

**CONSTRUCCIÓN DE UN MODELO PREDICTIVO PARA LA DETERMINACIÓN
DEL PORCENTAJE DE CARBÓN CONRADSON EN FONDOS DE VACÍO
USANDO REDES NEURONALES**

**KATHERINE CAMARGO REINA
STEFFANY OVIEDO HERNANDEZ**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE QUÍMICA
BUCARAMANGA**

2014

**CONSTRUCCIÓN DE UN MODELO PREDICTIVO PARA LA DETERMINACIÓN
DEL PORCENTAJE DE CARBÓN CONRADSON EN FONDOS DE VACÍO
USANDO REDES NEURONALES**

**KATHERINE CAMARGO REINA
STEFFANY OVIEDO HERNANDEZ**

Trabajo de Grado presentado para optar el título de Químico

**Director:
ENRIQUE MEJÍA OSPINO
Químico. Ph. D**

**Codirector:
JÁDER GUERRERO BERMÚDEZ**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE QUÍMICA
BUCARMANGA**

2014

DEDICATORIA

A mis padres Marta y Alberto por apoyarme, guiarme y estar conmigo en todo momento, en la peleas, en las alegrías; a ellos les debo la persona que soy hoy, les debo todos mis logros y todos mi triunfos, los amo con todo mi corazón.

A mi hermana Brigitte que siempre ha estado conmigo, en las buenas, en las malas, regañándome o dándome sus sabios consejos. Al amor de mi vida, el niño por el cual doy todo en este mundo y todo lo que hago lo hago por y para él, mi hermanito Juan David, te amo.

A mis compañeros y amigos de carrera, aquellos que de una u otra forma recorrieron este camino conmigo, que me ayudaron, me apoyaron, y me dieron ánimos para continuar cuando decaída, gracias por compartir momentos alegres, tristes, amargos siempre los recordare. También a mis amigos de otras carreras, que poco a poco fui conociendo, siempre ocuparan un lugar en mi corazoncito.

Los quiero mucho.

A mis amigos del trabajo, los cuales siempre me apoyaron y me dieron fuerzas, ánimos para poder cumplir esta meta en mi vida, a mis nenas locas y a una personita muy especial que me aguanto y trasnocho conmigo, los quiero y los llevo en mi corazón.

Y finalmente a mi compañerísima, amiga, hermana la Churca, la cual desde el inicio estuvo conmigo a lo largo de la carrera, compartiendo tantas cosas, tantas anécdotas, tantas locuras, tantas risas, llantos, amarguras, peleas, en fin gracias por todo señorita, porque gracias a su comprensión y paciencia logramos culminar satisfactoriamente esta meta.

Katherine

DEDICATORIA

*Esta dedicatoria la hago en nombre de Dios por su inmenso amor.
A mi gran familia, a mis tantas hermanas y hermanos. En especial a mi padre y mi
bella madre que siempre ha estado a mi lado, que ha dado mucho por mí, por
tener palabras de aliento en los momentos alegres y tristes.
A aquellas personas que han sabido ser un faro en mi vida. A la familia Camargo
Reina por hacerme sentir que hago parte de esta amorosa familia.
A los tantos compañeros de estudio que he tenido, a los muchos maestros del
ejercicio de la vida.
A mi gran amiga y compañera Kate la reina, por decidir ser mi apoyo en los
buenos y malos momentos, por la paciencia y buena actitud que siempre tiene, por
ser mi compañera de trabajo.*

Steffany

AGRADECIMIENTOS

Agradecemos de todo corazón a:

A Dios y a la Virgen del Carmen, que siempre han guiado nuestros pasos.

A nuestras familias, amigos, colegas y demás personitas que nos apoyaron durante la carrera.

A nuestro director el Dr. Enrique Mejía Ospino por brindarnos su apoyo en la realización de este trabajo.

Al codirector el Dr. Jader Guerrero por su inmensa colaboración, por sus enseñanzas, por su apoyo y por todas las reuniones que nos brindó.

Al calificador el Dr. Juan Carlos Poveda, por las sugerencias y correcciones en la escritura del trabajo.

Al Laboratorio de Espectroscopia Atómica y Molecular (LEAM) y a los miembros del mismo, por darnos la oportunidad de realizar este trabajo de pregrado.

A nuestros amigos y compañeros, los cuales aportaron un granito de arena para culminar con éxito este trabajo, gracias por su compañía y ayuda incondicional.

CONTENIDO

	Pág.
INTRODUCCIÓN	17
1. TEORÍA	19
1.1. PETRÓLEO	19
1.1.1. Generalidades	19
1.1.2. Fondos De Vacío	21
1.1.3. Carbón Conradson	22
1.2. ESPECTROSCOPIA EN EL INFRARROJO	23
1.2.1. Generalidades	23
1.2.2. Reflectancia Total Atenuada	25
1.3. QUIMIOMETRÍA	26
1.3.1. Construcción De Modelos Predictivos.	26
1.3.2. Análisis Por Componentes Principales.	28
1.4. REDES NEURONALES	31
1.4.1. Antecedentes.	31
1.4.2. Red Neuronal Biológica	32
1.4.3. Red Neuronal Artificial.	33
1.4.4. Elementos De Una Red Neuronal	33
1.4.5. Características De Las Redes Neuronales	35
1.4.6. Perceptrón Multicapa.	36
2. ESTADO DEL ARTE	38
3. METODOLOGÍA	42
3.1. OBTENCIÓN Y PROCESAMIENTO DE ESPECTROS	44
3.2. SELECCIÓN DE DATOS PARA LA RED NEURONAL	46
3.3 SELECCIÓN DE LA ARQUITECTURA DE LA RED NEURONAL	49
3.4 ENTRENAMIENTO DE LA RED NEURONAL	49

3.5 ESTIMACIÓN DEL DESEMPEÑO DE LA RED NEURONAL	50
4. RESULTADOS	51
4.1. ANÁLISIS ESPECTRAL	51
4.2. ANÁLISIS POR COMPONENTES PRINCIPALES (PCA)	52
4.3 REDES NEURONALES	55
5. CONCLUSIONES	62
6. RECOMENDACIONES	63
REFERENCIAS BIBLIOGRÁFICAS	64
BIBLIOGRAFÍA	69
ANEXOS	71

LISTA DE FIGURAS

	Pág.
Figura 1. Fenómenos de absorción, transmisión y reflexión de la radiación electromagnética al interaccionar con la materia. Tomada de la referencia [13]	24
Figura 2. Modo operación de un cristal ATR. Tomada de la referencia [15]	25
Figura 3. Neurona Biológica. Tomada de la referencia [26]	33
Figura 4. Esquema de una red neuronal. Tomada de la referencia [21]	34
Figura 5. Esquema de una red Perceptrón Multicapa. Tomada de la referencia [27]	37
Figura 6. Espectros de Calibración	43
Figura 7. Espectros de Validación	43
Figura 8. Espectros de las muestras asignadas	45
Figura 9. Frecuencias de vibración en el infrarrojo medio.	51
Figura 10. Reconstrucción de los espectros.	54
Figura 11. Varianza acumulativa.	55
Figura 12. Relaciones entre los valores conocidos y simulados por la red neuronal para la calibración, para el error más bajo y más alto.	58
Figura 13. Relaciones entre los valores conocidos y simulados por la red neuronal para la validación, para el error más bajo y más alto	59

LISTA DE TABLAS

	Pág.
Tabla 1. Composición del petróleo. Tomada de la referencia [4]	20
Tabla 2. Gravedad API de los crudos. Tomada de la referencia [3]	20
Tabla 3. Regiones de la Espectroscopia en el Infrarrojo. Tomada de la referencia [12]	24
Tabla 4. Porcentaje de CCR de las fracciones, crudos, condensados y residuos	46
Tabla 5. Porcentaje de CCR de los fondos de vacío	48
Tabla 6. Grupos funcionales en el IR Medio	52
Tabla 7. Mejores resultados de la Red Neuronal.	58
Tabla 8. . Resultados de la Red Neuronal	60

LISTA DE ANEXOS

	Pág.
ANEXO A. Rutina Matlab de los componentes principales	71
ANEXO B. Rutina de la Red Neuronal	75

RESUMEN

TÍTULO: CONSTRUCCIÓN DE UN MODELO PREDICTIVO PARA LA DETERMINACIÓN DEL PORCENTAJE DE CARBON CONRADSON EN FONDOS DE VACÍO USANDO REDES NEURONALES*

AUTOR: Katherine Camargo Reina, Steffany Oviedo Hernandez**

PALABRAS CLAVES: Fondos de Vacío, PCA, ATR-IR, Redes Neuronales

La espectroscopia infrarroja es una herramienta versátil que se aplica a la determinación cualitativa y cuantitativa de las especies moleculares de todo tipo. Una de las aplicaciones del infrarrojo medio es la espectroscopia de reflexión interna, dentro de estas aplicaciones se encuentra la Reflexion Total Atenuada (ATR), el principio de la técnica se basa en la medida de los cambios que ocurren en un haz infrarrojo reflejado internamente cuando interactúa con cualquier material absorbente; este haz infrarrojo se refleja desde la superficie interna del cristal y crea una onda evanescente, que se proyecta ortogonalmente a la muestra en íntimo contacto con el cristal ATR.

Las Redes Neuronales Artificiales (RNA) son técnicas sofisticadas capaces de modelar funciones complejas, hacen parte de una propuesta ágil ya que permiten el análisis en tiempo real, posee la capacidad de asociación, capacidad de aprendizaje y a su vez permite reducir el tiempo de análisis. Las RNA se están utilizando con éxito en un amplio campo de aplicación, en diversas áreas como ingeniería, física, finanzas, medicina y con fines predictivos, de clasificación y control.

Aquí se presenta un modelo predictivo para la determinación del porcentaje de carbón conradson en fondos de vacío de crudos colombianos a partir de espectros ATR, la utilización de RNA y el modelo de Perceptron Multicapa. Las entradas utilizadas en la red son las longitudes de ondas de todo el espectro. Se exploraron diferentes arquitecturas y en cada una de ellas se variaron la cantidad de capas ocultas y el número de iteraciones. Las métricas de evaluación del desempeño de la red en la predicción son: del error estándar de calibración (RMSEC) y el error estándar de predicción (RMSEP).

* Trabajo de Grado

** Facultad de Ciencias, Escuela de Química, Director: MEJÍA Enrique, Codirector: GUERRERO Jader

ABSTRACT

TITLE: CONSTRUCTION OF PREDICTIVE MODEL FOR THE DETERMINATION OF CARBON CONRADSON VACUUM USING NEURAL NETWORKS *

AUTHOR: Katherine Camargo Reina-Steffany Oviedo Hernandez **

KEY WORDS: Vacuum Residue, PCA, ATR-IR, Neural Networks.

The infrared spectroscopy is a versatile tool that is applied to the qualitative and quantitative determination of molecular species of all kinds. One application is mid-infrared spectroscopy internal reflection within these applications is the attenuated total reflection (ATR), the principle of the technique is based on measuring changes occurring in an infrared beam internally reflected when interacts with any absorbent material, that the infrared beam is reflected from the inner surface of the crystal and creates an evanescent wave that extends orthogonally to the sample in close contact with the ATR crystal.

Artificial Neural Networks (ANN) are capable of modeling techniques sophisticated complex functions are part of a proposal allowing agile and real-time analysis, has the ability to partner, ability to learn and in turn reduces the time of analysis. ANNs are being applied successfully in a wide range of applications in various areas such as engineering, physics, finance, medicine and predictive purposes, classification and control.

Here is a predictive model for determining the percentage of funds Conradson Carbon Colombian crude vacuum from ATR spectra and the use of RNAs and Multilayer Perceptron model. The inputs used in the network are the wavelengths of the spectrum. Were explored in different architectures where each is varied the amount of hidden layers, the number of iterations. The evaluation of the performance metrics of the network in the prediction are: the standard error calibration (RMSEC) and standard error of prediction (RMSEP).

* Trabajo de Grado

** Facultad de Ciencias, Escuela de Química, Director: MEJÍA Enrique, Codirector: GUERRERO Jáder

INTRODUCCIÓN

El acelerado desarrollo industrial y la sobre explotación de los combustibles fósiles ha llevado actualmente a tener menores reservas de crudos livianos, teniéndose que recurrir a la explotación de las reservas de crudos pesados, los cuales aportan mayor cantidad de fondos de vacío, que son catalogados como una fracción de poco interés, es así como se hace necesaria la caracterización de los respectivos residuos en la destilación del petróleo crudo. El aumento de este material ha llevado a plantear nuevas investigaciones que enfoquen hacia su aprovechamiento.

Las técnicas espectroscópicas como la espectroscopia infrarroja con reflectancia total atenuada, se presenta como una técnica ágil en la identificación y caracterización de los crudos, puesto que no requiere de prolongados tiempos de medición, permitiendo hacer análisis rápidos; otra cualidad a resaltar es que requiere de pocas cantidades de muestra e incluso la no utilización de reactivos, ya que la muestra no requiere ser tratada previamente a las mediciones, lo que la hace altamente atractiva para la realización de análisis sobre muestras complejas.

En este trabajo se propone un modelo predictivo para la determinación de Carbón Conradson (CCR) de algunos fondos de vacío de muestras Colombianas. El Carbón Conradson está definido como la cantidad de residuo carbonoso después de someter una muestra a un fuerte estrés térmico; el contenido de CCR puede correlacionar con otras propiedades del petróleo, es decir, también presenta indicaciones de volatilidad y formación de coque; además este residuo carbonoso se puede asociar a la relación atómica C/H, el contenido de nitrógeno, azufre, la viscosidad y también con los rendimientos de coque para las fracciones de materias primas como los saturados y aromáticos, las resinas y los asfaltenos. [1]

La quimiometría se ha convertido en una herramienta en la creación de modelos predictivos con datos obtenidos experimentalmente. El uso de herramientas computacionales como las Redes Neuronales Artificiales, permite hacer consideraciones en tiempo real, asimismo, las redes neuronales artificiales son un modelo que tiene la facilidad de aprendizaje sobre el sistema de estudio; una característica que ofrece esta herramienta es ser más rápida que los métodos de simulación convencionales.

El planteamiento del modelo de cálculo simulado requiere de la selección de variables y el mejoramiento de los datos iniciales, lo que se logra con el método de componentes principales, que es una reducción dimensional de la información representativa contenida en el rango espectral, convirtiendo este método de operación en una ventaja, ya que se encarga de la eliminación del ruido al utilizar el espectro completo en la construcción de un modelo predictivo.

El desarrollo del presente trabajo permitió analizar los diferentes espectros tomados mediante la técnica de espectroscopia infrarrojo con reflectancia total atenuada, clasificar las muestras teniendo en cuenta el valor del porcentaje de Carbón Conradson, utilizar técnicas quimiométricas como el análisis de componentes principales y las redes neuronales, con el fin de entrenar, validar y obtener la mejor arquitectura de la red neuronal; dando como resultado un buen desempeño tanto de los valores introducidos en la red como el error predicho por la misma.

1. TEORÍA

1.1. PETRÓLEO

1.1.1. Generalidades. El petróleo crudo es una mezcla compleja de hidrocarburos líquidos y sólidos, la mayoría de las veces contiene hidrocarburos gaseosos, además de pequeñas fracciones de compuestos orgánicos de azufre y nitrógeno; el oxígeno también está presente como ácidos orgánicos (ácidos nafténicos) en pequeñas cantidades de fenoles y dióxido de carbono disuelto. El petróleo crudo por lo regular se encuentra asociado con agua, el cual contiene disueltos materiales inorgánicos, en algunos casos, formando emulsiones relativamente estables. [2]

El petróleo es una sustancia aceitosa, generalmente de color oscuro, a la que por sus compuestos de hidrógeno y carbono se le denomina hidrocarburo. Ese hidrocarburo puede estar en estado líquido o en estado gaseoso. En el primer caso es un aceite al que también se le dice crudo. En el segundo se le conoce como gas natural.

En la composición del petróleo se encuentran fundamentalmente hidrocarburos gaseosos, líquidos y sólidos, estando estos últimos disueltos o suspendidos en los líquidos. En términos generales, el petróleo se compone de elementos como el carbón, hidrógeno, azufre y nitrógeno, que se presentan en la Tabla 1.

Tabla 1. Composición del petróleo. Tomada de la referencia [4]

Elemento	% Peso
Carbón	84-87
Hidrógeno	11-14
Azufre	0-2
Nitrógeno	0,2

Puede considerarse que un 95% del petróleo esta constituido por hidrocarburos, es decir, sustancias formadas por carbono e hidrógeno, formado principalmente por moléculas hidrocarbonadas. La densidad de una crudo será tanto menor cuanto mayor sea la relación H/C, esta densidad puede oscilar entre 0,7 y 1, expresándose con mucha frecuencia en grados API (American Petroleum Institute) cuyo valor varía entre 10 y 5, este parámetro internacional hace referencia a que cuantos más grados API tenga un petróleo, menos refinación se requiere para obtener de él los productos más costosos que son llamados “blancos”, entre ellos la gasolina. Por eso los petróleos con mayor API son considerados los de mejor calidad.

Tradicionalmente, los petróleos de mejor calidad son aquellos clasificados como “livianos”, los que tiene más de 26 grados API. Los “intermedios” se sitúan entre 20 y 26 grados API y los “pesados” por debajo de 20 grados API, como se puede observar en la Tabla 2. [3] [4]

Tabla 2. Gravedad API de los crudos. Tomada de la referencia [3]

Crudo	Densidad (g/cm ³)	API
Extra pesado	>1,000	>10,0
Pesado	1,000-0,934	10,0-20,0
Mediano	0,934-0,876	20,0-30,0
Livianos	<0,876	>30,0

Las tres series de la química orgánica: saturados o parafínicos, no saturados o alifáticos y los aromáticos o bencénicos, se encuentran presentes en el petróleo. Los saturados o parafínicos, denominados así por su poca reactividad, conforman la mayor parte de los hidrocarburos del petróleo. Dentro de un rango de variación dado por las distintas procedencias, puede decirse que un 50-70% está constituido por esta serie, correspondiendo la mayor parte a los nafténicos o cicloparafinas; los hidrocarburos no saturados, por su reactividad no suelen encontrarse en los petróleos naturales en cantidades apreciables, pero si adquieren gran importancia durante el procesamiento por ser formados en los procesos de refinación; por último esta la serie de los aromáticos o bencénicos, los cuales constituyen una familia numerosa por la formación de sistemas de anillos condensados, estando presente en la fracciones más pesadas del petróleo. [5]

1.1.2. Fondos De Vacío. La refinación del petróleo arroja productos y subproductos, dentro de los subproductos se hallan los fondos de vacío, estos son obtenidos por someter el crudo de carácter pesado y extra pesado a condiciones de presión menores de 20 mm Hg. Estos residuos han sido considerados como desechos de poco valor industrial; que debido al progresivo decaimiento en las reservas de crudo liviano se han tornado en potenciales fuentes de energía, así como en la producción de materiales carbonosos especializados como son las fibras carbonosas, ánodos industriales y carbones activados.

La composición química de los fondos de vacío es variada y muy compleja, en esencia son una mezcla de compuestos que varían en un amplio rango de pesos moleculares, desde hidrocarburos parafínicos y nafténicos hasta hidrocarburos aromáticos, también contienen en menor cantidad heteroátomos como azufre, nitrógeno, y oxígeno conformando grupos funcionales como ácidos carboxílicos, aldehídos, cetonas, tiofenos, anillos pirrólicos y bases nitrogenadas como piridina; además, incluyen trazas de metales como níquel, vanadio, hierro, magnesio y

calcio, presentes en forma de óxidos, sales en estructuras organometálicas tipo porfirina. [6]

La carbonización de los fondos de vacío de petróleo a temperaturas alrededor de 700°C produce un material carbonoso tipo coque. Durante este proceso, ocurre una disminución en el contenido de las fracciones de saturados, resinas y aromáticos, ya que estas sufren la conversión a gases como H₂, CH₄ y CO, en tanto que las moléculas de alto peso molecular como los asfáltenos producen un incremento en el contenido de insolubles en tolueno o semicoque, debido a las reacciones de condensación que promueven la formación del coque.

1.1.3. Carbón Conradson. El residuo de carbón está definido como la cantidad de residuo carbonoso después de someter la muestra de un producto del petróleo a un fuerte estrés térmico. Los métodos para evaluar la cantidad de carbón son: CCR (ASTM D189) [7], y la cantidad de Carbón Ramsbottom (CRB) (ASTM D524) [8], es una medida estándar de la tendencia a la formación de carbón para la industria del petróleo.

La prueba convencional de CCR, involucra el calentamiento de 10 gr de muestra en un crisol, calentado con un mechero Bunsen, luego se mide el residuo como un porcentaje de peso de la muestra inicial. Esta técnica para evaluar la cantidad carbón se puede remplazar por la medición de micro carbón (ASTM D4530), la prueba está diseñada para arrojar los mismos valores de CCR, pero la diferencia es que requiere menor cantidad de muestra. El porcentaje de la muestra no volatilizada en el crisol durante la prueba usando la norma ASTM D189 o la norma ASTM D4530 es considerada como residuo de CCR.

La cantidad de residuo carbonoso de un combustible depende de los procesos de refinación empleados. Para los combustibles por destilación directa, el valor típico de CCR es 10% a 12% en peso y para los combustibles del refinado secundario la

cantidad depende de la severidad de los procesos aplicados, en algunas zonas se puede alcanzar un 20% en peso. [9] [10]

1.2. ESPECTROSCOPIA EN EL INFRARROJO

1.2.1. Generalidades. La espectroscopia en el infrarrojo es una técnica que se aplica a la determinación cualitativa y cuantitativa de especies moleculares de todo tipo considerando sus características vibracionales, esta técnica se basa en la interacción entre la radiación electromagnética y las moléculas. Las moléculas absorben radiación del haz incidente a diferentes frecuencias, estas frecuencias son dependientes de las interrelaciones espaciales de los átomos contenidos en la unidad molecular. [11]

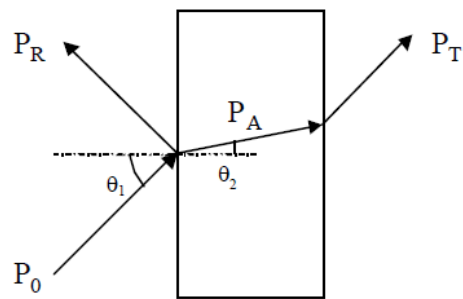
La espectroscopia en el infrarrojo se dividen en tres grandes categorías relacionadas con tres regiones espectrales: el Infrarrojo Lejano, donde se producen las absorciones debidas a cambios rotacionales y el cual se encuentra en una longitud de onda de 5×10^4 - 10^6 nm; el Infrarrojo Medio tienen lugar las vibraciones fundamentales y una longitud de onda de 2500- 5×10^4 nm y el Infrarrojo Cercano donde se producen absorciones debidas a sobretonos y combinaciones de las bandas fundamentales y una longitud de onda de 700-2500nm, también en la Tabla 3 podemos observar otras características de cada una de las regiones en el infrarrojo.

Tabla 3. Regiones de la Espectroscopia en el Infrarrojo. Tomada de la referencia [12]

Regiones espectrales	Longitud de onda (nm)	Número de onda (cm ⁻¹)	Tipo de medida
Infrarrojo Cercano (NIR)	700-2500	14300-4000	Reflectancia difusa Absorción
Infrarrojo Medio (IR)	2500-5x10 ⁴	4000-200	Absorción Reflectancia Emisión
Infrarrojo Lejano	5x10 ⁴ -10 ⁶	200-10	Absorción

Cuando la radiación incide en la muestra, como se puede ver en la Figura 1, ésta puede sufrir diferentes fenómenos: absorción, transmisión y reflexión. La intensidad de la luz transmitida a través de la muestra (P_T) es menor que la intensidad incidente (P_0). Una parte de esta intensidad incidente se ha reflejado (P_R), mientras que otra parte ha sido absorbida por la sustancia (P_A).

Figura 1. Fenómenos de absorción, transmisión y reflexión de la radiación electromagnética al interaccionar con la materia. Tomada de la referencia [13]



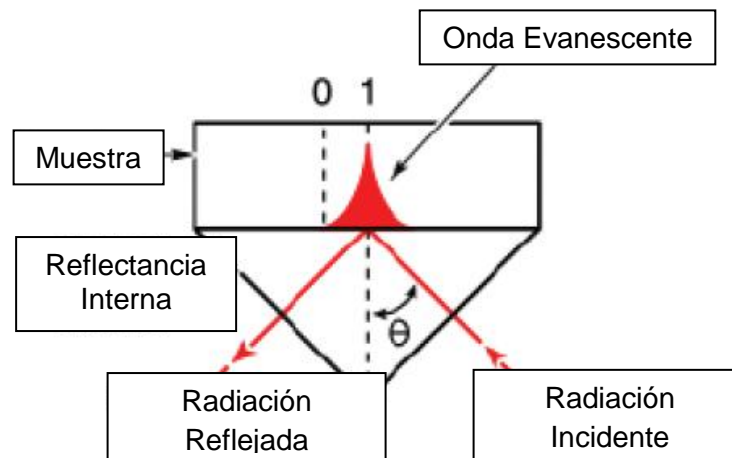
$$P_0 = P_A + P_T + P_R$$

1.2.2. Reflectancia Total Atenuada. El principio de la técnica de reflectancia total atenuada (ATR) se basa en la medida de los cambios que ocurren en un haz infrarrojo reflejado internamente cuando interactúa con cualquier material absorbente. Este haz infrarrojo se refleja desde la superficie interna del cristal y crea una onda evanescente, que se proyecta ortogonalmente a la muestra en íntimo contacto con el cristal ATR. [14]

Si el medio menos denso absorbe la radiación evanescente, se produce una atenuación del haz en las longitudes de onda de las bandas de absorción. La Figura 2, muestra el fenómeno donde se presenta solo reflexión ATR.

Para un uso adecuado de la celda ATR, es necesario tener en cuenta los factores experimentales que afectan el espectro final: índice de refracción del cristal ATR y de la muestra, profundidad de penetración, número de reflexiones dentro del cristal ATR y calidad del contacto entre la muestra y el cristal. [15]

Figura 2. Modo operación de un cristal ATR. Tomada de la referencia [15]



El aspecto más importante es el índice de refracción del cristal ATR y la muestra, esto en virtud de la siguiente Ecuación 1, donde n_2 es el índice de refracción de la muestra, n_1 es el índice de refracción del cristal y θ_c es el ángulo crítico.

Ecuación 1.

$$\theta_c = \sin^{-1} \frac{n_2}{n_1}$$

Al exceder el ángulo crítico, se observa un espectro ATR puro; cuando el ángulo crítico no es conocido, se observa un resultado combinado entre ATR y reflectancia externa. Esto se presenta cuando el ángulo de incidencia del rayo es muy bajo, el índice de refracción del cristal es muy bajo y el índice de refracción de la muestra es muy alto o cuando se presenta una combinación de estos tres factores.

Si un haz de radiación monocromática atraviesa un prisma de forma que el medio más denso está formado de algún material de elevado índice de refracción que transmite la radiación infrarroja, y el medio enrarecido por una muestra, entonces la muestra absorberá algo de la radiación infrarroja incidente. [16]

1.3. QUIMIOMETRÍA

1.3.1. Construcción De Modelos Predictivos. La norma ASTM E1655-05 define la calibración multivariante como un proceso para la creación de un modelo que relaciona concentración de los componentes o propiedades y las absorbancias de un conjunto de muestras de referencia en más de una longitud de onda. Esta norma sugiere tener en cuenta los siguientes pasos para la construcción de modelos de calibración multivariable con espectroscopia infrarroja:

- Selección del conjunto de muestras de calibración: las muestras deben tener la máxima variabilidad física y química y deben ser representativas.
- Caracterización de las muestras de calibración: mediante la determinación de las concentraciones o propiedades para las muestras de calibración.

- Toma de espectros infrarrojo: la toma de los espectros se hace teniendo en cuenta las condiciones experimentales de adquisición espectral y el tratamiento de las muestras.
- Calcular el modelo matemático: se lleva a cabo el pre-tratamiento de los espectros, seguido de la aplicación de técnicas de regresión o correlación, entre las cuales están las lineales y las no lineales.
- Validación del modelo de calibración: se desarrolla aplicando el modelo de calibración obtenido a un grupo de muestras diferentes (validación) para después comparar estadísticamente con los valores de referencia. La validación puede ser externa o cruzada; la externa es aquella en la cual se emplean muestras diferentes a las empleadas en la calibración, mientras que la cruzada es aquella en la cual se emplean muestras que están incluidas en la calibración del modelo.
- La selección del número óptimo de componentes requiere evaluar el comportamiento del error estándar de calibración (RMSEC) y el error estándar de predicción (RMSEP).
 - RMSEC: es una medida de la desviación estándar de los residuales obtenidos por la diferencia entre los valores observados y predichos por el modelo para las muestras de calibración; donde x es el valor predicho por la red neuronal, x el valor conocido de la propiedad y N el número de muestras.

Ecuación 2. Error estándar de calibración

$$RMSEC = \frac{\sqrt{\sum_{i=1}^N (x - \hat{x})^2}}{N}$$

- RMSEP: es una medida de la desviación estándar de los residuales obtenidos por la diferencia entre los valores observados y predichos por el modelo para las muestras de validación; donde \hat{y} es el valor predicho por la red neuronal, y el valor conocido de la propiedad y N el número de muestras.

Ecuación 3. Error estándar de predicción.

$$RMSEP = \frac{\sqrt{\sum_{i=1}^N (y - \hat{y})^2}}{N}$$

- Aplicación del modelo para el análisis de las muestras desconocidas: el modelo de calibración obtenido es utilizado en muestras desconocidas. [17] [18] [19]

1.3.2. Análisis Por Componentes Principales. Esta técnica fue inicialmente desarrollada por Pearson a finales del siglo XIX y posteriormente fue estudiada por Hotelling en los años 30 del siglo XX. Los componentes son combinaciones lineales de las variables originales y se espera que, solo unas pocas (las primeras) recojan la mayor parte de la variabilidad de los datos, obteniéndose una reducción de la dimensión en los mismos.

El propósito fundamental de la técnica consiste en la reducción de la dimensionalidad de los datos con el fin de simplificar el problema en estudio y las nuevas variables son combinaciones lineales de las anteriores que se van construyendo según el orden de importancia en cuanto a la variabilidad total que recogen de la muestra.

Suponemos que las p columnas de X generan un espacio p dimensional, de forma que los n individuos pueden representarse en dicho espacio en lo que llamaremos una hipernube. La transformación es de hecho, una rotación en el espacio p -dimensional. El espacio generado por la primera q componentes es entonces, un subespacio vectorial q -dimensional del espacio p -dimensional original.

Los componentes principales se pueden obtener por varios métodos alternativos, que son:

1. Buscando la combinación lineal de las variables que maximiza la variabilidad. (Hotelling)
2. Buscando el subespacio de mejor ajuste por el método de los mínimos cuadrados. (Pearson)
3. Minimizando la discrepancia entre las distancias euclídeas entre los puntos calculadas en el espacio original y en el subespacio de baja dimensión.
4. Mediante regresiones alternadas, métodos biplot.

Al utilizar métodos espectroscópicos se obtienen respuestas de cada muestra para cientos o miles de variables, en este caso longitudes de onda. La matriz $R_{I \times J}$ (I filas por J columnas) representa las respuestas de I muestras analizadas a J longitudes de onda diferentes. El elevado número de variables J que caracterizan cada muestra impiden un análisis y representación gráfica sencillos de las muestras.

El análisis en componentes principales proporciona una aproximación a la matriz R como un producto de dos matrices: la matriz de *scores* T y la matriz de *loadings* P , que captura la estructura de los datos de R . Los *scores* capturan la estructura de las filas o lo que es lo mismo, las relaciones entre objetos (muestras) y los *loadings* retienen la relación existente entre las variables.

$$R = TP^T + E$$

En este caso, el método de descomposición en componentes principales es muy útil, porque permite representar la variabilidad presente en R; este análisis se aplica frecuentemente cuando se trabaja con datos colineales, es decir, aquellos datos que se encuentran en la misma recta y los cuales se pueden condensar en un conjunto más pequeño de A variables, ya que poseen la información principal de las variables R. Cada una de las variables A se le denomina componente principal, entonces el nuevo conjunto de factores se representa como la suma de A términos de la forma $t_a p_a^T$, que corresponden a cada una de las columnas de las matrices T y P.

$$R = t_1 p_1^T + t_2 p_2^T + \dots + t_a p_a^T + E$$

El primer componente principal es aquél que explica una mayor cantidad de la información contenida en R (maximiza $p_1^T R^T R p_1 = t_1^T t_1$). El siguiente factor p_2 , explica la máxima información de R no contenida en p_1 (maximiza $p_2^T R^T R p_2 = t_2^T t_2$) es ortogonal al primer componente principal, esto es ($t_1^T t_2 = 0$). Los sucesivos factores explican cada vez menos información y son ortogonales a los anteriores. Las condiciones de ortogonalidad de *scores* y *loadings* se pueden asumir como:

$$P^T P = I$$

$$T^T T = \text{diag } t_a^T t_a \quad a = 1, 2 \dots A$$

Si la matriz R esta centrada, entonces $t_a^T t_a$ representa los valores propios de la matriz $R^T R$, y los vectores p_a representan sus vectores propios. Esto significa que los *loadings* satisfacen la ecuación:

$$R^T R p_a = p_a \hat{\sigma}_a$$

La magnitud de los valores propios indica la cantidad de variabilidad (información) que retiene cada uno de los componentes principales.

El análisis de componentes principales es muy útil para la interpretación de datos multivariantes. Por un lado, la representación de los scores permite establecer relaciones entre las muestras, permitiendo así la detección de muestras discrepantes y agrupaciones. Por otra parte, los loadings permiten comparar y estudiar la influencia de las distintas variables (longitudes de onda en este caso). [20]

1.4. REDES NEURONALES

1.4.1. Antecedentes. Las redes neuronales son una herramienta que simula el proceso cognitivo del cerebro y es de mucha utilidad para resolver aquellos problemas que no pueden ser solucionados por los métodos numéricos convencionales. Las RNA, funcionan como sistemas dinámicos no lineales que aprenden a reconocer patrones a través del entrenamiento.

Alan Turing, en 1936, fue el primero en estudiar el cerebro como una forma de ver el mundo de la computación; sin embargo, los primeros teóricos que concibieron los fundamentos de la computación neuronal fueron Warren McCulloch, un neurofisiólogo, y Walter Pitts, un matemático, quienes en 1943 propusieron una teoría acerca de la forma de trabajar de las neuronas. [21] [22]

En 1957, Frank Rosenblatt comenzó el desarrollo del Perceptrón, el perceptrón es la más antigua red neuronal, y se usa hoy en día de varias formas para la aplicación como reconocedor de patrones. Este modelo era capaz de generalizar; es decir, después de haber aprendido una serie de patrones era capaz de reconocer otros similares, aunque no se le hubieran presentado anteriormente. [23]

Actualmente, son numerosos los trabajos que se realizan y publican en pro del desarrollo de las redes neuronales teniendo en cuenta las nuevas aplicaciones y las empresas encargadas del diseño de Software especializados. [24]

1.4.2. Red Neuronal Biológica. El modelo neuronal artificial fue inspirado en la estructura y funcionamiento del sistema nervioso, donde la neurona es el elemento fundamental. El cerebro humano se compone principalmente de más de 10 mil millones de neuronas, cada una conectada a aproximadamente a otras 10,000 neuronas.

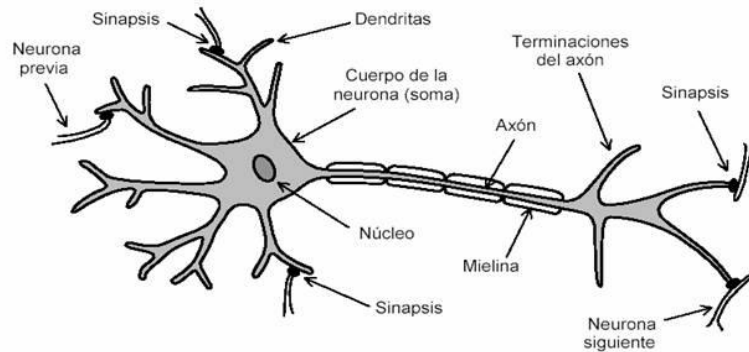
Cada neurona recibe entradas electroquímicas de otras neuronas, si la suma de estas entradas eléctricas es suficientemente potente para activar la neurona, que transmite una señal electroquímica a lo largo del axón y esta señal pasa a las otras neuronas.

Las neuronas están constituidas por un cuerpo celular más o menos esférico de 5 a 10 micras de diámetro, del que salen una rama principal, el *axón*, y varias ramas más cortas, llamadas *dendritas*. Las dendritas y el cuerpo celular reciben señales de entrada; el cuerpo celular las combina e integra y emite señales de salida; el axón transporta esas señales a los terminales axónicos que se encargan de distribuir información a un nuevo conjunto de neuronas, en la Figura 3 se observa la estructura de una neurona biológica. [25]

Sin embargo, debe tenerse en cuenta que las redes neuronales artificiales sólo representan simplificados modelos formales de las neuronas biológicas y sus interconexiones, sin hacer ningún intento de modelar el sistema biológico en sí.

Su importancia reside en el hecho de que las redes artificiales son herramientas computacionales inspiradas en el funcionamiento del cerebro para resolver problemas complejos. [26]

Figura 3. Neurona Biológica. Tomada de la referencia [26]



1.4.3. Red Neuronal Artificial. Una red neuronal artificial (RNA) es un paradigma computacional que intenta reproducir el comportamiento del sistema nervioso central, donde los dispositivos elementales de proceso son las neuronas.

1.4.4. Elementos De Una Red Neuronal

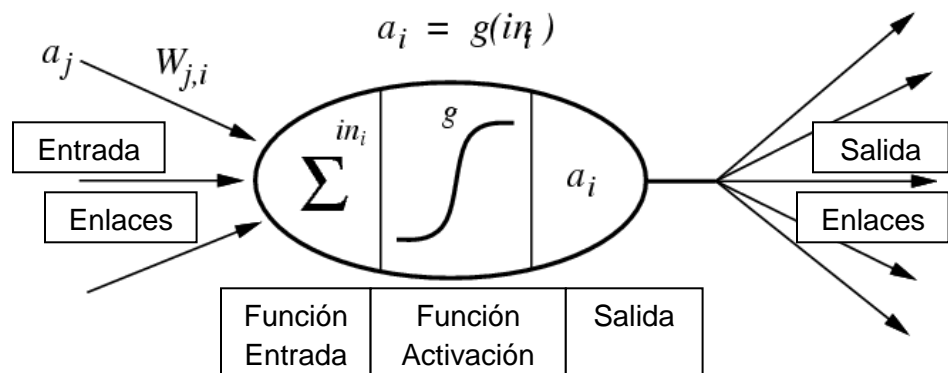
- Neuronas de entrada: aquellas que reciben estímulos externos, que toman la información de entrada; éstas se relacionan con el aparato sensorial.
- Neuronas de procesamiento: aquellas que transmiten la información captada a ciertos elementos internos que se ocupan de su procesamiento. Es la representación sináptica o interconexión interna de las neuronas, en este nivel se genera cualquier tipo de representación interna de la información; puesto que no tienen relación directa con la información de entrada ni con la salida, estos elementos se denominan unidades ocultas.
- Neuronas de salida: aquellas que una vez finalizado el periodo de procesamiento, la información llega a las unidades de salida, cuya misión es dar la respuesta del sistema.

- **Unidades de proceso**

El estudio detallado de las estructuras de las redes neuronales se centra en el modelado de la neurona artificial, ver Figura 4 donde a_j son los valores de entrada, $W_{j,i}$ son los pesos sinápticos, g es la función de activación utilizada para la red, a_i son los valores de salida.

Es útil caracterizar tres tipos de unidades: entradas (datos de entrada a la red), salidas (envían la señal fuera del sistema) y ocultas (cuyas entradas y salidas se encuentran dentro del sistema, no tiene contacto con el exterior).

Figura 4. Esquema de una red neuronal. Tomada de la referencia [21]



- **Estado de activación**

El estado de activación está influenciado por las interacciones entre las neuronas, ya que el efecto que producirá una neurona sobre otra será proporcional a la fuerza, peso o magnitud de la conexión entre neuronas, por otro lado la señal que envía cada una de las neuronas a las neuronas adyacentes dependerá de su propio estado de activación.

- **Función de salida o de transferencia**

En las neuronas artificiales existen un conjunto de conexiones que permite la unión de unas neuronas con otras. Cada unidad transmite señales a aquellas que

están conectadas con su salida, lo que transformara su estado de activación actual a una señal de salida.

Existen cuatro funciones de transferencia típicas que determinan los distintos tipos de neuronas: Función escalón, Función lineal y mixta, Función sigmoideal y Función gaussiana.

- **Conexiones entre neuronas**

Cada conexión entre neuronas (sinapsis), esta ponderada por un peso que es el que hace que la red adquiera conocimiento. En el conjunto de los pesos, se refleja la influencia de conexiones entre neuronas, esta influencia puede ser de carácter negativo (sinapsis inhibitoria), positivo (sinapsis excitativa) y nula (supone que no habrá conexión entre neuronas).

- **Regla de aprendizaje**

Dentro de las redes neuronales artificiales se considera que el conocimiento se encuentra representado en los pesos de las conexiones entre neuronas. Todo proceso de aprendizaje implica cierto número de cambios entre conexiones. Se puede decir que la red aprende modificando el valor de sus pesos.

1.4.5. Características De Las Redes Neuronales

- **Topología**

Consiste en la organización y disposición de las neuronas en la red formando capas o agrupaciones de neuronas más o menos alejadas de la entrada y salida de la red. Al momento de hacer una clasificación topológica de la red, se tiene en cuenta el número de capas o niveles de neuronas que conforman la red neuronal.

Dentro de las topologías más comunes encontramos:

- Redes mono capa: son utilizadas típicamente en tareas relacionadas con auto asociación; por ejemplo para generar informaciones de entrada que se presentan a la red incompleta o distorsionada.
- Redes multicapas: disponen de neuronas agrupadas en varios niveles o capas; comúnmente, todas las neuronas de una capa reciben señales de entrada de otra capa anterior y envían las señales de salida a una capa posterior, este tipo de conexiones se denomina conexión hacia adelante o *feedforward*. También es posible conectar las salidas de la neurona de capas posteriores a las entradas de las capas anteriores, a este tipo de conexiones se le denomina conexión hacia atrás o *feedback*.

- **Mecanismo de aprendizaje**

El aprendizaje es el proceso por el cual una red neuronal modifica sus pesos en respuesta a una información de entrada. En los modelos de redes neuronales artificiales, la creación de una nueva conexión implica que el peso de la misma pasa a tener un valor distinto de cero y de la misma forma, una conexión se destruye cuando su peso se hace cero.

Se pueden distinguir según el tipo de aprendizaje: redes neuronales con aprendizaje supervisado y redes neuronales con aprendizaje no supervisado; la diferencia fundamental se halla en la existencia de un agente externo (supervisor) el cual controlara el proceso de aprendizaje de la red.

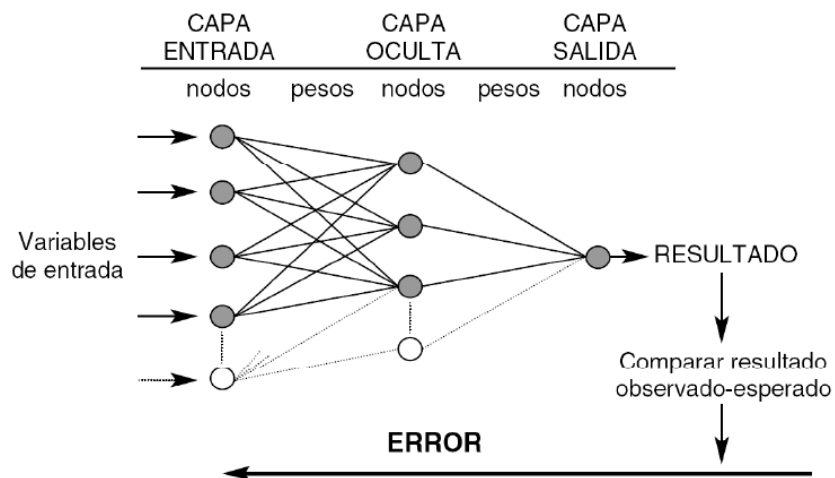
1.4.6. Perceptrón Multicapa. El perceptrón multicapa es una RNA formada por múltiples capas, lo cual le permite resolver problemas que no son linealmente separables. El perceptrón multicapa puede ser totalmente o localmente conectado; en el primer caso cada salida de una neurona de la capa "i" es entrada de todas las neuronas de la capa "i+1", mientras que en el segundo cada neurona de la capa "i" es entrada de una serie de neuronas (región) de la capa "i+1".

En la Figura 5 se pueden ver el esquema general de la red y además los tipos de capas que presentan, los cuales son:

- Capa de entrada: constituida por aquellas neuronas que introducen los patrones de entrada en la red.
- Capas ocultas: formada por aquellas neuronas cuyas entradas provienen de capas anteriores y cuyas salidas pasan a neuronas de capas posteriores.
- Capa de salida: neuronas cuyos valores de salida corresponden con las salidas de toda la red.

La propagación hacia atrás (también conocido como retropropagación del error o regla delta generalizada), es un algoritmo utilizado en el entrenamiento de estas redes, por ello, el perceptrón multicapa también es conocido como red de retropropagación (no confundir con la red de contrapropagación).

Figura 5. Esquema de una red Perceptrón Multicapa. Tomada de la referencia [27]



2. ESTADO DEL ARTE

La industria presenta un interés creciente por el desarrollo de métodos ágiles en la caracterización de materias primas en todas sus etapas de transformación, que a su vez sirvan como herramientas de control y optimización en los procesos de producción. Es por ello que se toma la química analítica como una herramienta en la predicción de parámetros fisicoquímicos y estructurales, además el enorme crecimiento en el desarrollo de investigaciones aplicadas a la industria del petróleo, requiere del procesamiento matemático de datos espectroscópicos, como lo es la aplicación de la espectroscopia IR, pues permite interpretar y procesar datos químicos, involucrándose así en la calibración y validación del desarrollo de modelos predictivos. [28]

El interés por determinar la composición química de los fondos de vacío del destilado del petróleo y la cantidad de los mismos, enfoca a los diferentes proyectos de investigación en usar la potencialidad de la técnicas espectroscópicas como es la Fluorescencia Inducida por Láser en la caracterización de muestras complejas como son los fondos de vacío, y en mostrar la elaboración de un modelo matemático que permita cuantificar las concentraciones de los compuestos involucrados en las muestras de estudio; así mismo como el uso de la técnica quimiométrica de componentes principales con el principal interés de reducir la dimensionalidad de los datos con información espectral. [29]

La espectroscopia infrarroja junto con el análisis multivariado han sido herramientas muy utilizadas para el estudio de múltiples propiedades que afectan la caracterización de productos y subproductos originados a partir del petróleo. Es por ello que en el laboratorio de espectroscopia atómica molecular LEAM, ha realizado investigaciones combinando los análisis espectroscópicos y el análisis

multivariable, con respuesta satisfactoria y con resultados versátiles como lo es menor tiempo de respuesta, al realizar el análisis. [30]

La técnica SARA consiste en fraccionar la muestra en Saturados, Aromáticos, Resinas y Asfáltenos, por la acción de solventes como n-heptano, tolueno o diclorometano. El análisis SARA requiere un tiempo promedio de dos días y el uso de grandes cantidades de solvente. El uso de técnicas analíticas como lo es el infrarrojo, la resonancia magnética nuclear (NMR), espectrometría de masas (MS) y la espectrometría de fluorescencia, Raman, han arrojado buenos resultados y se presentan como una excelente alternativa de análisis en corto tiempo de respuesta y análisis cada vez más limpios y amigables con el medio ambiente. [31]

La rutina en la medición de propiedades en los diferentes laboratorios del petróleo, son basados en métodos estándar muy bien establecidos, siguiendo procedimientos determinados por organizaciones nacionales e internacionales como la ASTM, IP, ISO entre otras. Aunque estos métodos son ampliamente aceptados, en muchos casos muestran desventajas, entre ellas el tiempo que consume, el requerimiento de grandes cantidades de muestra y el uso de químicos tóxicos y peligrosos. El método involucra operaciones manuales con evaluaciones subjetivas de la información experimental, además se expone a errores operacionales. Cuando se necesita un monitoreo rápido y confiable de las corrientes de producción en una refinería, los múltiples métodos estándar son difícilmente aplicables en línea de producción. Una alternativa a esta limitación es el uso de técnicas espectroscópicas en línea y fuera de línea en modo conjugado con técnicas de calibración multivariable. [32]

El uso de la espectroscopia infrarroja con reflectancia total atenuada (IR-ATR) presenta una gran aplicación en la industria alimenticia, es así como se ha utilizado en la determinación de la cantidad lactosa, proteína y grasa en muestras de suero lácteo y leches deshidratadas. La intensidad de absorción infrarroja

contribuye al análisis cuantitativo, puesto que las intensidades son proporcionales a la concentración de los grupos funcionales. El empleo de la espectroscopia FTIR mostró ser útil a la hora determinar los componentes principales de los productos lácteos. [33]

La utilización de la espectroscopia infrarroja en la región media junto con la reflectancia total atenuada (ATR-MIR), es una técnica potenciadora para la realización de análisis en tiempos menores exigidos por los análisis convencionales, pues esta técnica requiere de un tratamiento mínimo en la muestra, la toma de sus espectros consta de pocos minutos. La ventaja que tiene el trabajar en la región media del infrarrojo, es la disminución del solapamiento de las bandas en comparación a las otras regiones en el infrarrojo.

Las propiedades de los aceites combustibles son determinadas por normas ASTM, como son la densidad, índice de cetano, viscosidad, temperaturas de destilación y temperaturas de recuperación entre otras; todas estas mediciones pueden ser modeladas con la ayuda de métodos quimiométricos y técnicas espectroscópicas como la espectroscopia infrarroja acoplada con la técnica ATR. El modelo experimental explora las combinaciones de las técnicas FTIR-ATR, FTNIR y FT-Raman en la región espectral del infrarrojo medio, estas combinaciones demanda tecnología en fibras ópticas y sensores de alta sensibilidad en relación señal ruido y el uso de analizadores de señales espectrales como son: PLS y RNA. [34]

Otra determinación de relevante importancia es la curva de destilación y la curva de enfriamiento del combustible, fueron determinadas y medidas en la región espectroscópica del infrarrojo medio, usando RNA como herramienta de análisis de señales. Distintas configuraciones para las redes neuronales fueron usadas para predecir la curva de destilación con un 5% de incremento en cada recuperación. La propiedad fue medida experimentalmente siguiendo los protocolos estándar ASTM, el desarrollo del algoritmo de predicción fue evaluado

con datos escogidos aleatoriamente, y esto mostro un elevado nivel de rigurosidad al momento de predecir las curvas de destilación, en comparación con los métodos estándar de laboratorio. El algoritmo desarrollado puede ser usado como una herramienta confiable en el control de calidad, y es fácilmente aplicable a varias fracciones involucradas en la producción de combustible, y es posible determinar importantes propiedades del combustible basándose en una sencilla y rápida medición de IR.

Los combustibles son afectados por agentes ambientales y aquellos agentes que tienen contacto en el medio natural del combustible, un ejemplo de ello es la continua reducción por sulfuros en aceites combustibles que da como resultado una pobreza en la lubricidad y fallas en la bomba del motor, por tal razón se han desarrollado numerosos métodos para medir los niveles de lubricidad. No obstante, la medición de la lubricidad en aceites combustibles es comúnmente costosa, los métodos predictivos fueron desarrollados en el pasado, basados en las diferentes propiedades de los aceites combustibles. La investigación propone un modelo enfocado en la determinación de las propiedades de aceites combustibles usando RNA; teniéndose como entradas a la red las propiedades de los combustibles, lo que permite destacar de la investigación los valores muy pequeños para los errores de validación del modelo de predicción para la lubricidad de los aceites combustibles. [35]

3. METODOLOGÍA

Se utilizaron ochenta y cinco muestras de residuos de vacío de crudos colombianos para la calibración y cincuenta y cinco para la validación. Los espectros fueron tomados en el Laboratorio de Química de Producción del Instituto Colombiano del Petróleo, los datos de esos espectros fueron suministrados para la realización de este proyecto y así llevar a cabo cada una de las etapas metodológicas.

En la primera etapa se llevó a cabo el análisis de los espectros, teniendo en cuenta el intervalo de trabajo y los diferentes picos e intensidades. En la Figura 6 se pueden observar los espectros de calibración y en la Figura 7 los espectros de validación.

En la segunda etapa se seleccionaron los datos que serían ingresados a la red neuronal artificial, esto mediante el uso de PCA; en esta etapa se tuvieron en cuenta las diferentes muestras y los valores que presentaron para el % Carbón Conradson.

En las etapas siguientes se llevó a cabo la selección de la arquitectura, el entrenamiento y el desempeño de la red neuronal.

Figura 6. Espectros de Calibración

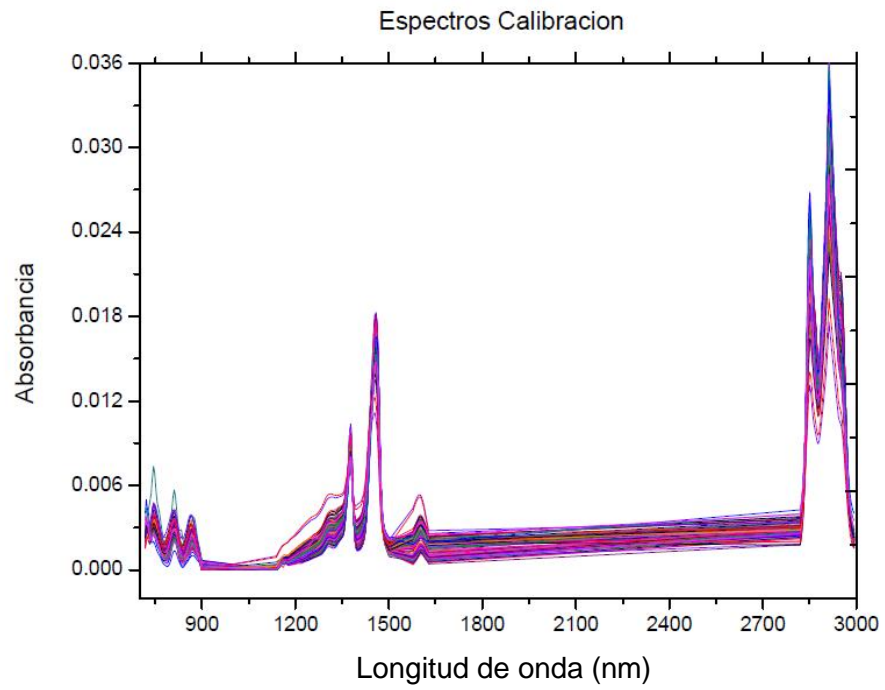
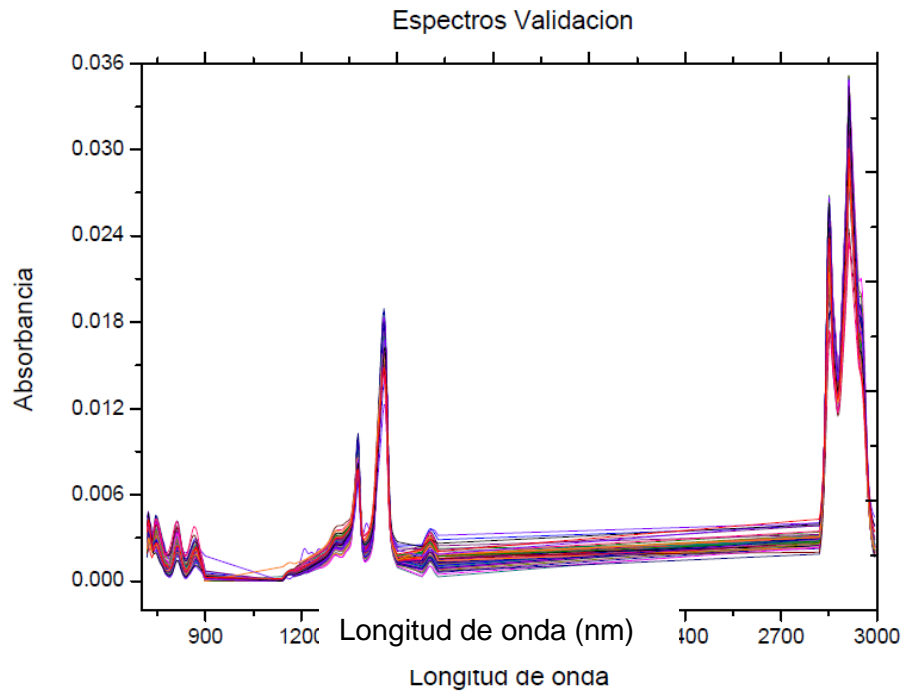


Figura 7. Espectros de Validación



3.1. OBTENCIÓN Y PROCESAMIENTO DE ESPECTROS

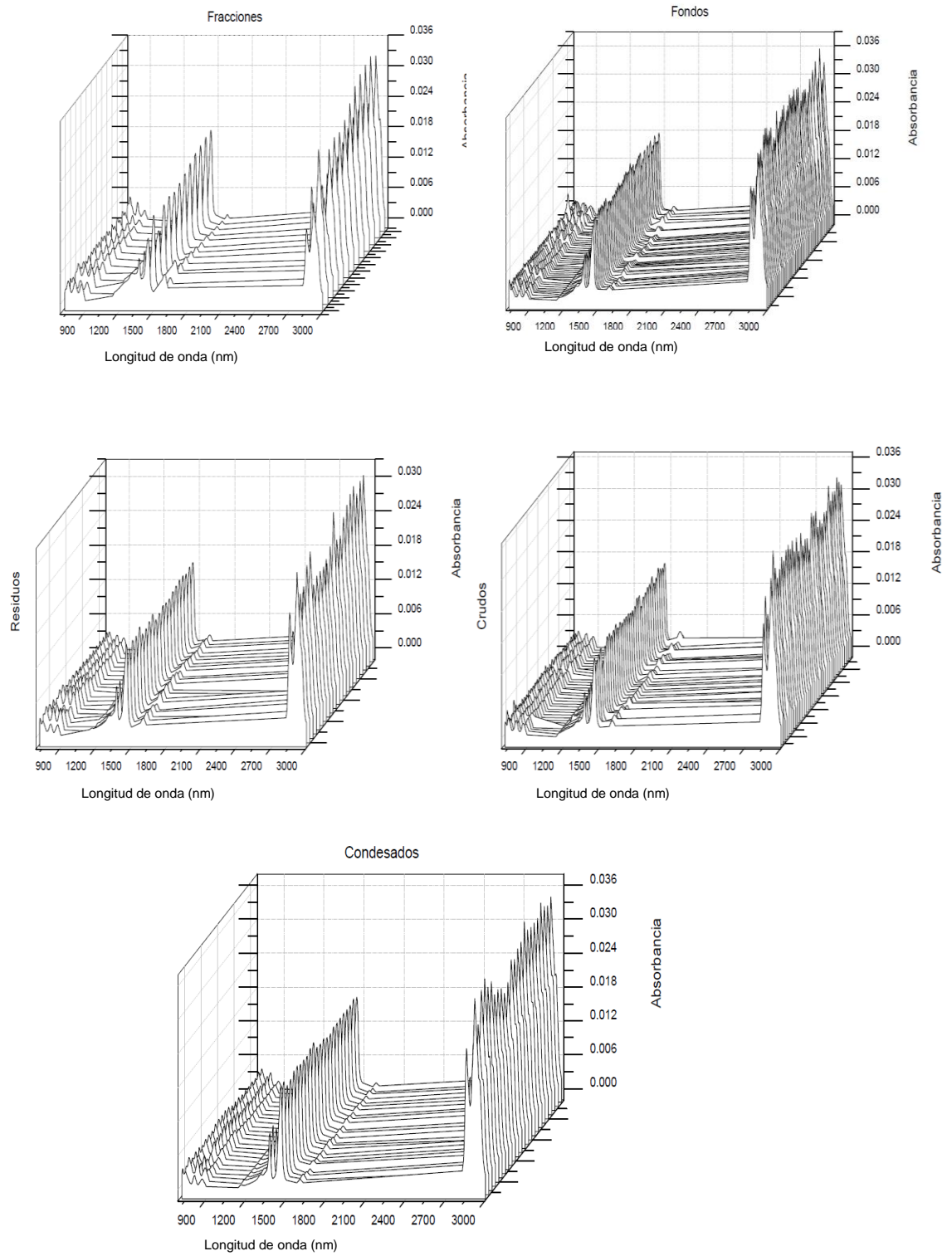
Se utilizaron muestras de fondos de vacío, condensados, residuos, crudos y fracciones provenientes de diferentes pozos: Rubiales, Castilla, Galán, Tibú entre otros. Los espectros fueron tomados usando la técnica de Espectroscopia FT-IR-ATR.

Para la obtención de los espectros se empleó una celda PIKE MIRacle, equipada con un cristal de diamante de reflexión simple con índice de refracción de 2,4; a esta celda se fijó una abrazadera de alta presión para proporcionar un contacto entre la muestra y el cristal ATR, la celda fue pre-alineada al sistema óptico del equipo IRPrestidge-21 de la marca Shimadzu. El procesamiento de los espectros inicialmente se somete a un tratamiento matemático, con el objetivo de mejorar la calidad de los datos:

- -Normalización: Permite conservar los datos dentro de la misma escala.
- -Suavizado: Reduce el ruido de los datos.
- -Derivadas: Extrae la información relevante del espectro.

La forma de los espectros de los fondos, residuos, condensados, crudos y fracciones se presenta en la Figura 8; los picos más intensos e importantes aparecen en las siguientes longitudes de onda 1377nm, 145nm, 2850nm, 2912nm tanto para los espectros de calibración como para los de validación.

Figura 8. Espectros de las muestras asignadas



3.2. SELECCIÓN DE DATOS PARA LA RED NEURONAL

Antes de ingresar los datos a la red, se realizó el análisis de componentes principales (PCA), con el uso del software Matlab con el fin de reducir la dimensionalidad del conjunto de datos.

Para llevar a cabo este análisis, se realizó una rutina en el software Matlab ver Anexo A, donde se utilizó una matriz de 69x197 para los espectros de calibración y una de 46X197 para los espectros de validación, donde las filas son las diferentes muestras y las columnas corresponden al número de onda del barrido espectral.

Las muestras tanto de calibración y validación (color azul) se presentan en la Tabla 4 y la Tabla 5.

Tabla 4. Porcentaje de CCR de las fracciones, crudos, condensados y residuos

Fracciones	CCR	Crudos	CCR
F1	1,04	CR1	19,85
F2	2,8	CR2	22,61
F3	4,83	CR3	16,13
F4	7,86	CR4	16,3
F5	13,43	CR5	20,73
F6	20,11	CR6	26,2
F7	25,93	CR7	15,8
F8	31,13	CR8	20,63
F9	36,65	CR9	21,58
F10	61,1	CR10	19,68
F11	28,77	CR11	19,85
F12	53,15	CR12	29,32

F13	30,74	CR13	18,32
		CR14	17,1
Condensados	CCR	CR15	21,46
C1	3,73	CR16	33,89
C2	2,43	CR17	20,14
C3	7,4	CR18	19,4
C4	9,9	CR19	21,4
C5	12,1	CR20	20,69
C6	10,4	CR21	20,78
C7	8,7	CR22	18,17
C8	9	CR23	22,07
C9	4,6	CR24	33,89
C10	7,21	CR1	17,5
C1	10,5	CR2	18,98
C2	5,3	CR3	26,61
Residuos	CCR	CR4	17,51
R1	21,79	CR5	20,26
R2	22,54	CR6	20,87
R3	27,97	CR7	8,19
R4	25,32	CR8	24,23
R5	30,48	CR9	22,12
R6	42,88	CR10	15,83
R7	39	CR11	22,54
R8	41,3		
R9	41		
R10	44,5		
R11	23,6		
R12	29,2		
R1	37,2		
R2	38		
R3	27,1		

Tabla 5. Porcentaje de CCR de los fondos de vacío

Fondos	CCR
FV1	35,11
FV2	3,4
FV3	9,3
FV4	16,7
FV5	16,2
FV6	2,58
FV7	17,04
FV8	22,86
FV9	4,54
FV10	16,67
FV11	1,36
FV1	4,34
FV2	13,2
FV3	21,4
FV4	2,65
FV5	4,17
FV6	6,5
FV7	11,2
FV8	20,7
FV9	4,24
FV10	4,24
FV11	8,5
FV12	11,83
FV13	18,29
FV14	5,32
FV15	8,39
FV16	10,46
FV17	15,2
FV18	17,74
FV19	15,69
FV20	22,73

Fondos	CCR
FV21	3,16
FV22	5,3
FV23	9,82
FV24	2,54
FV25	3,29
FV26	5,45
FV27	132
FV28	18
FV29	18,29
FV30	18,3

3.3 SELECCIÓN DE LA ARQUITECTURA DE LA RED NEURONAL

Para la selección de la arquitectura de la red se escogió el modelo de Perceptrón Multicapa. La optimización de la arquitectura en la red neuronal se llevó a cabo variando los valores de entrada de la red y cambiando el número de capas ocultas encargadas del procesamiento de los datos dentro de la red.

3.4 ENTRENAMIENTO DE LA RED NEURONAL

Este entrenamiento se llevó a cabo con un aprendizaje no supervisado, ya que no tiene ningún agente externo (supervisor) que lo controle. Durante este entrenamiento se varió el número de componentes principales, las capas ocultas, un valor constante de 1000 épocas y se utilizó el algoritmo de Backpropagation mediante el criterio del gradiente descendiente de la función de error, el cual ajusta los pesos y biases.

Los datos de entrada fueron los Score tanto de calibración como de validación, los cuales se obtuvieron del análisis de componentes principales, mientras que los datos de salida como respuesta al entrenamiento fueron los errores de predicción (RMSEP) y calibración (RMSEC).

3.5 ESTIMACIÓN DEL DESEMPEÑO DE LA RED NEURONAL

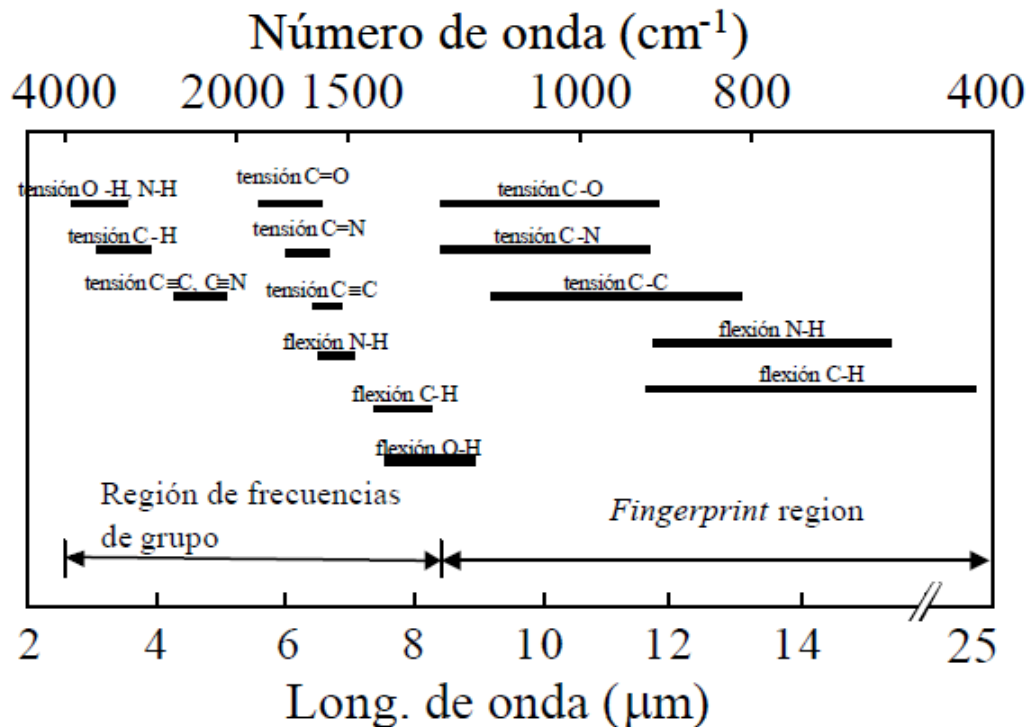
Se estimó el desempeño, haciendo uso de parámetros estadísticos como el RMSEP (Root Mean Square Error Prediction) y el RMSEC (Root Mean Square Error Calibration).

4. RESULTADOS

4.1. ANÁLISIS ESPECTRAL

Los espectros en el infrarrojo medio abarcan un intervalo en número de onda de 4000cm^{-1} hasta 200cm^{-1} como se puede observar en la Figura 9, en donde la región de frecuencias de grupo está en el intervalo de $4000\text{-}1300\text{ cm}^{-1}$ y la región de la huella dactilar en el intervalo entre $1300\text{-}400\text{cm}^{-1}$.

Figura 9. Frecuencias de vibración en el infrarrojo medio.



Se realizó la representación de los espectros contenidos en la data de información, haciendo uso del software Matlab y teniendo en cuenta las tablas de espectroscopia infrarrojo y haciendo una revisión bibliográfica se tiene el siguiente listado de las bandas más representativas dentro del espectro en la Tabla 6.

Tabla 6. Grupos funcionales en el IR Medio

Absorción [cm ⁻¹]	Grupo Funcional
3090-3030	Tensión (C-H) _{ar}
2924-2916	R ² CH ₂ Tensión asimétrica del metileno
2954-2926	RCH ₃ Tensión asimétrica metilo
2677-2858	RCH ₃ Tensión simétrica metilo
2858-2846	R ² CH ₂ Tensión simétrica metileno
1632-1700	Tensión C=O
1578-1626	Tensión (C-C) _{ar}
1423-1469	Flexión asimétrica metilo
1395-1390	Flexión simétrica metilo
1150-1070	Tensión (C-O-O) alifática
1060-970	Tensión S=O
844-898	Car-H(1H) Adyacente, fuera del plano
783-841	Car-H(2H Ó 3H) Adyacente, fuera del plano
740-780	Car-H(4H) Adyacente, fuera del plano
720-736	R(CH ₂) _n -R Deformaciones de balanceo metileno

De los espectros asignados y de la Tabla 6, se observan que las bandas más notorias son las que corresponden a las vibraciones de las bandas de tensión aromática (C-H) 3090-3030, tensión asimétrica del etileno 2924-2916, tensión simétrica del metileno 2858-2846, flexión asimétrica del metilo 1423-1469, tensión C-C aromática 1578-1628, deformación de balanceo del metileno 720-736, flexión fuera del plano C-H aromática que permite determinar el tipo de sustitución en el anillo 848-780.

4.2. ANÁLISIS POR COMPONENTES PRINCIPALES (PCA)

La matriz de datos utilizada para el análisis de componentes principales se compone de 197 columnas por 69 filas para la calibración y de 197 columnas por 45 filas para la validación; donde las columnas corresponden al número de onda

del barrido espectral de 700 a 3000 cm^{-1} y las filas corresponden a diferentes muestras de fondos de vacío.

Para empezar con el análisis, se realizó una rutina utilizando el Software Matlab, Anexo A, en la cual se obtuvieron los scores para 15 componentes principales; en esta rutina se utilizó la data de información de los espectros suministrada por el Laboratorio de Química de Producción del Instituto Colombiano del Petróleo, a continuación se utilizó la función *isnan* la cuál nos permitió descartar los valores de Carbón Conradson (para calibración y validación) que no presentaban ningún valor, posteriormente se prosiguió a realizar el cálculo de vectores y valores propios; todo estos valores obtenidos fueron finalmente utilizados, primero para obtener los scores de calibración y validación los cuales fueron las entradas de la red neuronal y segundo para realizar la reconstrucción de los espectros cómo podemos observar en la Figura 10 que al graficar el espectro original (en negro) y el espectro reconstruido con los componentes (en rojo) la diferencia es poca, los picos más representativas del espectro permanecen, lo cuál nos indica que el análisis por componentes principales y la rutina realizada son correctos y el comportamiento es satisfactorio, pues no fue necesario realizar el centrado de la media ya que no le sustraemos la media a ninguno de los valores iniciales.

Por otra parte debido a que los primeros componentes están asociados a la proporción de la varianza acumulativa, se puede decir que para el primer componente como es el que recoge la mayor información, se obtuvo un índice de varianza acumulativa de 99% lo cuál nos indica que entre las variables originales existe cierto grado de colinealidad o correlación, mientras que para los siguientes componentes este índice se acercó a un porcentaje de 1%, en la Figura 11 se puede observar esta tendencia y los porcentajes para los demás componentes.

El análisis de componentes principales se emplea sobre todo en un análisis exploratorio de datos y para la construcción de modelos predictivos, además comporta el cálculo de la descomposición en auto valores de la matriz de covarianza, lo cual centra los datos en la media de cada muestra; para este caso no se llevó a cabo este centrado, debido a que los valores presentan gran similitud y por tal motivo se utilizó el método basado en la matriz de covarianzas, que es utilizado cuando los datos son dimensionalmente homogéneos y presentan valores medios similares.

Figura 10. Reconstrucción de los espectros.

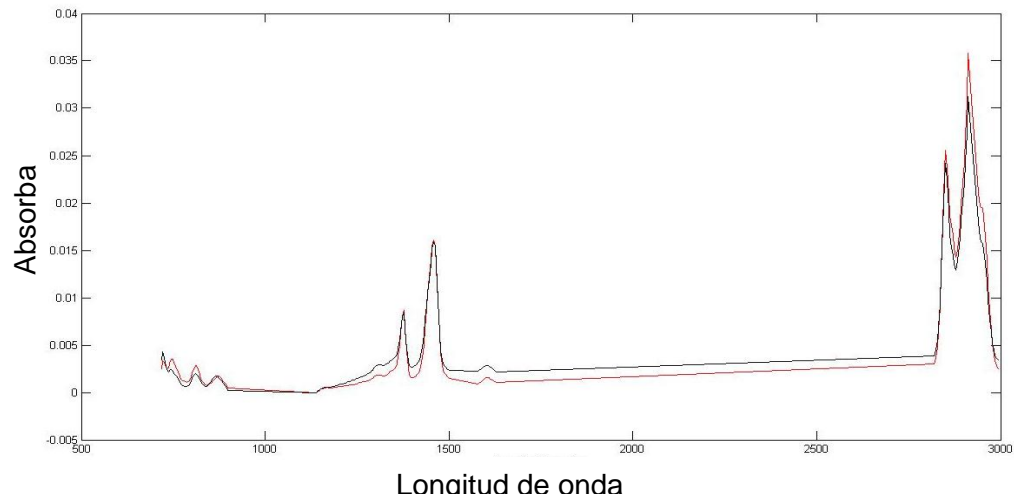
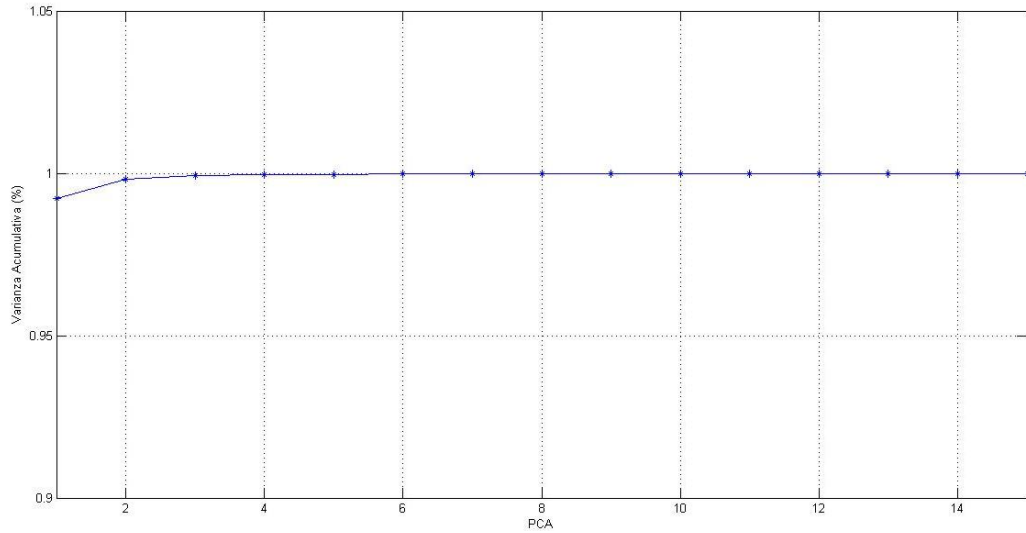


Figura 11. Varianza acumulativa.



4.3 REDES NEURONALES

El análisis fue realizado con el uso del software MATLAB 2008 (The MathWorks Inc.) junto con Neural Networks Toolbox V6.0. La rutina del modelo se encuentra en el Anexo B, en esta rutina se pueden ver los parámetros de la red neuronal los cuales son: *m* (componente principal), *nco* (número de capas ocultas), *minma* (valores mínimos y máximos, $[0\ 70]$ (rango de entrenamiento para los valores del CCR), *tansig* (función de transferencia sigmoide tangente hiperbólica, calcula la salida de la capa a partir de la entradas iniciales), *purelin* (función de transferencia lineal, calcula la salida de la capa a partir de la capa oculta), *trainlm* (algoritmo de propagación hacia atrás y función de la red la cual actualiza los pesos y bias de acuerdo a la optimización de Levenberg-Marquardt) y *learnngdm* (función de aprendizaje).

La calidad de los modelos en los conjuntos de calibración y predicción se evaluó sobre la base de los indicadores comunes: r^2 , RMSEC y RMSEP, donde r^2 es el coeficiente de correlación lineal, el cuál se determinó realizando una regresión

lineal múltiple entre los componentes principales y la propiedad en estudio CCR, el RMSEC hace referencia al error estándar de calibración y el RMSEP al error para el conjunto de predicción; estos errores están dados por las Ecuaciones 1 y 2, y en la rutina del Anexo B se presentan de la siguiente forma, Ecuación 4 y 5. Donde O_c y O_v son los valores de la propiedad (calibración y validación), estimados usando la simulación de la red neuronal, CCR_{cal} y CCR_{val} son los valores conocidos de la propiedad en estudio (Carbón Conradson), y N_c y N_v son los valores de la longitud del vector utilizado para desarrollar el modelo de calibración y validación.

Ecuación 4. Error de calibración.

$$SECPRESS = \frac{\sum_{i=1}^{N_c} (O_c - CCR_{cal'})^2}{N_c}$$

Ecuación 5. Error de predicción.

$$SEPPRESS = \frac{\sum_{i=1}^{N_v} (O_v - CCR_{val'})^2}{N_v}$$

Los valores de los coeficientes de correlación lineal, el error de calibración y predicción para los mejores resultados se presentan en la Tabla 7. El modelo de calibración y predicción desarrollado para las diferentes arquitecturas arrojó buenos resultados, con valores de r^2 que fluctuaron entre 0,94 y 0,99 para la calibración y con valores de 0,86 y 0,99 para la predicción. Las líneas de regresión entre los valores medidos y pronosticados para el conjunto de calibración y predicción se presentan en las Figuras 12 y 13. En general, las predicciones de la propiedad en cada tipo de arquitectura fueron satisfactorias, pero los valores del RMSEP obtenidos para los conjuntos de validación fueron mayores que los valores del RMSEC y es por tal razón que los valores de la predicción fueron los

más importantes para este estudio, aunque la red se entrenó varias veces el error más bajo se presentó en el componente 6, raíz 77 el cuál fue 2,15% con una capa oculta de 4 y el error más alto fue de 3,24% con el componente 14 y una raíz de 13 en la capa oculta 9; otros resultados para los 15 componentes se presentan en la Tabla 8 con sus respectivos errores de calibración y predicción.

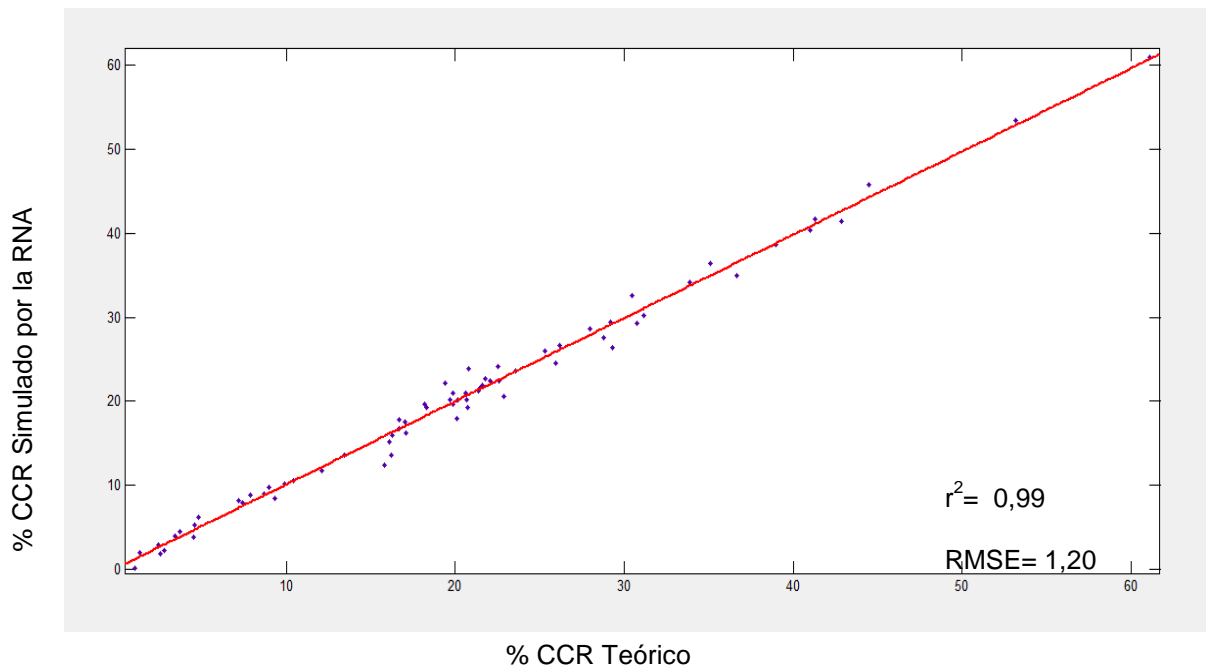
En particular, para predecir el porcentaje de CCR de los fondos de vacío, se adoptó una red neural artificial Perceptron Multicapa, donde los valores de entrada de la red fueron los scores de los componentes principales, los valores de salida los errores de calibración y predicción, número de capas ocultas 30, número de épocas de 1000, función de transferencia lineal para las capas de entrada y salida, mientras que para las capas ocultas función de transferencia sigmoide tangente; el entrenamiento de la red se logró con el algoritmo de retropropagación de Levenberg-Marquardt por medio del descenso de gradiente de la función de error el cual ajusta los pesos y bias, mientras que el aprendizaje utilizado fue el no supervisado.

La escogencia de la mejor arquitectura de la red neuronal, es un proceso dispendioso, que requiere que el analista desarrolle experiencia en el diseño y entrenamiento, dicho proceso junto a la evaluación de prueba y error, dependiendo de la complejidad del problema a desarrollar. Aunque la tarea de escoger la cantidad de neuronas y el número de capas es desconocida para la resolución de un problema, existen diferentes estructuras de la red que permiten hacer la misma tarea, hallándose la diferencia en la cantidad de error arrojado por tal combinación.

Tabla 7. Mejores resultados de la Red Neuronal.

Arquitectura Red Neuronal	N° de Componentes	Conjunto de Calibración		Conjunto de Predicción	
		r ²	RMSEC	r ²	RMSEP
PCA-Raíz-Capa Oculta					
6-77-4	6	0,99	1,19	0,99	2,15
9-53-4	9	0,99	1,15	0,93	2,37
10-51-3	10	0,99	0,7	0,92	2,38
5-51-6	5	0,98	1,6	0,92	2,43
7-25-4	7	0,98	1,27	0,92	2,45
11-55-2	11	0,99	0,94	0,92	2,53
12-13-1	12	0,98	1,44	0,91	2,57
8-79-7	8	0,99	0,98	0,91	2,64
4-71-3	4	0,98	1,66	0,92	2,73
15-95-1	15	0,99	1,16	0,89	2,87
13-63-1	13	0,98	1,3	0,88	2,91
3-73-1	3	0,96	2,39	0,89	3,05
2-13-15	2	0,94	2,93	0,88	3,08
14-13-9	14	0,99	0,32	0,86	3,24

Figura 12. Relaciones entre los valores conocidos y simulados por la red neuronal para la calibración, para el error más bajo y más alto.



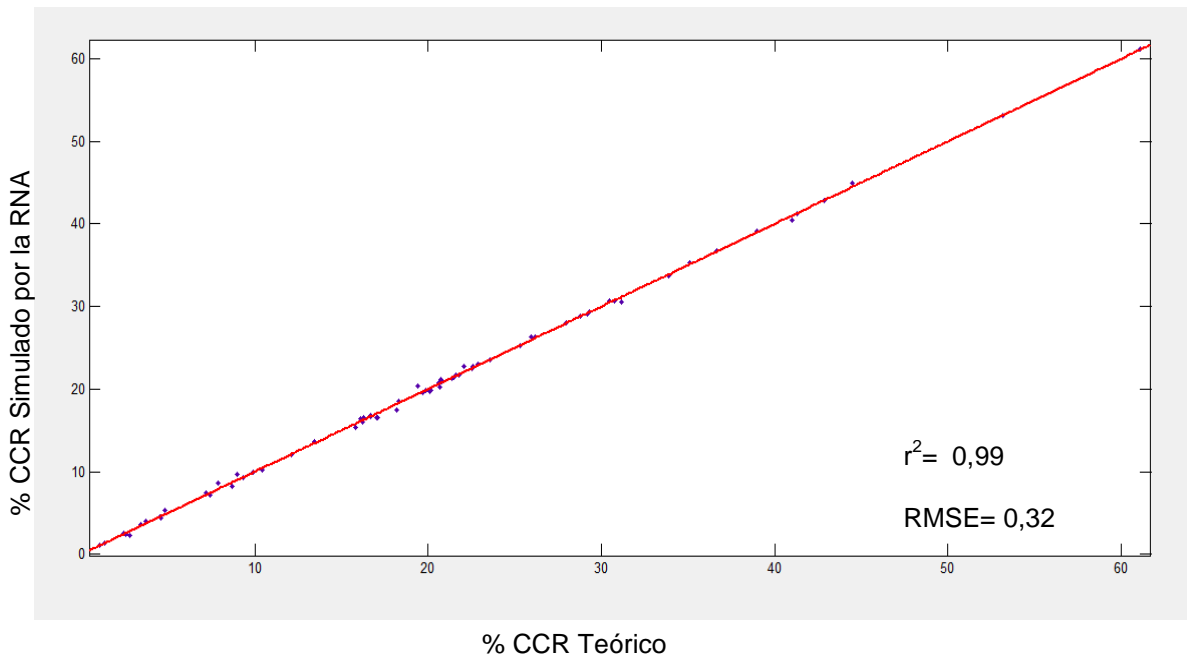
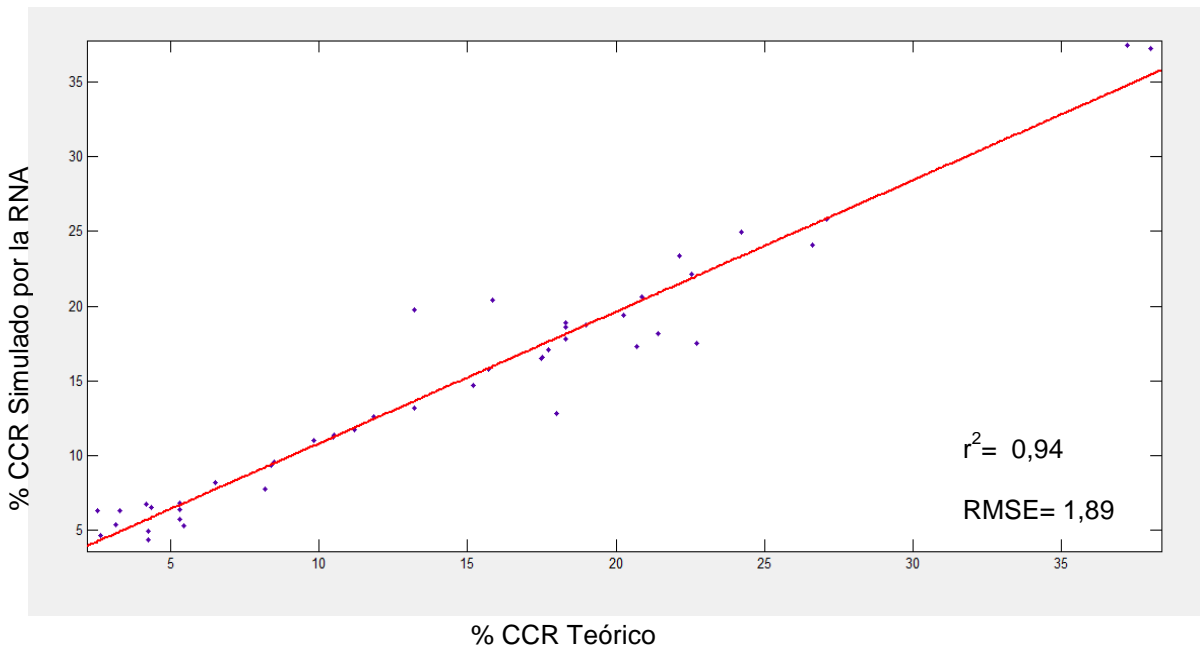


Figura 13. Relaciones entre los valores conocidos y simulados por la red neuronal para la validación, para el error más bajo y más alto



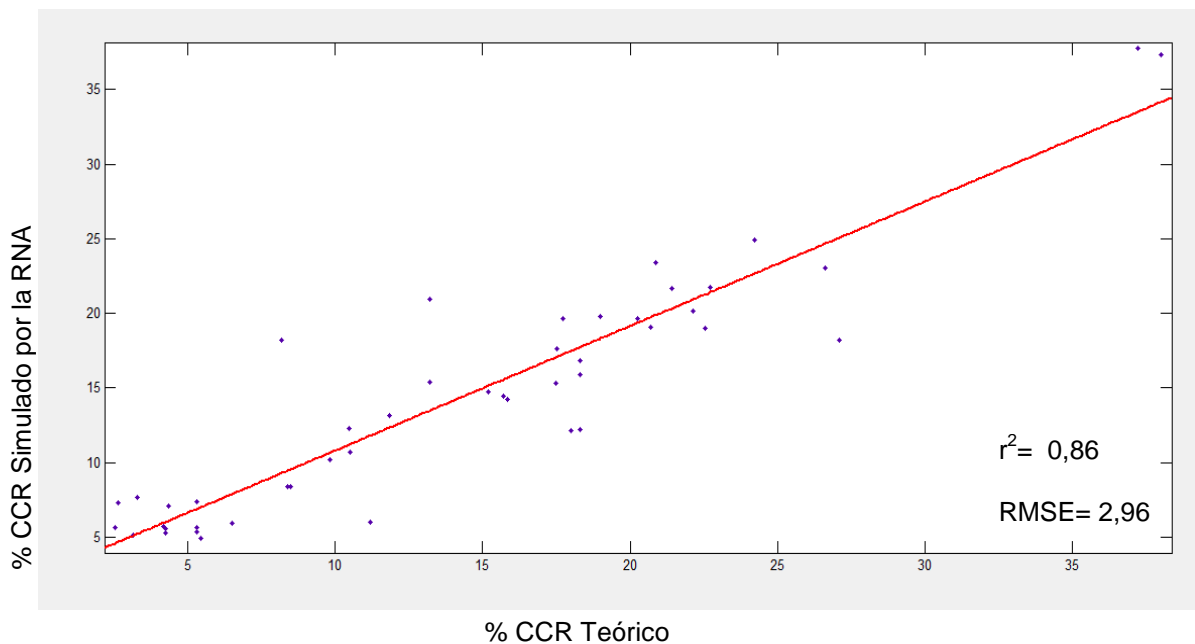


Tabla 8. . Resultados de la Red Neuronal

RNA Arquitectura (PCS-Raíz-CCR)	REC (%)	REP (%)	RNA Arquitectura (PCS-Raíz-CCR)	REC (%)	REP (%)
2-13-1	2,93	3,08	9-53-1	1,15	2,37
2-3-1	2,95	3,09	9-59-1	1,09	2,50
2-7-1	2,98	3,10	9-25-1	1,25	2,50
2-1-1	2,98	3,11	9-1-1	1,45	2,52
2-15-1	2,98	3,12	9-5-1	1,58	2,53
2-19-1	2,99	3,13	9-9-1	1,58	2,53
3-13-1	1,96	2,84	10-51-1	0,70	2,38
3-31-1	1,69	2,84	10-17-1	1,47	2,41
3-17-1	1,87	2,89	10-27-1	0,70	2,42
3-39-1	2,16	2,91	10-7-1	0,70	2,43
3-51-1	2,16	2,91	10-19-1	1,29	2,44
3-73-1	2,39	3,05	10-1-1	1,44	2,56
4-71-1	1,66	2,73	11-61-1	0,79	2,36
4-55-1	1,71	2,74	11-41-1	0,73	2,44
4-7-1	1,79	2,75	11-55-1	0,94	2,53

4-37-1	1,66	2,77	11-11-1	1,44	2,57
4-67-1	1,65	2,80	11-83-1	1,44	2,57
4-23-1	1,65	2,81	11-97-1	1,44	2,57
5-51-1	1,60	2,43	12-43-1	1,44	2,57
5-99-1	1,60	2,45	12-75-1	1,44	2,57
5-41-1	1,60	2,47	12-13-1	1,44	2,57
5-33-1	1,57	2,48	12-87-1	1,44	2,57
5-35-1	1,57	2,50	12-21-1	1,44	2,57
5-53-1	1,58	2,55	12-15-1	0,49	3,12
6-77-1	1,19	2,15	13-11-1	1,30	2,91
6-45-1	1,20	2,16	13-63-1	1,30	2,91
6-69-1	1,23	2,18	13-95-1	1,30	2,91
6-15-1	1,23	2,20	13-27-1	1,30	2,91
6-37-1	1,23	2,21	13-83-1	1,30	2,91
6-47-1	1,19	2,23	13-5-1	1,30	2,91
7-25-1	1,27	2,45	14-25-1	1,19	2,86
7-73-1	1,30	2,50	14-73-1	1,19	2,86
7-59-1	1,27	2,51	14-85-1	1,19	2,86
7-55-1	1,26	2,52	14-11-1	1,19	2,86
7-15-1	1,25	2,53	14-7-1	1,19	2,86
7-1-1	1,58	2,54	14-13-1	0,32	3,24
8-73-1	1,29	2,51	15-65-1	1,16	2,87
8-49-1	1,22	2,52	15-33-1	1,16	2,87
8-79-1	0,98	2,64	15-15-1	1,16	2,87
8-71-1	1,03	2,74	15-77-1	1,16	2,87
8-45-1	1,26	2,80	15-95-1	1,16	2,87
8-87-1	1,27	2,81	15-21-1	1,16	2,87

5. CONCLUSIONES

La espectroscopia infrarroja en la región media junto a las redes neuronales artificiales, demostró ser un método alternativo en el diseño del modelo predictivo de una propiedad físico química como el CCR en muestras de fondos de vacío, haciendo de este método una herramienta ágil en el procesamiento de información en tiempo real, logrando obtener valores de predicción en un rango de 2,1% a 3,5%.

Fue posible hacer un modelo predictivo de una propiedad físico química como el Carbón Conradson en muestras de fondos de vacío, con espectros obtenidos por técnicas como la espectroscopia infrarroja acoplada a reflectancia total atenuada y utilizando las redes neuronales artificiales.

Las mejores arquitecturas de RNAs que se obtuvieron en el entrenamiento de la red, presentan el componente principal, el número de raíz, los cuáles siempre fueron números impares y por último el número de la capa oculta, con valores menores que 10.

6. RECOMENDACIONES

Con la construcción de modelos predictivos se puede explorar y controlar procesos productivos en su línea de consecución, como se hace en la refinación del petróleo. Es por tal razón que se insta a el desarrollo de más modelos de predicción de propiedades físico químicas diferentes a la ya estudiada es este trabajo de investigación.

Se recomienda continuar con la exploración de nuevas topologías, dentro de la red neuronal con miras al hallazgo de combinaciones que permitan reducir los errores.

REFERENCIAS BIBLIOGRÁFICAS

[1] Speight, J.G. The chemistry and technology of petroleum. Third Edition. Editorial Marcel Dekker: New York, pp 333-335.

[2] Castillo, Alba. Análisis de fondos de vacío de crudo Colombiano por Ablación Láser acoplada a espectroscopia de emisión atómica con plasma generado por acoplamiento inductivo (LA-ICP-AES). Tesis de Grado (Física). Universidad Industrial de Santander. Facultad de Ciencias. Escuela de Física. Bucaramanga 2010.

[3] Waquier, J.P. El refinado del petróleo. Petróleo crudo, productos petrolíferos, esquemas de fabricación. Editorial Díaz de Santos: España, 2004, pp 1-15.

[4] Empresa Colombiana de Petróleos ECOPETROL. Acceso a internet: www.ecopetrol.com.co/especiales/elpetroleoysumundo/

[5] Universidad Nacional de Cuyo. La refinación del petróleo Tomo I. Facultad de ciencias aplicadas a la industria. Acceso a internet: <http://www.fcai.uncu.edu.ar/upload/LIBRO%20TOMO%201.pdf>.

[6] Altgelt, K.H.; Boduszynski, M.M. Composition and analysis of heavy petroleum fractions. Second Edition. Editorial Marcel Dekker: New York, 2002, pp 495.

[7] American Society for Testing and Materials. Standard Test Method for Conradson Carbon Residue of Petroleum Products. (ASTM D189-06). Acceso a internet: <http://www.astm.org/Standards/D189.htm>

[8] American Society for Testing and Materials. Standard Test Method for Ramsbottom Carbon Residue of Petroleum Products. (ASTM D524-10). Acceso a internet: <http://www.astm.org/Standards/D524.htm>

[9] Parker Kittiwake Ltd2011. Carbon Residue in Fuel oils. Acceso a internet: www.condition-monitoring.com/bunker_fuel_carbon_residue.htm.

[10] Meyers, R. Handbook of petroleum refining processes. Third edition. McGraw-Hill, 1986.

[11] Ortega, L.J.; Blanco, F. M. Identificación de compuestos orgánicos por espectroscopia infrarroja. Editorial M.O.P.V.: Madrid, 1982.

[12] Skoog, D.; Holler, J.; Nieman A, Timothy. Principios de análisis instrumental. Quinta edición. Editorial McGraw-Hill: pp 451-453.

[13] Macho-Aparicio, S. Metodologías analíticas basadas en espectroscopia de infrarrojo y calibración multivariante. Aplicación a la industria petroquímica. Tesis doctoral (doctor en Química). Univertat Rovira Virgili. Área de Química analítica. Departamento de Química analítica Química orgánica. Tarragona 2002.

[14] Application note. ATR-Theory and Applications. PIKE technologies. 2005. Acceso a internet: www.piketech.com/files/ATRAN611.pdf.

[15] Orrego R, J. Estudio de la estructura de cinco carbones colombianos por espectroscopia fotoacustica en el infrarrojo medio. Tesis de grado (Química). Universidad industrial de Santander. Facultad de ciencias. Escuela de química. Bucaramanga 2008.

[16] Espectroscopia IR. Acceso a internet: http://karin.fq.uh.cu/~cnv1/qf/docencia/pregrado/estruc_2/curso_08_09/3_1_IR1.pdf.

[17] American Society for Testing and Materials. Standard practices for infrared multivariate quantitative analysis. ASTM 212 (ASTM E1655-05).

[18] Parra, S.I. Predicción de la curva de destilación de fondos de vacío de crudos Colombianos usando espectroscopia fotoacústica con transformada de Fourier en el infrarrojo medio y quimiometría. Tesis de grado (Química). Universidad Industrial de Santander. Facultad de ciencias. Escuela de Química. Bucaramanga 2012.

[19] Aparicio, C.A. Predicción del número de bromo de naftas Colombianas utilizando espectroscopia infrarrojo en la región cercana y métodos quimiométricos. Tesis de grado (Química). Universidad Industrial de Santander. Facultad de ciencias. Escuela de Química. Bucaramanga 2013.

[20] Vicente, J.L. Análisis de Componentes Principales. Departamento de Estadística.

[21] Turing, A. Computing machinery and intelligence. Mind, New Series. Vol. 59, 1950, pp 433-460.

[22] Mcculloch, W.; Pitts, W. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics. Vol. 5, 1943, pp 115-133.

[23] Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review. Vol. 65, 1958, pp 386-408.

[24] Hilera, J.R.; Martinez, V.J. Redes Neuronales Artificiales, Fundamentos, modelos y aplicaciones. Alfaomega. 2000, pp 4-6, 45-46, 52-53, 72-75.

[25] Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA. Vol 79, April 1982.

[26] Chen, C.R.; Ramaswamg, H.S. Modeling Food Thermal Network. 2006. Taylor & Francis Group, LLC.

[27] Tarazona, J.L. Construcción de un modelo predictivo para la concentración de níquel y vanadio en fondos de vacío de crudos colombianos utilizando redes neuronales artificiales y espectroscopia de plasma inducido por laser. Tesis de Maestría (Física). Universidad Industrial de Santander. Facultad de ciencias. Escuela de Física. Bucaramanga 2012.

[28] Khanmohammadi, Bagheri.; De la Guardia, M. Characterization of petroleum-based products by infrared spectroscopy and chemometrics. Trends in Analytical Chemistry, Vol. 35, 2012, pp 135-149.

[29] Ballesteros-Lancheros, S.L.; Orrego, J.A.; Cabanzo, R.; Mejía-Ospino, E. Determinación de la composición de fondos de vacío de petróleos colombianos por espectroscopia de fluorescencia inducida por laser (LIF). Revista colombiana de física. Vol. 43, 2011, pp 648-652.

[30] Lamus, C. Uso de análisis multivariado en la determinación SARA de crudos mediante la espectroscopia NIR. Tesis de grado (magister químico). Universidad Industrial de Santander. Facultad de ciencias. Escuela de química. Bucaramanga 2010.

[31] Lache, A., Meléndez, L. Predicción de análisis SARA de crudos colombianos aplicando espectroscopia FTIR-ATR y métodos quimiométricos. Tesis de grado (químico). Universidad Industrial de Santander. Facultad de ciencias. Escuela de química. Bucaramanga 2010.

[32] Pasadakis, N.; Sourligas, S.; Foteinopoulos, Ch. Prediction of the distillation profile and cold properties of diesel fuels using mid-IR spectroscopy and neural networks. Science direct, Fuel. Vol. 85, 2005, pp 1131-1137.

[33] Teniza, O., Solis, M., Delgado, R., Rojas, M. "Espectroscopia infrarroja con reflectancia total atenuada aplicada a la determinación de lactosa y proteínas en productos lácteos". LI CONGRESO NACIONAL SMF/ XXI REUNIÓN ANUAL AMO. Zacatecas, 2008, pp 1-5.

[34] Santos, V.O.; Oliveira, F.; Lima, D.; Petry, A.; García, E.; Suárez, P.; Rubin, J. A comparative study of diesel analysis by FTIR, FTNIR and FT-Raman spectroscopy using PLS and artificial neural network analysis. Science Direct, Analytica chimica Acta 547, 2005, pp 188-196.

[35] Korres, D.M., Anastopoulos, G., Lois, E.; Alexandridis, A.; Sarimveis, H.; BAFAS, G. A neural network approach to the prediction of diesel fuel lubricity. Elsevier. Fuel 81, 2002, pp 1243-1250.

BIBLIOGRAFÍA

American Society for Testing and Materials. Standard Test Method for Conradson Carbon Residue of Petroleum Products. (ASTM D189-06). Acceso a internet: <http://www.astm.org/Standards/D189.htm>

American Society for Testing and Materials. Standard Test Method for Ramsbottom Carbon Residue of Petroleum Products. (ASTM D524-10). Acceso a internet: <http://www.astm.org/Standards/D524.htm>

Ballesteros-Lancheros, S.L.; Orrego, J.A.; Cabanzo, R.; Mejía-Ospino, E. Determinación de la composición de fondos de vacío de petróleos colombianos por espectroscopia de fluorescencia inducida por laser (LIF). Revista colombiana de física. Vol. 43, 2011, pp 648-652.

Chen, C.R.; Ramaswamg, H.S. Modeling Food Thermal Network. 2006. Taylor & Francis Group, LLC.

Empresa Colombiana de Petróleos ECOPETROL. Acceso a internet: www.ecopetrol.com.co/especiales/elpetroleoysumundo/

Hilera, J.R.; Martinez, V.J. Redes Neuronales Artificiales, Fundamentos, modelos y aplicaciones. Alfaomega. 2000, pp 4-6, 45-46, 52-53, 72-75.

Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA. Vol 79, April 1982.

Khanmohammadi, Bagheri.; De la Guardia, M. Characterization of petroleum-based products by infrared spectroscopy and chemometrics. Trends in Analytical Chemistry, Vol. 35, 2012, pp 135-149.

Lache, A., Meléndez, L. Predicción de análisis SARA de crudos colombianos aplicando espectroscopia FTIR-ATR y métodos quimiométricos. Tesis de grado (químico). Universidad Industrial de Santander. Facultad de ciencias. Escuela de química. Bucaramanga 2010.

Orrego R, J. Estudio de la estructura de cinco carbones colombianos por espectroscopia fotoacustica en el infrarrojo medio. Tesis de grado (Química). Universidad industrial de Santander. Facultad de ciencias. Escuela de química. Bucaramanga 2008.

Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review. Vol. 65, 1958, pp 386-408.

Waquier, J.P. El refino del petróleo. Petróleo crudo, productos petrolíferos, esquemas de fabricación. Editorial Díaz de Santos: España, 2004, pp 1-15.

ANEXOS

ANEXO A. Rutina Matlab de los componentes principales

```
clc %% borrar pantalla
clear all %%Borra todas las variables
close all %% Cierra todas las ventanas
format bank
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%Abrir archivos de Excel
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
cd C:\Users\usuario\Documents\Proyecto_FV\Archivos_Matlab\New_datos
esp_cal=xlsread('datosredes.xlsx','espectroscal','B3:GP87');
CCRcal=xlsread('datosredes.xlsx','propiedadcal','E3:E87');
esp_val=xlsread('datosredes.xlsx','espectrosval','B2:GP56');
CCRval=xlsread('datosredes.xlsx','propiedadval','E2:E56');
numero_onda=xlsread('datosredes.xlsx','espectroscal','B1:GP1');
L=numero_onda;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%
ICCR=isnan(CCRcal);
ICCRV=isnan(CCRval);
n=1;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for m=1:length(CCRcal)
    if ICCR(m)==1
```

```

        pm(n)=m;
        n=n+1;
    else
    end
end
CCRcal(pm)=[];
esp_cal(pm,:)=[];
ss=esp_cal;
n=1;
clear pm
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for m=1:length(CCRval)
    if ICCRV(m)==1
        pm(n)=m;
        n=n+1;
    else
    end
end
CCRval(pm)=[];
esp_val(pm,:)=[];
ssv=esp_val;
[NO NV]=size(ss);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C=(1/NO)*(ss'*ss); %%Matriz de covarianza
[VecP ValoresP]=eig(C); %% Cálculo de vectores y valores propios
VecP=flipr(VecP);
m=15;
VecP=VecP(:,1:m);
score_cal=ss*pinv(VecP)';

```

```

score_val=ssv*pinv(VecP)';
reconstruidos=score_val*VecP';
SRMS=100*mean(sqrt(sum((reconstruidos-ssv).*(reconstruidos-ssv))/length(ssv)));
reconstruidos=score_cal*VecP';
SRMS=100*mean(sqrt(sum((reconstruidos-ss).*(reconstruidos-ss))/length(ss)));
[NO NV]=size(ssv);
for i=1:NO
    drawnow
    plot(L,reconstruidos(i,:), 'r',L,ssv(i,:), 'k');pause(0.2)
end
xlabel('Longitud de onda')
ylabel('Espectros Validación')

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%Cálculo del índice de varianza acumulativo
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% ValoresP=fliplr(ValoresP);
% ValoresP=flipud(ValoresP);
% D=diag(ValoresP);
% D=D(1:m);
% c=0;
% for i=1:m
%     c=c+(D(i)/sum(D));
%     cv(i)=c;
%     nc(i)=i;
% end
% plot(nc,cv, '-*');axis([1 m 0.9 1.05]); grid on;

```

```
% % % %  
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%  
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%  
% xlswrite('pcs1.xlsx',score_cal,'score_cal');  
% xlswrite('pcs1.xlsx',score_val,'score_val');  
% xlswrite('pcs1.xlsx',CCRcal,'CCR','A1:A69');  
% xlswrite('pcs1.xlsx',CCRval,'CCR','B1:B47');  
% xlswrite('pcs1.xlsx',VecP,'loading');
```

ANEXO B. Rutina de la Red Neuronal

```
clc %% Borrar pantalla
clear all %%Borra todas las variables
close all %% Cierra todas las ventanas
%=====
cd C:\Users\usuario\Documents\Proyecto_FV\Archivos_Matlab\New_datos
%%%Trayectoria del archivo OJO ACTUALIZAR SIEMPRE
%=====
for n=3 %%%Número de componentes principales
    score_cal=xlsread('pcs1.xlsx','score_cal');
    score_cal=score_cal(:,1:n); %Varia el número de componentes principales
    score_val=xlsread('pcs1.xlsx','score_val');
    score_val=score_val(:,1:n);
    CCR_cal=xlsread('pcs1.xlsx','CCR','A1:A69');
    CCR_val=xlsread('pcs1.xlsx','CCR','B1:B46');
    [NFS NFC]=size(score_cal);
    [FCCR CCCR]=size(CCR_cal);
    minma= repmat([-1 1],n,1);
    for raiz=1:2:100
        m=1;
        for nco=1:30
            net = newff(minma,[0 70],nco,{'tansig' 'purelin'},'trainlm','learngdm');
            %=====
            %Parámetros de la red
            %=====
            net.trainParam.epochs =1000;
            net.trainParam.goal =0.1 ;
            net.trainParam.max_fail = 5;
```

```

net.trainParam.mem_reduc = 1;
net.trainParam.min_grad = 1e-10;
net.trainParam.mu = 0.01;
net.trainParam.mu_dec = 0.1;
net.trainParam.mu_inc = 10;
net.trainParam.mu_max = 1e10;
net.trainParam.show = 1;
net.trainParam.time = 240;
net.performFcn = 'mse';
net.divideFcn='';
%=====
rand('twister',raiz); net.IW{1,1}=rand(nco,NFC);
rand('twister',raiz); net.LW{2,1}=rand(1,nco);
rand('twister',raiz); net.b{1}=rand(nco,1);
rand('twister',raiz); net.b{2}=rand(1,CCCR);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
[net , tr] = train(net,score_cal',CCR_cal');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%Prueba o simulación de
desempeño%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Oc=sim(net,score_cal');
Ov=sim(net,score_val');
format bank

Nc=length(Oc);
Nv=length(Ov);

PRESS=sum((Oc-CCR_cal').^2); %Predictive residual sums of squares
SECPRESS=sqrt(PRESS/Nc); %Standar error of calibration

```

```

DESCAL=sqrt(sum((Oc-mean(Oc)).^2)/(Nc));

PRESSV=sum((Ov-CCR_val).^2); %Predictive residual sums of squares
SEPPRESS=sqrt(PRESSV/Nv); %Standar error of prediction
DESVAL=sqrt(sum((Ov-mean(Ov)).^2)/(Nv));
metricas(m,:)= [nco SECPRESS DESCAL SEPPRESS DESVAL]
m=m+1;
end
Hoja=['Raiz_' num2str(raiz)];
etiquetas={'nco' 'SECPRESS' 'DESCAL' 'SEPPRESS' 'DESVAL'};

Archivo=['C:\Users\usuario\Documents\Proyecto_FV\Archivos_Matlab\New_datos'
num2str(n) '.xlsx'];
xlswrite(Archivo, etiquetas, Hoja, 'A1:E1');
xlswrite(Archivo, metricas, Hoja, 'A2:E31');
clear metricas
raiz
end
end
save net
clear all
load net

```