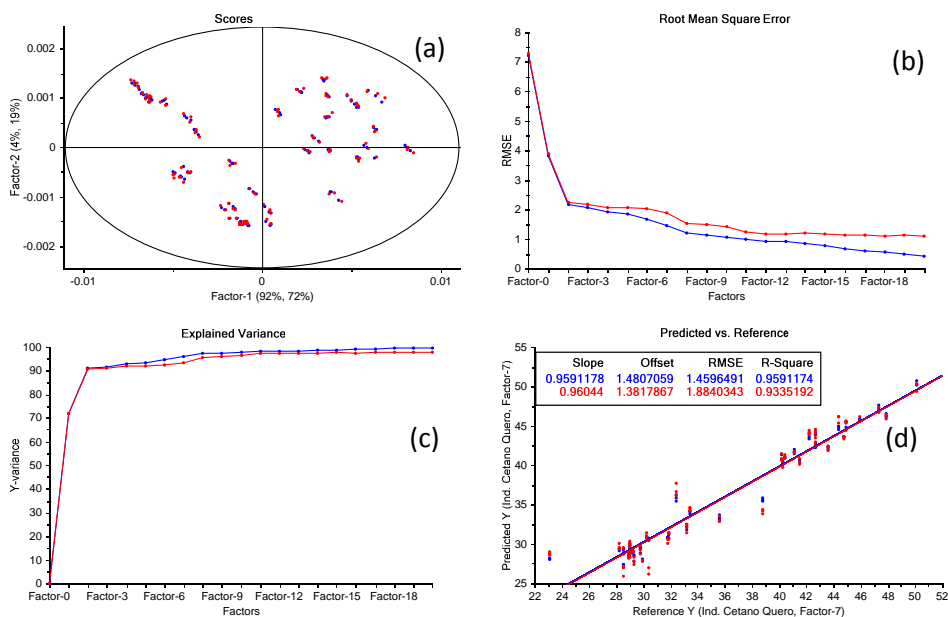


Figura 71. Principales resultados modelo PLS índice de cetano fracción queroseno.

Con dos factores se explica más del 90% de la varianza de calibración y de validación del modelo (figura 71 c). Las dos varianzas son muy similares, mostrando que el modelo de índice de cetano tiene la capacidad para describir correctamente nuevos datos. En la figura 71 (b), se observa el comportamiento de los errores RMSEC y RMSECV. Se seleccionaron siete variables como el número óptimo de variables latentes, con errores de 1.46 para RMSEC y 1.88 para RMSECV, en un intervalo de valores de índice de cetano entre 22 y 52. Los coeficientes de correlación

encontrados para las regresiones de calibración y validación fueron 0.959 y 0.933, con errores de calibración de 1.46 y validación de 1.88.

La validación externa del modelo se realizó con un set de 16 muestras no usadas para la calibración. La regresión para la validación externa se presenta en la figura 72, con coeficiente de correlación de 0.857 y error de predicción de 1.949.

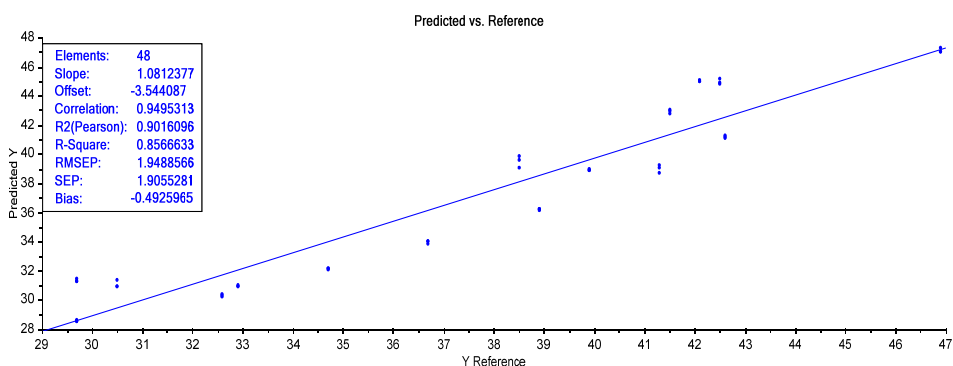


Figura 72. Validación externa modelo PLS índice de cetano fracción queroseno.

En la tabla 41 se presentan los resultados obtenidos en la predicción. El error máximo reportado fue del 10%, lo cual muestra un buen desempeño del modelo en la predicción de nuevas muestras.

El resumen de los parámetros estadísticos obtenidos para cada uno de los modelos PLS de propiedades, desarrollados para la fracción queroseno es presentado en la tabla 42. Los modelos de API e índice de cetano arrojaron resultados aceptables, de acuerdo con los resultados de calibración, validación y predicción. Los máximos errores relativos de predicción para muestras externas estuvieron entre 7.58% y 10.1%, es decir, fueron inferiores al 15%.

Tabla 41. Validación externa modelo PLS índice de cetano fracción queroseno

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|-----------|------------------|----------------|------------|--------------------|
| 5 | 29.7 | 31.3 | 1.6 | 5.55 |
| 8 | 46.9 | 47.1 | 0.2 | 0.44 |
| 10 | 38.5 | 39.5 | 1.0 | 2.64 |
| 13 | 30.5 | 31.1 | 0.6 | 1.81 |
| 14 | 34.7 | 32.1 | 2.6 | 7.42 |
| 15 | 39.9 | 38.9 | 1.0 | 2.50 |
| 23 | 32.6 | 30.3 | 2.3 | 6.98 |
| 25 | 41.3 | 39.0 | 2.3 | 5.54 |
| 26 | 42.1 | 45.0 | 2.9 | 6.88 |
| 28 | 29.7 | 28.6 | 1.1 | 3.85 |
| 38 | 36.7 | 34.0 | 2.7 | 7.44 |
| 42 | 42.6 | 41.2 | 1.4 | 3.36 |
| 46 | 42.5 | 44.9 | 2.4 | 5.74 |
| 51 | 41.5 | 43.0 | 1.5 | 3.51 |
| 52 | 32.9 | 30.9 | 2.0 | 5.97 |
| 52 | 32.9 | 36.2 | 3.3 | 10.1 |

Tabla 42. Resultados modelos de calibración y validación PLS para fracción queroseno (199-249°C)

| Modelo | Media Calibración | R ² Calibración | RMSEC | R ² Validación | RMSECV | VL | Media Val. externa | R ² Predicción | RMSEP |
|--------------------------|-------------------|----------------------------|-------|---------------------------|--------|----|--------------------|---------------------------|-------|
| API | 35.8 | 0.940 | 0.967 | 0.936 | 1.002 | 2 | 36.8 | 0.768 | 1.399 |
| Azufre ¹ (%p) | 0.095 | 0.895 | 0.018 | 0.872 | 0.020 | 4 | 0.087 | 0.944 | 0.010 |
| Azufre ² (%p) | 0.413 | 0.914 | 0.049 | 0.871 | 0.061 | 7 | 0.418 | 0.930 | 0.031 |
| Ind.Cetano | 36.4 | 0.959 | 1.460 | 0.934 | 1.884 | 7 | 37.6 | 0.857 | 1.949 |

Azufre^{1,2}: baja y alta concentración

Aunque el modelo de API arrojó un coeficiente de correlación para predicción inferior a 0.8 y errores equivalentes de calibración, validación y predicción de 10% al 14%, obtuvo el menor error relativo de predicción (7.58%) en muestras externas, con respecto a todos los modelos. Los modelos de azufre mejoraron sustancialmente al dividir las muestras en dos intervalos de calibración. Los errores de predicción arrojaron resultados satisfactorios.

3.6.4 Modelos PLS API, azufre e índice de cetano acpm liviano

Los resultados de las correlaciones (valores de referencia versus valores predichos) para los modelos PLS de API e índice de cetano para esta fracción se presentan en la figura 73.

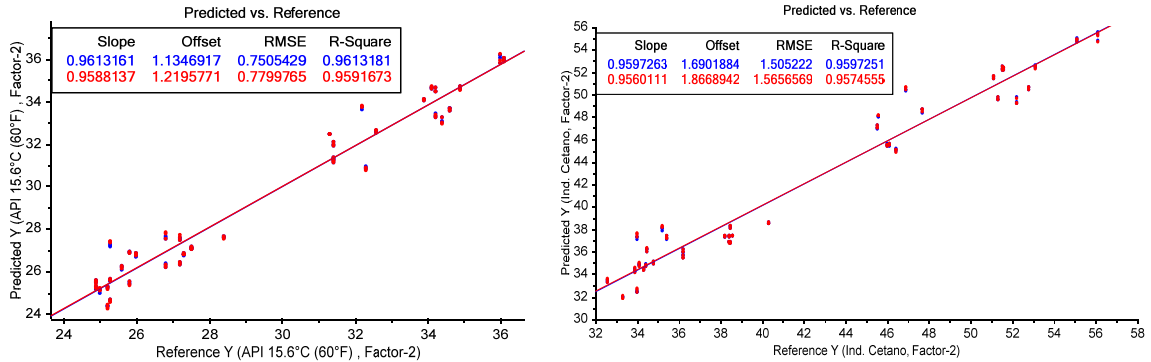


Figura 73. Calibración y validación cruzada modelos PLS de API e índice de cetano fracción acpm liviano.

Se obtuvieron correlaciones de calibración y validación cruzada similares para los dos modelos (R^2 de 0.96). De acuerdo con la figura 74, el 96% de la variación en las respuestas de calibración y validación (datos Y) se explica con dos componentes para los modelos de API e índice de cetano.

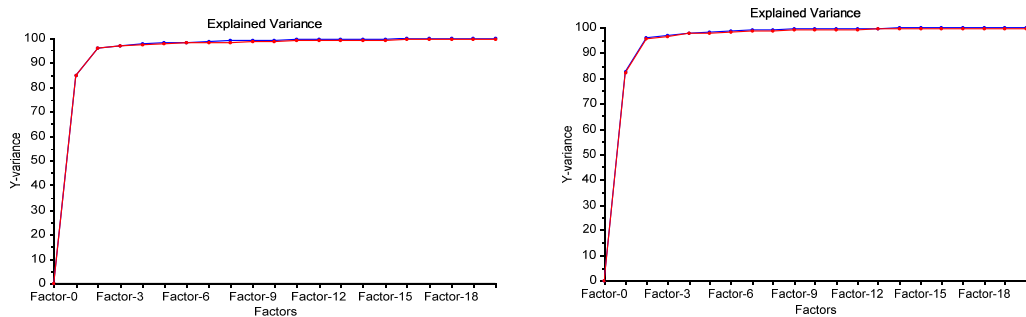


Figura 74. Varianza explicada de calibración y validación modelos PLS acpm liviano: (a) API, (b) índice de cetano.

Las correlaciones obtenidas para la validación externa de los modelos son mostradas en la figura 75. Los coeficientes de correlación variaron entre 0.90 y 0.94, con errores de predicción (RMSEP)

de 0.70 y 1.87 para API e índice de cetano, respectivamente. Los parámetros anteriores indican que los modelos tienen un desempeño aceptable para la predicción de nuevas muestras.

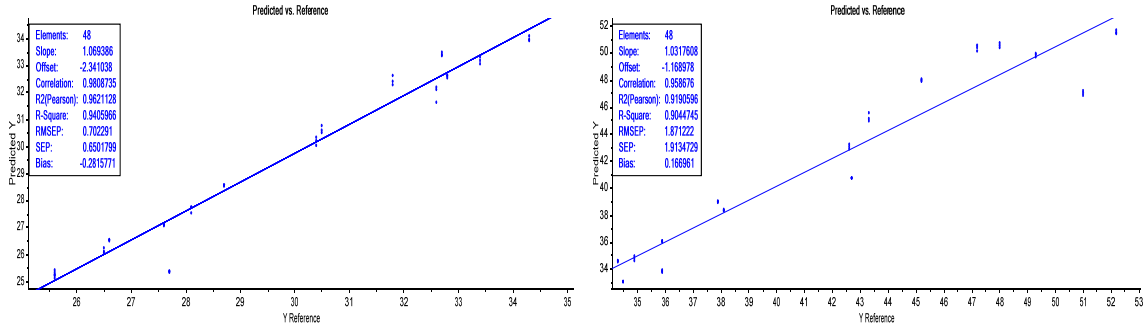


Figura 75. Validación externa modelos PLS de API e índice de cetano fracción acpm liviano.

La validación externa de los modelos PLS de API e índice de cetano se realizó con sets de 16 muestras. En las tablas 43 y 44 se presentan los resultados obtenidos.

Tabla 43. Validación externa modelo PLS API fracción acpm liviano

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|-----------|------------------|----------------|------------|--------------------|
| 5 | 26.6 | 26.5 | 0.1 | 0.25 |
| 8 | 34.3 | 34.0 | 0.3 | 0.90 |
| 10 | 30.5 | 30.6 | 0.1 | 0.42 |
| 13 | 25.6 | 25.3 | 0.3 | 1.16 |
| 14 | 26.5 | 26.1 | 0.4 | 1.34 |
| 15 | 30.4 | 30.2 | 0.2 | 0.67 |
| 23 | 25.6 | 25.3 | 0.3 | 1.36 |
| 25 | 32.6 | 32.0 | 0.6 | 1.89 |
| 26 | 32.8 | 32.6 | 0.2 | 0.62 |
| 28 | 27.6 | 27.1 | 0.5 | 1.83 |
| 38 | 28.1 | 27.7 | 0.4 | 1.45 |
| 42 | 33.4 | 33.2 | 0.2 | 0.63 |
| 46 | 32.7 | 33.4 | 0.7 | 2.19 |
| 51 | 31.8 | 32.4 | 0.6 | 1.99 |
| 52 | 27.7 | 25.4 | 2.3 | 8.42 |
| 53 | 28.7 | 28.6 | 0.1 | 0.45 |

Tabla 44. Validación externa modelo PLS índice de cetano fracción acpm liviano

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|-----------|------------------|----------------|------------|--------------------|
| 5 | 35.9 | 33.8 | 2.1 | 5.73 |
| 8 | 52.2 | 51.5 | 0.7 | 1.31 |
| 10 | 43.3 | 45.2 | 1.9 | 4.41 |
| 13 | 34.3 | 34.6 | 0.3 | 0.75 |
| 14 | 35.9 | 36.0 | 0.1 | 0.40 |
| 15 | 42.6 | 43.1 | 0.5 | 1.07 |
| 23 | 34.9 | 34.7 | 0.2 | 0.44 |
| 25 | 51.0 | 47.0 | 4.0 | 7.85 |
| 26 | 47.2 | 50.3 | 3.1 | 6.62 |
| 28 | 38.1 | 38.3 | 0.2 | 0.55 |
| 38 | 37.9 | 39.0 | 1.1 | 2.80 |
| 42 | 48.0 | 50.5 | 2.5 | 5.31 |
| 46 | 49.3 | 49.8 | 0.5 | 1.05 |
| 51 | 45.2 | 47.9 | 2.7 | 6.06 |
| 52 | 34.5 | 33.1 | 1.4 | 4.14 |
| 53 | 42.7 | 40.7 | 2.0 | 4.61 |

Los errores relativos para las predicciones de API e índice de cetano fueron inferiores al 9%, lo cual refleja un buen desempeño de estos dos modelos.

Para la propiedad azufre, el modelo inicial desarrollado no fue satisfactorio. El 30% de los errores relativos obtenidos para las muestras externas de predicción fueron superiores al 20%, en un caso hasta del 67%. De manera similar a lo sucedido con el modelo de azufre para la fracción queroseno, los valores de concentración de azufre en esta fracción variaron en dos órdenes de magnitud (desde 0.082% a 1.550%), por lo anterior, se desarrollaron modelos adicionales, agrupando las muestras en dos niveles de concentración: bajo azufre (0.082% – 0.389%) y alto azufre (0.389 – 1.550%).

Los gráficos de la figura 76 presentan los principales resultados obtenidos para los dos modelos PLS desarrollados. Con cinco componentes se explica más del 90% de la varianza de calibración y de validación para el modelo de bajo azufre, y con seis componentes más del 90% de las varianzas para el de alto azufre.

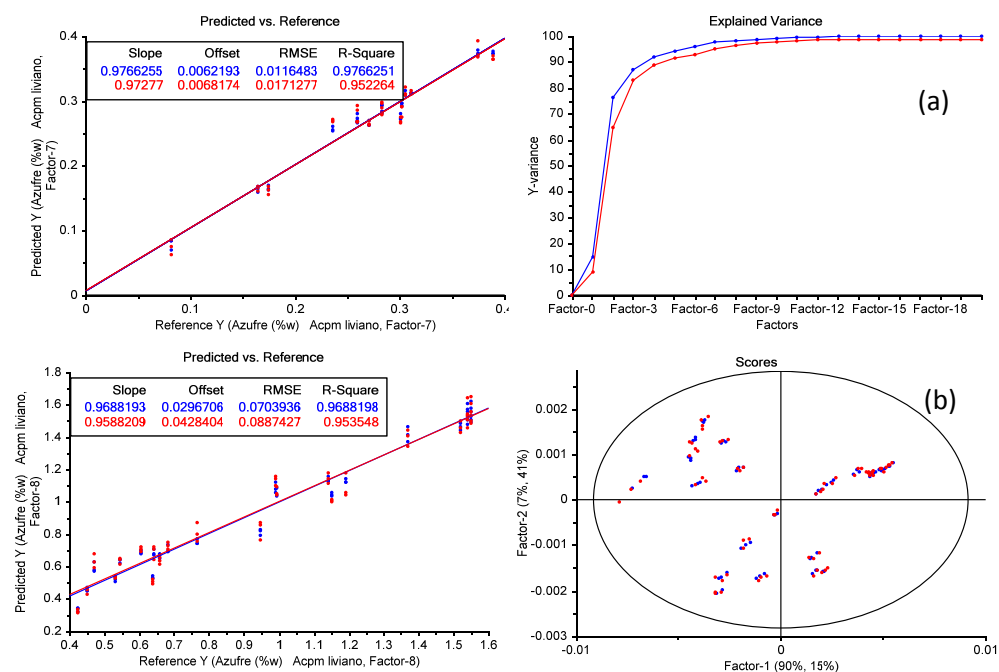


Figura 76. Principales gráficos modelos PLS fracción acpm liviano: (a) bajo azufre, (b) alto azufre.

Para el modelo de bajo azufre fueron seleccionadas siete variables latentes y ocho para el de alto azufre, como el número óptimo de componentes principales. Las regresiones de calibración y validación de los dos modelos (figura 76 superior izquierda e inferior izquierda) muestran coeficientes entre 0.964 y 0.983. Los errores de calibración (RMSEC) y validación (RMSECV) fueron de 0.0099% y 0.0150% para el modelo de bajo azufre, y de 0.056 y 0.075 para el de alto azufre.

La validación de los modelos PLS se realizó con dos sets de 4 y 8 muestras, no usadas en la calibración. Los errores de predicción RMSEP para bajo y alto azufre fueron de 0.017% y 0.081%, respectivamente (figura 74), en los intervalos de azufre de 0.2 a 0.4% y 0.4 al 1.5%.

En las tablas 45 y 46 se presentan los resultados de las predicciones realizadas en los dos sets de muestras externas.

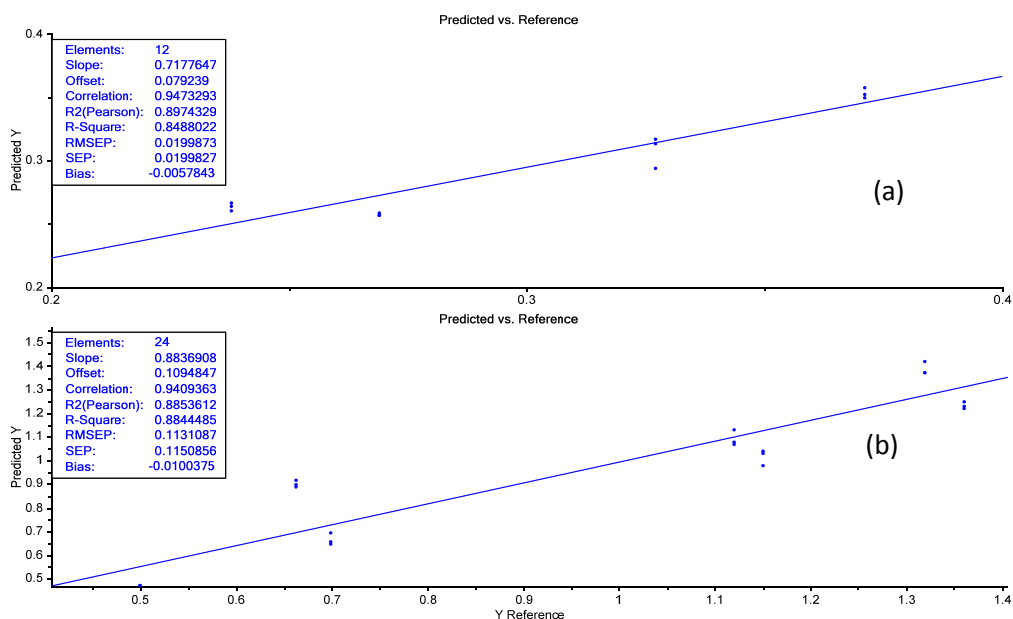


Figura 77. Validación externa modelos PLS fracción acpm liviano: (a) bajo azufre, (b) alto azufre.

Tabla 45. Validación externa del modelo PLS bajo azufre fracción acpm liviano

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|-----------|------------------|----------------|------------|--------------------|
| 5 | 0.371 | 0.353 | 0.018 | 4.83 |
| 10 | 0.327 | 0.308 | 0.019 | 5.81 |
| 28 | 0.269 | 0.257 | 0.012 | 4.41 |
| 42 | 0.238 | 0.264 | 0.026 | 10.8 |

Tabla 46. Validación externa del modelo PLS alto azufre fracción acpm liviano

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|-----------|------------------|----------------|------------|--------------------|
| 8 | 1.120 | 1.091 | 0.029 | 2.61 |
| 13 | 1.320 | 1.386 | 0.066 | 4.97 |
| 14 | 0.699 | 0.665 | 0.034 | 4.87 |
| 23 | 1.410 | 1.393 | 0.017 | 1.22 |
| 46 | 1.150 | 1.016 | 0.134 | 11.7 |
| 51 | 1.360 | 1.232 | 0.128 | 9.40 |
| 53 | 0.500 | 0.459 | 0.041 | 8.19 |

Los errores relativos máximos obtenidos fueron 10.8% para bajo azufre y 11.7% para alto azufre, lo cual demuestra un buen desempeño para estos dos modelos.

La tabla 47 presenta el resumen de los parámetros estadísticos obtenidos en cada uno de los modelos PLS de propiedades desarrollados para la fracción acpm liviano.

Tabla 47. Resultados modelos de calibración y validación PLS fracción acpm liviano (249-315°)

| Modelo | Media Calibración | R ² Calibración | RMSEC | R ² Validación | RMSECV | VL | Media Val. externa | R ² Predicción | RMSEP |
|--------------------------|-------------------|----------------------------|-------|---------------------------|--------|----|--------------------|---------------------------|-------|
| API | 29.4 | 0.961 | 0.751 | 0.959 | 0.780 | 2 | 29.7 | 0.941 | 0.702 |
| Azufre ¹ (%p) | 0.266 | 0.977 | 0.012 | 0.952 | 0.017 | 7 | 0.247 | 0.849 | 0.020 |
| Azufre ² (%p) | 0.952 | 0.969 | 0.070 | 0.954 | 0.089 | 8 | 0.948 | 0.884 | 0.113 |
| Ind. Cetano | 42.1 | 0.960 | 1.505 | 0.957 | 1.566 | 2 | 42.1 | 0.904 | 1.871 |

Azufre^{1,2}: baja y alta concentración

Los modelos de API, azufre e índice de cetano arrojaron parámetros estadísticos aceptables, de acuerdo con los resultados de calibración, validación y predicción obtenidos. Los máximos errores relativos de predicción para muestras externas estuvieron entre 7.34% y 12.3%, es decir, por debajo del 15%.

3.6.5 Modelos PLS API, azufre e índice de cetano acpm pesado

El desempeño de los modelos PLS de API, azufre e índice de cetano para la fracción acpm pesado se reflejan en los gráficos de valores predichos versus valores de referencia presentados en la figura 78.

Para los tres modelos de calibración y validación cruzada se obtuvieron coeficientes de regresión entre 0.954 y 0.982 (tabla 51). De acuerdo con la figura 79 (izquierda y derecha), para los modelos de API e índice de cetano, el 95% de la variación en las respuestas de calibración y validación se explica con 2 componentes.

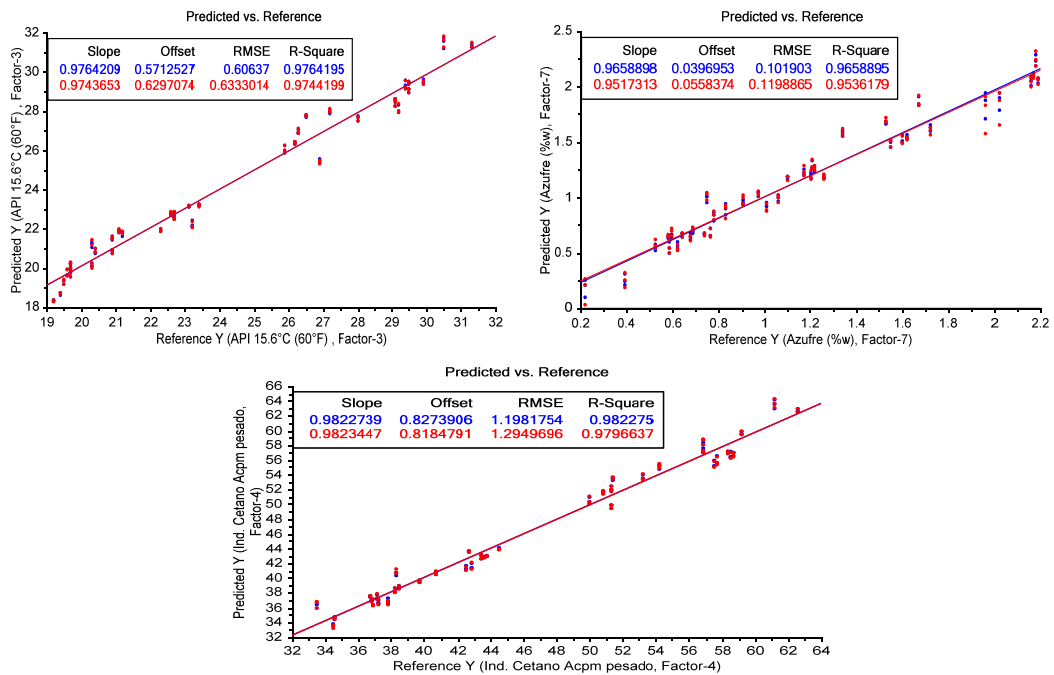


Figura 78. Calibración y validación cruzada modelos PLS de API, azufre e índice de cetano fracción acpm pesado.

En el modelo de azufre, se requieren cinco componentes para explicar el 90% de la varianza. Lo anterior permite concluir que los datos de calibración y de validación cruzada son representativos para todos los modelos.

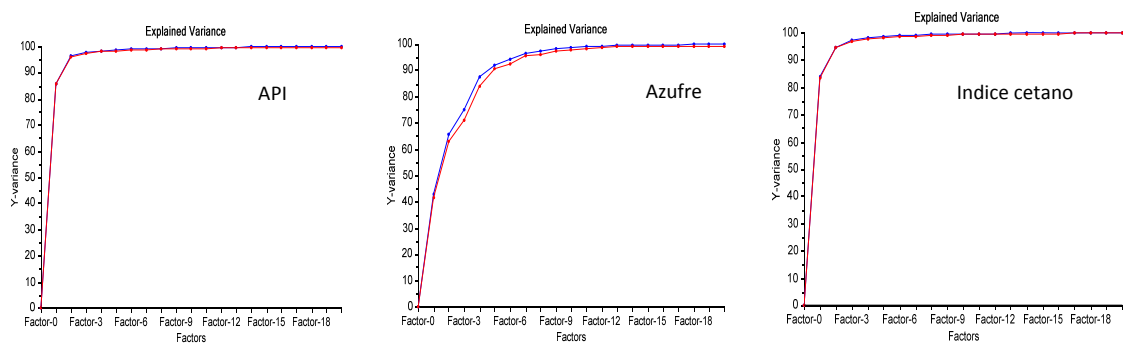


Figura 79. Varianza explicada de calibración y validación modelos PLS de API, azufre e índice de cetano fracción acpm pesado.

La validación externa de los modelos PLS de API, índice de cetano y azufre para la fracción acpm pesado se realizó con sets de entre 13 y 16 muestras. Las regresiones obtenidas para la validación externa de los tres modelos (figura 80), obtuvieron coeficientes de correlación entre 0.954 y 0.968.

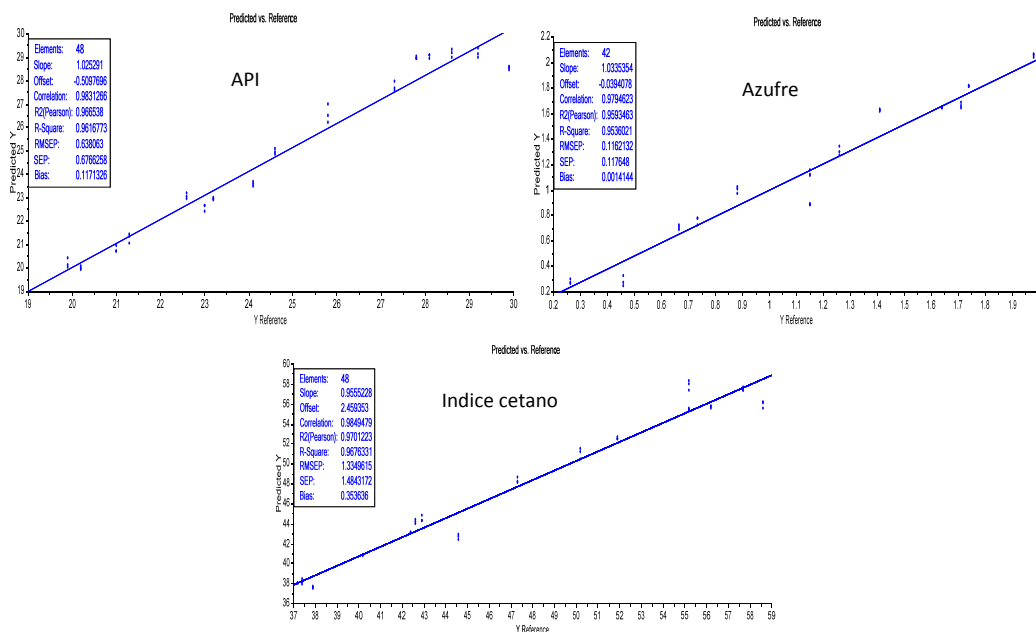


Figura 80. Validación externa modelos PLS de API, azufre e índice de cetano fracción acpm pesado.

En las tablas 48, 49 y 50 se presentan los resultados obtenidos para la predicción de muestras externas.

Los errores relativos para las predicciones de API e índice de cetano para la fracción acpm pesado fueron inferiores al 5%, reflejando un buen desempeño de estos dos modelos. Para la propiedad azufre, el error relativo máximo fue de 23.1%, mostrando un desempeño aceptable sin dividir el set de calibración/validación eterna en dos intervalos, como sucedió con los modelos de azufre para queroseno y acpm liviano.

Tabla 48. Validación externa del modelo PLS API fracción acpm pesado

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|-----------|------------------|----------------|------------|--------------------|
| 5 | 23.0 | 22.6 | 0.4 | 1.91 |
| 8 | 28.6 | 29.2 | 0.6 | 1.96 |
| 10 | 25.8 | 26.6 | 0.8 | 2.96 |
| 13 | 19.9 | 20.2 | 0.3 | 1.49 |
| 14 | 21.3 | 21.3 | 0.0 | 0.08 |
| 15 | 24.6 | 24.9 | 0.3 | 1.42 |
| 23 | 20.2 | 20.0 | 0.2 | 1.09 |
| 25 | 29.9 | 28.5 | 1.4 | 4.58 |
| 26 | 29.2 | 29.2 | 0.0 | 0.17 |
| 28 | 23.2 | 23.0 | 0.2 | 1.04 |
| 38 | 22.6 | 23.1 | 0.5 | 1.99 |
| 42 | 28.1 | 29.0 | 0.9 | 3.27 |
| 46 | 27.8 | 29.0 | 1.2 | 4.21 |
| 51 | 27.3 | 27.7 | 0.4 | 1.61 |
| 52 | 21.0 | 20.8 | 0.2 | 1.04 |
| 53 | 24.1 | 23.6 | 0.5 | 2.17 |

Tabla 49. Validación externa del modelo PLS azufre fracción acpm pesado

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|-----------|------------------|----------------|------------|--------------------|
| 5 | 0.733 | 0.756 | 0.023 | 3.07 |
| 8 | 1.710 | 1.662 | 0.048 | 2.81 |
| 13 | 1.980 | 2.054 | 0.074 | 3.74 |
| 14 | 1.260 | 1.305 | 0.045 | 3.56 |
| 15 | 1.150 | 1.127 | 0.023 | 1.99 |
| 23 | 2.000 | 1.931 | 0.069 | 3.43 |
| 25 | 0.263 | 0.275 | 0.012 | 4.49 |
| 28 | 0.666 | 0.702 | 0.036 | 5.45 |
| 38 | 1.150 | 0.884 | 0.266 | 23.1 |
| 46 | 1.640 | 1.648 | 0.008 | 0.47 |
| 51 | 1.740 | 1.814 | 0.074 | 4.27 |
| 52 | 1.410 | 1.623 | 0.213 | 15.1 |
| 53 | 0.882 | 1.001 | 0.119 | 13.5 |

Tabla 50. Validación externa del modelo PLS índice de cetano fracción acpm pesado

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|-----------|------------------|----------------|------------|--------------------|
| 5 | 44.6 | 42.6 | 2.0 | 4.38 |
| 8 | 55.2 | 55.4 | 0.2 | 0.42 |
| 10 | 50.2 | 51.0 | 0.8 | 1.64 |
| 13 | 37.9 | 37.6 | 0.3 | 0.86 |
| 14 | 40.2 | 40.8 | 0.6 | 1.50 |
| 15 | 47.3 | 48.3 | 1.0 | 2.15 |
| 23 | 37.4 | 38.2 | 0.8 | 2.05 |
| 25 | 58.6 | 55.9 | 2.7 | 4.58 |
| 26 | 57.7 | 57.5 | 0.2 | 0.43 |
| 28 | 42.4 | 43.1 | 0.7 | 1.63 |
| 38 | 42.6 | 44.2 | 1.6 | 3.76 |
| 42 | 55.2 | 57.9 | 2.7 | 4.82 |
| 46 | 56.2 | 55.7 | 0.5 | 0.92 |
| 51 | 51.9 | 52.5 | 0.6 | 1.20 |
| 52 | 37.2 | 38.0 | 0.8 | 2.18 |
| 53 | 42.9 | 44.5 | 1.6 | 3.63 |

La tabla 51 presenta el resumen de los parámetros estadísticos de la regresión para cada uno de los modelos PLS de propiedades desarrollados en la fracción acpm pesado.

Tabla 51. Resultados modelos de calibración y validación PLS fracción acpm pesado (315-371°C)

| Modelo | Media Calibración | R ² Calibración | RMSEC | R ² Validación | RMSECV | VL | Media Val. externa | R ² Predicción | RMSEP |
|-------------|-------------------|----------------------------|-------|---------------------------|--------|----|--------------------|---------------------------|-------|
| API | 24.3 | 0.976 | 0.606 | 0.974 | 0.633 | 3 | 24.8 | 0.962 | 0.638 |
| Azufre(%p) | 1.164 | 0.966 | 0.102 | 0.954 | 0.120 | 7 | 1.125 | 0.954 | 0.116 |
| Ind. Cetano | 46.7 | 0.982 | 1.198 | 0.980 | 1.295 | 4 | 47.3 | 0.968 | 1.335 |

Según los resultados de calibración, validación y predicción, los modelos de API e índice de cetano arrojaron los mejores parámetros estadísticos respecto a los modelos desarrollados para las fracciones queroseno y acpm liviano. Los errores relativos máximos de predicción para muestras externas fueron de 4.58% y 4.82%, es decir, inferiores al 5%.

En el modelo de azufre, se obtuvieron parámetros estadísticos aceptables. Sólo en un caso se obtuvo un error relativo de predicción superior al 20% (muestra 38, error del 23.1%), lo cual demuestra que este modelo arroja resultados satisfactorios. Lo anterior es lógico si se tiene en cuenta que el intervalo de concentraciones de azufre para esta fracción varió entre 0.2% y 2.2%, es decir, las concentraciones estuvieron del mismo orden de magnitud y sus valores fueron mayores que para las fracciones de queroseno y acpm liviano.

3.6.6 Modelos PLS API y azufre crudo reducido

El modelo de API (figura 81) presenta regresiones con coeficientes de correlación entre 0.979 y 0.989 para calibración y validación cruzada, con errores de calibración (RMSEC) y de validación (RMSECV) de 0.478 y 0.547, en una escala de 3 a 20. Con dos componentes se explica más del 90% de la varianza en los datos Y. Fueron seleccionadas seis variables latentes.

Los resultados obtenidos muestran que este modelo arroja muy buen desempeño en la predicción de nuevas muestras.

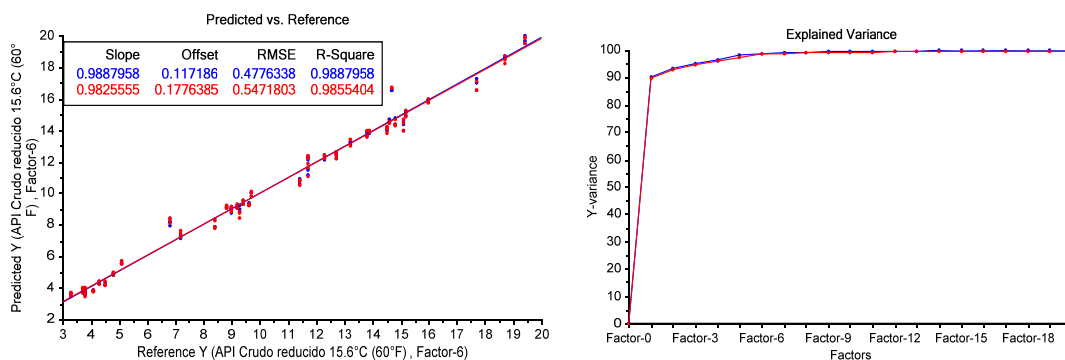


Figura 81. Principales resultados modelo PLS para API fracción crudo reducido.

Para el modelo de azufre (figura 82), los coeficientes de regresión para calibración y validación fueron de 0.984 y 0.979, con errores de calibración y validación de 0.112 y 0.128, respectivamente. Contrario a los modelos de azufre para las fracciones queroseno y acpm liviano,

el modelo para esta fracción arroja muy buen desempeño, de acuerdo con los parámetros estadísticos obtenidos.

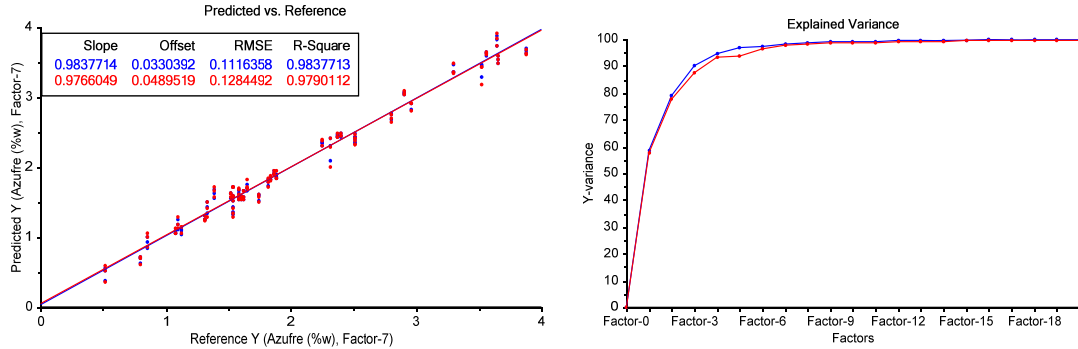


Figura 82. Principales resultados modelo PLS para azufre fracción crudo reducido.

La validación externa de los dos modelos anteriores se realizó con sets de entre 14 y 16 muestras de crudos, distribuidas en un intervalo de API entre 3 y 20. En la figura 83 se observan las correlaciones obtenidas.

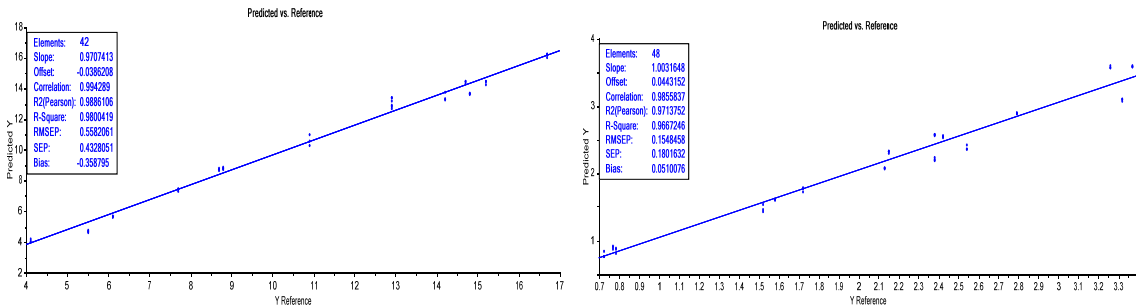


Figura 83. Validación externa modelos PLS de API y azufre fracción crudo reducido.

Los coeficientes de correlación obtenidos para la validación externa fueron de 0.980 para API y 0.967 para azufre, con errores de predicción de 0.558 para API y 0.155 para azufre. Los resultados de la validación externa mostrados en las tablas 52 y 53 indican que los errores relativos de predicción máximos no superaron el 14.5% y 16.3% para API y azufre del crudo reducido, respectivamente.

Tabla 52. Validación externa del modelo PLS API fracción crudo reducido

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|------------------|-------------------------|-----------------------|-------------------|---------------------------|
| 5 | 7.70 | 7.40 | 0.30 | 3.95 |
| 8 | 14.7 | 14.4 | 0.29 | 1.96 |
| 10 | 10.9 | 10.6 | 0.28 | 2.53 |
| 13 | 4.10 | 4.07 | 0.03 | 0.65 |
| 14 | 5.50 | 4.70 | 0.80 | 14.5 |
| 15 | 12.9 | 13.3 | 0.43 | 3.37 |
| 25 | 15.2 | 14.4 | 0.81 | 5.36 |
| 26 | 14.2 | 13.4 | 0.76 | 5.33 |
| 28 | 8.80 | 8.81 | 0.01 | 0.09 |
| 42 | 16.7 | 16.1 | 0.55 | 3.31 |
| 46 | 12.9 | 12.8 | 0.08 | 0.60 |
| 51 | 14.8 | 13.6 | 1.16 | 7.84 |
| 52 | 6.10 | 5.66 | 0.44 | 7.23 |
| 53 | 8.70 | 8.73 | 0.03 | 0.38 |

Tabla 53. Validación externa del modelo PLS azufre fracción crudo reducido

| Crudo No. | Valor Referencia | Valor Predicho | Diferencia | Error Relativo (%) |
|------------------|-------------------------|-----------------------|-------------------|---------------------------|
| 5 | 2.38 | 2.21 | 0.17 | 7.05 |
| 8 | 2.54 | 2.38 | 0.16 | 6.31 |
| 10 | 1.52 | 1.48 | 0.04 | 2.81 |
| 13 | 3.37 | 3.59 | 0.22 | 6.40 |
| 14 | 2.79 | 2.89 | 0.10 | 3.48 |
| 15 | 1.72 | 1.76 | 0.04 | 2.48 |
| 23 | 3.26 | 3.58 | 0.32 | 9.75 |
| 25 | 0.77 | 0.90 | 0.13 | 16.3 |
| 26 | 0.72 | 0.79 | 0.07 | 9.56 |
| 28 | 1.58 | 1.61 | 0.03 | 1.81 |
| 38 | 2.15 | 2.32 | 0.17 | 7.74 |
| 42 | 0.78 | 0.86 | 0.08 | 10.0 |
| 46 | 2.42 | 2.55 | 0.13 | 5.23 |
| 51 | 2.38 | 2.57 | 0.19 | 8.10 |
| 52 | 3.32 | 3.10 | 0.22 | 6.74 |
| 53 | 2.13 | 2.08 | 0.05 | 2.35 |

En la tabla 54 se presenta el resumen de los parámetros estadísticos obtenidos para los modelos de rendimientos, API y azufre de la fracción crudo reducido. Como era de esperarse, a medida que la fracción de destilado es más pesada, los modelos mejoraron los parámetros de predicción. Lo anterior debido a que los intervalos de los valores para las diferentes propiedades fisicoquímicas fueron más amplios que en las fracciones antecesoras al crudo reducido.

Tabla 54. Resultados modelos de calibración y validación PLS fracción crudo reducido (371°C+)

| Modelo | Media Calibración | R ² Calibración | RMSEC | R ² Validación | RMSECV | VL | Media Val. externa | R ² Predicción | RMSEP |
|------------|-------------------|----------------------------|-------|---------------------------|--------|----|--------------------|---------------------------|-------|
| API | 10.5 | 0.989 | 0.478 | 0.986 | 0.547 | 6 | 11.1 | 0.980 | 0.558 |
| Azufre(%p) | 2.036 | 0.984 | 0.112 | 0.979 | 0.128 | 7 | 2.115 | 0.967 | 0.155 |

Los modelos de API y azufre presentaron buen desempeño según los resultados obtenidos en los parámetros estadísticos de calibración, validación y predicción. Los máximos errores relativos de predicción para muestras externas estuvieron entre 14.5% y 16.3%, es decir, inferiores al 20%.

3.6.7 Reproducibilidad modelos PLS de propiedades

Los modelos desarrollados se evaluaron en términos de la reproducibilidad de las normas ASTM estándar empleadas en la caracterización de las propiedades, excepto los rendimientos de las fracciones atmosféricas.

La naturaleza del método de destilación atmosférica (uso de grandes cantidades de muestra y alto consumo de tiempo) y la naturaleza del producto obtenido (material altamente volátil e inestable para los cortes muy livianos), hacen muy difícil el uso de técnicas de control estadístico. Además, este método no produce un resultado sencillo, ni es una serie de resultados obtenidos bajo condiciones rígidamente definidas (American Society for Testing and Materials 2013b). Adicionalmente, la evaluación de la reproducibilidad para este método implicaría un costo adicional a la realización de los Assays Tipo III, razones por las cuales es poco viable la evaluación de la reproducibilidad del método en el laboratorio.

En las Tablas 55 a 58, se presentan los resultados de la comparación entre la reproducibilidad ASTM de los métodos estándar y los valores de RMSEP para los diferentes modelos desarrollados.

Tabla 55. Evaluación de la reproducibilidad modelos PLS fracción queroseno (199-249°C)

| Propiedad | Método de referencia | Fórmula Reproducibilidad (R) | Límites R | | Promedio | Valor R | RMSEP |
|--------------------------|----------------------|---------------------------------------|-----------|-------|----------|---------|-------|
| | | | Min | Máx | | | |
| API | ASTM D5002 | 0.5 | 35.30 | 36.30 | 35.8 | 0.5 | 1.399 |
| Azufre ¹ (%p) | ASTM D4294 | $(1.9182*((y*10000)^{0.6446}))/10000$ | 0.079 | 0.111 | 0.095 | 0.016 | 0.010 |
| Azufre ² (%p) | ASTM D4294 | $(1.9182*((y*10000)^{0.6446}))/10000$ | 0.372 | 0.454 | 0.413 | 0.041 | 0.031 |
| Ind. cetano | ASTM D4737 | NR | NR | NR | 36.39 | NR | 1.949 |

NR: no reportado.

Tabla 56. Evaluación de la reproducibilidad modelos PLS fracción acpm liviano (249-315°)

| Propiedad | Método de referencia | Fórmula Reproducibilidad | Límites R | | Promedio | Valor R | RMSEP |
|--------------------------|----------------------|---------------------------------------|-----------|-------|----------|---------|-------|
| | | | Min | Máx | | | |
| API | ASTM D5002 | 0.5 | 29.20 | 30.20 | 29.7 | 0.5 | 0.702 |
| Azufre ¹ (%p) | ASTM D4294 | $(1.9182*((y*10000)^{0.6446}))/10000$ | 0.217 | 0.277 | 0.247 | 0.030 | 0.020 |
| Azufre ² (%p) | ASTM D4294 | $(1.9182*((y*10000)^{0.6446}))/10000$ | 0.878 | 1.018 | 0.948 | 0.070 | 0.113 |
| Ind. cetano | ASTM D4737 | NR | NR | NR | 42.1 | NR | 1.871 |

NR: no reportado.

Tabla 57. Evaluación de la reproducibilidad modelos PLS fracción acpm pesado (315-371°C)

| Propiedad | Método de referencia | Fórmula Reproducibilidad | Límites R | | Promedio | Valor R | RMSEP |
|-------------|----------------------|---------------------------------------|-----------|-------|----------|---------|-------|
| | | | Min | Máx | | | |
| API | ASTM D5002 | 0.5 | 24.30 | 25.30 | 24.8 | 0.5 | 0.638 |
| Azufre(%p) | ASTM D4294 | $(1.9182*((y*10000)^{0.6446}))/10000$ | 1.047 | 1.203 | 1.125 | 0.078 | 0.116 |
| Ind. cetano | ASTM D4737 | NR | NR | NR | 47.3 | NR | 1.335 |

NR: no reportado.

Tabla 58. Evaluación de la reproducibilidad modelos PLS fracción crudo reducido (371°C+)

| Propiedad | Método de referencia | Fórmula Reproducibilidad | Límites R | | Promedio | Valor R | RMSEP |
|------------|----------------------|---------------------------------------|-----------|-------|----------|---------|-------|
| | | | Min | Máx | | | |
| API | ASTM D5002 | 0.5 | 10.60 | 11.60 | 11.1 | 0.5 | 0.558 |
| Azufre(%p) | ASTM D4294 | $(1.9182*((y*10000)^{0.6446}))/10000$ | 1.997 | 2.233 | 2.115 | 0.118 | 0.155 |

Los valores de RMSEP obtenidos para todos los modelos fueron superiores a la reproducibilidad de las normas ASTM, excepto para los modelos de azufre en queroseno en los dos intervalos de concentración, y para el primer intervalo de concentración de azufre en acpm liviano, en los cuales los RMSEP fueron menores a los reportados por la norma ASTM D4294.

En los modelos de API, se observa que el error RMSEP disminuye a medida que la fracción se hace más pesada, acercándose cada vez más al valor de reproducibilidad de la norma ASTM.

De acuerdo con los resultados anteriores y dado que este desarrollo se realizó a partir de los espectros IR de los crudos madre y no de los espectros IR de cada fracción, el desempeño de los modelos se considera aceptable para realizar seguimiento o evaluar tendencias.

4. CONCLUSIONES

Los modelos PLS y AG-PLS desarrollados para predecir los rendimientos y propiedades de las fracciones de destilación atmosférica a partir del espectro de sus crudos madre, muestran que esta puede ser una alternativa útil para obtener resultados preliminares de rendimientos y propiedades de las fracciones, con errores relativos de predicción máximos del 20%.

La metodología empleada permitió identificar las variables responsables de las diferencias espectrales entre las muestras de crudos. Estas variables, originadas por las diversas funcionalidades químicas presentes en las muestras, son las causantes de las diferencias espectrales distinguibles a la matemática multivariada, lo cual permitió asociar las propiedades de los destilados atmosféricos a los espectros de sus crudos madre mediante los modelos quimiométricos.

Los resultados de los parámetros estadísticos obtenidos para los ocho modelos PLS y AG-PLS propuestos para rendimientos de las fracciones atmosféricas, muestran que los modelos arrojan un desempeño aceptable y son equivalentes. Para los primeros, se obtuvieron coeficientes de correlación de calibración entre 0.81 y 0.98 y errores relativos máximos de predicción entre 10% y 18%, para los segundos, coeficientes entre 0.90 y 0.97, con errores relativos de predicción del 11% al 17%. Los mejores parámetros estadísticos se obtuvieron para los modelos PLS y AG-PLS de la fracción pesada (crudo reducido).

Se propusieron trece modelos para las propiedades de API, azufre e índice de cetano. La separación de los modelos de azufre en dos intervalos de concentración para quero y acpm liviano, mejoró las predicciones para esta propiedad. Se obtuvieron coeficientes de correlación entre 0.90 y 0.99, con errores relativos máximos de predicción entre 5 y 25%. Los mejores resultados se obtuvieron para los modelos de API y azufre de la fracción crudo reducido.

En todos los casos, los mejores parámetros estadísticos se obtuvieron para la fracción más pesada (crudo reducido). Lo anterior sugiere que se podrían obtener resultados satisfactorios en el

desarrollo de modelos PLS o AG-PLS de fracciones pesadas de destilación al vacío, a partir de sus crudos madre.

Es importante analizar los datos analíticos de los crudos y sus destilados atmosféricos de forma conjunta para asociar información analítica de propiedades con los espectros IR, como en el caso del número de ácido, parámetro que permitió explicar la inclusión de las variables asociadas al grupo carbonilo en la selección de variables mediante algoritmos genéticos.

5. RECOMENDACIONES

Los modelos pueden mejorarse incluyendo más muestras en el set de calibración, que cubran una amplia gama de composición química de los crudos, para construir un solo modelo o separarlos según sus características especiales como API, acidez o contenido de azufre y evaluar su desempeño.

Los modelos desarrollados aplican a muestras químicamente similares a las empleadas en la construcción de los modelos de calibración y en los intervalos de concentración de las propiedades incluidas. Teniendo en cuenta que los crudos base cargados a las refinerías pueden cambiar debido a la preparación de mezclas, inclusión de crudos importados o nuevos hallazgos, es necesario realizar la actualización constante de los modelos para garantizar resultados aceptables en las predicciones de nuevas muestras.

Con base en los parámetros del algoritmo genético optimizados en esta investigación, podría evaluarse el ajuste de otros ítems que fueron asignados por default, para determinar su influencia y mejorar los modelos desarrollados.

BIBLIOGRAFIA

- Abbas, O. et al., 2012. PLS regression on spectroscopic data for the prediction of crude oil quality: API gravity and aliphatic/aromatic ratio. *Fuel*, 98, pp.5–14.
- American Society for Testing and Materials, 2013a. ASTM D1160. Standard Test Method for Distillation of Petroleum Products at Reduced Pressure 1. , pp.1–20.
- American Society for Testing and Materials, 2013b. ASTM D2892. Standard Test Method for Distillation of Crude Petroleum (15-Theoretical Plates). , p.34.
- American Society for Testing and Materials, 2013c. ASTM D5236. Standard Test Method for Distillation of Heavy Hydrocarbon Mixtures (Vacuum Potstill Method)1. , p.18.
- American Society for Testing and Materials, 2012. ASTM E1655. Standard Practices for Infrared Multivariate Quantitative Analysis 1. , 5(Reapproved), p.29.
- American Society for Testing and Materials, 2003. *Significance tests for Petroleum Products* 7th ed. S. J. Rand, ed., Bridgeport, N.J.: ASTM International.
- Andreas A. Kardamakis, N.P., 2010. Autoregressive modeling of near-IR spectra and MLR to predict RON values of gasolines. *Fuel*, 89, pp.158–161.
- Aske, N., H., K. & J., S., 2001. Determination of saturate, aromatic, resin and asphaltenic (SARA) components in crude oils by means of infrared and Near-infrared spectroscopy. *Energy & Fuels*, 15, pp.1304–1312.
- Baldrich, C.-A., Novoa, L.-A. & Bueno, A., 2010. Comparison between NIR and UVVis Spectra Chemometrics for Predicting FCC Feedstocks Properties. *CT&F Ciencia, Tecnología y Futuro*, 4(1), pp.113–126.
- Baldrich F., C.-A. & Novoa M., L.-A., 2006. Rapid Characterization of Diesel Fuel by Infrared Spectroscopy. *CT&F Ciencia, Tecnología y Futuro*, 3, pp.171–182.
- Baldrich F., C.-A. & Novoa M., L.-Á., 2007. Detail Chemical Characterization of Petroleum Middle Fractions by Chemometrics Analysis of their Ultraviolet Spectrum. *CT&F Ciencia, Tecnología y Futuro*, 3, pp.173–190.
- Barry M. Wise, Neal B. Gallager, Rasmus Bro, Jeremy M. Shaver, Willem Windig, R.S.K., 2006. *Chemometrics Tutorial for PLS _ Toolbox and Solo*, Wenatchee: Eigenvector Research, Inc.
- Bougeard, D. et al., 1995. *Infrared and Raman Spectroscopy. Methods and Applications* B. Schrader, ed., Weinheim: VHC Verlagsgesellschaft mbH.
- Bro, R. & Smilde, A.K., 2014. Principal component analysis. *Analytical Methods*, 6(9), p.2812.
- Bruker OptiK International, 2012. Surface Analytics and Sample Preparation - Training Courses. , pp.1–122.

- Castro, L. V & Vazquez, F., 2009. Fractionation and Characterization of Mexican Crude Oils. *Energy & Fuels*, 23(3), pp.1603–1609.
- Chu, X.L. et al., 2011. Rapid identification and assay of crude oils based on moving-window correlation coefficient and near infrared spectral library. *Chemometrics and Intelligent Laboratory Systems*, 107(1), pp.44–49.
- Donald L. Pavia, Gary M. Lampman, G.S.K., 2001. *Introduction to Spectroscopy* 3rd. T. Learning, ed., Singapore: Brooks Cole.
- Douglas A. Skoog, F. James Holler, T.A.N., 2001. *Análisis instrumental* 5a. C. F. Madrid, ed., Madrid: Mac Graw Hill Inc.
- Durand, A. et al., 2007. Genetic algorithm optimisation combined with partial least squares regression and mutual information variable selection procedures in near-infrared quantitative analysis of cotton-viscose textiles. *Analytica Chimica Acta*, 595, pp.72–79.
- Ferrer, C.-A.B., 2008. Prediction of physicochemical properties of FCC Feedstock by chemometric analysis of their ultraviolet spectrum. *CT&F Ciencia, Tecnología y Futuro*, 3(4), pp.143–156.
- Filgueiras, P.R. et al., 2014. Determination of API gravity, kinematic viscosity and water content in petroleum by ATR-FTIR spectroscopy and multivariate calibration. *Fuel*, 116, pp.123–130.
- Gaweł, B., Eftekhardakhah, M. & Øye, G., 2014. Elemental Composition and Fourier Transform Infrared Spectroscopy Analysis of Crude Oils and Their Fractions. *Energy & Fuels*, 28(2), pp.997–1003.
- Geladi, P., 2003. Chemometrics in spectroscopy. Part 1. Classical chemometrics. *Spectrochimica Acta Part B*, 58, pp.767–782.
- Ghasemi, J., Niazi, A. & Leardi, R., 2003. Genetic-algorithm-based wavelength selection in multicomponent spectrophotometric determination by PLS: application on copper and zinc mixture. *Talanta*, 59, pp.311–317.
- Hongfu Y., X.C., L., H. & X., Y., 2006. Determination of multi-properties of residual oils using mid-infrared attenuated total reflection spectroscopy. *Fuel*, 85, pp.1720–1728.
- Hopke, P.K., 2003. The evolution of chemometrics. *Analytica Chimica Acta*, 500(1–2), pp.365–377.
- Jaime, B.H. et al., 2013. *Cadena del Petróleo*, Bogotá.
- Khanmohammadi, M., Garmarudi, A.B. & De la Guardia, M., 2012. Characterization of petroleum-based products by infrared spectroscopy and chemometrics. *TrAC - Trends in Analytical Chemistry*, 35, pp.135–149.
- Kuptsov, a. K. & Arbusova, T. V., 2011. A study of heavy oil fractions by Fourier-transform near-infrared Raman spectroscopy. *Petroleum Chemistry*, 51(3), pp.203–211.
- Lavine, B.K. & Workman, J., 2013. Chemometrics. *Analytical Chemistry*, 85, pp.705–714.

- Laxalde, J. et al., 2011. Characterisation of heavy oils using near-infrared spectroscopy: Optimisation of pre-processing methods and variable selection. *Analytica Chimica Acta*, 705, pp.227–234.
- Leardi, R. & Lupiáñez González, A., 1998. Genetic algorithms applied to feature selection in PLS regression: How and when to use them. *Chemometrics and Intelligent Laboratory Systems*, 41, pp.195–207.
- Leardi, R., Seasholtz, M.B. & Pell, R.J., 2002. Variable selection for multivariate calibration using a genetic algorithm: Prediction of additive concentrations in polymer films from Fourier transform-infrared spectral data. *Analytica Chimica Acta*, 461, pp.189–200.
- Li, K. et al., 2013. Correlation of chemical and physical properties of an Alaska heavy oil from the Ugnu formation. *Fuel*, 103, pp.843–849.
- Liliana de Fátima Bezerra de Lira et al, 2010. Prediction of properties of diesel/biodiesel blends by infrared spectroscopy and multivariate calibration. *Fuel*, 89, pp.405–409.
- Lindon, J.C., 2000. *Encyclopedia of Spectroscopy and Spectrometry* J. L. Lindon, John C., Tranter, George E., Holmes, ed., London: Academic Press.
- Lluch Urpí, J., 2004. *Tecnología y margen de refino del petróleo* D. de Santos, ed., España.
- M.P. Gómez-Carracedo, J.M. Andrade, M. Calviño, E. Fernández, D. Prada, S.M., 2003. Multivariate prediction of eight kerosene properties employing vapour-phase mid-infrared spectrometry. *Fuel*, 82, pp.1211–1218.
- Marshall, A.G. & Rodgers, R.P., 2004. Petroleomics : The Next Grand Challenge for Chemical Analysis Petroleum and Mass Spectrometry : Divergent. *Accounts of chemical research*, 37(1), pp.53–59.
- Martens, H. & Naes, T., 1992. *Multivariate Calibration* 1st ed. M. B. S. Wiley-interscience, ed., Oslo.
- Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y., Kaufman, L., 2003. *Chemometrics: a textbook* R. S. C. Vandeginste, B.G.M., ed., Amsterdam: Elsevier B.V.
- Mehmood, T. et al., 2012. A review of variable selection methods in Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory Systems*, 118, pp.62–69.
- Meléndez, L. V. et al., 2012. Prediction of the SARA analysis of Colombian crude oils using ATR-FTIR spectroscopy and chemometric methods. *Journal of Petroleum Science and Engineering*, 90–91, pp.56–60.
- Mohammad Al-Ghouti, Yahya Al-Degs, F.M., 2010. Determination of hydrogen content, gross heat of combustion, and net heat of combustion of diesel fuel using FTIR spectroscopy and multivariate calibration. *Fuel*, 89, pp.193–201.
- Molina, D. V, Uribe, U.N. & Murgich, J., 2007. Partial least-squares (PLS) correlation between refined product yields and physicochemical properties with the ¹H nuclear magnetic

- resonance (NMR) spectra of Colombian crude oils. *Energy & Fuels*, 21(3), pp.1674–1680.
- Orrego-Ruiz, J. a et al., 2011. Mid-infrared attenuated total reflectance (MIR-ATR) predictive models for asphaltene contents in vacuum residua: Asphaltene structure–functionality correlations based on partial least-squares regression (PLS-R). *Energy & Fuels*, 25(8), pp.3678–3686.
- Pasquini, C. & Bueno, A.F., 2007. Characterization of petroleum using near-infrared spectroscopy: Quantitative modeling for the true boiling point curve and specific gravity. *Fuel*, 86(12–13), pp.1927–1934.
- de Peinder, P. et al., 2009. Partial least squares modeling of combined infrared, ¹H NMR and ¹³C NMR spectra to predict long residue properties of crude oils. *Vibrational Spectroscopy*, 51(2), pp.205–212.
- Pemex, Consultado el 23 de mayo de 2014. *Pemex*. Available at: http://www.pemex.com/productos/petroleo_crudo/Paginas/default.aspx#.U4EncXJ5OYQ.
- Riazi, M.R., 2005. *Characterization and Properties of Petroleum Fractions* 1st ed. ASTM International, ed., Philadelphia.
- Rocha, J.T.C. et al., 2016. Sulfur Determination in Brazilian Petroleum Fractions by Mid-infrared and Near-infrared Spectroscopy and Partial Least Squares Associated with Variable Selection Methods. *Energy and Fuels*, 30, pp.698–705.
- Roman M. Balabin, S.V.S., 2011. Variable selection in near-infrared spectroscopy: Benchmarking of feature selection methods on biodiesel data. *Analytica Chimica Acta*, 692, pp.63–72.
- Sarkhosh, M. et al., 2014. Application of genetic algorithms for pixel selection in multivariate image analysis for a QSAR study of trypanocidal activity for quinone compounds and design new quinone compounds. *Chemometrics and Intelligent Laboratory Systems*, 139, pp.168–174.
- Satya, S. et al., 2007. Estimation of properties of crude oil residual fractions using chemometrics. *Energy & Fuels*, 21(2), pp.998–1005.
- Silva, M.P.F. da et al., 2014. Classification of gasoline as with or without dispersant and detergent additives using infrared spectroscopy and multivariate classification. *Fuel*, 116, pp.151–157.
- Stuart, B., 2004. *Infrared Spectroscopy: Fundamentals and Applications* D. J. Ando, ed., Chichester, England: John Wiley & Sons, Ltd.
- Vandeginste, B.G.M., Massart, D.L., Buydens, L.M.C., De Jong, S., Lewi, P.J., Smeyers-Verbeke, J., 2003. *Handbook of Chemometrics and Qualimetrics: Part B* 2nd ed. S. C. Vandeginste, B.G.M., Rutan, ed., Amsterdam: Elsevier B.V.
- Waquier, J.P., 2004. *El Refino del Petróleo* D. de Santos, ed., Madrid: Instituto Superior de la Energía.

- Wold, S., Sjoström, M. & Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58, pp.109–130.
- Wörner, H.J. & Merkt, F., 2011. Fundamentals of Electronic Spectroscopy. In *Handbook of High-resolution Spectroscopy*. John Wiley & Sons, Ltd, p. 262.
- Xiaobo, Z. et al., 2010. Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta*, 667, pp.14–32.