

**Contenido de Azufre y RVP de las gasolinas producidas en la GRB: predicción con
espectroscopia NIR y redes neuronales artificiales**

Julián Alberto Zarate Gualdrón

Proyecto de grado para optar al título de Magister en Ingeniería Química

Director:

Giovanni Morales Medina

Doctor en Ingeniería Química

Universidad Industrial de Santander

Facultad de Ingenierías Fisicoquímicas

Escuela de Ingeniería Química

Maestría en Ingeniería Química

Bucaramanga

2019

Contenido

	Pág.
Introducción	12
1. Objetivos	15
1.1 Objetivo General	15
1.2 Objetivos Específicos.....	15
2. Marco Teórico	16
2.1 Preparación de Gasolina Comercial.....	16
2.2 Contenido de Azufre en las gasolinas	18
2.3 Vapor de Presión Reid (RVP).....	19
2.4 Espectroscopía del Infrarrojo Cercano (NIR)	20
2.5 Modelos de predicción basados en espectros IR.....	22
2.6 Redes Neuronales Artificiales	24
2.7 Estado del Arte	27
3. Metodología	29
3.1 Recopilación de espectros NIR y propiedades de las gasolinas.	29
3.2 Pretratamientos y análisis PCA de la base de espectros NIR.....	30
3.3 Proposición de líneas de código para entrenamiento y validación de modelos RNA.	31
3.4 Aplicación del algoritmo para modelos RNA con componentes principales.....	32
3.5 Entrenamiento y validación de las RNA con el área en las absorbancias pretratadas	32

3.6. Selección de los modelos RNA.....	33
4. Análisis de resultados.....	33
4.1 Espectros NIR y propiedades de las naftas de la GRB.	33
4.2 Análisis PCA de la base de espectros NIR.....	39
4.3 Código de entrenamiento y validación de modelos RNA.....	43
4.4 Aplicación del código considerando componentes principales	45
4.5 Aplicación del algoritmo considerando absorbancias pretratadas	55
4.6 Selección de los modelos RNA.....	57
5. Conclusiones	65
6. Recomendaciones	66
Referencias Bibliográficas	67

Lista de Figuras

	Pág.
<i>Figura 1.</i> Esquema ilustrativo de los flujos de naftas y el <i>blending</i> de gasolinas en la GRB.	18
<i>Figura 2.</i> Diagrama del espectro electromagnético.	21
<i>Figura 3.</i> Ejemplo de espectros de absorción NIR para de gasolinas.	22
<i>Figura 4.</i> Pretratamientos utilizados en análisis de espectros IR.	23
<i>Figura 5.</i> Ilustración arquitectura y parámetros que definen una RNA.	25
<i>Figura 6.</i> Funciones de activación habitualmente utilizadas en RNA.	26
<i>Figura 7.</i> Detalle de los espectros en diferentes rangos de números de onda.	35
<i>Figura 8.</i> Modificación de los espectros con los pretratamientos.	38
<i>Figura 9.</i> Propiedades de las muestras de naftas y gasolinas.	39
<i>Figura 10.</i> Resultados del análisis PCA aplicado a los espectros pretratados.	41
<i>Figura 11.</i> Variación de los estadísticos de desempeño con el número de neuronas de la capa interna (x) en la red 6:x:1–Log en la predicción del contenido de Azufre.	47
<i>Figura 12.</i> Variación de los estadísticos de desempeño con el número de neuronas de la capa interna (x) en la red 6:x:1–TanH en la predicción del contenido de Azufre.	47
<i>Figura 13.</i> Variación de los estadísticos de desempeño con el número de neuronas de la capa interna (x) en la red 6:x:1–Log en la predicción del RVP.	48
<i>Figura 14.</i> Variación de los estadísticos de desempeño con el número de neuronas de la capa interna (x) en la red 6:x:1–TanH en la predicción del RVP.	48

<i>Figura 15.</i> Variación de los estadísticos de desempeño total con el número de neuronas de la capa interna (x) y el número de PC de entrada (y) para la red y:x:1–Log en la predicción del contenido de azufre.	50
<i>Figura 16.</i> Variación de los estadísticos de desempeño total con el número de neuronas de la capa interna (x) y el número de PC de entrada (y) para la red y:x:1–TanH en la predicción del contenido de azufre.	51
<i>Figura 17.</i> Variación de los estadísticos de desempeño total con el número de neuronas de la capa interna (x) y el número de PC de entrada (y) para la red y:x:1–Log en la predicción del RVP.	52
<i>Figura 18.</i> Variación de los estadísticos de desempeño total con el número de neuronas de la capa interna (x) y el número de PC de entrada (y) para la red y:x:1–TanH en la predicción del RVP.	53
<i>Figura 19.</i> Variación de los estadísticos de desempeño total con el número de neuronas en la segunda capa interna (x) para la predicción del contenido de azufre.	54
<i>Figura 20.</i> Variación de los estadísticos de desempeño total con el número de neuronas en la segunda capa interna (x) para la predicción del RVP.	54
<i>Figura 21.</i> Valores de la integral cada 20 cm ⁻¹ en los rangos pretratados 4800–4000 cm ⁻¹ y 6200–5400 cm ⁻¹	55
<i>Figura 22.</i> Mejores desempeños totales obtenidos con las arquitecturas según las entradas definidas.	58
<i>Figura 23.</i> Desempeños de las redes para las mejores entradas según el número de neuronas de la capa interna.	59

<i>Figura 24. Desempeños considerando una segunda capa oculta en las estructuras de mejor</i>	
<i>desempeño.....</i>	<i>60</i>

Lista de Tablas

	Pág.
Tabla 1. <i>Algunos requisitos de calidad de la gasolina regular y extra.</i>	17
Tabla 2. <i>Regiones de absorción en el infrarrojo.</i>	21
Tabla 3. <i>Frecuencias de absorción de algunos grupos funcionales en hidrocarburos.</i>	22
Tabla 4. <i>Valores de los primeros seis componentes principales para las muestras (scores) con una varianza acumulada del 99.62%.</i>	42
Tabla 5. <i>Resumen de las mejores estructuras de red en el ajuste del contenido de Azufre y RVP de las muestras de nafta de la GRB.</i>	61

Resumen

Titulo. Contenido de Azufre y RVP de las gasolinas producidas en la GRB: predicción con espectroscopia NIR y redes neuronales artificiales*

Autor: Julián Alberto Zarate Gualdrón**

Palabras clave: Contenido de azufre, RVP, NIR, redes neuronales artificiales, modelo circunstancial, PCA.

Descripción

La Gerencia Refinería Barrancabermeja, GRB, genera gasolinas con los parámetros de calidad según la Resolución 1180 de 21 de junio de 2006. La preparación de gasolinas requiere la validación por laboratorio de los parámetros de calidad; sin embargo, los tiempos de medición pueden demorar *ca.* 6 horas. Una alternativa ante los tiempos de validación corresponde a la aplicación de modelos de redes neuronales artificiales, RNA, desarrollados con base en espectros NIR. En el presente trabajo diferentes arquitecturas RNA fueron entrenadas y validadas con espectros NIR para el ajuste del contenido de azufre y la presión de vapor Reid, RVP, de gasolinas de la GRB. Según los resultados, los rangos de mayor contribución en la señal están entre $4000 - 4800 \text{ cm}^{-1}$ y $5400 - 6200 \text{ cm}^{-1}$. Asimismo, según el análisis PCA, el valor de la segunda derivada en los números de onda entre $4309 - 4316 \text{ cm}^{-1}$ y $4326 - 4338 \text{ cm}^{-1}$ define las tendencias en los espectros de las gasolinas nacionales. También, los mejores desempeños en términos de RMSE y R^2 fueron exhibidos por las redes 30:12:1-TanH y 6:12:1-TanH para el ajuste del contenido de Azufre y el RVP, respectivamente. La red 30:12:1-TanH utilizó las 30 primeras áreas de mayor valor, mientras que la red 6:12:1-TanH requirió de los 6 primeros componentes principales. El RMSE de la red 30:12:1-TanH de ajuste del contenido de Azufre se encontró entre la repetibilidad y la reproducibilidad promedios (ASTM D4294-16). Respecto al RVP, el RMSE fue menor que la repetibilidad y la reproducibilidad promedios. La comparación de los mejores desempeños obtenidos en el presente trabajo con los reportados en la literatura abierta soporta la capacidad de predicción cuantitativa de las redes neuronales artificiales 30:12:1-TanH y 6:12:1-TanH.

* Proyecto de grado

** Facultad de Ingenierías Fisicoquímicas Escuela de Ingeniería Química Maestría en Ingeniería Química Director: Giovanni Morales Medina Doctor en Ingeniería Química

Abstract

Title. Sulphur content and RVP for gasolines produced by GRB refinery: Prediction using NIR spectra and artificial neural networks *

Author: Julián Alberto Zarate Gualdrón**

Keywords: Sulphur content, RVP, NIR, artificial neural network, circunstancial model, PCA.

Description:

Gerencia Refinería Barrancabermeja”, GRB, produces gasolines according to Colombian regulation “Resolución 1180 de 21 de junio de 2006”. Gasoline preparation by blending requires validation of properties in the lab; however, measurement protocols can last for *ca.* 6 hours before calculating property values. As an alternative, artificial neural network models (ANN) based on NIR spectra can predict properties, reducing the number of lab validation tests. Herein, different ANN architectures were trained and validated by the use of NIR spectra for adjusting sulphur content and Reid vapor pressure, RVP, of GRB gasolines. According to the results, the major contribution ranges for the spectroscopical signal were found to be $4000 - 4800 \text{ cm}^{-1}$ and $5400 - 6200 \text{ cm}^{-1}$. Likewise, from PCA it follows that second derivatives in the regions of wavenumbers $4309-4316 \text{ cm}^{-1}$ and $4326-4338 \text{ cm}^{-1}$ define tendencies for Colombian gasolines. Also, the best performances according to RMSE and R^2 are exhibited by 30:12:1-TanH and 6:12:1-TanH nets for adjusting sulphur content and RVP, respectively. The 30:12:1-TanH net reads the 30 major values of areas, whereas the 6:12:1-TanH requires the first six principal components. The RMSE of the 30:12:1-TanH net is between repeatability and reproducibility for the sulphur content (ASTM D4294-16). With respect to RVP, the RMSE of the 6:12:1-TanH is less than repeatability and reproducibility for the corresponding norm. Comparison of the best performances of RNA trained in the current document with those reported in the open literature supports the application of the 30:12:1-TanH and 6:12:1-TanH nets in the prediction of sulphur content and RVP, respectively.

* Proyecto de grado

** Facultad de Ingenierías Fisicoquímicas Escuela de Ingeniería Química Maestría en Ingeniería Química Director: Giovanni Morales Medina Doctor en Ingeniería Química

Introducción

Los productos de la refinación del petróleo corresponden a un importante renglón de la economía colombiana (OCDE, 2017; UPME, 2015); según MinCIT, 2018, los hidrocarburos aportaron cerca del 31% del PIB industrial en el 2017. Los ingresos por combustibles dependen del cumplimiento de las especificaciones comerciales de los productos finales, así como de la elevación en la producción (ACIP, 2017). La Gerencia Refinería Barrancabermeja, GRB, de Ecopetrol debe generar gasolinas que cumplan con parámetros de calidad como el contenido de azufre (ASTM D4294 o ASTM D2622) y la presión de vapor Reid, RVP (ASTM D4953, ASTM D5191 o ASTM D323). Los valores de los parámetros de calidad para la gasolina y otros combustibles de origen fósil se encuentran especificados en la Resolución 1180 de 21 de junio de 2006 del Ministerio del Ambiente, Vivienda y Desarrollo Territorial y el Ministerio de Minas y Energía de Colombia. Estos valores se expresan en términos de límites máximos y mínimos que buscan la alineación con regulaciones internacionales.

Respecto al contenido de azufre en la gasolina, las repercusiones ambientales por el efecto invernadero han promovido la disminución en el límite máximo estipulado para esta propiedad en las normas (*e.g.* Fenger, 2009); la directiva Euro 3/III (año 2000), para combustibles a gasolina, define el límite superior de especificación para azufre en 150 ppm, mientras que la especificación de gasolina Euro 4/IV (año 2005) limita el máximo en 50 ppm y la Euro 5 (año 2009) lo establece en 10 ppm (*EU fuel quality monitoring* – 2015, EEA Report No 36/2016). A nivel local, aunque la Resolución 1180 de 21 de junio de 2006 establece un límite máximo de

300 ppm en el contenido de azufre, la GRB tiene la meta a corto plazo de producir gasolinas con contenido de azufre según la especificación Euro 4/IV (Fedebio combustible).

Por su parte, el parámetro RVP es una medida del grado de volatilidad de los compuestos livianos en la gasolina; esta propiedad es de importancia para la combustión en motores debido a que este fenómeno se realiza en fase gaseosa. Aunque, una gasolina debe evaporarse con facilidad, un valor de RVP elevado conlleva a un aumento en la emisión de vapores al ambiente (Andersen *et al.*, 2010), con la consecuente contaminación ambiental y la disminución en la economía por pérdida de material; una disminución en 1.0 en el valor de RVP disminuye en un 4.5% las emisiones provenientes de los motores (Reuter *et al.*, 1992). La Resolución 1180 de 21 de junio de 2006 establece un valor de RVP de 8 psia a 100 °F a partir del año 2010; este valor de presión es satisfecha por las gasolinas preparadas en la GRB.

La exigencia en el cumplimiento de estas especificaciones comerciales conduce a la mezcla o *blending* de los diferentes tipos de naftas generadas en una refinería; entre las naftas generadas se encuentran nafta virgen, nafta de craqueo catalítico y nafta de aromáticos. El *blending* en la GRB envuelve un proceso iterativo, en donde se realiza una estimación para las proporciones de cada nafta para el cumplimiento de los parámetros comerciales en la preparación del lote de gasolina comercial en turno. Con la preparación, una muestra es tomada con el fin de validar en laboratorio los parámetros de calidad; los tiempos de medición experimental pueden demorar *ca.* 6 horas, según la prueba y la cantidad de muestras en espera. Ante la carencia de visto bueno por incumplimiento en algún parámetro, el lote de gasolina debe ser reprocesado hasta la consecución de la calidad requerida, lo cual conduce a retrasos en la producción, incumplimientos de entrega, reducción en el margen de refinación y detrimento de la imagen corporativa de Ecopetrol.

Una alternativa ante los tiempos considerables de validación experimental de los parámetros comerciales corresponde a la aplicación de modelos de predicción, a partir de mediciones instrumentales (Tan *et al.*, 2013; Mahji *et al.*, 2012; Chung, 2007; Ferreiro-González *et al.*, 2014). Las técnicas instrumentales aplicadas en la industria generan una vasta cantidad de información, con lo cual su análisis por regresión lineal múltiple puede conllevar a resultados fútiles. En este sentido, las redes neuronales artificiales, RNA, se presentan como una herramienta que facilita la obtención de tendencias y modelos de predicción de propiedades del analito (Blanco *et al.*, 2001; Martínez *et al.*, 2008). El desarrollo de una RNA para la predicción de propiedades basadas en datos instrumentales, como el espectro NIR (*Near InfraRed*), requiere de un conjunto de muestras, que puede comprender más de 50 espectros. Con la RNA desarrollada, la generación del estimado de la propiedad o parámetro de calidad necesita de la medición del espectro NIR del analito problema y unos cuantos segundos para la ejecución del algoritmo; por lo anterior, un modelo de predicción RNA corresponde a una opción atractiva para la reducción en los tiempos de verificación del visto bueno en la preparación de gasolinas.

El presente documento analiza los principales resultados obtenidos durante el desarrollo de modelos de predicción RNA para las propiedades contenido de azufre y RVP a partir de los espectros NIR de las naftas usadas y las gasolinas preparadas en la sección de *blending* de la GRB. Con estos modelos RNA será posible aplicar, en un desarrollo posterior, una herramienta para el verificación del visto bueno en línea con el respectivo hardware de medición de espectros NIR (*e.g.* Chung, 2007; Litani-Barzilai, 1997).

1. Objetivos

1.1 Objetivo General

Desarrollar modelos de predicción de las propiedades contenido de azufre y presión de vapor Reid en las gasolinas generadas por la GRB, a partir de espectros del infrarrojo cercano (NIR) y redes neuronales artificiales.

1.2 Objetivos Específicos

- Analizar mediante el método de componentes principales una base de datos de espectros NIR de gasolinas utilizadas en la sección de *blending* de productos comerciales en la GRB.
- Definir un algoritmo de entrenamiento y validación de modelos RNA de predicción de propiedades de gasolinas a partir de los correspondientes espectros NIR.
- Proponer modelos con desempeños cuantitativos de predicción del contenido de azufre y el RVP de gasolinas, a partir de la aplicación del protocolo de desarrollo de modelos RNA utilizando la base de datos de espectros NIR.

2. Marco Teórico

2.1 Preparación de Gasolina Comercial

La gasolina se encuentra constituida principalmente por una mezcla de hidrocarburos de las clases olefinas, parafinas, naftenos y aromáticos; el rango de número de carbonos para estas clases se encuentra entre C_4 y C_{12} , con lo cual el intervalo de destilación se encuentra entre 30 y 220°C (Pradelle *et al.*, 2015). Debido a la variabilidad en su composición y en sus propiedades, las gasolinas de venta comercialmente deben cumplir con parámetros de calidad exigidos según la normatividad; algunos de estos parámetros son presentados en la Tabla 1.1. Según esta tabla, el requisito de contenido de azufre debe ser inferior a 300 ppm (0.03 %wgt) según la prueba ASTM D4294; asimismo, la presión de vapor Reid se estipula en un máximo de 8.0 psia según el método ASTM D4953. El límite máximo para el azufre se debe a que su presencia en las gasolinas conduce a la generación de compuestos corrosivos, capas de ensuciamiento en los motores y emisiones perjudiciales para el ambiente (Gary *et al.*, 2007; Fenger, 2009). Por su parte, el límite para el RVP corresponde al balance entre evaporación de la gasolina y la consecución de la ignición del motor en frío (Andersen *et al.*, 2010).

Tabla 1.

Algunos requisitos de calidad de la gasolina regular y extra.

Característica	Unidad	Especificación	Métodos ASTM
Índice Antidetonante, mínimo	----		D2699 y D2700
Gasolina corriente		84	
Gasolina extra		89	
RVP, máximo	kPa (psi)	55 (8.0)	D4953
Azufre, máximo	% wgt	0.03	D4294 o D2622

Nota. Tomado de: Resolución 1180 de 2006.

En la GRB, la producción de gasolina diaria es *ca.* 30%vol de la carga de crudo alimentada a la Refinería. La GRB comercializa dos tipos de gasolina, la gasolina motor regular (GMR) y la gasolina motor extra (GME), que se obtiene luego del proceso de mezclado (*blending*) de varios tipos de naftas o gasolinas no comerciales con diferentes propiedades (Figura 1). Los componentes usuales de preparación utilizados son nafta virgen, nafta craqueada, nafta hidrotratada, nafta de aromáticos y alquilato; la nafta virgen es generada en la destilación primaria del crudo, mientras que la nafta craqueada resulta del proceso de craqueo catalítico de fracciones pesadas del petróleo. Debido a impurezas como el azufre, una parte de la nafta producida es procesada en las unidades de hidrotratamiento, conduciendo a una nafta de calidad superior o comparable con las exigencias comerciales. Por otra parte, la nafta de aromáticos se obtiene de la mezcla de varias corrientes con contenidos de benceno, tolueno, xileno y orthoxileno, mientras que el alquilato o nafta de alquilación se encuentra constituida por

hidrocarburos ramificados de elevado número de octano; esta última nafta es utilizada para preparación de la GME.

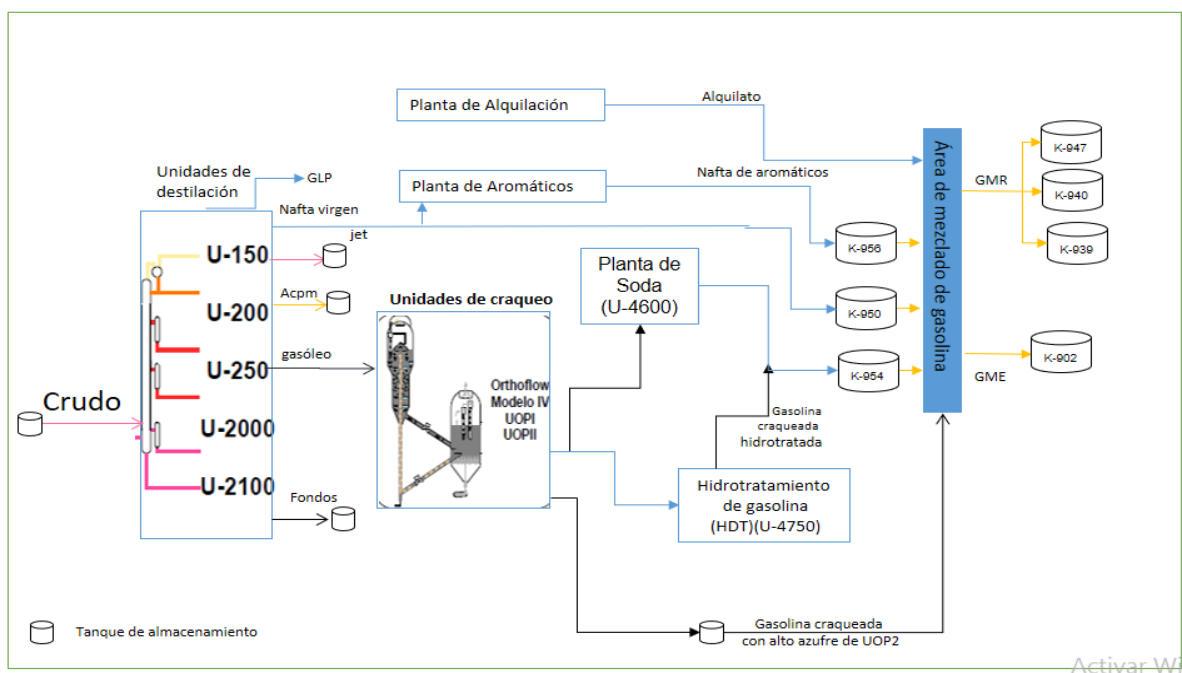


Figura 1. Esquema ilustrativo de los flujos de naftas y el *blending* de gasolinas en la GRB.

Nota. Tomado de: Ecopetrol, GRB.

2.2 Contenido de Azufre en las gasolinas

El azufre se encuentra presente en las gasolinas y otras fracciones livianas del petróleo en compuestos como mercaptanos, sulfuros, disulfuros y compuestos heterociclos (Breitkreitz, 2003). Estos compuestos generan dióxido de azufre en los motores de combustión interna, contribuyendo con la contaminación y con la frecuencia de precipitaciones de lluvia ácida (Breitkreitz, 2003; Han *et al.*, 2018). De igual manera, la formación de ácido sulfuroso y sulfúrico por contacto entre la humedad y los óxidos de azufre pueden ocasionar desgastes

importantes al interior de los motores. Por su parte, los compuestos heterociclos contenidos en los gases de combustión pueden contribuir con la aparición de enfermedades de origen cancerígeno y mutaciones (Han *et al.*, 2018).

Debido a su peligrosidad, los contenidos de compuestos azufrados en la gasolina son regulados a nivel internacional. La agencia de protección ambiental de Estados Unidos (EPA) reguló en 2017 el contenido de azufre en las gasolinas a un límite superior de 10 ppm; esta medida ya había sido aplicada 9 años atrás en Europa con la regulación denominada Euro V.

El protocolo consignado en la norma ASTM D4294 (2006) conlleva a la determinación del contenido de azufre en combustibles derivados del petróleo; la reproducibilidad reportada por esta norma corresponde a 0.05% o 500 ppm. En la norma se utiliza la espectrometría de fluorescencia de rayos X como método de referencia para la cuantificación de azufre; la medición se realiza al vacío y aunque los valores de desviación estándar de las mediciones son bajos, los costos del equipo y su mantenimiento restringen su aplicación directa en diversos laboratorios (Rocha *et al.*, 2016).

2.3 Vapor de Presión Reid (RVP)

Por norma, el RVP se define como la presión total de 1 mL de gasolina con aire saturado a 37.8 °F en una cámara de volumen 5 mL. La muestra es enfriada entre 0 y 1 °C y subsecuentemente, inyectada a la cámara. El incremento en la presión es medido por un sensor hasta que la temperatura llega a 37.8 °C; el RVP puede ser reportado en kPa o en psi. El desempeño de los motores depende de un adecuado valor de esta presión (Mendes *et al.*, 2017). Una gasolina con alto valor de RVP conduce a una gran generación de vapores de gasolina, lo cual puede inhibir la

combustión por un bajo valor de la relación aire/vapores; por su parte, un bajo valor de RVP puede también inhibir la combustión por una vaporización baja del combustible.

De otro lado, los vapores de gasolina pueden contener compuestos aromáticos con alta toxicidad para la salud humana (Page & Nehlman, 1989). Asimismo, el RVP provee información sobre el riesgo de fuego durante el transporte y la manipulación de la gasolina. Por otra parte, Takeshita *et al.*, 2008, apuntan sobre la medición del RVP en la evaluación de no conformidades por adulteración de las gasolinas en las estaciones de servicio; una gasolina adulterada presenta un menor valor de RVP.

2.4 Espectroscopía del Infrarrojo Cercano (NIR)

El infrarrojo (IR, *InfraRed*) corresponde a la región del espectro electromagnético comprendida entre 760 y 10^6 nm (entre 12800 y 10 cm^{-1} , en número de onda); esta región se encuentra ubicada después del espectro visible, como es presentado en la Figura 2. En la Tabla 1.2 se desglosan las subregiones que conforman el espectro IR; también, esta tabla describe los movimientos moleculares asociados con cada subregión; particularmente, la subregión NIR sobretonos y combinaciones entre vibración y rotación molecular (Skoog *et al.*, 2008). A pesar de los sobretonos y las combinaciones, la identificación de algunos grupos funcionales puede ser efectuada en la región NIR; la Tabla 1.3 identifica grupos funcionales, que pueden estar en las fracciones del petróleo, con sus respectivas bandas de absorción. Un espectro NIR típico para la gasolina es presentado en la Figura 3; en esta figura se aprecian las principales bandas aportantes, las cuales se encuentran entre las regiones $5000 - 4000\text{ cm}^{-1}$ y 6000 y 5500 cm^{-1} ; estas bandas se

encuentran asociadas a combinaciones entre las bandas vibracionales de los enlaces C-H (Falla *et al.*, 2006).



Figura 2. Diagrama del espectro electromagnético.

Nota. Tomado de: Atlas of Science, 2016.

Tabla 2.

Regiones de absorción en el infrarrojo.

Región Infrarroja	Longitud de onda	Número de onda, (cm ⁻¹)	Transición característica
Cercana (NIR)	780 a 2500 nm	12800 a 4000	Sobretonos y combinaciones
Media (MIR)	2,5 a 50 μm	4000 a 200	Vibraciones fundamentales
Lejana (FIR)	50 a 1000 μm	200 a 10	Rotaciones

Nota. Tomado de: García, 2014.

Tabla 3.

Frecuencias de absorción de algunos grupos funcionales en hidrocarburos.

Grupo funcional	Frecuencia (cm ⁻¹)
Acetilenos	9800 – 9430
	6580 – 6400
Alcoholes	7140 – 7010
Aldehídos	4640 – 4520
Alifáticos	8000
Aromáticos	4525
	4445
Formaldehído	4775 – 4630

Nota. Tomado de: Patnaik, 2004.

2.5 Modelos de predicción basados en espectros IR

La construcción de un modelo matemático de predicción, basado en mediciones espectrales de un conjunto de muestras de interés puede ser guiada por la norma ASTM E-1655 (2017). Según esta norma las muestras utilizadas para el desarrollo del modelo de calibración deben ser representativas de la variabilidad que pueda darse durante el proceso para obtener un modelo robusto. Respecto a la propiedad que se desea predecir, el método analítico de medición experimental debe ser confiable y preciso para tener un buen resultado en el modelo.

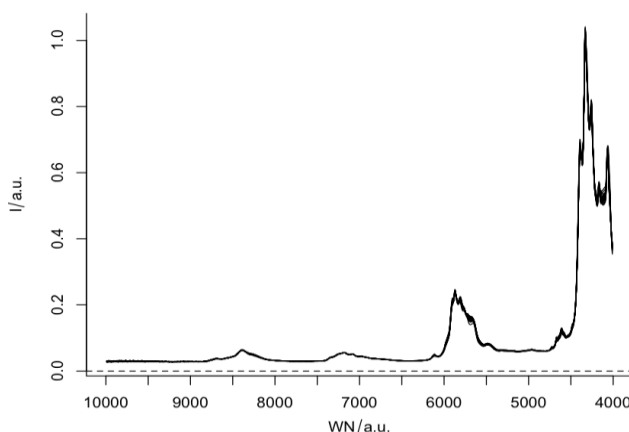


Figura 3. Ejemplo de espectros de absorción NIR para de gasolinas.

Por otra parte, la norma ASTM E-1655 recomienda aplicar una serie de pretratamientos a los espectros, con lo cual se puede minimizar errores de medición, ruidos y contribuciones espectrales no deseadas. La Figura 4 ilustra diferentes tipos de pretratamientos que pueden ser aplicados a las señales espectrales. Dentro de estos se tienen suavizado espectral, el cual que busca filtrar matemáticamente el ruido aleatorio, y las derivadas, las cuales reducen el solapamiento de las señales y los desplazamientos de la línea base. Asimismo, la aplicación de la variable normal estándar (SNV, *Standard Normal Variate transformation*) elimina los efectos de dispersión (*scattering*) del haz de luz, que se pueden confundir con los cambios de concentración. Otra herramienta matemática corresponde a la reducción en la dimensionalidad del espectro utilizando el análisis de componentes principales (PCA); con PCA la variabilidad del espectro inherente a los compuestos es concentrada en un número reducido de nuevas variables constituidas por combinaciones lineales de números de onda de ciertas regiones del espectro.



Figura 4. Pretratamientos utilizados en análisis de espectros IR.

Nota. Tomado de: García, 2014.

Después de la aplicación de pretratamientos, el modelo de predicción puede ser desarrollado o calibrado con base en técnicas de regresión lineal, mínimos cuadrados parciales o redes neuronales artificiales. La capacidad predictiva del modelo o validación puede efectuarse por medio de la técnica de validación interna o cruzada, la cual utiliza las mismas muestras de generación del modelo. El modelo final a proponer debe exhibir una capacidad predictiva alta, cercana al desempeño mostrado con la calibración del modelo.

2.6 Redes Neuronales Artificiales

La teoría y el diseño de redes neuronales artificiales (RNA) están basados en la estructura y funcionamiento de los sistemas nerviosos, en donde la neurona es la unidad más simple o elemental de una red. Una red neuronal está organizada en función de (Figura 5): número de niveles o capas, número de neuronas por nivel, patrones de conexión y flujo de información. El arreglo o distribución de las neuronas dentro de la red se realiza formando niveles o capas de un número determinado de neuronas. Partiendo de su ubicación dentro de la red se pueden distinguir tres tipos de capas diferentes:

- **Capa de entrada:** Es la capa expuesta que recibe la información que proviene del entorno o fuentes externas a la red.
- **Capa oculta:** Pertenecen al conjunto de niveles formados por neuronas que no tienen contacto directo con el entorno exterior, el número de estas capas puede variar entre cero y un número considerable determinando así las distintas tipologías de las redes neuronales (Isazi & Galván, 2004; Del Brío & Sanz, 2002).

- **Capa de salida:** Es la capa encargada de realizar la transferencia de información desde la RNA hacia los dispositivos asociados a esta.

En la funcionalidad, cuando determinada neurona recibe una cantidad de información de entrada, cada señal de entrada es multiplicada por el peso asociado a cada conexión. Posteriormente, las entradas ponderadas con los pesos (W) son sumadas y transformadas por la función de activación para generar la señal de salida (Figura 5). Estas funciones de activación deben ser continuas y derivables en todos los puntos; en la Figura 6 se exhiben las funciones más utilizadas.

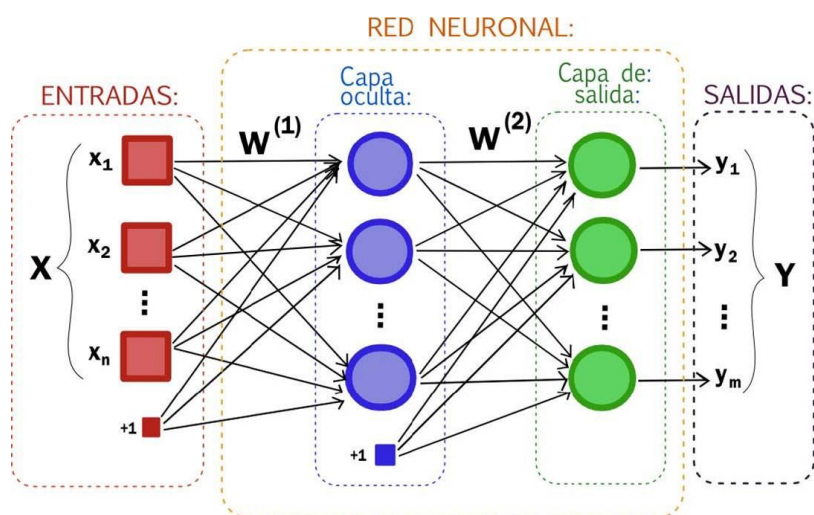


Figura 5. Ilustración arquitectura y parámetros que definen una RNA.

Fuente: Hilera, 2000.

Por otra parte, el modelo de RNA se obtiene mediante entrenamiento o aprendizaje en una estructura de prueba que conlleve a la reproducción de datos experimentales (aprendizaje supervisado); el aprendizaje consiste en aplicar un algoritmo que defina los pesos que afectan las señales de conexión que transfieren los conocimientos. El algoritmo de entrenamiento

denominado *backpropagation* es el más utilizado, ya que presenta información sobre la sensibilidad de las predicciones de la red respecto a los pesos de las señales; el algoritmo actualiza los pesos con base en el gradiente de una función de costo –que indica que tan eficiente es la predicción, por ejemplo, el error cuadrático medio–. Las ecuaciones derivadas del algoritmo *backpropagation* corresponden a: una ecuación de error en la capa de entrada, una ecuación de error en términos de los errores de la capa anterior y dos ecuaciones diferenciales para el cambio de la función de costo –respecto a los pesos y a los sesgos– (Isazi & Galván, 2004; Del Brío & Sanz, 2002). Las funciones son resueltas durante la minimización de la función de costo, por medio de un método como el de *Levenberg-Marquardt*.

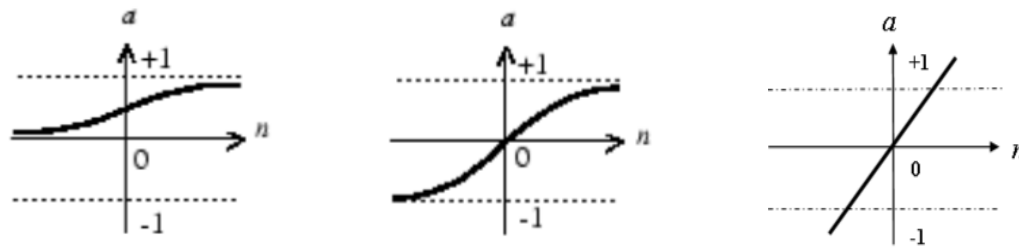


Figura 6. Funciones de activación habitualmente utilizadas en RNA.

Nota. Tomado de: Del Brío & Sanz, (2002).

El algoritmo de aprendizaje puede utilizar como función de error la suma del error cuadrático medio (MSE: *mean squared error*), según la expresión,

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2 \quad \begin{matrix} a=\text{logsig}(n) \\ a=\text{tanh}(n) \\ a=\text{purelin}(n) \end{matrix} \quad (1)$$

Donde, y_i , $\tilde{y}_i$, n corresponden al valor experimental, al valor predicho y al número de datos, respectivamente. Asimismo, un método de regularización aplica y detiene el entrenamiento de las RNA en el punto mínimo de la curva de aprendizaje (*early stopping*) para evitar el sobreajuste

(*overfitting*); en el programa de uso libre R, la función *neuralnet* aplica este método de regularización (Zhang, 2016). Para el entrenamiento y la validación de las redes, usualmente los datos son divididos en dos conjuntos; el 70% de los datos se destina al entrenamiento y el 30% restante a la validación. Los desempeños de las estructuras de red se determinan por medio de los estadísticos $RMSE$ y R^2 , definidos como,

$$RMSE = \sqrt{\frac{\sum (y_i - \tilde{y}_i)^2}{n}} \quad (2)$$

$$R^2 = 1 - \frac{\sum (y_i - \tilde{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

Estos estadísticos son determinados en las etapas de entrenamiento y validación y para el total de estas dos. Los valores $RMSE$ y R^2 reportados corresponden al promedio calculado sobre la ejecución de la arquitectura de red un determinado número de veces (Swingler, 1996).

2.7 Estado del Arte

Los desarrollos en los modelos de predicción de las propiedades contenido de azufre y RVP de gasolinas de origen fósil basadas en espectros IR han sido direccionados en los siguientes tópicos:

- **Aplicación de técnicas multivariadas:** Reboucas *et al.*, 2011, reportaron la combinación entre el diseño experimental Doehlert y el método PLS para la generación de modelos de predicción de propiedades de la gasolina. El diseño consideró los factores: rango de números de onda, pretratamientos y técnicas de calibración y validación. Según los autores, el diseño de Doehlert conduce a la obtención de parámetros adecuados para la predicción de las propiedades de las gasolinas.

- **Aplicación de RNA:** Martínez *et al.*, 2008, aplicaron el modelado por redes neuronales artificiales para la predicción de propiedades de gasolinas. Estos autores utilizaron los componentes principales derivados del análisis PCA de los espectros NIR de gasolinas; estos componentes constituyeron las variables de entrada a las respectivas neuronas. Según los autores, los modelos RNA presentan desempeños comparables con los métodos de predicción estadística multivariables.
- **Selección de regiones espectrales:** Lee *et al.*, 2006, seleccionaron rangos espectrales óptimos en el espectro NIR para la predicción de las propiedades RON y RVP en gasolinas. Estos autores aplicaron el método de estadística multivariada denominado MW PLS, con lo cual exploraron diferentes rangos del espectro. Según Lee *et al.*, el rango espectral entre 4550 y 4040 cm^{-1} conlleva a la generación de los mejores modelos de regresión para estas propiedades.
- **Proposición de hardware de medición:** Brouillette *et al.*, 2016, desarrollaron una analizador IR portable. Según los autores, el analizador reporta el espectro en 5 s, en el rango entre 5–6 nm de resolución; con esto las predicciones realizadas basadas en el analizador reportan valores que se ajustan a las repetibilidades reportadas en las respectivas normas.

Según la revisión de literatura, la predicción de las propiedades contenido de azufre y RVP de gasolinas reporta factibilidad por medio de la medición espectral correspondiente y la aplicación de redes neuronales artificiales. Las entradas a las redes pueden corresponder a los componentes principales derivados de una análisis PCA. Reportes para otras propiedades de las gasolinas y de propiedades de fracciones del petróleo utilizan los valores de las segundas

derivadas o de sus áreas bajo la curva como entradas a las RNA (Li *et al.*, 2005; Han *et al.*, 2011).

3. Metodología

3.1 Recopilación de espectros NIR y propiedades de las gasolinas.

Diferentes muestras de naftas y gasolinas fueron recolectadas y sus espectros medidos para formar la base de datos NIR. El espectro NIR fue determinado en la región espectral que comprende las longitudes de onda entre 1000 y 2500 nm (en números de onda entre 10000 y 4000 cm^{-1} , respectivamente) con un barrido cada 1.21 nm –1.93 cm^{-1} para 3102 datos totales de absorbancias–, utilizando el equipo ABB modelo *FPLA2000* disponible en el laboratorio de la Refinería de Barrancabermeja. La medición de los espectros IR fue guiada por la norma EN 14078, la cual incluye la selección de condiciones experimentales óptimas de adquisición espectral y tratamientos previos de acondicionamiento de las muestras. Asimismo, el contenido de azufre y la presión de vapor Reid (RVP) fueron medidas según los protocolos consignados en las normas ASTM D4294 y ASTM D4953, respectivamente.

3.2 Pretratamientos y análisis PCA de la base de espectros NIR

Los espectros NIR fueron pretratados, con lo cual se disminuyó la variabilidad en las mediciones del equipo, a la vez que se incrementó la relación señal/ruido (Rinnan, 2014). Los pretratamientos involucraron: suavizado según el procedimiento de Savitzky-Golay, primera y segunda derivadas por medio de la aproximación de Savitzky-Golay, y normalización SNV. Los pretratamientos fueron aplicados según los códigos de las funciones *savitzkyGolay* y *standardNormalVariate* del paquete *prospectr* versión 0.1.3 codificado para su aplicación en el programa de uso libre denominado R versión 3.4.3 (<https://www.r-project.org>) –R es un lenguaje y un ambiente para el desarrollo de cálculos y gráficas, que se encuentra diseñado sobre el sistema operativo GNU (<http://www.gnu.org>); R fue conceptualizado bajo la filosofía de desarrollo de librerías y funciones que pudieran ser compartidas por diferentes usuarios a nivel mundial; una gran variedad de procedimientos de análisis y procesamiento de datos pueden ser ejecutados en R, entre los que se tienen, análisis estadísticos univariados y multivariados, regresiones lineales, regresiones logísticas, análisis PCA, regresiones PLS, regresiones PCR, ajustes por RNA y análisis de agrupaciones por *kmean*, entre otros—.

Asimismo, una matriz en donde las filas representan las muestras de gasolina y las columnas los aportes en cada número de onda fue definida para la aplicación del análisis por componentes principales (PCA). Con este análisis la varianza de los espectros NIR fue concentrada en un número reducido de variables denominadas componentes principales. Estas variables son obtenidas por medio de la descomposición de los espectros en *scores* y *loadings*, los primeros representando a las muestras, mientras que los segundos representando a los valores de intensidad pretratada en los correspondientes números de onda (ver Wold *et al.*, 1987). Muestras

atípicas y regiones espectrales con contribuciones importantes fueron determinadas por medio de los *scores* y los *loadings* de los componentes principales con mayor varianza. Por otro lado, el algoritmo PCA aplicado correspondió al codificado en la función *prcomp* de R.

3.3 Proposición de líneas de código para entrenamiento y validación de modelos RNA.

En esta etapa, las líneas de código requeridas para la aplicación de la función *neuralnet* del paquete *Neuralnet* (Zhang, 2016; Berthold & Hann, 2007; Blanco et al., 2000) fueron desarrolladas en el programa R. La función *neuralnet* cuenta con las siguientes entradas, *neuralnet(f, data, hidden, linear.output, err.fct, act.fct, stepmax, threshold, algorithm, rep)*

Donde *f* corresponde a la definición de las entradas y las salidas, mientras *data* contiene la matriz de información del sistema. Asimismo, *hidden* define el número de capas y neuronas internas, mientras que *linear.output*, *err.fct* y *act.fct* establecen la función de activación de la capa de salida, la función de error en entrenamiento y validación y, la función de activación de las neuronas internas, respectivamente. Del mismo modo, *stepmax*, *threshold*, *algorithm* y *rep* definen los parámetros de ejecución de la función.

El código desarrollado consideró la variación en los componentes principales o en el área bajo la curva de la segunda derivada, el número de neuronas internas y las funciones de activación disponibles; las nomenclaturas definidas para las diferentes arquitecturas fueron “y:x:1-Log” y “y:x:1-TanH”, para las cuales y y x representan a las variaciones en el número de entradas y en el número de neuronas de la capa interna, respectivamente.

3.4 Aplicación del algoritmo para modelos RNA con componentes principales

En esta etapa, el código desarrollado fue ejecutado en el entrenamiento y la validación de diferentes estructuras de red, considerando los componentes principales como las respectivas entradas; en la nomenclatura “y:x:1–”, los componentes principales correspondieron a y. La ejecución de este código enmarcó el número de neuronas ocultas, la variación del número de capas y las funciones de activación, así como la repetición de 20 veces para cada estructura según lo recomendado en la literatura (Piloto-Rodríguez *et al.*, 2013). El número de neuronas ocultas fue modificado desde 1 hasta el doble del número de componentes principales en cada estructura, de acuerdo a lo recomendado por Berry & Linoff, 2000; con esto, en la nomenclatura, x varió desde 1 hasta 2y. Los desempeños fueron evaluados en términos de los estadísticos RMSE y R^2 .

3.5 Entrenamiento y validación de las RNA con el área en las absorbancias pretratadas

El código desarrollado para el entrenamiento y la validación de diferentes estructuras de red fue ejecutado utilizando las áreas bajo la curva en los números de onda según los pretratamientos definidos en la etapa 2.2. Las áreas bajo la curva fueron obtenidas cada 20 cm^{-1} y posteriormente ordenadas decrecientemente; en la nomenclatura “y:x:1–”, las áreas correspondieron a y. El número de neuronas de la capa oculta fue modificado desde 1 hasta el doble del número de regiones con contribuciones importantes, según las áreas (Berry & Linoff, 2000); según la nomenclatura, x tomó valores desde 1 hasta 2y. La ejecución del código consideró además la variación del número de capas ocultas y las funciones de activación, así como la repetición de 20

veces para cada estructura según lo recomendado en la literatura (Piloto-Rodríguez *et al.*, 2013). Los desempeños fueron evaluados en términos de los estadísticos RMSE y R^2 .

3.6. Selección de los modelos RNA.

Los desempeños obtenidos con la aplicación de las anteriores dos etapas fueron analizados según el número de capas, el número de neuronas y los desempeños reportados por cada estructura de red según los estadísticos RMSE y R^2 .

4. Análisis de resultados

4.1 Espectros NIR y propiedades de las naftas de la GRB.

Las muestras tomadas de diferentes puntos en el sistema de producción de naftas de la GRB fueron enviadas al laboratorio; las muestras también son diferentes en la fecha tomada. En total, 80 espectros IR fueron medidos; asimismo, las propiedades azufre total y RVP fueron determinadas para cada muestra. La Figura 7 1a muestra los espectros NIR de las muestras de gasolina en el intervalo de números de onda entre 10000 y 4000 cm^{-1} . Con esta figura es posible identificar diferencias en las líneas base ocasionadas principalmente por cambios en el instrumento –los espectros fueron obtenidos en diferentes fechas en una ventana de tiempo de 6 meses–. Asimismo, según en esta figura se aprecian espectros con diferentes tendencias. Las

Figuras 7 b-c detallan los espectros en los rangos de mayor contribución, *i.e.* entre los intervalos $4800 - 4000 \text{ cm}^{-1}$ y $6200 - 5400 \text{ cm}^{-1}$, respectivamente. En estas figuras es posible distinguir dos espectros que se diferencian del comportamiento del colectivo; el primer espectro corresponde a la nafta de alta calidad importada por la GRB y almacenada en el tanque K-942; la muestra de esta gasolina fue tomada el día 13 de mayo de 2018. El segundo espectro corresponde a una nafta craqueada de la GRB con una cantidad importante de compuestos insaturados (Sandoval, 2017).

En términos globales, los espectros poseen regiones de absorbancias con igual tendencia. En las regiones con números de onda mayores a 6200 cm^{-1} y en $5400-4800 \text{ cm}^{-1}$ (Figura 7a), los espectros exhiben absorbancias con valores inferiores de un 5% de la intensidad máxima de la señal; en estas regiones se presentan los segundos sobretonos de las bandas de estiramiento del grupo C-H (Reboucas & Neto, 2001; Kardamakis & Pasadakis, 2010). Por su parte, en la región entre $6200-5400 \text{ cm}^{-1}$ (Figura 7b) las absorbancias son mayores que en la primera región, exponiendo la contribución de los primeros sobretonos de las bandas de estiramiento del grupo C-H. También, las absorbancias son apreciables en la región entre los números de onda $4800-4000 \text{ cm}^{-1}$ (Figura 7c), región característica de combinaciones de las bandas de estiramiento para el grupo funcional C-H (Reboucas & Neto, 2001; Kardamakis & Pasadakis, 2010; Reboucas *et al.*, 2011).

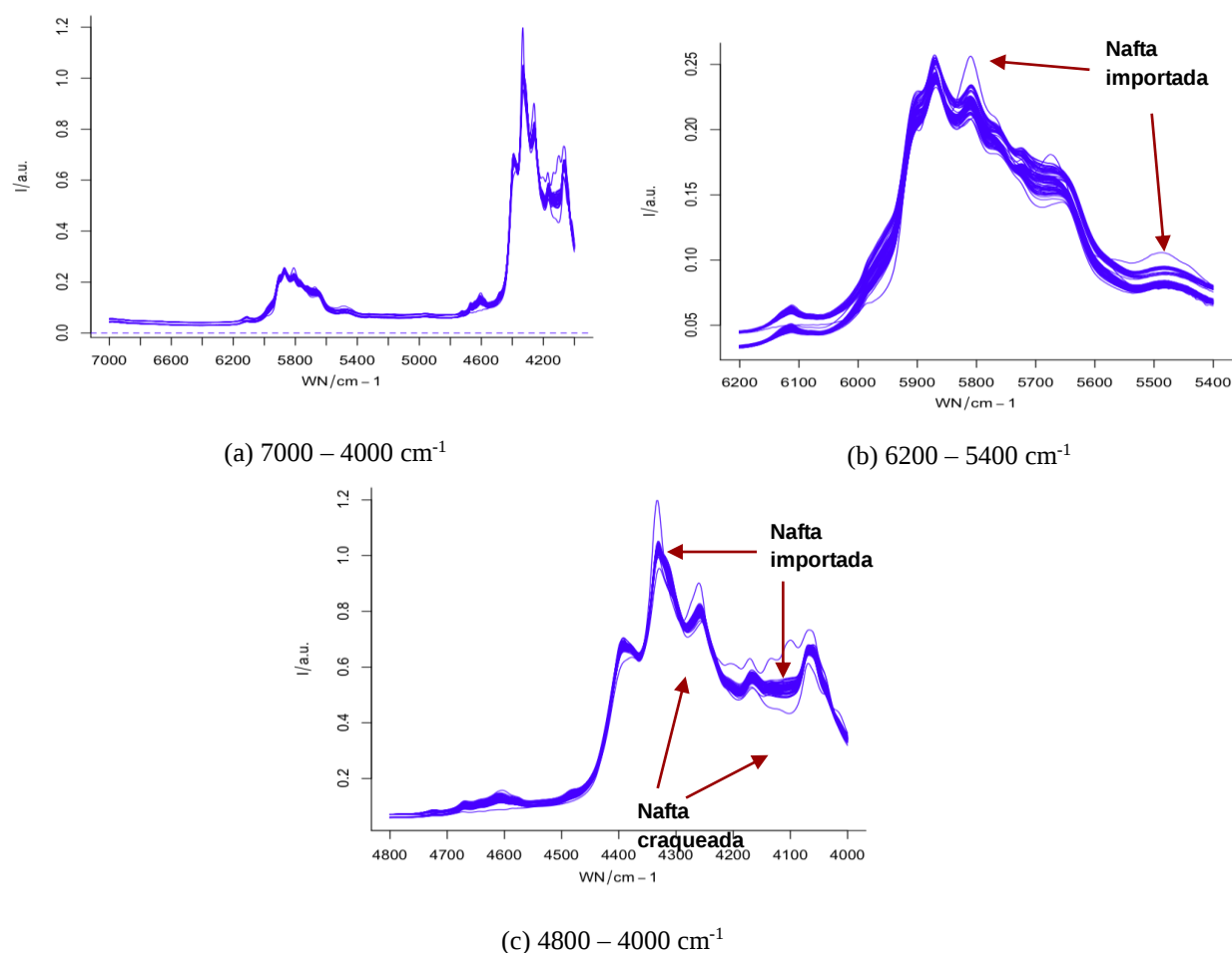


Figura 7. Detalle de los espectros en diferentes rangos de números de onda.

La Figura 8 presenta los pretratamientos realizados a los espectros NIR; el pretratamiento de suavizado fue aplicado como requisito para la continuidad de los espectros, previo a la aplicación de las derivadas y la transformación SNV. En las Figuras 8a-b se pueden detallar los pretratamientos de primera y segunda derivadas. Según estas figuras, las variaciones de las señales espectroscópicas en los rangos 6200–5400 cm⁻¹ y 4800–4000 cm⁻¹, aparecen con mayor detalle con la aplicación de la segunda derivada; el número de puntos críticos generados con la aplicación de la segunda derivada (Figuras 8b), aumenta a más del doble de los obtenidos con la primera derivada (Figuras 8a). Los valores de la señal pretratada con la primera y segunda

derivadas disminuyen respecto a los valores de la señal original; la aplicación de la transformación SNV, mostrada en la Figura 8c, normaliza la señal con la desviación estándar de los espectros. Con esto, la señal pretratada aumenta exhibiendo valores en el intervalo $[-18; 8]$ (Figuras 8d-e), con una distribución equitativa para cada espectro, representando la variación contenida en los espectros originales. Esta transformación también acentúa las diferencias en los espectros de la nafta importada y de la nafta craqueada, lo cual puede ser verificado en las Figuras 8d-e; las diferencias en las tendencias de estas naftas respecto a las restantes muestras son más acentuadas en el rango $4800\text{--}4000\text{ cm}^{-1}$, sugiriendo su relación con las bandas C-H de los compuestos aromáticos y tipo alqueno.

Por otra parte, las propiedades medidas para las muestras de nafta recolectadas pueden ser analizadas en la Figura 9. Es importante mencionar que el número de muestras con propiedades contenido de Azufre y RVP corresponden a 80 y 73, respectivamente; con esto 7 muestras carentes de la propiedad RVP fueron omitidas en el respectivo entrenamiento. En las Figuras 9a-c,e se detallan las dispersiones para las propiedades Aromáticos en volumen, contenido de Azufre, RVP; en la Figura 9b, además, se presenta un gráfico de dispersión considerando las propiedades Aromáticos, Azufre y número de octano de investigación (RON). Con las anteriores figuras se aprecia una variabilidad en las propiedades mencionadas, representando los diferentes tipos de corrientes utilizadas en la sección de mezcla de gasolinas comerciales (*blending* de gasolinas). La posición relativa de la nafta importada es detallada en estas figuras, con lo cual es posible afirmar que corresponde a una corriente con especificaciones de calidad según la norma internacional Euro 4/IV. Esta nafta es utilizada en la GRB como flujo corrector para alcanzar los niveles especificados en la Resolución 1180 de 21 de junio de 2006.

Referente al contenido de Azufre y al RVP, los histogramas de las Figuras 9d,f muestran que una gran cantidad de naftas exhiben valores para estas propiedades en los intervalos entre 100–250 ppm y 7.4–8.0 psi, respectivamente. Estos valores son característicos de las naftas producidas en las unidades de destilación de crudos de la GRB; las naftas con contenidos menores de Azufre son generadas en la unidad de hidrotratamiento. Los valores poco frecuentes en los histogramas representan la variabilidad de los crudos procesados y los eventos operacionales acontecidos durante el periodo de muestreo.

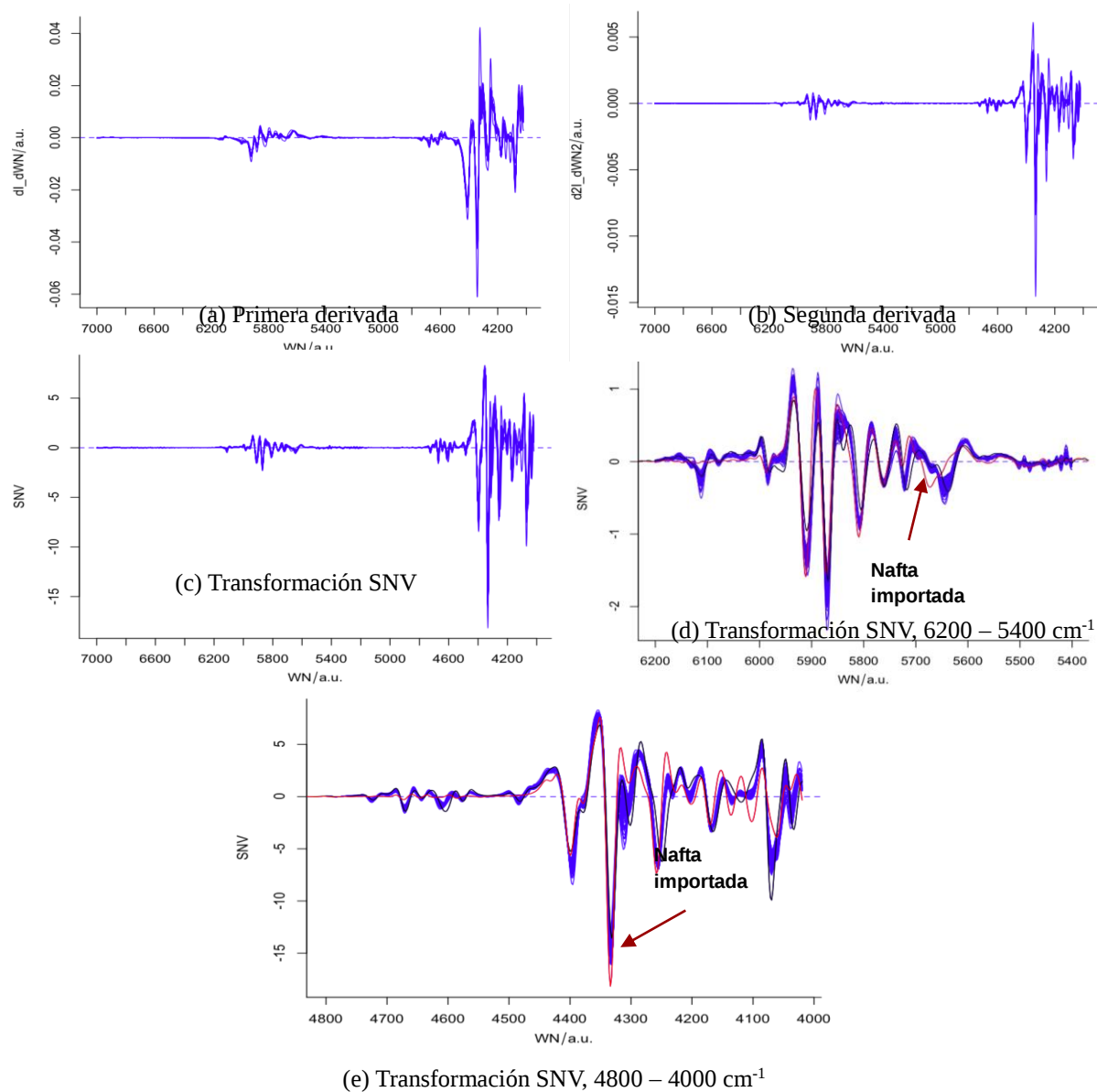


Figura 8. Modificación de los espectros con los pretratamientos.

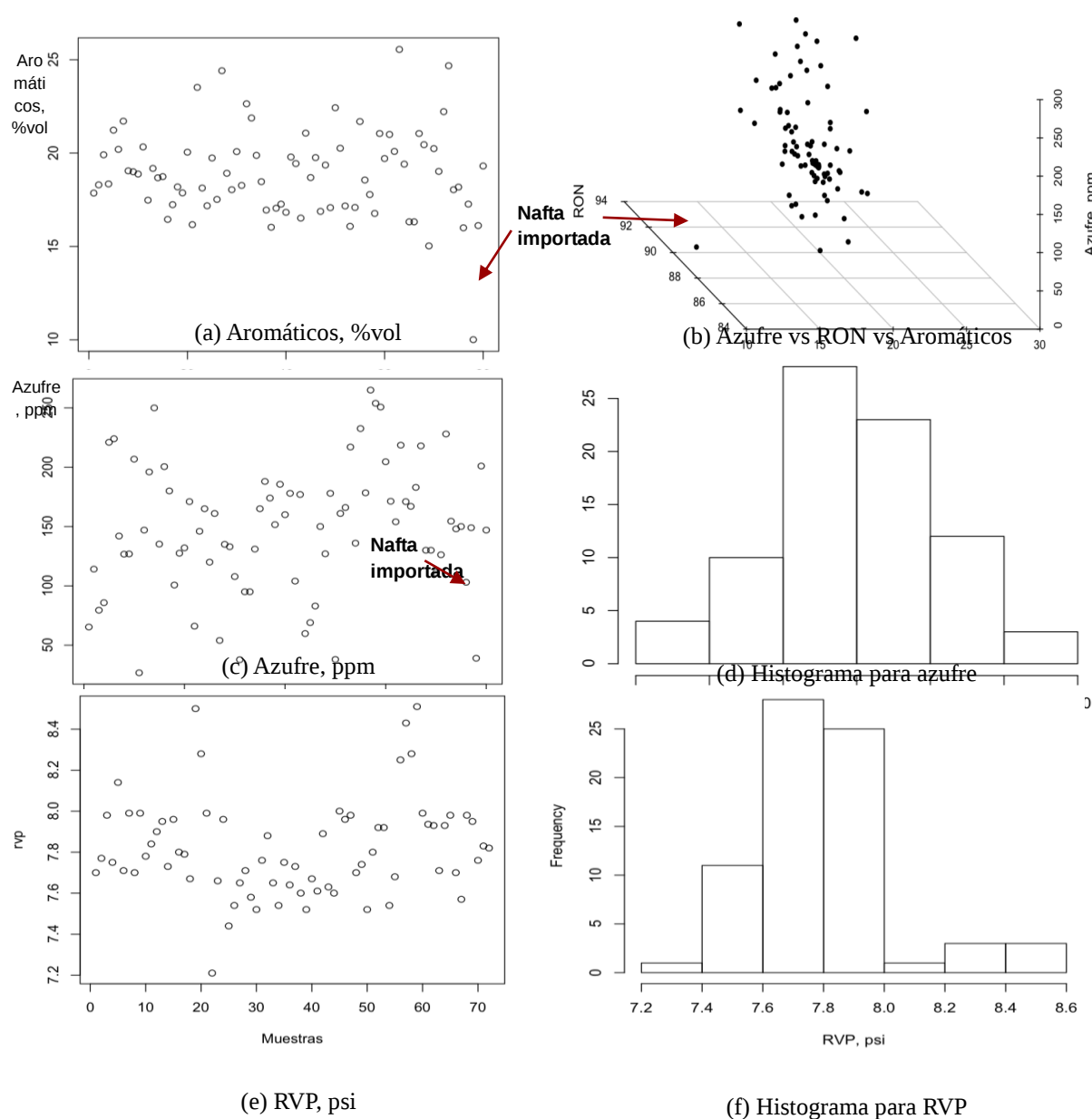


Figura 9. Propiedades de las muestras de naftas y gasolinas.

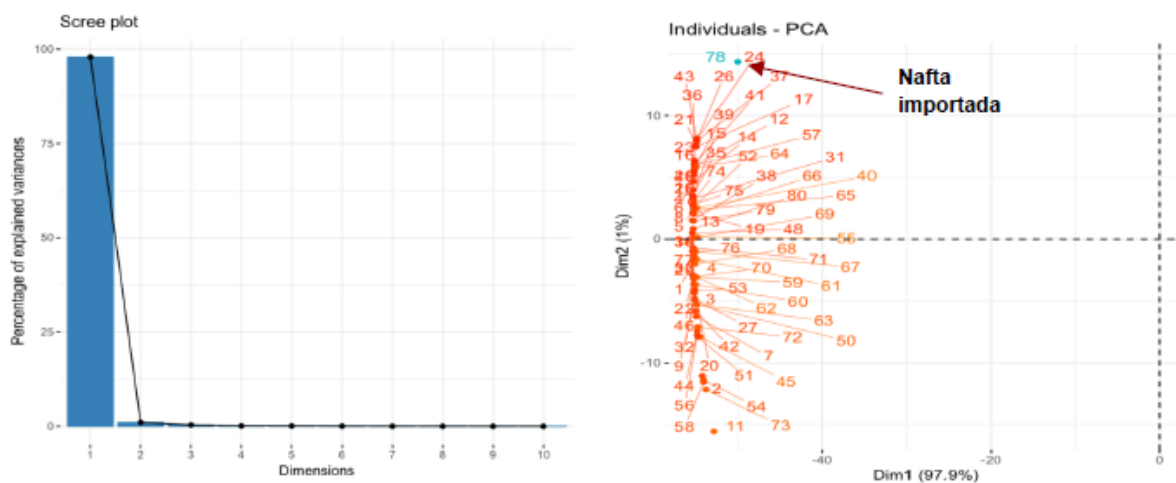
4.2 Análisis PCA de la base de espectros NIR

La base de datos de espectros pretratados fue dispuesta en forma matricial para el análisis PCA; los 80 crudos con propiedades contenido de Azufre fueron considerados en la matriz de análisis.

La Figura 10 presenta los resultados de este análisis con las gráficas de la varianza explicada (Figura 3.4a) y, las gráficas de los *scores* (Figura 10b) y de los *loadings* (Figura 10c) para los dos primeros componentes principales (98.9% varianza). Según la figura de la varianza (Figura 10a), el primer componente principal, PC1, explica la mayor parte de la variabilidad exhibida por los espectros pretratados (97.9%); el PC2 contribuye con sólo el 1% de la varianza. La menor proporción del PC2 influye en la formación de dos grupos de muestras, un grupo con valores positivos y el otro con valores negativos para este componente; los grupos pueden ser analizados en el gráfico de los *scores* (Figura 10b). En este gráfico, el número de muestras se reparte entre los valores positivos y negativos del PC2, según su valor en la segunda derivada del espectro (Figura 10). El gráfico de los *loadings* (Figura 10c) permite identificar la región espectral que tiene la mayor influencia en la agrupación según el PC2. El valor de la segunda derivada en los números de onda entre 4309–4316 cm^{-1} (*loadings* con mayor valor positivo de PC2, Figura 10c) define los espectros con valores positivos en el PC2 de los *scores*, mientras que el valor de la segunda derivada en los números de onda entre 4326–4338 cm^{-1} (*loadings* con menor valor negativo de PC2, Figura 10c) define los espectros con valores negativos en el PC2 de los *scores*; en la Figura 10b se detalla la nafta importada, cuyo espectro presenta el mayor valor de los *loadings* respectivos (las muestras fueron reenumeradas por omisión de las muestras sin la propiedad contenido de Azufre). Con lo anterior, es de esperarse que estas regiones puedan tener una influencia mayoritaria en los procesos de regresión de propiedades utilizando estas naftas.

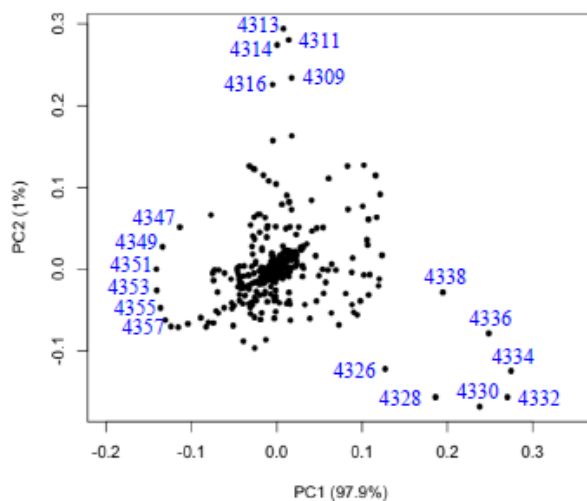
Los valores de los *scores* de las 80 muestras utilizadas en el PCA son presentados en la Tabla 3.1; en esta tabla se listan los primeros 6 componentes principales, los cuales reproducen el 99.62% de la varianza contenida en los espectros pretratados de las respectivas muestras de

nafta. Esta reducción en el número de variables que representan a los espectros hace del PCA una herramienta que facilita el desarrollo de modelos de regresión de propiedades.



(a) Varianza explicada con el número de PC

(b) Gráfico de los *scores* PC2 vs PC1 (98.9%)



(c) Gráfico de los *loadings* (98.9%)

Figura 10. Resultados del análisis PCA aplicado a los espectros pretratados.

Tabla 4.

Valores de los primeros seis componentes principales para las muestras (scores) con una varianza acumulada del 99.62%.

	PC1	PC2	PC3	PC4	PC5	PC6		PC1	PC2	PC3	PC4	PC5	PC6
1	-55,32	-1,70	2,99	2,33	-1,87	-0,70	41	-55,18	6,07	0,75	1,50	-0,50	0,14
2	-54,24	-11,04	2,49	1,93	-1,09	-2,33	42	-55,01	-5,78	3,60	1,61	-0,84	-2,52
3	-55,15	-4,06	4,34	0,53	1,93	1,52	43	-54,91	8,16	1,22	0,83	0,11	0,22
4	-55,22	-1,38	4,86	0,31	0,47	2,49	44	-54,84	-7,41	2,54	1,88	-0,81	-2,75
5	-55,33	2,21	-3,15	0,82	1,58	-1,42	45	-54,48	-7,88	6,26	1,85	-1,94	-1,96
6	-55,22	5,18	-2,09	0,86	0,21	-0,80	46	-55,28	-3,20	3,12	1,89	-0,72	-1,67
7	-54,97	-6,23	1,53	1,10	3,94	1,61	47	-55,22	5,50	1,04	1,08	0,17	-1,07
8	-55,38	2,77	1,42	0,80	0,31	1,06	48	-55,44	0,50	1,97	1,55	-0,06	-1,19
9	-55,16	-4,88	1,44	1,04	3,12	1,95	49	-55,20	5,69	-0,19	0,49	0,57	-1,54
10	-55,23	5,53	0,15	1,19	-0,65	0,49	50	-54,92	-5,25	-1,81	2,38	3,25	2,79
11	-52,82	-15,54	5,56	2,27	3,19	2,09	51	-54,72	-7,92	-3,73	-1,61	-0,29	-0,71
12	-55,13	5,93	1,01	1,48	-0,92	1,50	52	-55,33	2,56	0,46	-1,88	-2,13	1,44
13	-55,40	2,81	-1,20	1,15	1,00	-0,08	53	-55,22	-4,18	0,62	-1,32	-3,15	0,02
14	-55,24	4,67	-2,85	0,79	0,49	-0,40	54	-54,09	-11,38	1,70	-0,12	-4,09	0,23
15	-55,38	3,50	-0,02	1,39	0,52	0,82	55	-54,81	0,16	-8,11	-1,85	0,94	-0,52
16	-55,18	6,06	-1,15	1,06	-0,16	0,10	56	-54,84	-7,77	0,16	-1,82	-2,52	-0,43
17	-54,98	7,47	-1,28	1,14	-0,90	0,15	57	-55,06	5,97	-2,30	-1,76	-1,14	0,90
18	-55,39	-1,00	-0,76	1,82	1,41	1,58	58	-54,07	-11,55	2,21	-1,10	-3,65	-0,36
19	-55,41	2,09	-0,68	1,39	0,60	1,02	59	-55,06	-3,05	-5,31	-2,10	0,60	-0,19
20	-55,26	-3,60	-1,83	1,22	2,37	0,07	60	-55,03	-3,65	-4,86	-1,80	-1,62	-1,72
21	-54,99	7,58	0,12	0,82	-1,24	0,71	61	-54,91	-1,60	-6,82	-3,21	1,09	-1,74
22	-55,30	-3,16	-0,70	1,22	2,48	1,64	62	-54,90	-3,06	-6,23	-2,74	0,62	-1,83
23	-55,17	6,28	-0,19	1,31	-0,19	0,68	63	-54,90	-5,26	-4,78	-2,20	-0,17	-1,76
24	-54,94	8,01	-0,29	1,03	-0,75	0,55	64	-55,03	5,80	-2,41	-2,65	-1,16	0,75
25	-55,33	3,98	-1,78	1,47	0,85	0,46	65	-55,27	2,35	3,25	-1,74	-1,15	2,31
26	-54,92	8,01	-0,87	1,26	-1,09	0,82	66	-55,26	2,02	3,12	-2,12	-1,13	2,22
27	-55,14	-4,80	-0,94	0,85	2,98	1,39	67	-55,05	-1,00	-6,38	-1,72	0,21	-0,71
28	-55,16	5,88	-1,15	1,70	0,04	0,82	68	-55,11	-1,98	-5,68	-1,00	-0,13	-0,17
29	-55,33	-2,90	0,32	0,84	0,61	-2,22	69	-55,23	0,09	3,88	-1,70	-2,84	1,97
30	-55,42	-0,94	1,62	2,14	-0,22	-1,73	70	-55,12	-3,01	-5,10	-0,88	-0,45	-0,85
31	-55,37	3,27	1,33	1,41	-0,42	-0,80	71	-55,31	-0,70	1,12	-3,93	-0,35	1,15
32	-55,22	-4,34	-1,80	2,04	1,04	-1,81	72	-54,76	-7,08	-0,68	-3,83	0,92	1,15
33	-55,37	-2,91	0,26	0,70	0,42	-2,38	73	-53,79	-12,15	2,72	-3,66	2,62	1,49
34	-55,40	-1,01	2,50	2,06	0,01	-1,33	74	-55,27	3,12	-2,88	-1,13	-0,85	0,39
35	-55,40	3,50	-0,18	1,67	0,07	-1,09	75	-55,36	2,81	0,07	-1,81	-0,76	0,83
36	-55,14	6,35	0,79	1,39	-0,80	-0,27	76	-55,32	-1,07	1,55	-1,72	-3,04	1,81
37	-54,93	7,90	0,37	1,59	-0,65	-0,06	77	-55,31	-1,46	2,80	-1,53	-0,66	1,32
38	-55,42	1,52	1,53	1,62	-0,98	-1,34	78	-50,03	14,35	16,02	-8,44	4,96	-4,48
39	-55,16	5,82	-0,88	1,58	-0,24	-0,81	79	-55,32	1,47	0,35	-1,95	-2,97	0,89
40	-54,89	2,54	-2,88	-1,15	7,09	-0,54	80	-55,34	0,83	0,47	-2,78	-0,98	2,34

4.3 Código de entrenamiento y validación de modelos RNA.

Las líneas de código desarrolladas en R que utiliza la función *neuralnet* descrita en la metodología son presentadas a continuación:

```
p<-6 # Definición número de variables para el modelo

data<-cbind(PC[,1:p],rvp) # Definición de la matriz de ajuste para la propiedad RVP en este
caso.

n <- names(data) # Lectura nombre de las columnas de la matriz de ajuste.

set.seed(nrow(data)) # Iniciación números aleatorios.

maxs <- apply(data, 2, max) # Valores máximos en las columnas de la matriz de ajuste.

mins <- apply(data, 2, min) # Valores mínimos en las columnas de la matriz de ajuste.

scaled01 <- as.data.frame(scale(data, center = mins, scale = maxs – mins)) # Escalado de las
columnas entre 0 y 1.

MSEnn<-matrix(nrow=2*p,ncol=20) # Definición matriz de errores validación.

MSEnnt<-matrix(nrow=2*p,ncol=20) # Definición matriz de errores entrenamiento.

MSEnntot<-matrix(nrow=2*p,ncol=20) # Definición matriz de errores totales.

R2_<-matrix(nrow=2*p,ncol=20) # Definición matriz de coeficiente de determinación en
validación.

R2_t<-matrix(nrow=2*p,ncol=20) # Definición matriz de coeficiente de determinación en
entrenamiento.

R2_tot<-matrix(nrow=2*p,ncol=20) # Definición matriz de coeficiente de determinación totales.

for (i in 1:(2*p)) {# Ciclo incremento número de neuronas ocultas.

  for (j in 1:20) {# Ciclo ajuste redes neuronales.
```

```
index = sample(seq_len(nrow(scaled01)), size = samplesize) # División datos escalados para  
entrenamiento y validación.
```

```
Dtrain<- scaled01[index,] # Matriz de entrenamiento.
```

```
Dtest <- scaled01[-index,] # Matriz de validación.
```

```
colnames(Dtrain) <- n # Nombre columnas matriz de entrenamiento.
```

```
colnames(Dtest) <- n # Nombre columnas matriz de validación.
```

```
f <- as.formula(paste("rvp ~", paste(n[!n %in% "rvp"], collapse = " + ")) # Definición variable  
dependiente.
```

```
model0 <- neuralnet(f,data=Dtrain,hidden=c(i),linear.output=T,  
err.fct="sse",act.fct="logistic",stepmax=100000,threshold=0.1, algorithm="rprop+",rep=10) #  
Entrenamiento y validación de la estructura de red neuronal con el algoritmo backpropagation y  
10 repeticiones.
```

```
pr.nn <- compute(model0,Dtest[,1:p]) # Valores predichos para la validación.
```

```
pr.nn_ <- pr.nn$net.result*(max(data$rvp)-min(data$rvp))+min(data$rvp) # Valores predichos  
para el entrenamiento.
```

```
test.r <- Dtest[,p+1]*(max(data$rvp)-min(data$rvp))+min(data$rvp) # Valores predichos para la  
validación en variables originales.
```

```
ym_ <- mean(Dtest[,p+1]) # Valor promedio para los datos utilizados en validación.
```

```
MSEnn[i,j]<-sum((test.r-pr.nn_)^2)/nrow(Dtest) # Error cuadrático medio en validación.
```

```
R2_[i,j]<-1-sum((test.r-pr.nn_)^2)/sum((test.r-ym_)^2) # Coeficiente de determinación en  
validación.
```

```
test.rt<- Dtrain[,p+1]*(max(data$rvp)-min(data$rvp))+min(data$rvp) # Valores predichos para  
el entrenamiento en variables originales.
```

```

ym_t<-mean(Dtrain[,p+1]) # Valor promedio para los datos utilizados en entrenamiento.
MSEnnt[i,j]<-sum((test.rt-pr.nn_t)^2)/nrow(Dtrain) # Error cuadrático medio en entrenamiento.
R2_t[i,j]<-1-sum((test.rt-pr.nn_t)^2)/sum((test.rt-ym_t)^2) # Coeficiente de determinación en
entrenamiento.

MSEnntot[i,j]=(sum((test.r-pr.nn_)^2)+sum((test.rt-pr.nn_t)^2))/(nrow(data)) # Error cuadrático
medio total.

R2_tot[i,j]<-1-(sum((test.r-pr.nn_)^2)+sum((test.rt-pr.nn_t)^2))/(sum((test.r-
ym_)^2)+sum((test.rt-ym_t)^2)) # Coeficiente de determinación total.

}

}

```

Estas líneas consideran el uso de los PC descritos en la Tabla 3.1 como entradas a las RNA. El número de PC seleccionados para las neuronas de entrada (p) es variable, al igual que el número de neuronas definidas para la capa oculta (k); el promedio se efectúa sobre la ejecución del entrenamiento y la validación un número de 20 veces (j); con este código, la ejecución de diferentes valores de p se efectúa manualmente. La propiedad objetivo corresponde al RVP; los estadísticos MSE de validación, entrenamiento y total corresponden a MSEnn, MSEnnt y MSEnntot, respectivamente, mientras para R^2 los estadísticos son R2_, R2_t y R2_tot, respectivamente.

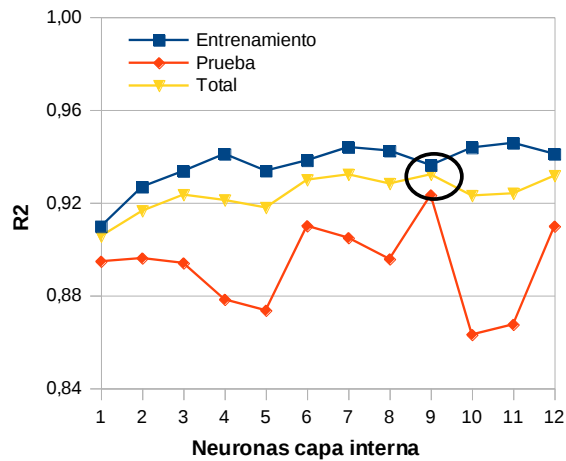
4.4 Aplicación del código considerando componentes principales

Los seis primeros componentes principales, con una varianza total del 99.62%, fueron utilizados para el entrenamiento de las arquitecturas de red utilizando el código presentado en la sección

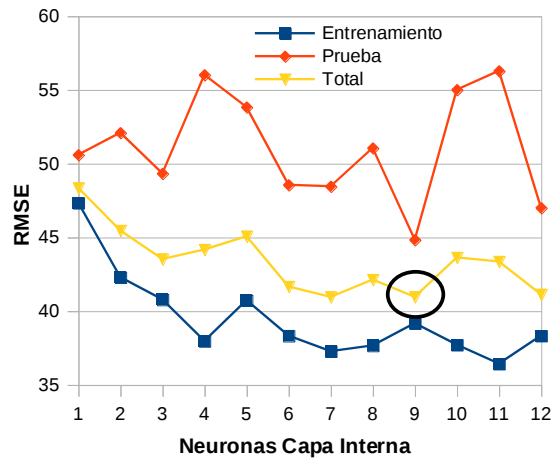
anterior. La Figura 11 exhibe los desempeños obtenidos en entrenamiento, validación y total de las diferentes arquitecturas utilizando la función de activación Log; estas gráficas representan los desempeños de las redes definidas como 6:x:1–Log, en donde 6 corresponde al número de PC de entrada, x al número de neuronas en la capa oculta, 1 a la propiedad en cuestión y Log a la función logística utilizada en las neuronas ocultas. Cada valor de desempeño fue obtenido de un promedio de 20 veces la ejecución de la correspondiente RNA. La Figura 11a presenta la variación de R^2 con el número de neuronas de la capa interna, mientras que la Figura 11b presenta lo propio para el RMSE. Según estas figuras, considerando los 6 PC la mejor red corresponde a la 6:9:1–Log para la regresión del contenido de Azufre; los desempeños alcanzados con esta estructura son mostrados con un círculo negro.

Asimismo, la Figura 12 presenta los resultados de los procesos con la red 6:x:1–TanH para el ajuste del contenido de Azufre; en esta red se aplica la función de activación tangencial hiperbólica en lugar de la logística. Según las Figuras 12a-b, los mejores desempeños se obtienen con la red 6:9:1–TanH. Comparando los desempeños con los obtenidos en la Figura 11, es posible inferir que la red 6:9:1–TanH es superior a la red 6:9:1–Log en el ajuste de la propiedad contenido de Azufre; la primera red exhibe un valor de $RMSE_{Total}$ de 37.5 ppm, mayor que el de la segunda red ($RMSE_{Total}=41.8$ ppm).

También, las Figuras 13 y 14 presentan los resultados para los procesos de ajuste del RVP con las estructuras de red 6:x:1–Log y 6:x:1–TanH, respectivamente. Los mejores desempeños para cada tipo de estructura son señalados con círculos negros, en las respectivas figuras. El mejor desempeño para la red 6:8:1–Log corresponde a un $RMSE_{Tot}$ de 0.21 psi (Figura 13), mientras que el mejor desempeño para la red 6:12:1–TanH es $RMSE_{Tot}$ de 0.19 psi (Figura 14).

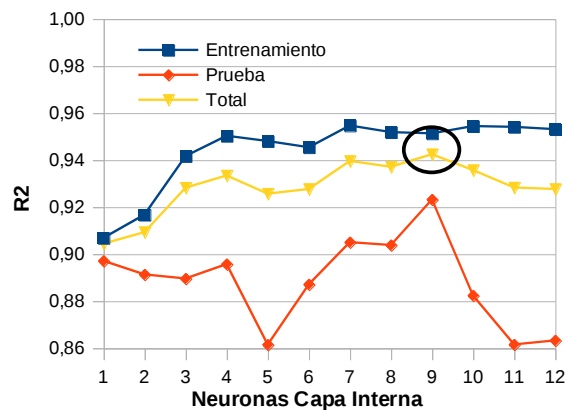


(a) R^2 vs neuronas capa interna

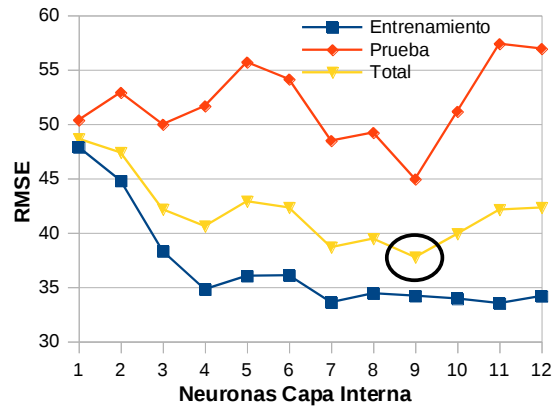


(b) RMSE vs neuronas capa interna

Figura 11. Variación de los estadísticos de desempeño con el número de neuronas de la capa interna (x) en la red 6:x:1-Log en la predicción del contenido de Azufre.



(a) R^2 vs neuronas capa interna



(b) RMSE vs neuronas capa interna

Figura 12. Variación de los estadísticos de desempeño con el número de neuronas de la capa interna (x) en la red 6:x:1-TanH en la predicción del contenido de Azufre.

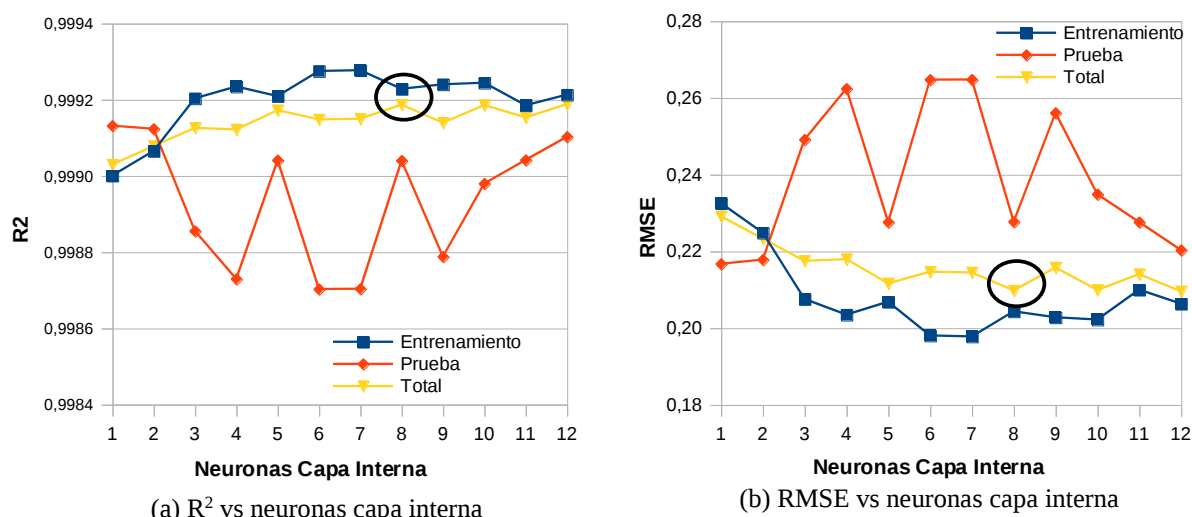


Figura 13. Variación de los estadísticos de desempeño con el número de neuronas de la capa interna (x) en la red 6:x:1-Log en la predicción del RVP.

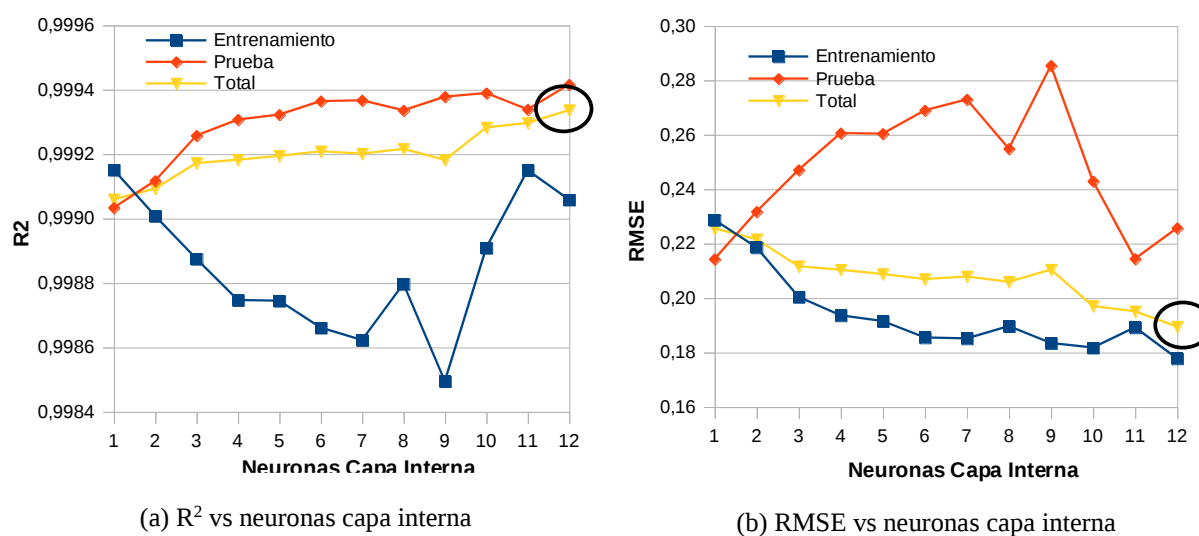


Figura 14. Variación de los estadísticos de desempeño con el número de neuronas de la capa interna (x) en la red 6:x:1-TanH en la predicción del RVP.

Las Figuras 15–18 muestran los desempeños en los procesos de ajustes totales de las propiedades con el número de neuronas de la capa interna y el número de PC utilizados en las

entradas; las redes son representadas con $y:x:1$ -Log o $y:x:1$ -TanH, según el caso, en donde y hace referencia al número de PC de entrada, mientras que x considera el número de neuronas en la capa oculta. Respecto al contenido de Azufre, las Figura 15 y 16 presentan los desempeños totales mostrados por las redes con $y:x:1$ -Log y $y:x:1$ -TanH, respectivamente. Según estas figuras, los mejores desempeños son obtenidos con las estructuras 5:9:1-Log y 6:9:1-TanH con valores de $RMSE_{Tot}$ correspondientes a 39.2 ppm y 37.5 ppm, respectivamente. Con lo anterior, una reducción en el número de PC de entrada (5 PC) mejora el desempeño de la red con función de activación Log (Figura 16), mientras que desmejora los desempeños de las redes con función de activación TanH (Figura 16). En referencia al RVP, los mejores desempeños totales son exhibidos por las estructuras considerando 6 PC de entrada, para las dos funciones de activación (Figuras 17 y 18); *i.e.* la disminución en el número de entradas desmejora los desempeños de los ajustes en las diferentes arquitecturas de red.

Por otra parte, pruebas adicionales considerando el aumento a dos capas internas fueron realizadas para determinar una mejora en los desempeños en el ajuste de las propiedades. La Figura 19 presenta los desempeños totales de las estructuras de red 5:9: x :1-Log y 6:9: x :1-TanH; según esta figura, los mejores desempeños totales se obtienen con las redes 5:9:7:1-Log y 6:9:6:1-TanH con valores de $RMSE_{Tot}$ de 44 ppm y 37.2 ppm, respectivamente. La mejora en el desempeño la presenta la red con función de activación TanH, sin embargo, esta mejora es poco significativa ya que corresponde a solamente una reducción en 0.3 ppm; con esto, la opción más adecuada corresponde a una sólo capa interna; una RNA con menor cantidad de parámetros obtiene mejor capacidad predictiva (ver Wythoff, 1993; Glorfeld, 1996).

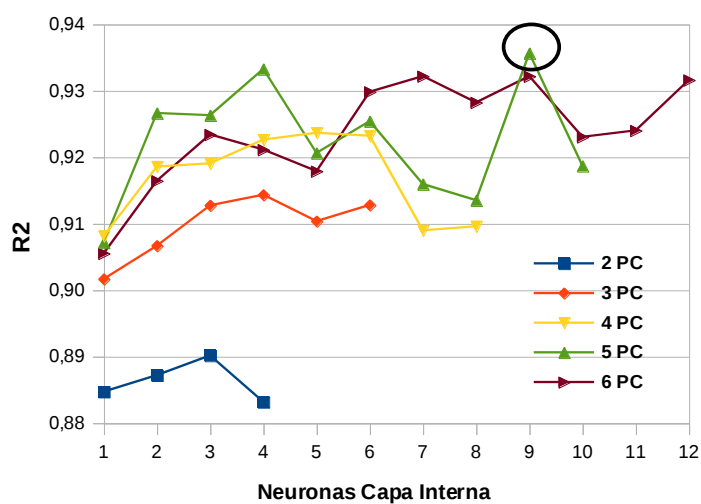
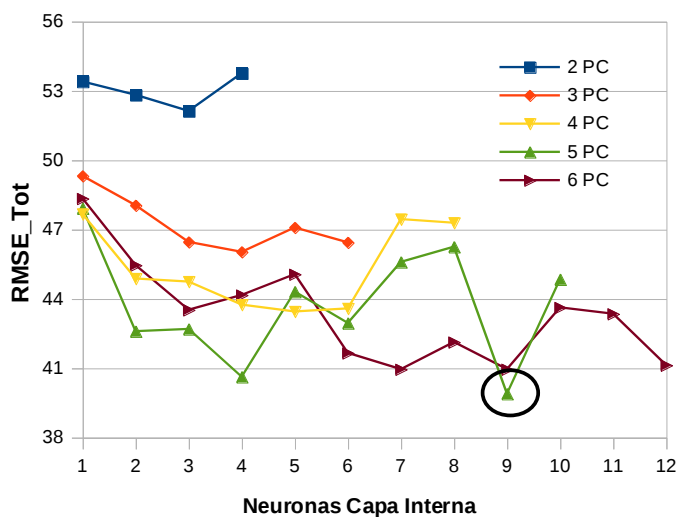

(a) R^2_{Tot} vs neuronas capa interna

(b) $RMSE_{Tot}$ vs neuronas capa interna

Figura 15. Variación de los estadísticos de desempeño total con el número de neuronas de la capa interna (x) y el número de PC de entrada (y) para la red $y:x:1$ -Log en la predicción del contenido de azufre.

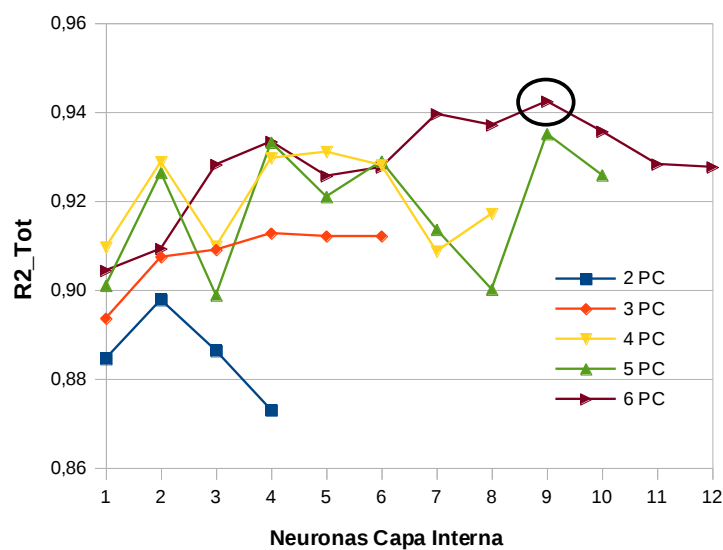
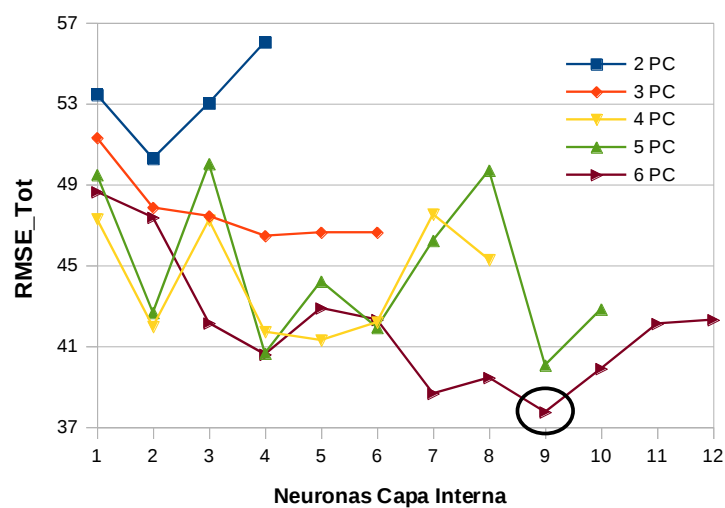
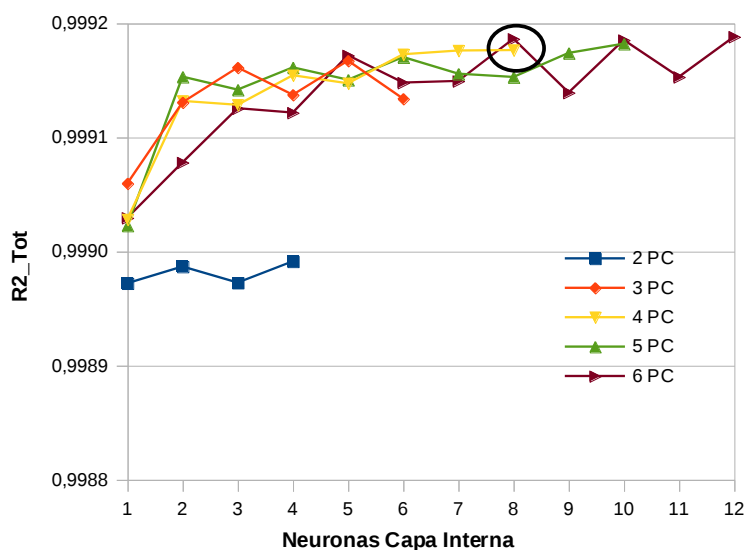
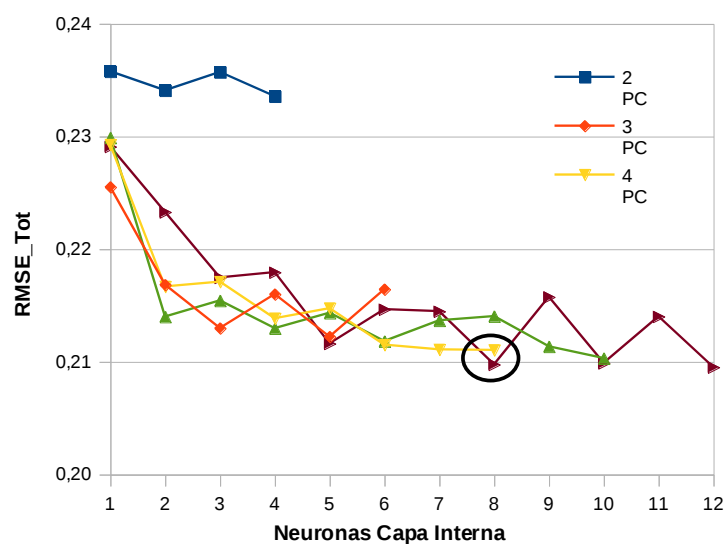
(a) R^2_{Tot} vs neuronas capa interna(b) RMSE_{Tot} vs neuronas capa interna

Figura 16. Variación de los estadísticos de desempeño total con el número de neuronas de la capa interna (x) y el número de PC de entrada (y) para la red $y:x:1$ -TanH en la predicción del contenido de azufre.



(a) R^2_{Tot} vs neuronas capa interna



(b) $RMSE_{Tot}$ vs neuronas capa interna

Figura 17. Variación de los estadísticos de desempeño total con el número de neuronas de la capa interna (x) y el número de PC de entrada (y) para la red $y:x:1$ -Log en la predicción del RVP.

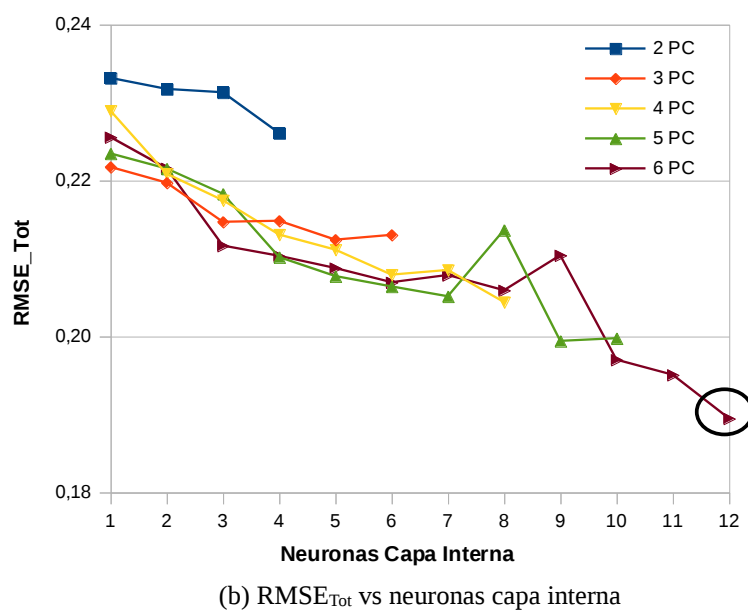
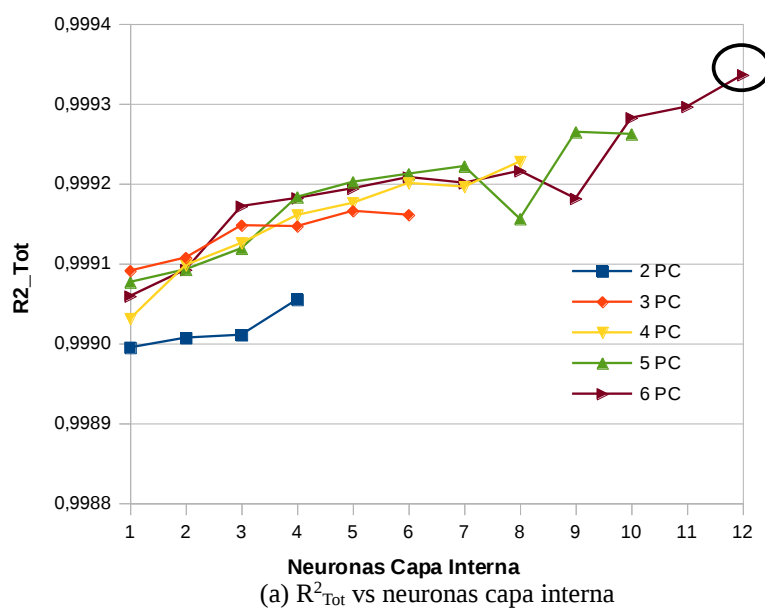


Figura 18. Variación de los estadísticos de desempeño total con el número de neuronas de la capa interna (x) y el número de PC de entrada (y) para la red $y:x:1$ -TanH en la predicción del RVP.

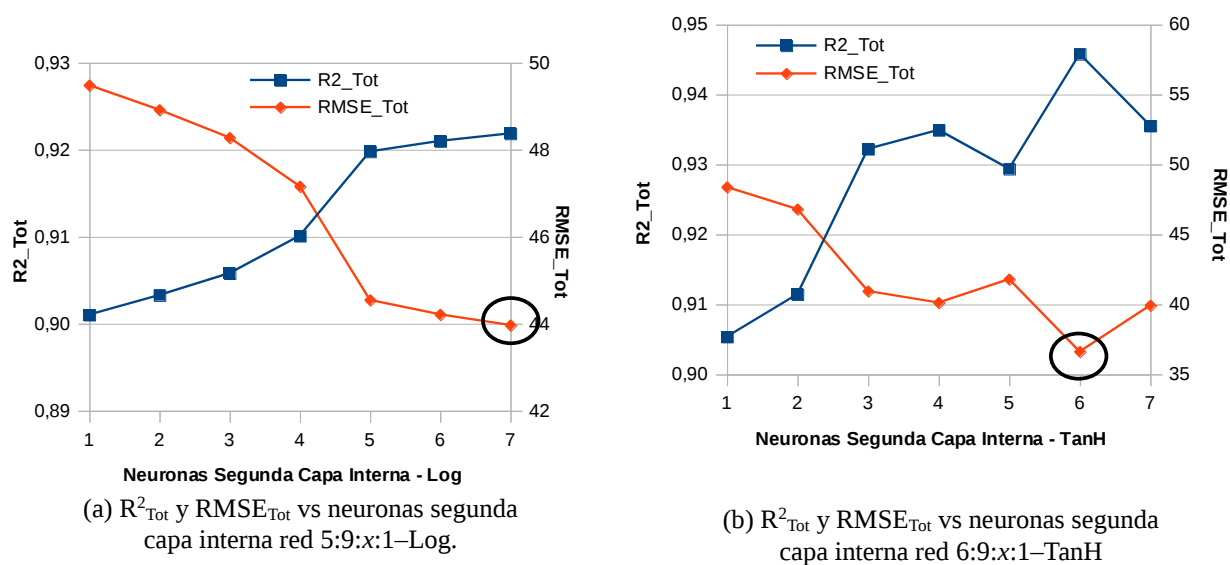


Figura 19. Variación de los estadísticos de desempeño total con el número de neuronas en la segunda capa interna (x) para la predicción del contenido de azufre.

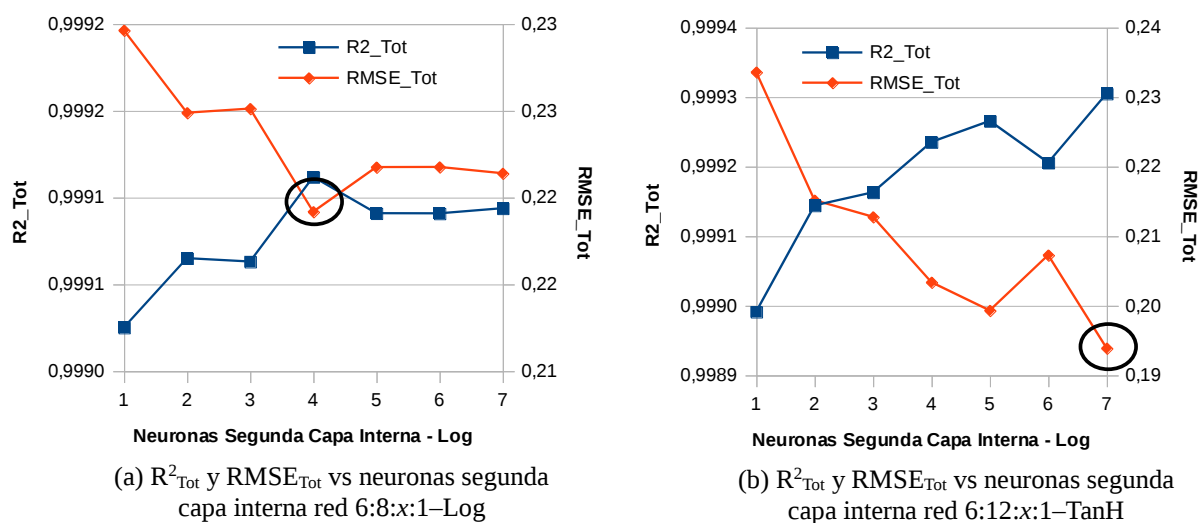


Figura 20. Variación de los estadísticos de desempeño total con el número de neuronas en la segunda capa interna (x) para la predicción del RVP.

4.5 Aplicación del algoritmo considerando absorbancias pretratadas

Los valores extremos obtenidos con los pretratamientos (Figuras 8d-e) fueron utilizados en el entrenamiento y la validación de estructuras de RNA para la propiedad contenido de Azufre, según lo definido en la metodología. Un total de 80 intervalos entre los rangos 4800–4000 cm^{-1} y 6200–5400 cm^{-1} después de la transformación SNV fueron definidos para la obtención de las respectivas integrales impropias. La Figura 21 expone los valores de las respectivas integrales en los 80 intervalos; según esta figura, los valores absolutos de integración son mayores en el primer rango, 4800–4000 cm^{-1} , lo cual concuerda con los valores de las bandas después del pretratamiento SNV (Figura 8c-e) y con lo obtenido en el análisis PCA.

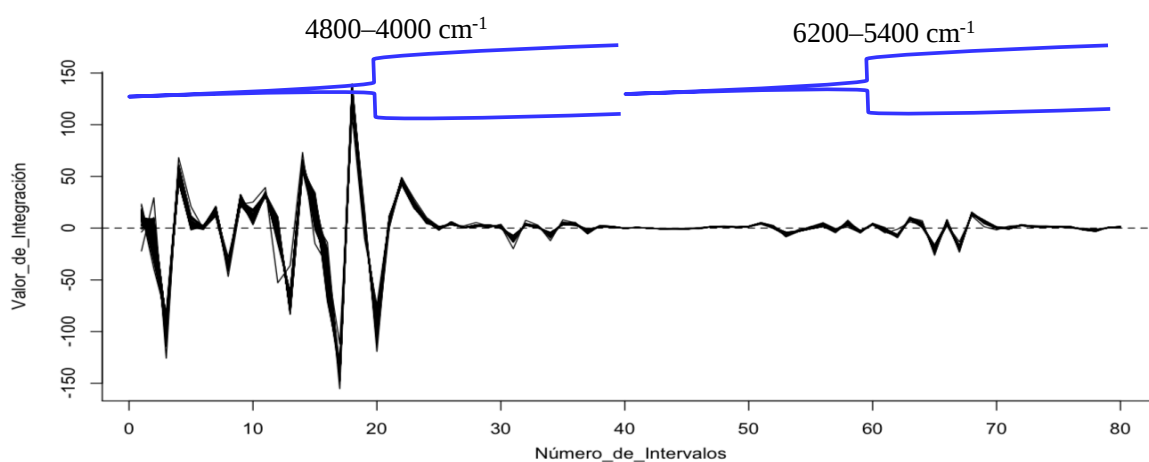


Figura 21. Valores de la integral cada 20 cm^{-1} en los rangos pretratados 4800–4000 cm^{-1} y 6200–5400 cm^{-1} .

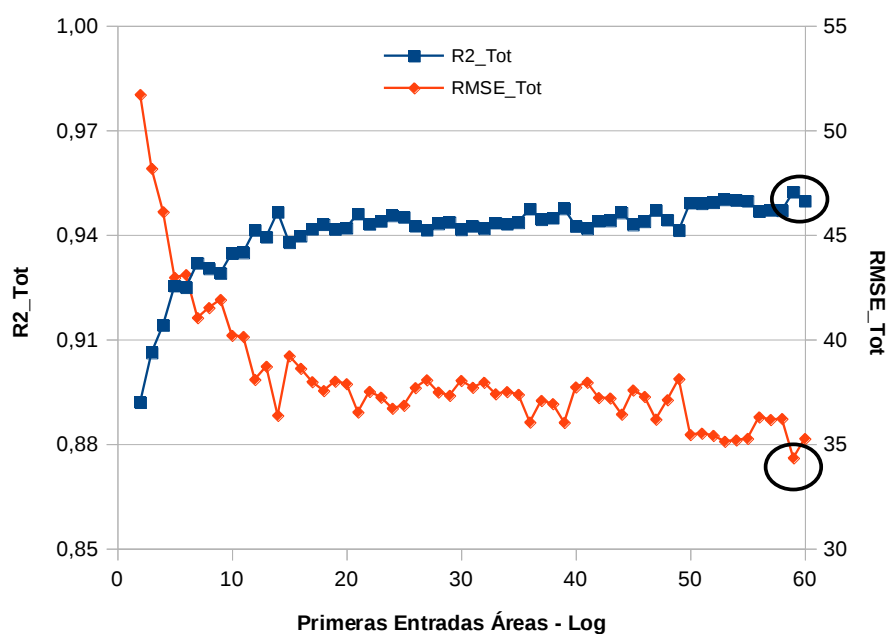
Los ajustes fueron realizados considerando las áreas como entradas, con un paso previo de ordenamiento de estas según su valor absoluto descendente. La presentación de los resultados de los ajustes con variación en la entradas (áreas) y neuronas de la capa oculta se realiza, primero

considerando los mejores desempeños por variación del número de entradas (sin describir a que número de neuronas ocultas corresponde) y posteriormente, considerando la variación del ajuste con el número de neuronas ocultas para el número de entradas con el mejor desempeño. La Figura 22 exhibe los mejores desempeños totales obtenidos de las diferentes arquitecturas de RNA según las entradas definidas, utilizando la función de activación Log (Figura 22a) y TanH (Figura 22b) para el ajuste del contenido de Azufre; cada valor de desempeño fue obtenido de un promedio de 20 veces la ejecución de la correspondiente RNA. Según la figura, los mejores desempeños se presentan considerando los primeros 59 valores de áreas con mayor valor absoluto (Figura 22a), para la función de activación Log, y los primeros 30 valores con mayor valor absoluto para la función de activación TanH (Figura 22b). La Figura 23 presenta los desempeños para las anteriores entradas según la variación en el número de neuronas de la capa interna. Según esta figura, los mejores desempeños se obtienen con las redes 59:13:1–Log y 30:12:1–TanH con valores de $RMSE_{Tot}$ de 34.3 ppm (Figura 23a) y 33.6 ppm (Figura 3.23b), respectivamente. Las Figuras 24a-b presentan los desempeños de los ajustes considerando una nueva capa oculta adicionada en las redes con mejores desempeños. Según estas figuras, las redes 59:13:17:1–Log alcanza un valor de $RMSE_{Tot}$ de 46.9 ppm (Figura 24a), mientras que la red 30:12:9:1–TanH reporta un $RMSE_{Tot}$ de 47.5 ppm (Figura 24b). Con lo anterior, la adición de una segunda capa interna no reporta mejoras en los ajustes para el contenido de Azufre.

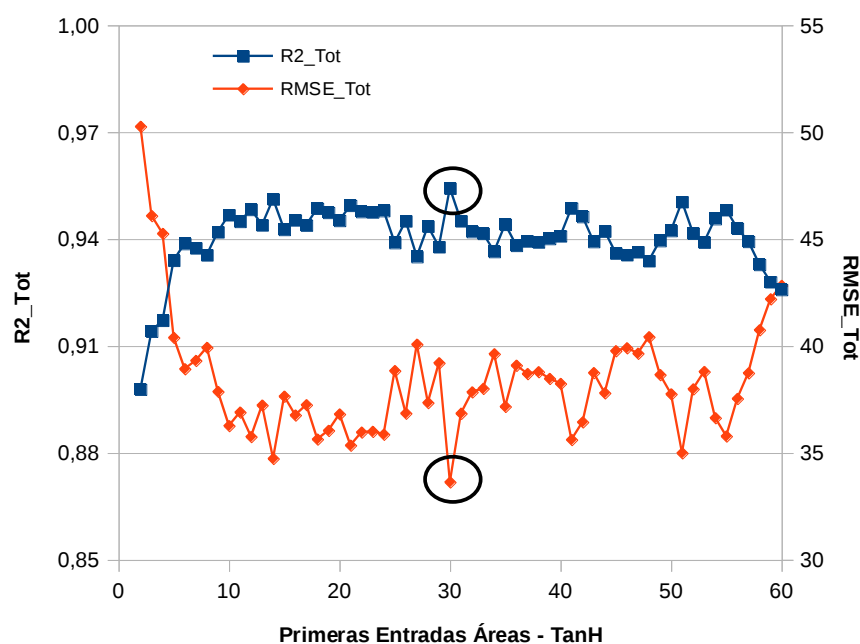
Por otra parte, las pruebas para el ajuste del RVP utilizando las áreas no presentaron mejoras significativas respecto a los desempeños obtenidos utilizando los PC.

4.6 Selección de los modelos RNA

La Tabla 3.2 resume los puntos de mejor desempeño total de las RNA con las dos funciones de activación y las capas ocultas, considerando como entradas los PC y las áreas bajo la curva de las absorbancias pretratadas. Según esta tabla, los mejores desempeños en términos del menor valor de $RMSE_{Tot}$ y el mayor valor de R^2_{Tot} corresponden a las redes 30:12:1–TanH y 6:12:1–TanH para el ajuste del contenido de Azufre y el RVP, respectivamente. El RMSE de la red 30:12:1–TanH de ajuste del contenido de Azufre corresponde a 33.6 ppm, el cual se encuentra entre la repetibilidad y la reproducibilidad promedios de 10 ppm y 45 ppm, respectivamente, según la norma ASTM D4294–16. Respecto al RVP, la repetibilidad y la reproducibilidad promedios son 0.21 y 0.36 psi, respectivamente. Con lo anterior, los modelos RNA para estas propiedades cuentan con capacidad predictiva a nivel de laboratorio.

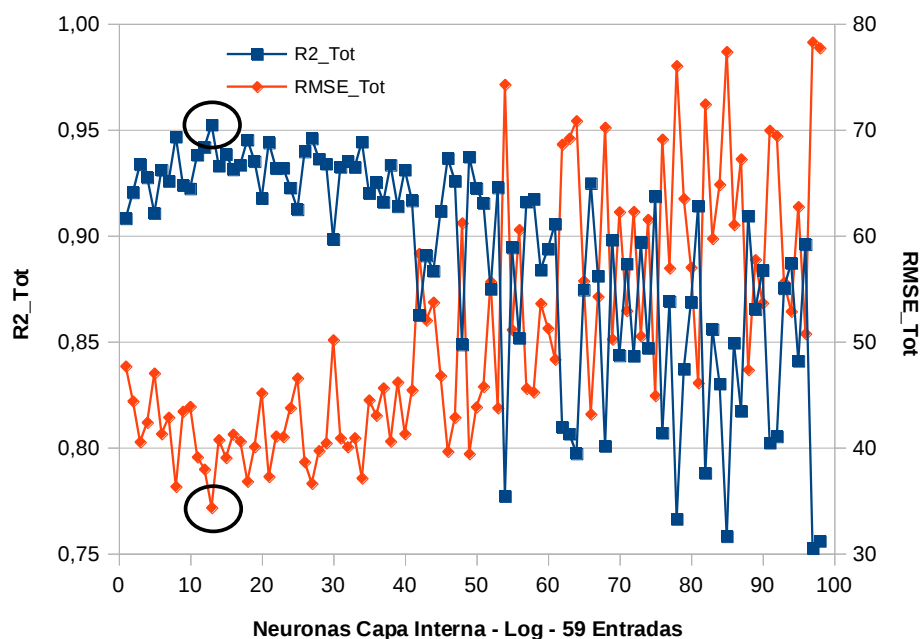


(a) Redes con neuronas internas con la función de activación Log.

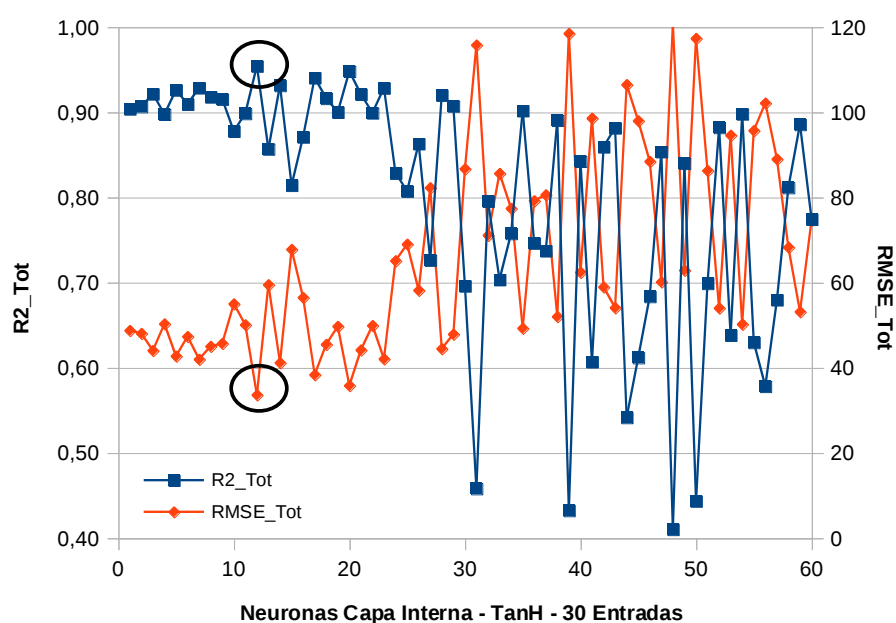


(b) Redes con neuronas internas con la función de activación TanH.

Figura 22. Mejores desempeños totales obtenidos con las arquitecturas según las entradas definidas.

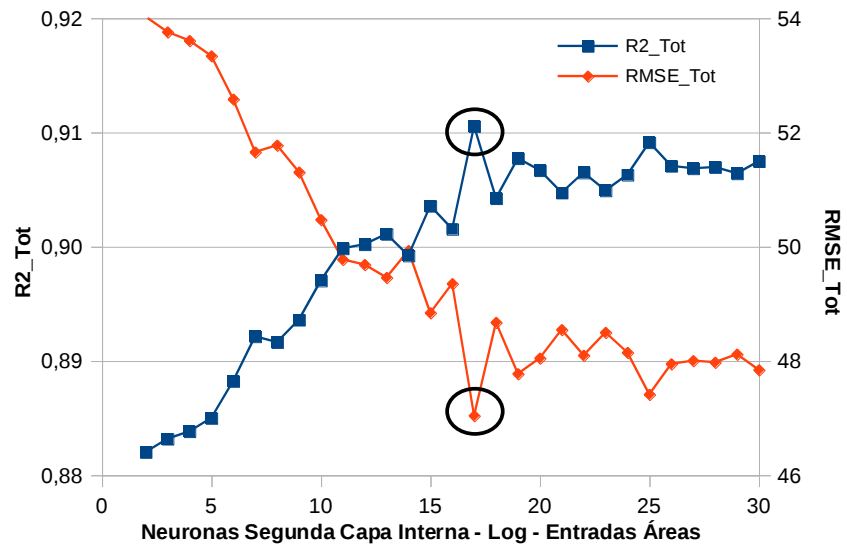


(a) Red de estructura 59:x:1-Log.

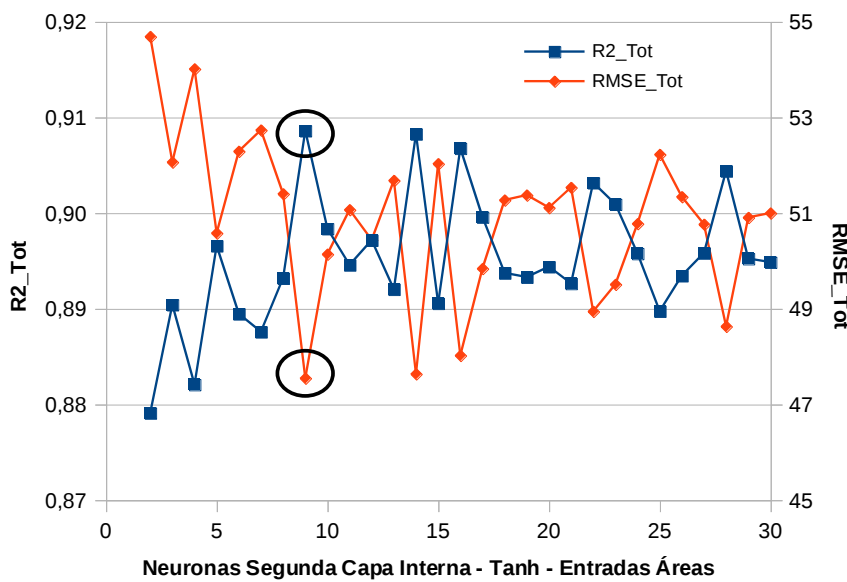


(b) Red de estructura 30:x:1-TanH.

Figura 23. Desempeños de las redes para las mejores entradas según el número de neuronas de la capa interna.



(a) Red de estructura 59:13:x:1-Log.



(b) Red de estructura 30:12:x:1-TanH.

Figura 24. Desempeños considerando una segunda capa oculta en las estructuras de mejor desempeño.

Tabla 5.

Resumen de las mejores estructuras de red en el ajuste del contenido de Azufre y RVP de las muestras de nafta de la GRB.

Estructura Red		R²_{Tot}	RMSE_{Tot}
Azufre, PC	6:9:1–Log	0.94	41.8 ppm
Azufre, PC	6:9:1–TanH	0.94	37.5 ppm
Azufre, PC	5:9:7:1–Log	0.90	44.0 ppm
Azufre, PC	6:9:6:1–TanH	0.91	37.2 ppm
Azufre, Área	59:13:1–Log	0.95	34.3 ppm
Azufre, Área	30:12:1– TanH	0.95	33.6 ppm
Azufre, Área	59:13:17:1–Log	0.91	49.9 ppm
Azufre, Área	30:12:9:1– TanH	0.91	47.5 ppm
RVP, PC	6:8:1–Log	0.99	0.211 psi
RVP, PC	6:12:1–TanH	0.99	0.190 psi
RVP, PC	6:8:4:1–Log	0.99	0.218 psi
RVP, PC	6:12:7:1– TanH	0.99	0.194 psi

Los desempeños obtenidos en el presente trabajo resultan superiores en comparación con los desempeños de diferentes manuscritos de literatura para el petróleo y sus fracciones medias. Barbosa *et al.*, 2016, reportan valores de RMSE_{val}=510 ppm en su modelo PLS de determinación de azufre basado en espectros de resonancia magnética nuclear de campo bajo (LF-NMR) de crudos de origen brasileño (°API entre 16.9 y 28.9, azufre entre 1000 y 30000 ppm). Rocha *et al.*, 2016, obtuvieron valores de RMSE_{val}=419 ppm en su mejor modelo PLS basado en espectros MIR de muestras de crudos brasileños con °API entre 12.3 y 57.7 y contenidos de azufre entre 0.5 y 9400 ppm.

Satya *et al.*, 2007, desarrollaron un modelo PLS basado en el espectro NIR para la estimación en crudos con contenido de azufre en el rango 500 – 25000 ppm, reportando un RMSEP de 350 ppm. Breitzkreitz *et al.*, 2003, en su modelo PLS basado en espectros NIR de fracciones de diésel con contenidos de azufre entre 700 – 3300 ppm reportan un RMSEP de 310 ppm. de Oliveira *et al.*, 2017, entrenaron RNA basadas en las propiedades destilación por la norma ASTM D86, gravedad específica, número de cetano y *flash point* para la predicción de contenidos de azufre en mezclas de diésel y biodiésel. Los contenidos de azufre estuvieron entre 4 – 1831 ppm. El valor de $RMSE_{total}$ reportado por de Oliveira *et al.* fue de 60.3 ppm.

Por otro lado, Kramer *et al.*, 2008, reportaron que las propiedades azufre, junto con la conductividad térmica, mostraron un bajo nivel de confianza (<70%) en los respectivos modelos basados en 43 espectros NIR y Raman de diferentes gasolinas tipo jet; el contenido de azufre del conjunto se encontraba entre 13 y 2453 ppm. De igual forma, Morris *et al.*, 2009, reportaron un bajo nivel de confianza en la reproducción del contenido azufre al aplicar regresiones PLS utilizando los espectros NIR de 800 muestras de gasolinas tipo jet.

Balabin & Lomakina, 2011, compararon los desempeños de modelos de regresión PLS, RNA y SVR (*support vector machine*) para diferentes propiedades de gasolinas, biogasolinas y diésel, entre otros. Los modelos fueron desarrollados utilizando los espectros NIR de las respectivas muestras. Respecto al azufre, la base de datos de espectros NIR fue conformada con 125 muestras con valores de contenido de azufre entre 303 y 5100 ppm. Estos autores mencionan que los modelos RNA y SVR obtuvieron los mejores desempeños con valores de RMSEP de 155 y 131 ppm, respectivamente, cercanos al valor de reproducibilidad máximo de 165 ppm según norma.

Por su parte, Qin *et al.*, 2014, aplicaron regresión PLS a un conjunto de espectros terahertz en el dominio del tiempo (*terahertz time-domain spectroscopy*, *Thz-TDS*) de 20 tipos de gasolinas con contenidos de azufre entre 0.2 y 50 ppm. Estos autores reportaron un valor de 1.2 ppm para el RMSE del modelo en validación, lo cual supera el trabajo del presente documento.

Por otro lado, Flecher *et al.*, 1997, aplicaron la regresión PLS sobre un conjunto de 205 espectros Raman de fibra óptica dispersiva (*dispersive fiber-optic Raman spectroscopy*) de gasolinas con valores de RVP entre 7.14–14.8 psi. Estos autores reportaron un RMSEP de 0.87 psi, el cual es mayor que el obtenido en el presente trabajo de 0.2 psi. Cooper *et al.*, 1995, desarrollaron un modelo de predicción PLS basado en 208 espectros Raman de muestras de gasolinas con un rango de RVP entre 7.07–14.74 psi. Estos autores reportan un modelo compuesto por 8 variables latentes, con lo cual obtienen un RMSEP de 0.57 psi. Los modelos RNA desarrollados en el presente trabajo alcanzan un nivel de predicción mejor (Tabla 3.2) que el reportado por Cooper *et al.*

Côcco *et al.*, 2005, aplicaron RNA en la predicción utilizando 32 muestras de gasolinas con valores de RVP entre 3.5 y 8 psi. Estos autores reportaron un error máximo de 3.11% y un error promedio de 1.67% en sus predicciones con una RNA de 3 neuronas internas, basada en 32 datos cromatográficos de composiciones; la función de activación utilizada por Côcco *et al.* para las neuronas internas fue la logística, mientras que la lineal fue utilizada para la neurona de salida. Los errores obtenidos en el presente trabajo (máximo 3,01% y promedio 1.5%) concuerdan con los reportados por Côcco *et al.* Estos autores también reportan que las RNA de mayor número de capas internas muestran desempeños inferiores que las de una capa interna.

Mendes *et al.*, 2017, utilizaron los datos de destilación según la norma ASTM D86 de 80 muestras de mezclas gasolina-etanol (3:1 en volumen) en la construcción de un modelo PLS (53

y 27 para calibración y validación, respectivamente) de predicción del RVP en el rango entre 7.1–9.8 psi. Estos autores reportan un valor RMSEP de 0.098 psi para su modelo de predicción. El valor de RMSEP obtenido en el presente trabajo corresponde al doble del reportado por Mendes *et al.*; sin embargo, la mayor incertidumbre en el presente trabajo se debe a que las muestras no fueron mezcladas con etanol, con lo cual, el valor del RVP proviene exclusivamente del combustible fósil.

Como se aprecia en los análisis anteriores, los desempeños obtenidos en el presente documento presentan desempeños superiores a varios reportes de la literatura. Sin embargo, las diferencias entre los desempeños presentados en literatura se deben a características especiales de estas propiedades. El contenido de azufre y el RVP son propiedades que no muestran causalidad con los espectros NIR o Raman (Trygstad *et al.*, 2015). Las moléculas conteniendo azufre, aunque poseen bandas distintivas, estas aparecen muy ténues o imperceptibles respecto a la cantidad de bandas referentes a los hidrocarburos. Para el caso del RVP, Trygstad *et al.*, 2015, afirman que los cambios en las composiciones de las sustancias responsables de esta propiedad, no aparecen en los espectros NIR ni Raman. Asimismo, Côcco *et al.*, 2004, Morales-Medina & Guzman, 2012, y Nascimento *et al.*, 2018, reportaron que los picos iniciales de la cromatografía resultaron con baja resolución, por lo cual, los cambios en la concentración de los compuestos volátiles responsables del RVP en fracciones del petróleo son también invisibles a nivel de cromatografía de gases. Con esto, los modelos desarrollados corresponden a una relación circunstancial entre las propiedades contenido de azufre y RVP con los espectros NIR de las gasolinas (Trygstad *et al.*, 2015). Sin embargo, los modelos circunstanciales RNA desarrollados en el presente documento exhiben un desempeño apreciable debido a que en su mayoría, las muestras pertenecen a la misma refinería y se encuentran sujetas a un conjunto de variables

operacionales con valores similares. Entrenamientos y validaciones adicionales a las RNA desarrolladas son requeridas con periodicidad debido a su característica circunstancial para las propiedades mencionadas.

De otro lado, la característica circunstancial de las RNA ajustadas restringe un análisis de sensibilidad de las redes respecto a la variación en sus respectivas entradas. Un análisis de sensibilidad con esta clase de modelos carecería de una base fisicoquímica para la asociación de las tendencias de las propiedades con fluctuaciones en la señal espectroscópica.

5. Conclusiones

Según el análisis de los espectros, los rangos de mayor contribución a la señal espectroscópica se encuentran entre $4000 - 4800 \text{ cm}^{-1}$ y $5400 - 6200 \text{ cm}^{-1}$. En estas regiones es posible diferenciar las tendencias de los espectros de la nafta de alta calidad importada y una nafta craqueada de la GRB con una cantidad importante de compuestos insaturados. Asimismo, según el análisis PCA, el valor de la segunda derivada en los números de onda entre $4309-4316 \text{ cm}^{-1}$ $4326-4338 \text{ cm}^{-1}$ define las contribuciones de importancia en los componentes principales, con lo cual pueden tener una influencia mayoritaria en los procesos de regresión de propiedades utilizando estas naftas.

Según los ajustes de diferentes estructuras de RNA, los mejores desempeños en términos del menor valor de RMSE_{Tot} y el mayor valor de R^2_{Tot} están exhibidos por las redes 30:12:1–TanH y 6:12:1–TanH para el ajuste del contenido de Azufre y el RVP, respectivamente. La red 30:12:1–

TanH utiliza las primeras 30 áreas de mayor valor de la señal espectral NIR pretratada, mientras que la red 6:12:1–TanH requiere de los 6 primeros componentes principales derivados del PCA. El $RMSE_{Tot}$ de la red 30:12:1–TanH de ajuste del contenido de Azufre corresponde a 33.6 ppm, el cual se encuentra entre la repetibilidad y la reproducibilidad promedios de 10 ppm y 45 ppm, respectivamente, según la norma ASTM D4294–16. Respecto al RVP, el $RMSE_{Tot}$ de la red 6:12:1–TanH es menor que la repetibilidad y la reproducibilidad promedios de 0.21 y 0.36 psi, respectivamente.

La comparación de los mejores desempeños obtenidos en el presente trabajo con los reportados en la literatura abierta, para las propiedades contenido de azufre y RVP, soporta la capacidad de predicción cuantitativa de las redes neuronales artificiales 30:12:1–TanH y 6:12:1–TanH. Esta capacidad de predicción cuantitativa es derivada de un modelo circunstancial, basado en la similitud entre las condiciones de operación de los crudos nacionales en la GRB.

6. Recomendaciones

Entrenamientos y validaciones adicionales a las RNA desarrolladas son requeridas con periodicidad debido a su característica circunstancial para las propiedades mencionadas.

Continuar con los desarrollos de modelos RNA en la predicción de otras propiedades para las naftas de la GRB.

Referencias Bibliográficas

- ACIP. (2017). Mercado de combustibles en Colombia: así avanzan las importaciones y el consumo de gasolina, diésel y jet fuel en 2017. Informe Económico No 6. Vicepresidencia de Asuntos Económicos. Asociación Colombiana del Petróleo. <https://acp.com.co/web2017/es/asustos/economicos/125-informe-economico-octubre-mercado-de-combustibles-en-colombia-asi-avanzan-las-importaciones-y-el-consumo-de-gasolina-diesel-y-jet-fuel-en-2017/file>
- Andersen, V.F., Anderson, J.E., Wallington, T.J., Mueller, S.A., Nielsen, O.J. (2010) Vapor Pressures of Alcohol-Gasoline Blends. *Energy & Fuels*, 24, 3647–3654.
- Balabin, R. M., & Lomakina, E. I. (2011). Support vector machine regression (SVR/LS-SVM)—an alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data. *The Analyst*, 136(8), 1703.
- Barbosa, L.L., Sad, C.M.S., Morgan, V.G., Figueras, P.R., Castro, E.R.V. (2016). Application of low field NMR as an alternative technique to quantification of total acid number and sulphur content in petroleum from Brazilian reservoirs. *Fuel*, 176, 146-152.
- Berthold, M. & Hand, D.J. (2007). Intelligent Data Analysis: An Introduction. *Springer-Verlag*, 2nd edition, Berlin.
- Berry, M.J.A., Linoff, G. (1987). Data mining techniques. *John Wiley & Sons*. Nueva York.
- Blanco, M., Maspoch, S., Villarroja, I., Peralta, X., González, J.M., Torres, J. (2000). Determination of the penetration value of bitumens by near infrared spectroscopy. *Analyst*, 125, 1823–1828.
- Blanco, M., Maspoch, S., Villarroja, I., Peralta, X., González, J.M., Torres, J. (2001). Determination of physical properties of bitumens by use of near-infrared spectroscopy with neural networks. Joint modelling of linear and non-linear parameters. *Analyst*, 126, 378–382.

- Breitkreitz, M.C., Raimundo, Jr, I.M., Rohwedder, J.J.R., Pasquini, C., Dantas Filho, H.A., José, G.E., & Araújo, M.C.U. (2003). Determination of total sulfur in diesel fuel employing NIR spectroscopy and multivariate calibration. *The Analyst*, 128, 1204–1207.
- Brouillette, C., Smith, W., Shende, C., Gladding, Z., Farquharson, S., Morris, R.E., Cramer, J.A., Schmitgal, J. (2016). Analysis of Twenty-Two Performance Properties of Diesel, Gasoline, and Jet Fuels Using a Field-Portable Near-Infrared (NIR) Analyzer. *Applied Spectroscopy*, 70, 746–755
- Chung, H. (2007). Applications of Near-Infrared Spectroscopy in Refineries and Important Issues to Address. *Applied Spectroscopy Reviews*, 42, 251–285.
- Côcco, L.C., Yamamoto, C.I., von Meien, O.F. (2005). Study of correlations for physicochemical properties of Brazilian gasoline. *Chemometrics and Intelligent Laboratory Systems*, 76, 55-63.
- Cooper, J.B., Wise, K.L., Groves, J., Weltch, W.T. (1995). Determination of Octane Numbers and Reid Vapor Pressure of Commercial Petroleum Fuels Using FT-Raman Spectroscopy and Partial Least-Squares Regression Analysis. *Anal. Chem.*, 67, 4096-4100.
- De Oliveira, F.M., de Carvalho, L.S., Teixeira, L.S.G., Fontes, C.H., Lima, K.M.G., Câmara, A.B.F., Sales, R.V. (2017). Predicting Cetane Index, Flash Point, and Content Sulfur of Diesel–Biodiesel Blend Using an Artificial Neural Network Model. *Energy & Fuels*, 31, 3913–3920.
- Empresa Colombiana de Petróleos. (2009) Manual de operación de área externa. Refinería de Barrancabermeja. ECOPETROL,.
- Falla, F. S., Larini, C., Le Roux, G. A. C., Quina, F. H., Moro, L. F. L. y Nascimento, C. A. O. (2006). Characterization of crude petroleum by NIR. *Journal of Petroleum Science and Engineering*, 51, 127–137.
- Fenger, J. (2009). Air pollution in the last 50 years – From local to global. *Atmospheric Environment*, 43 13-22.

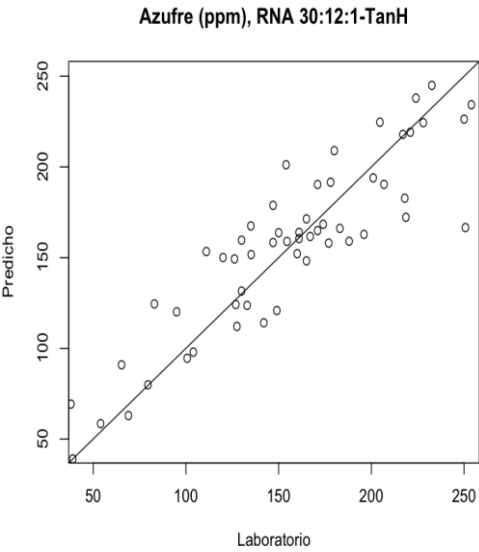
- Ferreiro-González, M., Ayuso, J., Álvarez, J.A., Palma, M., Barroso, C.G. (2014). New Headspace-Mass Spectrometry Method for the Discrimination of Commercial Gasoline Samples with Different Research Octane Numbers. *Energy Fuels*, 28, 6249–6254.
- Flecher, P.E., Welch, W.T., Albin, S., Cooper, J.B. (1997). Determination of octane number and reid vapor pressure in commercial gasoline using dispersive fiber-optic Raman spectroscopy. *Spectrochimica Acta Part A*, 53, 199-206.
- Gary, G.E.; Handwerk, M.; Kaiser, J. (2007). *Petroleum Refining technology and economics*, fifth edition, CRC press.
- García, A.P. (2014). Desarrollo de nuevas metodologías espectrales para el control analítico de productos y procesos petroquímicos y farmacéuticos. Tesis doctoral. Bellaterra: Universidad Autónoma de Barcelona.
- Glorfeld, L.W. (1996). *Expert Syst. Appl.*, 10, 37 – 54.
- Han, X., Wang, J.X., Liu, J. (2011). Quantitative Prediction of Total Acid Number in Aviation Kerosene by BP-ANN and NIR Spectroscopy. *Journal of Analytical Science*, 06.
- Han, Y., Zhang, Y., Xu, Ch., Hsu, Ch.S. (2018). Molecular Characterization of Sulphur-containing Compounds in Petroleum. *Fuel*, 221, 144-159.
- Hilera Gonzáles, J. R., & Martinez , V. J. (2000). *Redes Neuronales Artificiales Fundamentos, Modelos y Aplicaciones*. Madrid España: ALFAOMEGA GRUPO EDITOR, S.A.
- Kramer, K.E., Morris, R.E., Rose-Pehrsson, S.L., Cramer, J., Johnson, K.J. (2008). Statistical Significance Testing as a Guide to Partial Least-Squares (PLS) Modeling of Nonideal Data Sets for Fuel Property Predictions. *Energy & Fuels*, 22, 523–534.
- Isasi, P. & Galvan, I. (2004). *Redes neuronales artificiales enfoque práctico*. Madrid: *Pearson*.
- Lee, Y., Chung, H., Kim, N. (2006). Spectral Range Optimization for the Near-Infrared Quantitative Analysis of Petrochemical and Petroleum Products: Naphtha and Gasoline. *Applied Spectroscopy*, 60, 892-897.

- Li, W.M., Wu, G.Y., Lin, X.P. (2005). Study of the Gasoline Octane Number Model Using ANN Method with near Infrared Spectra Analysis. *Journal of Jiangsu Polytechnic University*, 03.
- Litani-Barzilai, I., Sela, I., Bulatov, V., Zilberman, I., Schechter, I. (1997). On-line remote prediction of gasoline properties by combined optical methods. *Analytica Chimica Acta*, 339 (1,2), 193.
- Liu, G., Wang, L., Qu, H., Shen, H., Zhang, X., Zhang, S., Mi, Z. (2007). Artificial neural network approaches on composition–property relationships of jet fuels based on GC–MS. *Fuel*, 86, 2551-2559.
- Martínez, E., Huertas, S., Ménez, H., Peña, J.L., Alcalde, A. (2008). Comparison of chemometric techniques applied to near infrared spectra for a gasoline blending control. *J. Near Infrared Spectrosc.* 16, 297–303.
- Majhi, A., Kukerti, V. S., Sharma, Y. K., Khanna, R., Datta, A. (2012). The detection of kerosene as an adulterant in gasoline. *Pet. Sci. Technol.* 2012, 30, 271–277.
- Mendes, G., Aleme, H.G., Barbeira, P.J.S. (2017). Reid vapor pressure prediction of automotive gasoline using distillation curves and multivariate calibration. *Fuel*, 187, 167-172.
- MinCIT. (2018). Informe de Gestión 2017. Sector Comercio, Industria y Turismo. Ministerio de Comercio, Industria y Turismo. Colombia. http://www.mincit.gov.co/loader.php?lServicio=Documentos&lFuncion=verPdf&id=83077&name=Informe_de_Gestion_2017_consolidado_sector_CIT_VF24012018.pdf&prefijo=file
- Morales-Medina, G. & Guzmán, A. (2012). Prediction of density and viscosity of Colombian crude oils from chromatographic data. *CT&F-Cienc. Tecn. Fut.*, 5, 57-74.
- Morris, R.E., Hammond, M.H., Cramer, A.C., Johnson, K.J., Giordano, B.C., Kramer, K.E., Rose-Pehrsson. (2009). Rapid Fuel Quality Surveillance through Chemometric Modeling of Near-Infrared Spectra. *Energy & Fuels*, 23, 1610 – 1618.

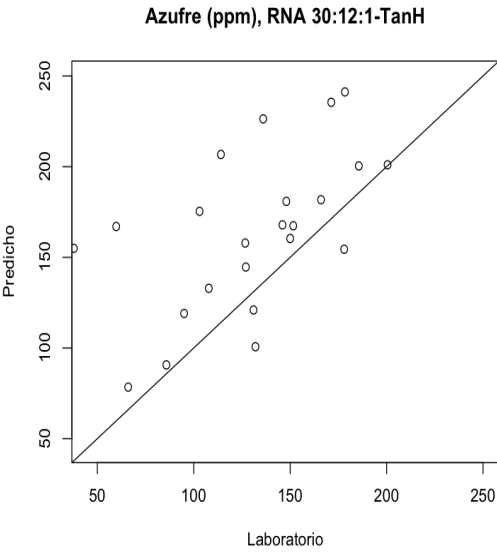
- Nascimento, M.H.C., Oliveira, B.P., Rainha, K.P., Castro, E.V.R., Silva, S.R.C., Filgueiras, P.R. (2018). Determination of flash point and Reid vapor pressure in petroleum from HTGC and DHA associated with chemometrics. *Fuel*, 234, 643-649.
- OCDE. (2017). Estudios Económicos de la OCDE: Colombia, Visión General. www.oecd.org/eco/surveys/economic-survey-colombia.htm
- Page, N.P., & Mehlman, M. (1989). Health Effects of Gasoline Refueling Vapors and Measured Exposures At Service Stations. *Toxicology and Industrial Health*, 5, 869–890.
- Patnaik, P. (2004). Dean's Analytical Chemistry Handbook, second ed., McGraw-Hill, New York, 2004, p. 7.23.
- Piloto-Rodríguez, R., Sánchez-Borroto, Y., Lapuerta, M., Goyos-Pérez, L. (2013). Prediction of the cetane number of biodiesel using artificial neural networks and multiple linear regression. *Energy Conversion and Management*, 65, 255–261.
- Pradelle, F., Braga, S.L., Martins, A.R.F.A., Turkovics, F., Pradelle, R.N.C. (2015). Gum Formation in Gasoline and Its Blends. *Energy & Fuels*, 29, 7753 – 7770.
- Qin, F., Li, Q., Zhan, H., Jin, W., Liu, H., Zhao, K. (2014). Probing the sulfur content in gasoline quantitatively with terahertz time-domain spectroscopy. *Science China Physics, Mechanics & Astronomy*, 57, 1404–1406.
- Reboucas, M.V., Santos, J.B. Pimentel, M.F., Teixeira, L.S.G. (2011). A novel approach for development of a multivariate calibration model using a Doehlert experimental design: Application for prediction of key gasoline properties by Near-infrared Spectroscopy. *Chemometrics and Intelligent Laboratory Systems*, 107, 185–193.
- Reuter, R.M., Benson, J.D., Burns, V.R., Gorse, R.A.Jr., Hochhauser, A.M., Koehl, W.J., Painter, L.J., Rippon, B.H., Rutherford, J.A. (1992). *Effects of Oxygenated Fuels and RVP on Automotive Emissions - AutoIOil Air Quality Improvement Program*. Auto/oil Air Quality Improvement Research Program (SP-920). *SAE TECHNICAL PAPER SERIES*, 920326.
- Rinnan, A. (2014). Pre-processing in vibrational spectroscopy – when, why and how. *Anal. Methods*, 6, 7124–7129.

- Rocha, J.T.C., Oliveira, L.M.S.L., Dias, J.C.M., Pinto, U.B., Marques, M. de L.S.P., Oliveira, B.P., de Oliveira, M.A.L. (2016). Sulfur Determination in Brazilian Petroleum Fractions by Mid-infrared and Near-infrared Spectroscopy and Partial Least Squares Associated with Variable Selection Methods. *Energy & Fuels*, 30, 698–705.
- Sandoval, D.J. (2017). *Modelo predictivo para la estabilidad a la oxidación de las gasolina de la gerencia refinería Barrancabermeja (GRB)*. Maestría en ingeniería química, Universidad Industrial de Santander. Bucaramanga, Colombia.
- Satya, S., Roehner, R.M., Deo, M.D., Hanson, F.V. (2007). Estimation of Properties of Crude Oil Residual Fractions Using Chemometrics. *Energy & Fuels*, 21, 998–1005.
- Skoog, D.A., Crouch, S.R., Holler, F.J. (2008). *Principios de análisis instrumental*. 6ta ed. Cengage Learning.
- Swingler, K. (1996). Applying Neural Networks, A Practical Guide. *Press Limited Oval Road* London NW1.
- Takeshita, E.V., Rezende, R.V.P., de Souza, S.M.A.G.U., de Souza, A.A.U. (2008). Influence of solvent addition on the physicochemical properties of Brazilian gasoline. *Fuel*, 87, 2168-77.
- Tan, K.M., Barman, I., Dingari, N.C., Singh, G.P., Chia, T.F., Tok, W.L. (2013). Toward the Development of Raman Spectroscopy as a Nonperturbative Online Monitoring Tool for Gasoline Adulteration. *Anal. Chem.*, 85, 1846–1851.
- Trygstad, W.M., Pell, R., Roberto, M. (2015). ISA 60th Analysis Division Symposium 2015, Galveston, TX USA. International Society of Automation. <https://chemometrix.files.wordpress.com/2015/05/inferential-motor-fuel-property-predictions-trygstad-pell-roberto.pdf>
- UPME. (2015). Evaluación de la contribución económica del sector de hidrocarburos colombiano frente a diversos escenarios de producción. http://www1.upme.gov.co/Hidrocarburos/Estudios%202014-2016/resumen_fedesarrollo_mme_final.pdf
- Wythoff, B.J. (1993). *Chem. Intel. Lab. Syst.*, 18, 115 – 155.

Zhang, Z. (2016). Neural Networks: further insights into error function, generalized weights and others. *Ann Transl Med*, 4, 300.



(a) Entrenamiento



(b) Validación

