

**APLICACIÓN DE LA REGRESIÓN LINEAL MÚLTIPLE Y LAS REDES
NEURONALES ARTIFICIALES PARA LA PREDICCIÓN DEL ÍNDICE DE
CETANO EN EL DIÉSEL UTILIZANDO PROPIEDADES MACROSCÓPICAS Y
ESPECTROS NIR**

JULIETH ANDREA DIAZ SERRANO

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISCOQUÍMICAS
ESCUELA DE INGENIERÍA QUÍMICA
BUCARAMANGA**

2015

**APLICACIÓN DE LA REGRESIÓN LINEAL MÚLTIPLE Y LAS REDES
NEURONALES ARTIFICIALES PARA LA PREDICCIÓN DEL ÍNDICE DE
CETANO EN EL DIÉSEL UTILIZANDO PROPIEDADES MACROSCÓPICAS Y
ESPECTROS NIR**

JULIETH ANDREA DIAZ SERRANO

Trabajo de grado presentado como requisito para optar al título de

INGENIERA QUIMICA

Director

GIOVANNI MORALES MEDINA

Ingeniero Químico, Ph.D

UNIVERSIDAD INDUSTRIAL DE SANTANDER

FACULTAD DE INGENIERÍAS FISICOQUÍMICAS

ESCUELA DE INGENIERÍA QUÍMICA

BUCARAMANGA

2015

DEDICATORIA

A Dios por estar presente en cada paso que he emprendido, por fortalecer mi espíritu, llenando de luz mi vida y cada nuevo paso que doy.

A mis padres, por enseñarme que con entrega y dedicación todo es posible, por su infinita confianza, motivación y apoyo durante el transcurso de esta grandiosa aventura.

A mi hermano, por ser una fuente de inspiración, de apoyo y amor, a quien le demuestro que la disciplina es una de las mejores virtudes que una persona puede tener y añorar para sí misma, ya que no hay cosas imposibles.

A Pérez, quien a pesar del inherente y confuso comienzo, fue mi compañero de carrera y logro convertirse en un hermano, con quien compartí muchas tardes de tertulias eternas en las que la mayoría de las veces no llegamos a ningún acuerdo.

A Andrea, Melisa, Oscar, Diego, Daniel y Oscar por acompañarme desde el comienzo de esta hermosa aventura, por brindarme su apoyo incondicional.

A Diana, Ángela, Jesús, Francy por ser llegar en el momento preciso, por su compañía, apoyo, confianza y compañerismo.

Y a todos y a cada una de las personas que he conocido en todo este caminar, lleno de muchas historias que guardan: alegrías, llanto, tristezas, traspasos, prisas, desilusiones, amores, desencantos y demás; que me recuerdan día con día que de esta historia solo culmina un capítulo, de un libro que se permite ser un lienzo.

AGRADECIMIENTOS

A Dios por su infinito amor y permitirme alcanzar esta meta.

A la Universidad Industrial de Santander y a la escuela de Ingeniería Química por su continuo apoyo durante todo el proceso de formación.

Dr. Giovanni Medina Morales, mi director, por haber confiado en mí para realizar este trabajo de investigación, con el cual tuve la oportunidad de conocer un poco más de cerca el fascinante mundo de la estadística multivariada y las redes neuronales artificiales. También quiero agradecerle por su paciencia, dedicación, entrega y por su continua invitación a recorrer el camino sin prisa y pausadamente, y el lema a recordar será: “el indio y su flecha”.

A Jorge y a Manuel ‘los analistas’ quienes se encargaron de sembrar, regar y germinar la semilla de la alegría o como ustedes mismos afirman: le adicionaron ‘el sabor’ a este último tramo de esta etapa.

A profesor Daniel Sierra, del grupo de investigación CEMOS, por abrirme sus puertas.

A todos y cada uno de los docentes que durante el transcurso de todos estos años han estado dispuestos a formar personas que serán profesionales íntegros y competentes.

A la unidad de blending de la refinería de Barrancabermeja, ECOPETROL S.A.

A mi familia y amigos quienes me acompañaron durante el recorrido de este arduo y fabuloso camino siendo fuente de motivación y apoyo.

TABLA DE CONTENIDO

	Pág.
INTRODUCCIÓN	14
1 MARCO TEÓTICO	16
1.1 PROPIEDADES DEL DIESEL	17
1.1.1 Número de cetano e índice de cetano.....	18
1.1.2 Temperatura de ebullición.	19
1.1.3 Punto de inflamación.	20
1.2 ESPECTROSCOPIA INFRAROJA	20
1.3 REGRESIÓN LINEAL MÚLTIPLE	21
1.4 REDES NEURONALES ARTIFICIALES	22
1.1.4 Componentes básicos de una RNA.	24
1.1.5 Características.	24
2 METODOLOGÍA	26
2.1..... ÍNDICE DE CETANO BASADO EN PROPIEDADES MACROSCÓPICAS Y RLM	26
2.1.1 Consecución y análisis de la información para la regresión.	26
2.1.2 Análisis y tratamiento de datos.....	26
2.1.3 Proposición y análisis de un modelo inicial	27
2.1.4 Generación y análisis de modelos derivados del inicial	27

2.1.5	Validación de los modelos derivados del inicial.....	28
2.1.6	Proposición del modelo de regresión para el índice de cetano: El s.	28
2.2	ÍNDICE DE CETANO BASADO EN RNA.....	28
2.2.1	Pretratamiento de datos.	30
2.2.2	Selección de la arquitectura de RNA.....	30
2.2.3	Entrenamiento de las RNA.	31
2.2.4	Desempeños y validación de la RNA	31
3	RESULTADOS Y ANÁLISIS DE RESULTADOS	32
3.1	PROPIEDADES MACROSCÓPICAS.....	32
3.2	REDES NEURONALES	46
3.2.1	Red neuronal con propiedades macroscópicas.....	47
3.2.2	Red neuronal basada en datos espectroscópicos	48
4	CONCLUSIONES	51
5	RECOMENDACIONES.....	52
	REFERENCIAS	53
	BIBLIOGRAFIA.....	57

LISTA DE TABLAS

Pág.

Tabla 1. Rangos de temperaturas de ebullición para las fracciones obtenidas en una destilación típica del petróleo (incluyendo destilación al vacío) [13].	17
Tabla 2. Propiedades características del diésel.	18
Tabla 3. Estadísticos de las muestras de diésel.	33
Tabla 4. Regresión de las propiedades fisicoquímicas con el IC	34
Tabla 5. Resumen ajuste ($R^2=0.9895$, $R^2_{ajust.}=0.9883$, $p\text{-value}<2.2e-16$).	35
Tabla 6. Prueba de modelos Anova caso Inicial*	35
Tabla 7. Nueva matriz de correlación	37
Tabla 8. Resultados de pruebas estadísticas de los modelos de regresión.	39
Tabla 9. Posibles puntos atípicos de los modelos de regresión.	39
Tabla 10. Modelo de regresión del Caso III.	41
Tabla 11. Prueba de modelos Anova caso III.	42
Tabla 12. Intervalos de confianza modelo caso III.	42
Tabla 13. Estudio del peso de las variables del modelo de regresión caso III	43
Tabla 14. Pruebas estadísticas de los modelos reducidos del caso III	44
Tabla 15. Intervalos de confianza modelo caso V.	44
Tabla 16. Resultados de las redes con funciones de activación I F.	47
Tabla 17. Resultados de las redes con funciones de activación II F	48
Tabla 18. Topología de las redes con funciones de transferencia I F NIR	48
Tabla 19. Topología de las redes con funciones de transferencia I F NIR (Continuación)	49
Tabla 20. Topología de las redes con funciones de transferencia II F NIR	49
Tabla 21. Resultados estadísticos con los mejores modelos de RLM y RNA	50
Tabla 20. Propiedades fisicoquímicas de las muestras del diésel de estudio.	65

Tabla 21. Propiedades fisicoquímicas de las muestras del diésel de estudio (Continuación).....	66
Tabla 22. Propiedades fisicoquímicas de las muestras del diésel de estudio (Continuación).....	67
Tabla 23. Propiedades fisicoquímicas de las muestras del diésel de estudio (Continuación).....	68
Tabla 24. Propiedades fisicoquímicas de las muestras del diésel de estudio (Continuación).....	69
Tabla 25. Propiedades fisicoquímicas de las muestras del diésel de estudio (Continuación).....	70
Tabla 26. Propiedades fisicoquímicas de las muestras del diésel de estudio (Continuación).....	71
Tabla 27. Correlaciones entre las variables	72
Tabla 28. Resumen ajuste caso I ($R^2=0.9832$, $R^2_{ajust.}=0.9817$).....	77
Tabla 29. Resumen ajuste caso II ($R^2=0.9884$, $R^2_{ajust.}=0.9874$).....	78
Tabla 30. Resumen ajuste caso III ($R^2=0.9892$, $R^2_{ajust.}=0.9884$).....	80
Tabla 31. Resumen ajuste caso IV ($R^2=0.9827$, $R^2_{ajust.}=0.9813$).	81
Tabla 32. Resumen ajuste caso V ($R^2=0.9783$, $R^2_{ajust.}=0.9771$).	83
Tabla 33. Resumen ajuste caso VI ($R^2=0.943$, $R^2_{ajust.}=0.9405$).	84
Tabla 34. Resumen ajuste caso VI ($R^2=0.9434$, $R^2_{ajust.}=0.941$).	86
Tabla 35. Correlaciones para las constantes E y F [16].....	89
Tabla 36. Datos CI ASTM D976 y ASTM D4737	91
Tabla 37. Datos CI ASTM D976 y ASTM D4737	91
Tabla 38. Datos CI ASTM D976 y ASTM D4737	92
Tabla 39. Datos CI ASTM D976 y ASTM D4737	92
Tabla 40. Datos CI ASTM D976 y ASTM D4737	93
Tabla 41. Desviación estándar y promedio de las propiedades autoescaladas.....	95

LISTA DE FIGURAS

	Pág.
Figura 1. Unidad de destilación atmosférica [13]	16
Figura 2. Configuración de un espectrofotómetro.	21
Figura 3. Modelo neurona artificial. Imagen tomada de [30].	23
Figura 4. Funciones de activación.	25
Figura 5. Espectros de calibración.....	29
Figura 6. Espectros de validación	29
Figura 7. Densidad vs Índice de refracción	37
Figura 8. Peso molecular vs Temperatura ebullición 50%	38
Figura 9. Leverage para el caso I	40
Figura 10. Distancia de Cook, caso I. Figura reportada por R.	40
Figura 11. Leverage para el caso I	77
Figura 12. Distancia de Cook caso I	78
Figura 13. Leverage para el caso II	79
Figura 14. Distancia de Cook caso II	79
<i>Figura 15. Leverage para el caso III</i>	<i>80</i>
<i>Figura 16. Distancia de Cook caso III</i>	<i>81</i>
Figura 17. Leverage para el caso IV	82
Figura 18. Distancia de Cook caso IV	82
Figura 19. Leverage para el caso V	83
Figura 20. Distancia de Cook caso V	84
Figura 21. Leverage para el caso VI	85
Figura 22. Distancia de Cook caso VI	85
Figura 23. Leverage para el caso VII	86
Figura 24. Distancia de Cook caso VII	87
Figura 25. Correlación IC ASTM 4737 vs IC ASTM D976.....	93

RESUMEN

TÍTULO: APLICACIÓN DE LA REGRESIÓN LINEAL MÚLTIPLE Y LAS REDES NEURONALES ARTIFICIALES PARA LA PREDICCIÓN DEL ÍNDICE DE CETANO EN EL DIÉSEL UTILIZANDO PROPIEDADES MACROSCÓPICAS Y ESPECTROS NIR.*

AUTOR: Julieth Andrea Díaz Serrano.**

PALBRAS CLAVE: Regresión lineal múltiple, Redes neuronales artificiales, Diésel, Espectroscopia infrarroja cercana, Índice de cetano.

DESCRIPCIÓN:

Son múltiples las ventajas de desarrollar un modelo de regresión basado en datos. Entre ellas es posible mencionar: comprensión de las variables influyentes en los procesos, determinación de los estadísticos que presentan los datos, análisis de datos atípicos e influyentes y predicción de nuevos valores de propiedades. Dadas las ventajas, el desarrollo de un modelo de regresión para la estimación del índice de cetano en el diésel permitiría descartar variables sin importancia para la determinación de esta propiedad así como proponer estructuras matemáticas alternativas a las descritas en las normas ASTM D 4737 y ASTM D 976.

En este trabajo se desarrollaron diferentes modelos de regresión para la predicción del índice de cetano (CI) en el diésel, por medio del uso de métodos de regresión lineal múltiple (RML) y las redes neuronales artificiales (RNA); los procedimientos matemáticos fueron aplicados según su codificación en el software R versión libre y en el *toolbox* respectivo de MATLAB. Para el desarrollo de los modelos de regresión fueron utilizados dos tipos de datos: el primero representado por propiedades macroscópicas como la densidad, la temperatura de ebullición y la viscosidad, entre otras, mientras que el segundo tipo corresponde a los derivados de la aplicación de técnicas instrumentales de espectroscopia infrarroja. Los resultados de este trabajo demuestran que los modelos propuestos predicen con alta exactitud el CI obtenido por la aplicación de la norma ASTM D-4737. Sin embargo por la facilidad de disposición y obtención de los datos se prefiere el modelo de predicción basado en las redes neuronales a partir de espectroscopia infrarroja cercana.

* Trabajo de grado

** Facultad de Ingenierías Físicoquímica. Escuela de Ingeniería Química. Director: Giovanni Morales, Dr. Ing. Químico

ABSTRACT

TÍTULO: APPLICATION OF MULTIPLE LINEAR REGRESSION AND ARTIFICIAL NEURAL NETWORKS FOR THE CETANE INDEX PREDICTION OF DIESEL-FUEL USING MACROSCOPIC PROPERTIES AND SPECTRA NIR *

AUTHOR: JULIETH ANDREA DÍAZ SERRANO.**

KEY WORDS: Multiple linear regression; Artificial neural networks; Diesel; Near infrared spectroscopy; Cetane index

DESCRIPCIÓN:

There are some advantages to developing a regression model based on data. These include: understanding of the influential variables on the process, calculation of statistics from data, analysis of outlier and influential points and, prediction of new values for the properties. Accordingly, development of a regression for the cetane index of diesel fuels will allow to disregard unimportant variables for the model as well as to propose alternative mathematical equations to that reported in the ASTM D 4737 and ASTM D 976 protocols.

In this document, some regression models were developed to predict the cetane index (CI) of diesel fuels through the application of multiple linear regression (MLR) and artificial neural networks (ANN) as codified in R and MATLAB computational programs, respectively. The regression models were developed considering two kind of data: macroscopic properties such as density, boiling temperature, viscosity, among others and, those data obtained from near infrared spectra. The results of the present work showed that the developed models predicted the CI (ASTM D-4737) with high accuracy. In terms of time and costs of the data, the best model for the prediction of CI corresponded to that obtained by the application of artificial neural networks using near infrared spectroscopy information.

* Degree Work

** Faculty of Phisico-Chemical Engineering. School of Chemical Engineerign. Director: Giovanni Morales Medina, Dr. Ing.Chemical

INTRODUCCIÓN

Los avances tecnológicos orientados al desarrollo de sistemas inteligentes en la industria del petróleo y la petroquímica son cada vez más numerosos, ya que estos proporcionan información de calidad sin requerir de un modelo determinístico riguroso, utilizando un tiempo de cálculo reducido. Adicionalmente, la aplicación de los sistemas inteligentes, en particular las redes neuronales artificiales, se ve beneficiada por la necesidad de realizar constantes monitoreos y controles de los procesos basados en la información que suministran los sistemas de medición en línea y los análisis realizados en el laboratorio (fuera de línea) [1,2].

A pesar de que los procedimientos de análisis fuera de línea, han sido estandarizados por organizaciones nacionales e internacionales (ASTM¹, ISO², etc.), varios de estos procedimientos requieren de un tiempo apreciable para generar datos confiables; adicionalmente, algunos procedimientos estándar requieren de sustancias ambientalmente peligrosas y ensayos laboriosos [3–5]. Por estas razones, diferentes modelos de regresión han sido desarrollados como alternativa a la estimación analítica. Los modelos de regresión pueden ser agrupados, en general, en dos tipos de datos [6]: el primero está representado por propiedades macroscópicas como la densidad, la temperatura de ebullición y la viscosidad, entre otras (*i.e.* mediciones que generan un solo dato por muestra), mientras que el segundo tipo se obtiene por la aplicación de técnicas instrumentales (*e.g.* espectroscopia infrarroja) que generan una gran cantidad de información por cada muestra medida. Comparativamente, las técnicas instrumentales requieren de una menor cantidad de muestra, menor tiempo de ensayo y un menor costo relativo [7,8].

¹ ASTM: American Society for Testing and Materials

² ISO : [International Organization for Standardization](#)

Son múltiples las ventajas de desarrollar un modelo de regresión basado en datos. Entre ellas es posible mencionar: comprensión de las variables influyentes en los procesos, determinación de los estadísticos que presentan los datos, análisis de datos atípicos e influyentes y predicción de nuevos valores de propiedades [9–11]. Dadas las ventajas que presenta un modelo de regresión, la estimación del índice de cetano antes de su verificación en el laboratorio permitiría la toma de decisiones y validaría la ejecución de correcciones sobre la marcha.

El propósito del presente trabajo se enfoca en proponer un modelo de regresión para la predicción del índice de cetano del diésel. La metodología utilizada consiste en tomar una serie de datos provenientes de laboratorio con los cuales se desarrollaran un modelo de regresión lineal múltiple y otro no lineal basado en redes neuronales artificiales. En primer lugar, los dos métodos de regresión son aplicados a un conjunto de datos macroscópicos (tipo I) reportados en la literatura para diferentes muestras de diésel; entre las propiedades macroscópicas se encuentran densidad, viscosidad, T10%, T50%, T90%, peso molecular, índice de refracción, punto de anilina, contenido de H₂, y punto de inflamación.

Posteriormente el método de redes neuronales artificiales es aplicado para el desarrollo de un modelo de predicción utilizando una base de datos de espectros infrarrojo (NIR) (datos tipo II) suministrados por el departamento de *blending* de la refinería de Barrancabermeja. Finalmente, los desempeños de los diferentes modelos de regresión son comparados en términos de estadísticos relacionados con el error de calibración o entrenamiento y validación.

1 MARCO TEÓTICO

El petróleo es una de las sustancias más importantes y consumidas en el mundo. Es una de las principales fuentes de energía útil para las industrias, la calefacción y el transporte, entre otras actividades. El petróleo puede ser definido como una mezcla compleja natural de hidrocarburos, generalmente en estado líquido, que puede incluir compuestos de azufre, nitrógeno, oxígeno, metales y otros elementos [12]. Esta mezcla de compuestos también conocida como crudo, se transfiere a una refinería para ser procesado y convertido en diversos productos. Los procesos que se llevan a cabo se pueden clasificar en tres tipos: separación (destilación), conversión (coquización y craqueo catalítico) y acabado (hidrotratamiento para la eliminación de azufre) [13]. En la destilación fraccionada atmosférica y al vacío (Figura 1), el crudo es separado en diferentes mezclas de compuestos o fracciones debido a la diferencia entre sus puntos de ebullición (Tabla 1).

Figura 1. Unidad de destilación atmosférica [13]

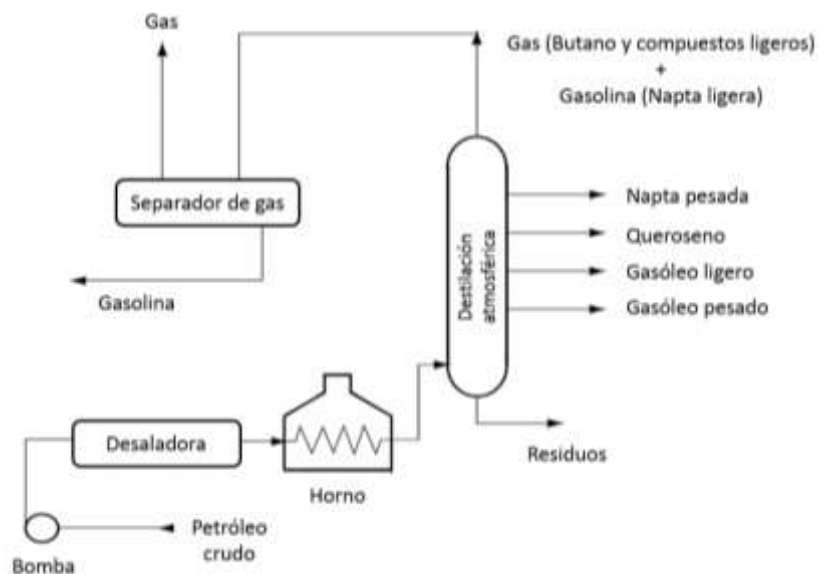


Tabla 1. Rangos de temperaturas de ebullición para las fracciones obtenidas en una destilación típica del petróleo (incluyendo destilación al vacío) [13].

Fracción	Rango de temperatura de ebullición [°C]
Gases	< 25
Gasolina	25 – 75
Gasolina pesada	75 – 190
Keroseno	190 – 250
Diésel	250 – 350
Gasóleo pesado (lubricantes)	300 – 370
Gasóleo de vacío	370 – 600
Residuo	>600

El diésel es uno de los productos obtenidos mediante la destilación atmosférica del petróleo (Tabla 1). El diésel es una mezcla compleja de cientos de compuestos individuales, generalmente conformados por cadenas de hidrocarburos en un rango de C9–C20 [13]. Se compone de aproximadamente del 75% de hidrocarburos saturados (parafinas normales, isoparafinas y cicloparafinas), e hidrocarburos aromáticos en un 25% (en los que se están incluidos los naftalenos y alquilbencenos) [14]. El diésel es principalmente utilizado como combustible en máquinas de combustión interna de alta eficiencia. La concentración de los componentes de los gases de la combustión corresponde a: 75,2% de Nitrógeno (N₂), 15% de Oxígeno (O₂), 7,1% de Dióxido de Carbono (CO₂), 2,6% de Agua (H₂O) y 0,09% de sustancias nocivas.

1.1 PROPIEDADES DEL DIESEL

En Colombia el Ministerio de Minas y Energía es el encargado de establecer la estructuración de los programas de aseguramiento y control de calidad (QA/QC) de los biocombustibles y sus mezclas con combustibles fósiles en Colombia, algunos de los parámetros se presentan a continuación:

Tabla 2. Propiedades características del diésel.

Parámetro	Unidad	Especificación fecha de vigencia	Método de ensayo
Azufre, máximo	% masa	0,005	ASTM D4294
Aromáticos, máximo	% Vol.	35 32 (Jul 31 de 2016 en adelante)	ASTM D5186 ó ASTM D1319
Número de Cetano, mínimo	—	43 45 Resto País Bogotá	ASTM D613
Índice de Cetano, mínimo	—	45	ASTM D976 ó ASTM D4737
Contenido de Biocombustible (máximo)	% Volumen	10	EN 14078
Corrosión al cobre, 3h a 50°C, máximo	Clasificación	2	ASTM D130
Punto de inflamación, mínimo	°C	52	ASTM D93

1.1.1 Número de cetano e índice de cetano.

El número de cetano es una medida de la calidad de ignición de un combustible e influye en las emisiones de humo y en la calidad de la combustión; su determinación se encuentra especificada en la norma ASTM D-613 [15]. En la práctica la determinación del número de cetano es costosa y laboriosa, por lo cual el instituto americano del petróleo propone una ecuación de dos parámetros (ASTM D-976) para la estimación del número de cetano por medio del índice de cetano [16]. El índice de cetano (IC) corresponde al porcentaje de cetano en una mezcla cetano – heptametilnonano con las mismas características de ignición que la muestra de diésel analizada.

De otro lado, el índice de cetano se incrementa a medida que aumenta la longitud de la cadena. En general, los aromáticos y los alcoholes tienen un índice de cetano bajo. Por ello el porcentaje de gasóleos desintegrados en el diésel, se ve limitado por su contenido de aromáticos. La norma ASTM D-4737 reporta una regresión basada en cuatro parámetros para calcular el IC (índice de cetano calculado) la cual se presenta a continuación:

$$CCI = 45.2 + (0.0892)(T_{10N}) + [0.131 + (0.901)B][T_{50N}] \\ + [0.0523 - (0.420)B][T_{90N}] + [0.00049][(T_{10N})^2 - (T_{90N})^2] + 107B \\ + 60B^2 \quad (1)$$

Donde: $B = \exp[-3,5(DN)] - 1$ $DN = \rho_{15} - 0,85$ $T_{10N} = T_{10} - 215$

$T_{50N} = T_{50} - 260$ $T_{90N} = T_{90} - 310$

T_{10} temperatura de ebullición recuperada al 10%, T_{50} temperatura de ebullición recuperada al 50%, T_{90} temperatura de ebullición recuperada al 90% por la norma ASTM D-86.

Sin embargo, la norma ASTM D 976 también reporta una regresión basada en dos parámetros para calcular el IC (índice de cetano calculado), la cual se presenta a continuación:

$$CCI = 454.74 - 1641.416 \rho_{15} + 774.74 \rho_{15}^2 - 0.554B + 97.803(\log B)^2 \quad (2)$$

Donde, ρ_{15} corresponde a la densidad a 15 °C, g/mL y B a la temperatura media de ebullición (T50%), °C.

1.1.2 Temperatura de ebullición. Las curvas de destilación (True Boiling Point) es una medida de evaluación del crudo, y es una herramienta muy útil para la caracterización, el diseño, operación y optimización de los procesos de la refinería, para la clasificación del petróleo, y para el desarrollo de correlaciones de propiedades del petróleo, las cuales se han utilizado en todo el mundo [17]. Para los productos derivados del petróleo se utiliza el proceso de destilación molecular, el cual está definido por ASTM D 86 como la medida en porcentaje de un líquido en la que la presión de vapor es igual a la presión atmosférica y el líquido se convierte en gas, con la cual es posible determinar TBP [18].

1.1.3 Punto de inflamación. El punto de inflamación (en Inglés *flash point*) es una característica muy importante de los líquidos y sustancias de bajo punto de ebullición. El *flash point* es definido por la ASTM D93 como la temperatura más baja, corregida a una presión de 101,3 kPa, en el que la aplicación de una fuente de ignición causa los vapores de un espécimen de muestra para inflamarse por debajo de las condiciones específicas [19].

Los valores para las anteriores propiedades macroscópicas varían en un determinado rango para los diferentes diésel del petróleo y corresponden a una consecuencia de su composición. Por lo tanto, y como Riazi y otros autores lo anotan en sus modelos predictivos [16], el índice de cetano puede ser predicho a través de un modelo de regresión utilizando otras propiedades del diésel como variables independientes.

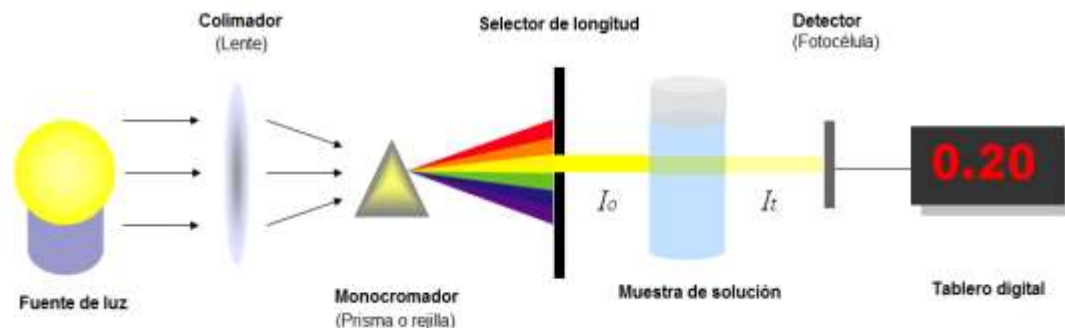
1.2 ESPECTROSCOPIA INFRAROJA

La espectroscopia infrarroja cercana (NIR) es una técnica que en los últimos años, ha sido cada vez más adaptada como herramienta analítica en diferentes campos como el ambiental, farmacéutico, alimentos, petroquímica entre muchos otros; esto se debe a que presenta ventajas al realizar un análisis (cualitativo o cuantitativo), tales como: la toma de espectros tarda unos pocos minutos, es posible realizar un análisis de muestras multicomponentes, y las muestras requieren muy poca o nula preparación [20,21].

La técnica NIR reporta la intensidad relativa de la muestra en el rango de números de onda entre 12800 y 4000 cm^{-1} . La espectrometría es un método que permite medir la cantidad de luz que absorbe una sustancia química por medio de la

intensidad de luz, como un haz de luz pasa por la solución de la muestra. Las partes del equipo se pueden apreciar en la siguiente figura [22]

Figura 2. Configuración de un espectrofotómetro.



Al realizar un análisis espectral, la información presente en un espectro NIR puede ser empleada en diferentes casos como lo es en la estimación de la concentración de un componente o para la estimación de una propiedad fisicoquímica. Para alcanzar esto, es posible hacer uso de diferentes métodos multivariados de análisis como lo son: PLS y las redes neuronales artificiales [22–26].

1.3 REGRESIÓN LINEAL MÚLTIPLE

Los modelos de regresión lineal múltiple (RLM) expresan la medida de una variable dependiente Y como una función de más de una variable de predicción o independiente (x_i), según:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i + \epsilon \quad (3)$$

Donde, Y es la variable de respuesta que se desea predecir, $\beta_0, \beta_1, \dots, \beta_i$ son los coeficientes de regresión, $x_1, x_2, x_3, \dots, x_i$ corresponden a las variables independientes y ϵ es el error de la estimación o residuo. Los supuestos del modelo

RLM respecto a ϵ son [9,10]: 1) los residuos de cada estimación son independientes entre ellos; 2) la distribución de los residuos presenta una media en 0 y una varianza constante de σ^2 para cualquier conjunto $x_1, x_2, x_3, \dots, x_i$, y 3) los residuos están normalmente distribuidos.

La solución del modelo RLM involucra el uso del algoritmo de mínimos cuadrados para obtener el mejor conjunto de valores para los coeficientes que minimicen la sumatoria del cuadrado de los residuos [10,11,27]. Según el algoritmo de mínimos cuadrados, los coeficientes pueden ser calculados según la fórmula matricial:

$$\beta = (X^T X)^{-1} X^T Y \quad (4)$$

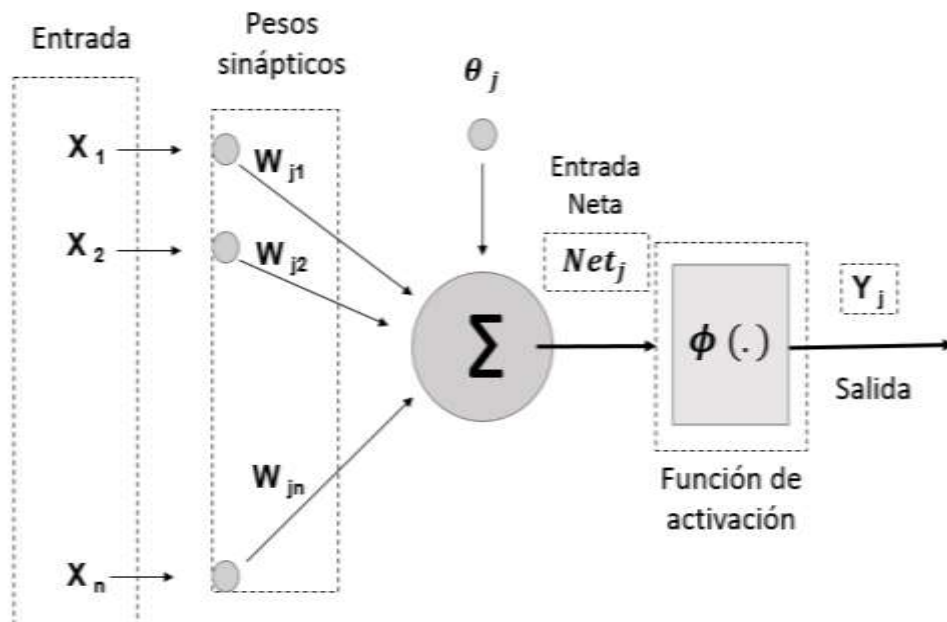
Donde β es el vector de coeficientes, X^T corresponde a la matriz transpuesta de la matriz de datos iniciales X y Y corresponde al vector de variables dependientes [9]. La aplicación del algoritmo de mínimos cuadrados no conlleva al cumplimiento de los supuestos sobre los residuos del modelo, por lo que se deben efectuar diferentes análisis a la solución obtenida. Estos análisis corresponden a: prueba de significancia del modelo y de sus coeficientes por medio de la tabla de análisis de varianza (ANOVA) y distribución de los residuos por medio del análisis de su tendencia gráfica y su desviación respecto a la normalidad.

1.4 REDES NEURONALES ARTIFICIALES

El cerebro humano contiene aproximadamente mil millones de neuronas. Las neuronas constituyen elementos simples en un gran sistema interconectado, por medio de un mecanismo de comunicación bien diferenciado. En términos generales, las dendritas reciben las señales de entrada, el cuerpo celular combina, integra y emite la señal de salida hacia los axones y de ahí a otras neuronas (sinapsis). Por lo general una neurona recibe información de miles de otras neuronas y, a su vez, envía información a otras miles. Biológicamente, el cerebro aprende mediante la modificación de esas conexiones sinápticas en el sistema neuronal [28].

Las redes neuronales artificiales (RNA) son algoritmos que imitan el comportamiento del sistema neuronal biológico a través de flujos de información entre diferentes modelos matemáticos [29]. En las RNA la reorganización de las conexiones sinápticas se da mediante un mecanismo de pesos, los cuales son ajustados durante la etapa de aprendizaje o entrenamiento. La Figura 3 ilustra las partes de una RNA, en donde la información que recibe la neurona se define con el vector $X = [x_1, x_2, \dots, x_n]$. La información recibida por la neurona es modificada por un vector w de pesos sinápticos cuyo papel es el de emular el proceso sináptico biológico. El parámetro θ_j se conoce como el umbral de una neurona. Los diferentes valores que recibe la neurona, modificados por los pesos sinápticos, son sumados para obtener la entrada neta al sistema. Esta entrada es la que determina si la neurona está activa. La entrada neta se evalúa en esta función y se obtiene la salida de la neurona [30].

Figura 3. Modelo neurona artificial. Imagen tomada de [30].



1.1.4 Componentes básicos de una RNA. La arquitectura de una RNA está compuesta por niveles o capas de un número determinado de neuronas; las neuronas de una misma capa reciben simultáneamente el mismo tipo de información. En una RNA se identifican tres niveles [29,30]:

1. Entrada: Corresponde al conjunto de neuronas que reciben la información de forma directa.

2. Niveles ocultos: Conjunto de neuronas internas de la red que no tienen contacto directo con el exterior. El conjunto de neuronas en cada nivel oculto puede interconectarse de diferentes maneras.

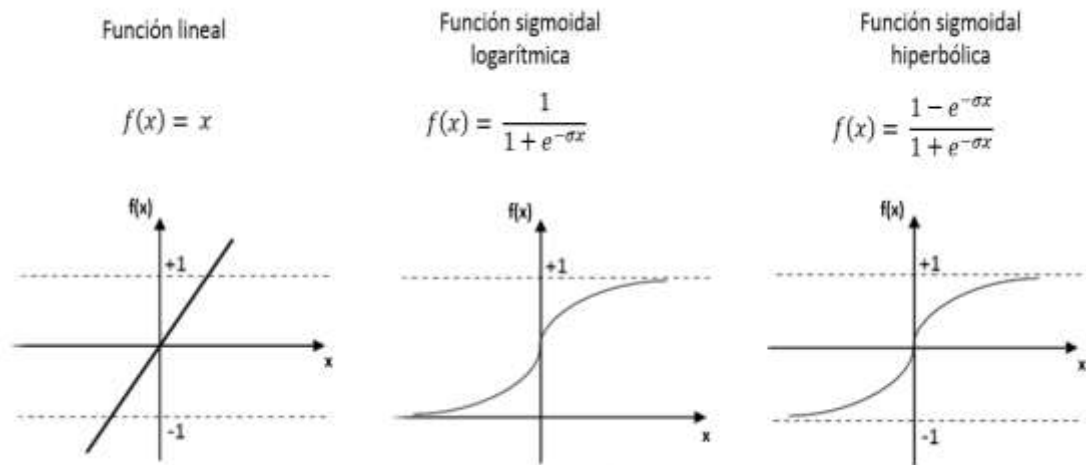
3. Salida: Son el conjunto de neuronas que transfieren la información de salida producto del procesamiento en la red.

1.1.5 Características. Existen tres aspectos que caracterizan una red neuronal: su topología o arquitectura, el mecanismo de aprendizaje, el tipo de asociación realizada. La arquitectura consiste en la organización de las neuronas formando capas o agrupaciones de neuronas más o menos alejadas de la entrada y salida de la red. Por ello deben identificarse los parámetros fundamentales que son [11]: número de capas (monocapa o multicapa), número de neuronas por capa, grado de conectividad y tipo de conexión entre neuronas.

El mecanismo de aprendizaje es definido por Mendel McClare como: “ un proceso mediante el cual los parámetros libres de una red neuronal artificial (pesos sinápticos o biases) son adaptados a través de un proceso de estimación del ambiente en el cual está embebida la red. El tipo de aprendizaje está determinado

por la forma como se cambian los parámetros en el proceso”, donde dentro de los más comunes se encuentran: el aprendizaje supervisado y aprendizaje no-supervisado. Los tipos de asociación están definidos por funciones de activación o también llamadas de transferencia, que corresponden a funciones matemáticas, utilizadas para modelar el comportamiento de una neurona artificial [29,30]. Existen varios tipos de estas funciones, las cuales se pueden apreciar en la siguiente figura:

Figura 4. Funciones de activación.



2 METODOLOGÍA

2.1 ÍNDICE DE CETANO BASADO EN PROPIEDADES MACROSCÓPICAS Y RLM

Las siguientes actividades fueron desarrolladas para la regresión del índice de cetano utilizando regresión lineal múltiple (RLM):

2.1.1 Consecución y análisis de la información para la regresión: En esta actividad se efectúa una búsqueda en la literatura abierta para conformar una base de datos de muestras de diésel con propiedades macroscópicas. Para el desarrollo de un modelo de RLM basado en propiedades macroscópicas, la base de datos conformada se divide en datos de calibración (70% de los datos totales) y datos de validación (30% de los datos totales); los modelos se construyen con los datos de calibración y se determina el desempeño de la misma con el conjunto de validación [9,10].

2.1.2 Análisis y tratamiento de datos: En esta actividad se aplican los métodos propios de la estadística univariada y bivariada a la base de datos recolectada [10,31]. Asimismo, los datos son transformados matemáticamente mediante el autoescalado; este tratamiento consiste en dos pasos: centrado de datos utilizando la media de cada variable y la normalización mediante la desviación estándar. Con el autoescalado, las nuevas variables obtenidas resultan ser adimensionales y presenta un similar rango de valores, con lo cual se garantiza que cada variable tenga la misma influencia en un modelo.

2.1.3 Proposición y análisis de un modelo inicial: La proposición de un modelo inicial se efectúa por medio del método de todas las regresiones posibles [10,11], según el algoritmo codificado en STATGRAPHIS Centurion VII. De acuerdo al método de las regresiones posibles, los modelos desarrollados contenían combinaciones desde una hasta las once variables, determinando los mejores modelos mediante los criterios de R^2 y $R^2_{ajustado}$. El modelo inicial obtenido (mayores valores de R^2 y $R^2_{ajustado}$) es analizado utilizando las pruebas estadísticas: tabla de análisis de varianza (ANOVA) y normalidad de los residuos.

2.1.4 Generación y análisis de modelos derivados del inicial: La prueba ANOVA que envuelve la significancia estadística de las variables en el modelo inicial corresponde al punto de partida para la generación de modelos derivados. Las variables que no muestran significancia estadística son retiradas de una en una del modelo generando modelos derivados a los cuales también se les aplica el análisis ANOVA y las pruebas para la normalidad de los residuos. Adicionalmente, en los modelos derivados se buscan puntos atípicos e influyentes que puedan sesgar los resultados; las pruebas estadísticas para estos corresponden al *Leverage* y la distancia de Cook. Los diferentes procedimientos para el ajuste de las RLM y las pruebas estadísticas son aplicadas según los algoritmos codificados en el paquete estadístico libre denominado *R* (Anexo C).

Para cada modelo derivado se calculan los estadísticos de desempeño: R^2 , $R^2_{ajustado}$ y la raíz del promedio del cuadrado de los errores (RMSE: *root mean square error*); el estadístico RMSE se define según (1) [32,33]

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N_t} (CI_i - \hat{CI}_i)^2}{N_t}} \quad (3)$$

Donde $CI_i - \widehat{CI}_i$ son los índices de cetano medidas y predichos, respectivamente; N_t es el número de muestras de prueba.

2.1.5 Validación de los modelos derivados del inicial: Los modelos derivados del inicial son validados con el conjunto del 30% de los datos totales (validación externa) y el desempeño de cada uno es verificado por medio de los estadísticos Q^2 y RMSEV. Q^2 corresponde al coeficiente de determinación para la validación y es determinado como $1 - \text{PRESS}/\text{SST}$, donde PRESS es la suma de los cuadrados de los residuos de predicción según el conjunto externo y SST representa la varianza total del conjunto. Asimismo, RMSEV corresponde a la sumatoria promedio de los cuadrados de los errores para la validación. Adicionalmente, cada modelo es validado utilizando el método de validación interna o cruzada, con lo cual se obtienen los estadísticos de validación cruzada R^2_{cv} y RMSECV [9–11].

2.1.6 Proposición del modelo de regresión para el índice de cetano: El modelo de regresión para el índice de cetano basado en propiedades macroscópicas se selecciona de acuerdo al valor de los estadísticos generados en cada etapa del desarrollo de los modelos de las actividades anteriores.

2.2 ÍNDICE DE CETANO BASADO EN RNA

En esta sección se realizaron dos modelos no lineales para la predicción del índice de cetano (IC) a partir de dos fuentes de bases de datos diferentes, el primero a partir de propiedades macroscópicas y el segundo basado en espectros del infrarrojo cercano (NIR). El primer modelo se basa en los datos recolectados en la sección 2.1. Para el segundo paquete de datos se utilizaron cincuenta y nueve (59) espectros NIR de diésel colombianos (Figuras 5 y 6); estos espectros fueron suministrados por el departamento de *blending* de la gerencia refinera de Barrancabermeja.

Figura 5. Espectros de calibración

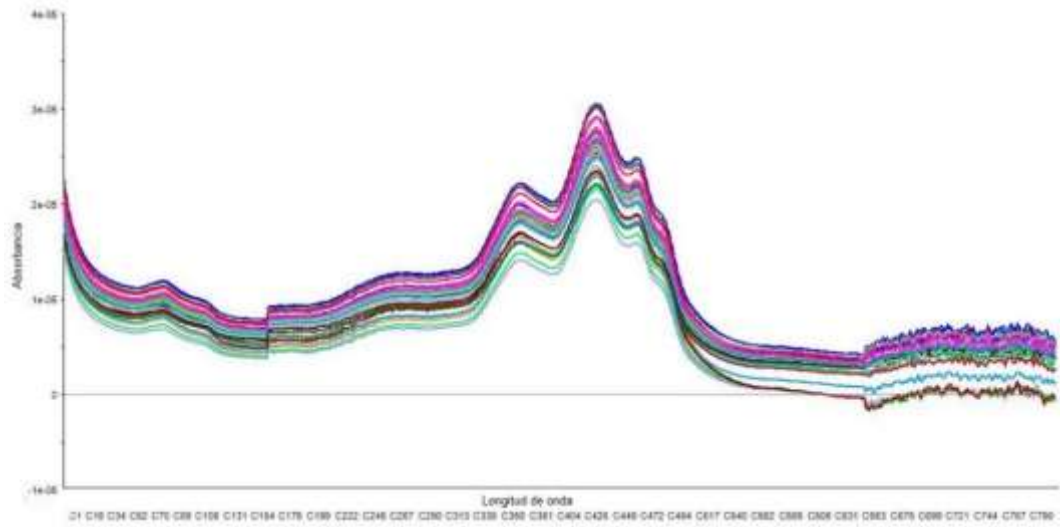
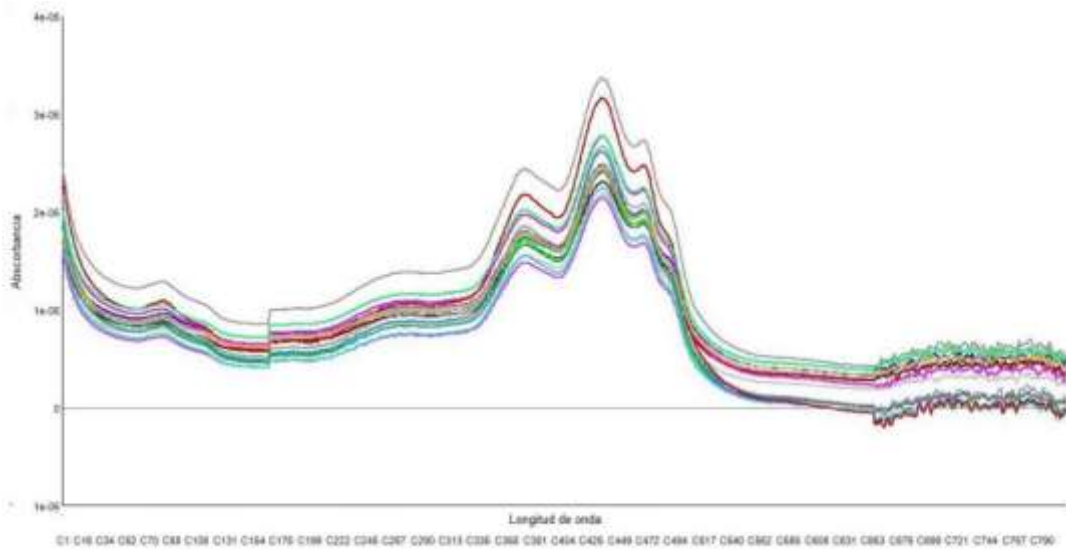


Figura 6. Espectros de validación



Del mismo modo que la sección anterior, se destina el 70% de los datos para el entrenamiento de la RNA respectiva y el 30% complementario para la validación de

los modelos obtenidos. Los dos modelos de RNA para la predicción del IC fueron desarrollados siguiendo las mismas actividades, las cuales se detallan a continuación.

2.2.1 Pretratamiento de datos. Debido a que los espectros NIR poseen variaciones entre las diferentes muestras analizadas (*i.e.* variaciones en las líneas base, cambios en las intensidades relativas, cercanía de los picos y resolución del espectro), se requiere la aplicación de un pretratamiento de los datos como etapa previa al entrenamiento de la RNA. Con el objetivo de mejorar la calidad de los datos, los espectros NIR se pueden tratar utilizando la normalización por rangos, en el cual los espectros se normalizan con el valor del área bajo la curva del espectro.

Los números de onda utilizados para el desarrollo del modelo RNA basado en NIR corresponde a los rangos: 4440-4760 cm^{-1} , 5280-6220 cm^{-1} y 8130-8430 cm^{-1} , según lo recomendado por Santos y colaboradores [22].

2.2.2 Selección de la arquitectura de RNA. La selección de la arquitectura de la red se basa en reportes de aplicación de las RNA en la predicción de propiedades para los productos del petróleo. Según el análisis de manuscritos [34–38], el modelo de RNA denominado *feed-forward* corresponde a uno de los más utilizados en la literatura. En esta RNA la información fluye entre capas de neuronas en una dirección siempre hacia adelante, con lo cual no se presenta reciclo de información a capas de neuronas previas. Asimismo, la optimización de la arquitectura en la red neuronal es posible considerando la variación en el número de neuronas presentes en la capa oculta, así como cambiando las funciones de activación (tansig-tansig y tansig-purelin) las cuales son las encargadas del procesamiento de los datos en el interior de la red [34–37,39].

2.2.3 Entrenamiento de las RNA. El algoritmo de entrenamiento denominado *backpropagation* es utilizado para el desarrollo de las RNA. En este entrenamiento se busca minimizar la sumatoria del cuadrado de los errores en el espacio descrito por los pesos que se le asigna a cada dato de entrada y salida de las neuronas, por medio de un método de gradiente descendente como el de Levenberg-Marquardt [22,24,34].

2.2.4 Desempeños y validación de la RNA. Con el fin de caracterizar la capacidad de predicción y precisión de los modelos, se calcularon los parámetros estadísticos como: R^2 y el RMSE para cada etapa [32,33]. La red con mejores estadísticos de desempeño es seleccionada como modelo de regresión para el índice de cetano en cada caso (propiedades macroscópicas o espectros NIR).

3 RESULTADOS Y ANÁLISIS DE RESULTADOS

3.1 PROPIEDADES MACROSCÓPICAS

La información para la regresión de propiedades macroscópicas del diésel fue tomada del manuscrito publicado por Stratiev y colaboradores [19]. La información consta de 140 muestras de diésel caracterizados con las propiedades: densidad, T10%, T50%, T90%, viscosidad, peso moléculas, índice de refracción, punto de anilina y contenido de hidrogeno. Esta información es resumida en la Tabla 2 y puede ser analizada en el Anexo A.

Adicional a las propiedades reportadas por Stratiev y colaboradores, se adicionaron la temperatura promedio de ebullición (T_p) y el punto de inflamación (PI); estas propiedades fueron consideradas ya que pueden influir en las características de ignición del diésel [16]. Las ecuaciones utilizadas para calcular T_p y PI fueron [5]:

$$T_p = \sum_{i=1}^n \vartheta_i T_i \quad (4)$$

Donde T_p es la temperatura promedio de ebullición en K, ϑ_i es el coeficiente de la temperatura y T_i corresponde a la temperatura para un determinado rendimiento volumétrico durante la ebullición del diésel (en K).

$$\begin{aligned} & \frac{1}{(1.8 * (PI) - 0.33)} \\ &= - 0.014568 + \frac{2.84947}{(1.8 * (T_p) - 0.33)} \\ &+ 1.903 * 10^{-3} \ln(1.8 * (T_p) - 0.33) \quad (5) \end{aligned}$$

Donde PI es la temperatura del punto de inflamación en K.

Tabla 3. Estadísticos de las muestras de diésel.

Propiedad	Símbolo	Unidades	Valor mínimo	Valor promedio	Valor máximo	Desviación estándar
CI ASTM D-4737	IC	-----	13,0000	48,6388	74,8000	10,2755
Densidad 15°C	ρ	kg/m ³	790,2000	848,1237	1013,0000	34,2848
T 10% (v/v)	T10	K	401,1500	505,5644	615,1500	41,0413
T 50% (v/v)	T50	K	437,1500	541,6601	618,6500	36,5331
T 90% (v/v)	T90	K	471,1500	584,6665	643,1500	43,6525
Viscosidad 40°C	ϑ	m ² /s	9,50E-07	3,54E-06	1,08E-05	1,90E-06
Peso molecular	PM	kg/kmol	128,0000	194,6978	259,0000	26,1241
Índice de refracción 20°C	IR	---	274,5910	274,6244	274,7359	0,0224
Punto de anilina	PA	K	294,1500	339,6464	368,3500	11,7129
Contenido H2 (%)	H2	%	8,3900	12,9353	14,7500	0,9662
T ebul. prom.	T_p	K	452,4833	543,9637	621,6500	35,4462
Punto de inflamación	PI	K	291,7791	361,4564	413,1181	22,8350

Como punto inicial para la generación del modelo de regresión, la RLM fue aplicada utilizando el algoritmo *stepwise* para disminuir el número de variables independientes de importancia para el modelo (Tabla 2) (Previamente las variables de la base de datos fueron autoescaladas, ver método en anexo G; este proceso de calibración del IC utilizó 98 muestras de la base de datos seleccionadas aleatoriamente. Algunos de los resultados del algoritmo *stepwise* son mostrados en la Tabla 3. Según los resultados de la Tabla 3, el punto de anilina es una de las propiedades con mayor peso en la explicación de la varianza residual del IC ($R^2=0,693$). A pesar de este resultado, se decidió partir con todas las variables independientes ya que explican aproximadamente el 99% de la varianza residual.

La tabla ANOVA fue aplicada el modelo de regresión inicial (todas las variables) para determinar su significancia estadística y el aporte estadístico de cada variable al modelo propuesto. Las Tablas 4 y 5 presentan el resultado del análisis de

varianza para el modelo inicial. Según la Tabla 4, el modelo de regresión lineal tiene significancia estadística al 95% de confianza ($p < 0.05$). De otro lado, las variables densidad (D), índice de refracción (IR) y viscosidad (PM) no aportan a la explicación de la varianza del modelo al 95% de confianza ($Pr > 0,05$). En forma de aclaración, la temperatura promedio de ebullición (Tp) fue excluida por los programas utilizados debido a su colinealidad con las temperaturas T10, T50 y T90.

Tabla 4. Regresión de las propiedades fisicoquímicas con el IC

Variable	R ² , %
Densidad	19,4
T10	9,3
T50	6,7
T90	4,9
Viscosidad	1,8
Peso molecular	18,5
Índice de refracción	24,4
Punto de anilina	69,3
Contenido de H2	40,7
Temperatura media ebullición	8,8
Punto de inflamación	9,3

Tabla 5. Resumen ajuste ($R^2=0.9895$, $R^2_{ajust.}=0.9883$, $p\text{-value}<2.2e-16$).

	Coeficiente	Error estándar	t - student	Pr(> t)
Densidad	6,5623	5,8261	1,1260	0,2631
T10	1,3794	0,2149	6,4190	6,73E-09
T50	-3,6347	1,2591	-2,8870	0,0049
T90	0,2223	0,0419	5,3080	8,20E-07
Viscosidad	0,0448	0,0657	0,6810	0,4975
Peso Molecular	-3,1478	0,4966	-6,3380	9,65E-09
Índice Refracción	-5,7924	4,9451	-1,1710	0,2446
Punto Anilina	8,4163	1,0183	8,2650	1,33E-12
Contenido H2	-6,4060	0,9356	-6,8470	9,74E-10
P. inflamación	-0,8656	0,2030	-4,2630	5,05E-05

Tabla 6. Prueba de modelos Anova caso Inicial*

	Grados de libertad	Suma de cuadrados	Media cuadrática	F estadístico	Pr(>F)
Densidad	1	27,5550	27,5550	2371,0950	< 2,2e-16
T10	1	49,3760	49,3760	4248,7143	< 2,2e-16
T50	1	15,9760	15,9760	1374,7287	< 2,2e-16
T90	1	1,4910	1,4910	128,3394	< 2,2e-16
Viscosidad	1	0,0930	0,0930	8,0350	5,69E-03
Peso Molecular	1	0,0190	0,0190	1,6294	2,05E-01
Índice Refracción	1	0,4600	0,4600	39,5786	1,19E-08
Punto anilina	1	0,1510	0,1510	12,9690	5,23E-04
Contenido H2	1	0,6440	0,6440	55,4550	6,16E-11
Punto de inflamación	1	0,2110	0,2110	18,1760	5,05E-05
Residuos	88	1	0,012		

* Esta tabla compara la adición sucesiva de las variables para la formación del modelo completo.

Los resultados del ajuste MLR (Tabla 4) y la tabla ANOVA de significancia de formación del modelo inicial (Tabla 5) son visualizados en la matriz de correlación del sistema (Anexo B). Con esta matriz se puede apreciar una correlación elevada entre la densidad y el IR y entre el PM y la T50% (las Figuras 7 y 8 muestran el gráfico de dispersión para estas variables), razón por la cual, las variables densidad e IR no explican la varianza residual del modelo al 95% de confianza (Tabla 4). De acuerdo a esto y a la tabla ANOVA (Tabla 5) que muestra que el PM no aporta a la explicación de la varianza con la adición de variables a la regresión, el modelo inicial fue reducido a 8 variables independientes (se excluyeron IR y PM).

La matriz de correlación del modelo con 8 variables independientes puede ser analizada en la Tabla 6. De acuerdo a esta matriz, las propiedades ϑ y T90 presentan baja correlación con el IC. De acuerdo a los resultados descritos anteriormente, tres casos de regresión fueron propuestos para el IC considerando diferentes variables independientes, según:

- Caso I: Densidad, T10, T50, T90, viscosidad, punto de anilina, contenido de H₂, y punto de inflamación.
- Caso II: Densidad, T10, T90, peso molecular, índice de refracción, punto de anilina, contenido de H₂, y punto de inflamación.
- Caso III: T10, T50, T90, peso molecular, contenido de H₂, punto de anilina, y punto de inflamación.

Tabla 7. Nueva matriz de correlación

	ρ	T10	T50	T90	ϑ	PA	H2	PI
ρ	1							
T10	0,4946	1						
T50	0,6467	0,8081	1					
T90	0,5059	0,4082	0,8362	1				
ϑ	0,6696	0,8632	0,8724	0,6053	1			
PA	-0,4759	0,3290	0,3614	0,3468	0,1890	1		
H2	-0,9497	-0,2733	-0,3802	-0,2722	-0,4601	0,7247	1	
PI	0,4995	0,9944	0,8167	0,4213	0,8332	0,3335	-0,2742	1
IC	-0,5330	0,3565	0,2655	0,2171	0,1471	0,9628	0,7571	0,3552

Figura 7. Densidad vs Índice de refracción

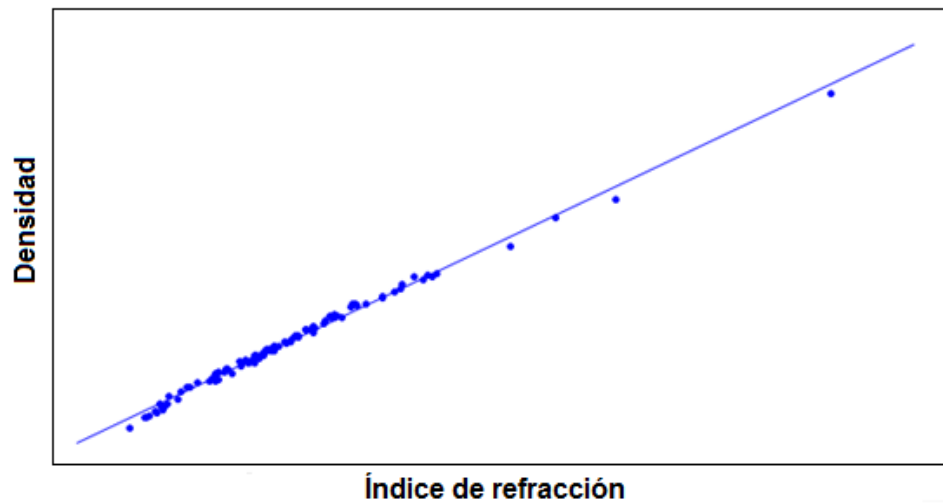
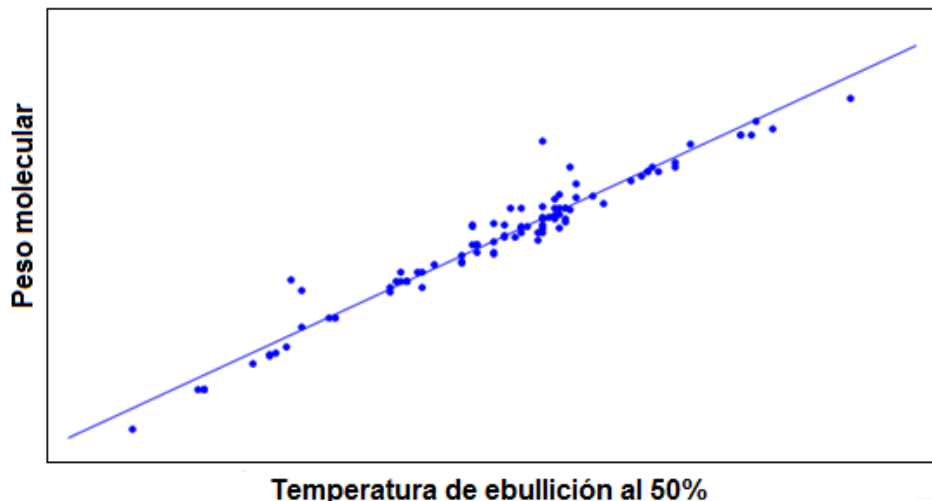


Figura 8. Peso molecular vs Temperatura ebullición 50%



Para cada caso se aplicó la regresión RLM, junto con los diferentes análisis descritos en el capítulo de metodología. Los coeficientes para cada uno de los casos y sus respectivas gráficas de análisis estadístico pueden ser consultados en el Anexo D. Según los resultados consignados en las tablas del Anexo D, los tres casos de regresión lineal presentan significancia estadística al 95% de confianza. Un cuarto caso fue derivado del caso I debido a que el contenido de H2 no reportó significancia estadística en la regresión; con esto el caso IV corresponde a la regresión con las mismas variables del caso I exceptuando el contenido de H2. Para los casos II, III y IV las variables independientes utilizadas presentan significancia estadística al 95% de confianza para explicar la varianza de la regresión respectiva.

La Tabla 7 presenta un resumen de algunos estadísticos obtenidos para los casos analizados.; la prueba de Shapiro-Wilk valida el supuesto de normalidad para los residuos de cada caso. Asimismo, las gráficas representativas de las pruebas estadísticas denominadas *Leverage* y la distancia de Cook para el caso I son

mostradas en las Figuras 9 y 10. Según estas gráficas, la base de datos presenta algunos puntos que pueden considerarse influyentes y atípicos para los modelos de regresión del IC. Estos puntos (Tabla 8) fueron omitidos para verificar su influencia en los modelos, sin embargo, el efecto de las observaciones influyentes y atípicas resultó no significativa para los casos planteados.

Tabla 8. Resultados de pruebas estadísticas de los modelos de regresión

	CASO I*	CASO II	CASO III	CASO IV
R ²	0,9832	0,9884	0,9892	0,9827
R ² ajustado	0,9817	0,9874	0,9884	0,9813
Leverage máximo	0,1837	0,1837	0,1633	0,1633
Distancia de Cook máxima	0,0452	0,0341	0,0350	0,0460
Prueba de Shapiro-Wilk	0,0122	0,0212	0,0212	0,0212
AIC	-105,2556	-141,9061	-150,5632	-104,3227
RMSE Calibración	1,4260	0,1070	0,1034	0,1310
RMSE Predicción	--	0,1323	0,1255	0,1480
RMSE Cross Validation	--	0,1203	0,1151	0,1495

* Este modelo no fue validado debido a la no significancia del contenido de H2.

Tabla 9. Posibles puntos atípicos de los modelos de regresión

	CASO I	CASO II	CASO III
Puntos atípicos e influyentes	45,48	46,48	45,48

Figura 9. Leverage para el caso I

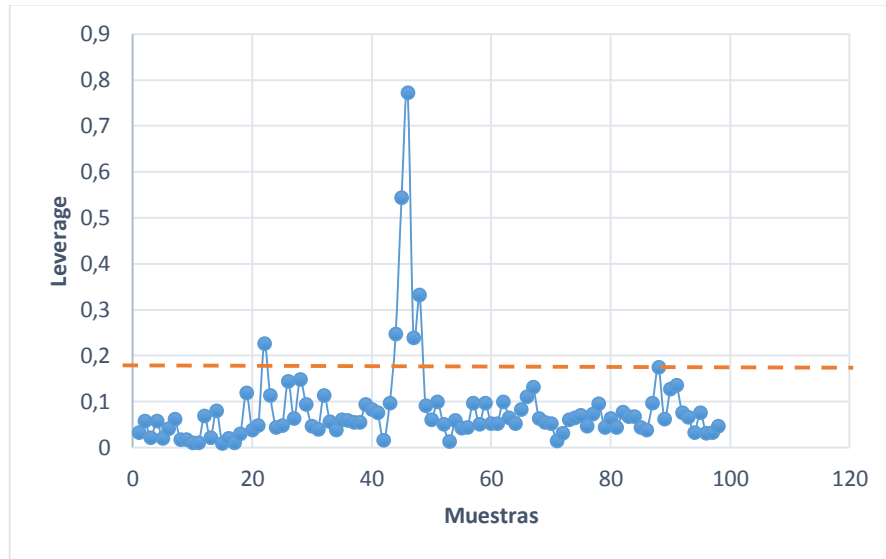
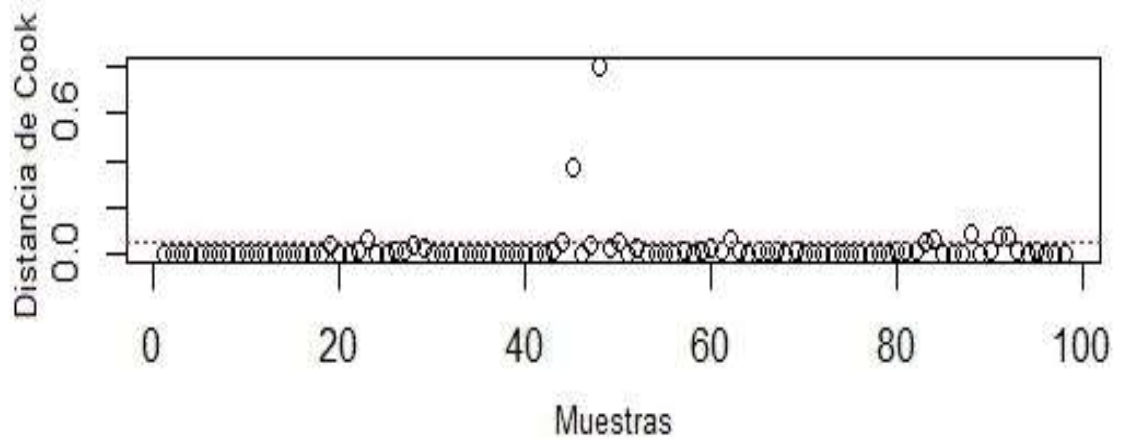


Figura 10. Distancia de Cook, caso I. Figura reportada por R.



Para analizar el conjunto de los cuatro casos de regresión lineal múltiple, se consideraron los siguientes criterios para descartar los diferentes modelos:

- Incumplimiento de la prueba estadística para alguna variable independiente ($p < 0,05$) (Anexo D): Se descarta el caso I.
- Se descartan los valores altos de la prueba de AIC: casos II y III.

Del análisis anterior, es posible concluir que el modelo más oprobado para la regresión del IC basado en propiedades macroscópicas corresponde al caso III; este modelo presenta le mejor desempeño de acuerdo criterio de información Akaike, así como para los otros estadísticos mostrados en la Tabla 7. Las Tablas 9, 10 y 11 muestran los coeficientes del modelo, los resultados de la prueba ANOVA y los intervalos de confianza respectivamente, para la regresión del Caso III.

Tabla 10. Modelo de regresión del Caso III.

	Coeficiente de estimación	Std. Error	t value	Pr(> t)
T10	1,4130	0,1598	8,8430	6,72E-14
T50	-2,3592	0,4220	-5,5900	2,35E-07
T90	0,2191	0,0406	5,3910	5,48E-07
Peso Molecular	-3,0907	0,3492	-8,8520	6,45E-14
Punto anilina	7,6422	0,8013	9,5370	2,37E-15
Contenido H2	-5,9835	0,7493	-7,9850	4,11E-12
P. inflamación	-0,8966	0,1652	-5,4260	4,73E-07

Tabla 11. Prueba de modelos Anova caso III.

	Grados de libertad	Suma de cuadrados	Media cuadrática	F estadístico	Pr(>F)
T10	1	12,3250	12,3250	1069,5850	< 2,2e-16
T50	1	0,1420	0,1420	12,3170	7,00E-04
T90	1	11,2300	11,2300	974,5280	< 2,2e-16
Peso Molecular	1	64,3610	64,3610	5585,3520	< 2,2e-16
Punto de anilina	1	7,0250	7,0250	609,6640	< 2,2e-16
Contenido H2	1	0,5290	0,5290	45,9170	1,20E-06
P. inflamación	1	0,3390	0,3390	29,4410	4,73E-04
Residuos	91	1,0490	0,0120		

Tabla 12. Intervalos de confianza modelo caso III.

	2.5 %	97.5 %
T10	1,0998	1,7262
T50	-3,1864	-1,5321
T90	0,1394	0,2988
Peso Molecular	-3,7750	-2,4063
Punto de anilina	6,0718	9,2127
Contenido H2	-7,4521	-4,5149
P. inflamación	-1,2204	-0,5727

El modelo definitivo se presenta a continuación:

$$\begin{aligned}
 CI = & 1,4130 * T10 - 2,3592 * T50 + 0,2191 * T90 - 3,0907 * PM \\
 & + 7,6422 * PA - 5.9835 * H2 - 0,8966 * FP \quad (6)
 \end{aligned}$$

El modelo de regresión propuesto para el IC (caso III) fue analizado para determinar la influencia de cada variable con el objetivo de obtener una regresión reducida. Debido a que las variables están autoescaladas, la influencia de las mismas puede obtenerse utilizando los coeficientes. La Tabla 12 muestra la influencia de cada variable en el modelo de regresión propuesto.

Tabla 13. Estudio del peso de las variables del modelo de regresión caso III

	Coeficiente	Valor absoluto coeficiente	Peso de la variable	Porcentaje en P. V.
Punto de anilina	7,6422	7,6422	0,3537	35,3738
Contenido H2	-5,9835	5,9835	0,2770	27,6958
Peso Molecular	-3,0907	3,0907	0,1431	14,3058
T50	-2,3592	2,3592	0,1092	10,9202
T10	1,4130	1,4130	0,0654	6,5404
P. inflamación	-0,8966	0,8966	0,0415	4,1499
T90	0,2191	0,2191	0,0101	1,0142

Según los resultados mostrados en la Tabla 13, es posible plantear dos nuevos casos reducidos a partir del modelo propuesto (caso III). Estos son:

- Caso V: T10, T50, peso molecular, contenido de H₂, punto de anilina.
- Caso VI: T50, peso molecular, contenido de H₂, punto de anilina.

El caso V contiene las variables que aportan hasta el 94% en el modelo de regresión, mientras que el caso VI considera las variables que aportan hasta el 88% según el peso de los coeficientes (Tabla 12). De acuerdo a los valores estadísticos de las regresiones de los dos casos anteriores, el caso V (7) corresponde a la regresión reducida de mejor desempeño (Tabla 13); *i.e.* la disminución en dos

variables (T90 y PI) sólo desmejora en un 1% la explicación de la varianza obtenida con la regresión propuesta (caso III).

Modelo de regresión lineal caso V se presenta a continuación y sus respectivos intervalos de confianza (Tablas 13 y 14)

$$CI = 0,3413 * T10 - 2,9751 * T50 - 2,1772 * PM + 7,5983 * PA - 6,1432 * H2 \quad (7)$$

Tabla 14. Pruebas estadísticas de los modelos reducidos del caso III

	CASO V	CASO VI
R ²	0,9783	0,9434
R ² ajustado	0,9771	0,9405
Leverage máximo	0,1224	0,1020
Distancia de Cook máxima	0,0286	0,0220
Prueba de Shapiro-Wilk	2,67 E-04	3,191E-04
AIC	-86,2795	6,4388
RMSE Calibración	0,1466	0,2376
RMSE Predicción	0,1466	0,2199
RMSE Cross Validation	0,1564	0,2480

Tabla 15. Intervalos de confianza modelo caso V.

	2.5 %	97.5 %
T10	0,2870	0,3957
T50	-4,1046	-1,8457
Peso Molecular	-2,8256	-1,5288
Punto de anilina	5,5736	9,6229
Contenido H2	-8,0770	-4,2094

La regresión propuesta en el presente trabajo (Caso III) se acerca con alto grado de exactitud a la predicción del índice de cetano que se realiza a través de la norma ASTM D-4737 mediante cuatro parámetros (densidad, T10%, T50% y T90%). Stratiev y colaboradores reportan que el número de cetano obtenido mediante la norma ASTM D-4737 presenta una varianza de 91,5% obtenida con la base de 140 datos experimentales [40]. De igual forma, Stratiev *et al.* proponen una regresión que alcanza una explicación del 95,6% de la varianza empleando los mismos parámetros utilizados por la norma ASTM D-4737.

Sin embargo, aunque el 95.6% de explicación de la varianza del número de cetano sea obtenido utilizando sólo cuatro parámetros, es deseable reducir aún más la dependencia de la regresión debido a los costos de los respectivos análisis. La norma ASTM D-976 presenta la ventaja de requerir sólo dos parámetros para la predicción del número de cetano (densidad y T50); desde luego, el desempeño de esta regresión de dos parámetros es inferior al obtenido por la norma ASTM D-4737. Según los resultados de Stratiev *et al.*, la regresión propuesta en la ASTM D-976 explica cerca del 80% de la varianza del número de cetano, mientras que la ASTM D-4737 explica 91.5% de la varianza de los mismos datos.

Con el objetivo de proponer una regresión que dependa sólo de dos parámetros (densidad y T50%), el modelo reducido (caso V) se puede expresar solo en función de la densidad y T50% por medio de ecuaciones reportadas en la literatura, las cuales se presentan en el Anexo E. Con lo anterior, la explicación de la varianza (97,83%) obtenida con el modelo caso V (Tabla 13) propuesta en el presente trabajo de grado, mejora la predicción del índice de cetano a partir de la norma ASTM D-976; la ecuación propuesta en la norma ASTM D-976 alcanza una explicación del 88.8% del índice de cetano calculado a partir de la norma ASTM D-4737 (Anexo F,

Figura 25). La regresión del caso V está basada en los 140 datos experimentales descritos en el Anexo A, por lo cual esta debe ser validada en un conjunto con un mayor número de datos.

3.2 REDES NEURONALES

El entrenamiento y la validación de las redes neuronales fueron aplicados por medio del *Toolbox ANN* de MATLAB 2015 (The MathWorks Inc.). La calidad de los modelos en los conjuntos de entrenamiento y validación fue evaluada sobre la base de los indicadores estadísticos: R^2 , RMSE y RMSEP, donde R^2 es el coeficiente de correlación lineal, el cual se determinó realizando una regresión lineal múltiple entre los componentes principales y la propiedad en estudio CI, el RMSE hace referencia a la raíz del promedio del cuadrado de los errores de entrenamiento y el RMSEP a la raíz del promedio del cuadrado de los errores de predicción.

La clase de red seleccionada corresponde a la *feed-forward* con el algoritmo de entrenamiento *Backpropagation* utilizando el método de optimización de Levenberg-Marquardt. Asimismo, la arquitectura seleccionada para la red contiene dos capas ocultas, en la que se realizó el estudio con dos tipos de funciones de transferencia:

- Función de activación I F: Tansig - Purelin
- Función de activación II F: Tansig - Tansig

3.2.1 Red neuronal con propiedades macroscópicas. La topología de las redes para la predicción del IC basada en datos de propiedades macroscópicas corresponde a 08:X:01, donde el 08 es el número de variables de entrada a la red, X corresponde al número de neuronas que se encuentran en cada una de las capas intermedias y 01 corresponde a la variable de salida, es decir la predicción del índice de cetano. En el presente trabajo se evaluaron diferentes valores para X considerando las dos funciones de activación IF y IIF.

Las Tablas 15 y 16 presentan los resultados para los diferentes valores de X en la arquitectura de red seleccionada. Según los resultados consignados, la red denominada I F B6, la cual contiene 8 neuronas por cada capa interna con función de activación I F, presentó los menores valores de RMSE y RMSEP (Tabla 16). Asimismo, para la función de activación II F, la red con cuatro neuronas por capa intermedio reportó los menores valores para los estadísticos RMSE y RMSEP (Tabla 16); para esta misma función de activación, las redes II F B2 y II F B3 no presentan resultados debido a inconvenientes de convergencia con el método de optimización seleccionado. De acuerdo a los desempeños de las redes aplicadas, la red que mejor se ajusta a los datos del índice de cetano utilizando los valores de las propiedades macroscópicas corresponde a la denominada I F B6.

Tabla 16. Resultados de las redes con funciones de activación I F.

Red	Topología	R2	RMSE	RMSEP
I F B1	08:05:01	0,9960	0,6543	0,5105
I F B2	08:10:01	0,9943	0,8268	0,4982
I F B3	08:15:01	0,9965	0,6118	0,4782
I F B4	08:20:01	0,9938	0,8190	0,9905
I F B5	08:12:01	0,9128	3,2846	2,5002
I F B6	08:08:01	0,9984	0,4099	0,5668
I F B7	08:06:01	0,9973	0,5516	0,8860

Tabla 17. Resultados de las redes con funciones de activación II F

Red	Topología	R2	RMSE	RMSEP
II F B1	08:05:01	0,9969	0,5822	1,2529
II F B2	08:10:01	---	---	---
II F B3	08:15:01	---	---	---
II F B4	08:08:01	0,9968	0,5771	1,1919
II F B5	08:06:01	0,9964	0,6277	1,2629
II F B6	08:03:01	0,9951	0,7300	1,2403
II F B7	08:09:01	0,9759	1,6314	1,1969

3.2.2 Red neuronal basada en datos espectroscópicos. La topología de las redes para la predicción del IC basada en datos de espectroscopia corresponde a 811:X:01, donde el 811 es el conjunto de valores de número de onda (rango preestablecido en la sección 2.2) de cada uno de los espectros para la entrada a la red, X corresponde al número de neuronas que se encuentran en cada una de las capas intermedias y 01 corresponde a la variable de salida, es decir la predicción del índice de cetano. En el presente trabajo se evaluaron diferentes valores para X considerando las dos funciones de activación IF y IIF.

Tabla 18. Topología de las redes con funciones de transferencia I F NIR

	Número de neuronas	R2	RMSECV	RMSEP
I F NIR A1	5	0,7963	0,7773	0,4872
I F NIR A2	10	0,0641	1,7187	1,1459
I F NIR A3	15	0,6763	0,8630	1,0839
I F NIR A4	8	0,2786	0,9858	0,6354
I F NIR A5	13	0,8837	0,6212	0,7962

Tabla 19. Topología de las redes con funciones de transferencia I F NIR (Continuación)

	Número de neuronas	R2	RMSECV	RMSEP
I F NIR A6	14	0,9093	0,5987	1,2598
I F NIR A7	12	0,53	1,0231	0,7721
I F NIR A8	3	0,7503	0,6677	0,8354
I F NIR A9	4	0,4161	1,5717	1,1819
I F NIR A10	18	0,6504	0,9319	0,6047

Tabla 20. Topología de las redes con funciones de transferencia II F NIR

	Número de neuronas	R2	RMSECV	RMSEP
II F NIR A1	5	0,74	0,8998	0,8008
II F NIR A2	10	0,8189	0,8926	0,6523
II F NIR A3	15	0,7679	0,7778	0,6031
II F NIR A4	13	0,9922	0,6562	0,8157
II F NIR A5	8	0,9771	0,8659	0,8928
II F NIR A6	18	0,5968	0,9478	1,1535

Según los resultados del entrenamiento de las redes con los espectros NIR de los diésel (Tablas 17 y 18), la función de activación I F presenta los menores desempeños para la predicción del índice de cetano; esto se debe probablemente a las grandes variaciones de intensidad encontradas en los números de onda utilizados para el entrenamiento. De otro lado, la función de activación II F muestra un mejor acople a los datos espectroscópicos, particularmente la red II F NIR A4 la cual presenta los mejores valores de los estadísticos calculados.

Tabla 21. Resultados estadísticos con los mejores modelos de RLM y RNA

	R ²	RMSECV	RMSEP
Caso V	0,9783	0,1466	0,1466
I F B6	0,9984	0,4099	0,5668
II F NIR A4	0,9922	0,6562	0,8157

Comparando los resultados de los ajustes estadísticos y por redes neuronales desarrollados en el presente trabajo, es posible afirmar que la predicción del índice de cetano puede efectuarse con buena exactitud utilizando tanto datos macroscópicos como resultados espectroscópicos (Tabla 19). La ventaja de utilizar las regresiones obtenidas con datos macroscópicos radica en que sólo se requieren cuatro valores experimentales (dos con la regresión reducida) para una predicción válida (basada en la base de datos correspondiente). De otro lado, aunque el uso de la red neuronal basada en espectros NIR requiere una mayor información, los tiempos y los costos de obtención de estos análisis son menores a los respectivos para las propiedades macroscópicas por medio de las normas ASTM. Por último, los resultados obtenidos en el presente trabajo deben ser probados en un conjunto de datos mayor para que puedan ser generalizados como herramientas útiles a nivel industrial.

4 CONCLUSIONES

Los resultados del presente proyecto de grado muestran que el uso de RLM y las RNA son de utilidad para la predicción exacta del índice de cetano a partir de propiedades macroscópicas y de datos de espectroscopia infrarroja NIR.

La aplicación de la RLM mediante el uso del software versión libre R y el método de las redes neuronales (RNAs) utilizando MATLAB para los dos conjuntos de bases de datos, condujo al desarrollo de nuevos procedimientos que lograron reproducir los valores reportados por la normas ASTM D-4737 para el índice de cetano del diésel.

Los resultados obtenidos en el presente proyecto con las regresiones desarrolladas reportaron mejores desempeños que la norma ASTM D-976 para la predicción del índice de cetano.

5 RECOMENDACIONES

Después del desarrollo de este trabajo de grado se plantean las siguientes recomendaciones:

- ❖ Utilizar los resultados obtenidos en el presente documento como una predicción útil en evaluaciones preliminares del índice de cetano.
- ❖ Emplear un conjunto amplio de muestras en lo posible certificadas, con el fin de corroborar la significancia estadística presentada por cada uno de los modelos propuestos.
- ❖ Continuar con el desarrollo de regresiones para otras propiedades de los diésel del petróleo.

REFERENCIAS

1. KARONIS D, LOIS E, ZANNIKOS F, ALEXANDRIDIS A, SARIMVEIS H. A Neural Network Approach for the Correlation of Exhaust Emissions from a Diesel Engine with Diesel Fuel Properties. *Energy & Fuels*. 2003;17(9):1259–65.
2. BOMBERGER JD, SEBORG DE. Determination of model order for NARX models directly from input-output data. *J Process Control*. 1998;8:459–68.
3. ERDEM S, KOCAKOC ID. A new ANN training approach for efficiency evaluation. *Hacettepe J Math Stat*. 2010;39:439–47.
4. QINQIN C, QIAN J, ZENGWEI Y, LEI H. Environmental risk source management system for the petrochemical industry. *Process Saf Environ Prot. Institution of Chemical Engineers*; 2014;92:251–60.
5. SUTTON I. Technical information and industry standards. In: Ltd E, editor. *Process Risk and Reliability Management*. First edit. Oxford: Ian Sutton; 2010. p. 277–300.
6. SMILDE A, BRO R, GELADI P. *Multi-Way Analysis with Applications in the Chemical Sciences*. Third edit. England: John Wiley & Sons; 2004. 1-371 p.
7. JAMIESON G, VICENTE KJ. Ecological interface design for petrochemical applications: supporting operator adaptation, continuous learning, and distributed, collaborative work. *Comput Chem Eng*. 2001;25:1055–74.
8. DUEZ P, VICENTE KJ. Ecological interface design and computer network management: The effects of network size and fault frequency. *Int J Hum Comput Stud* [Internet]. 2005;63(6):565–86. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S1071581905001084>
9. DRAPER NR, SMITH H. *Applied Regression Analysis*. Third edit. USA: John Wiley & Sons; 1998. 736 p.
10. HAIR JJ, BLACK W, BABIN B, ANDERSON R. *Multivariate Data Analysis*. 7th editio. USA: Prentice Hall; 2009.
11. HILERA J, MARTÍNEZ V. *Análisis estadístico multivariante: un enfoque teórico y práctico*. McGraw-Hill. 2000. 712 p.

12. AMERICAN SOCIETY FOR TESTING MATERIALS. Standard Terminology Relating to Petroleum, Petroleum Products, and Lubricants. ASTM D4175. Pennsylvania, EU; p. 1–63.
13. SPEIGHT JG. The Chemistry and Technology of Petroleum, Fifth Edition. Cuarta. Taylor & Francis Group; 2014. 953 p.
14. ARDILA A. J V., PINILLA T. AM. Optimización de un proceso limpio para la desulfurización de una muestra de diésel colombiano enriquecida utilizando el líquido iónico tetrafluoroborato de 1-Butil-3-Metilimidazolio [BMIM][BF₄]. Universidad Industrial de Santander; 2014.
15. AMERICAN SOCIETY FOR TESTING MATERIALS. Standard Test Method for Cetane Number of Diesel Fuel Oil . ASTM D613 – 10a. Pennsylvania, EU; 2012. p. 1–16.
16. RIAZI MR. Characterization and properties of petroleum fractions. First edit. ASTM, editor. USA: ASTM International Standards Worldwide; 2005. 407 p.
17. WINTER A, BATISTLLA CB, WOLF MMR, MACIEL F. R, LOPES MS. A true boiling point curve through molecular distillation using framol correlation. ... Resour Available. 2007;1–6.
18. AMERICAN SOCIETY FOR TESTING MATERIALS. ASTM D86 Standard Test Method for Distillation of Petroleum Products at Atmospheric Pressure. Pennsylvania, EU; p. 27.
19. XINSHUAI L, ZHENYI L. Research progress on flash point prediction. J Chem Eng Data. 2010;55(9):2943–50.
20. RAGHAVACHARI R. Near-Infrared Applications in Biotechnology. Nueva York: Marcel Decker, Inc; 2010. 397 p.
21. FALLA FS, LARINI C, LE ROUX G a. C, QUINA FH, MORO LFL, NASCIMENTO C a. O. Characterization of crude petroleum by NIR. J Pet Sci Eng [Internet]. 2006;51(1-2):127–37.
22. SANTOS VO, OLIVEIRA FCC, LIMA DG, PETRY AC, GARCIA E, SUAREZ P a. Z, et al. A comparative study of diesel analysis by FTIR, FTNIR and FT-Raman spectroscopy using PLS and artificial neural network analysis. Anal Chim Acta [Internet]. 2005;547(2):188–96.

23. GONZAGA FB, PASQUINI C. A low cost short wave near infrared spectrophotometer: Application for determination of quality parameters of diesel fuel. *Anal Chim Acta* [Internet]. Elsevier B.V.; 2010;670(1-2):92–7.
24. BALABIN RM, LOMAKINA EI, SAFIEVA RZ. Neural network (ANN) approach to biodiesel analysis: Analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (NIR) spectroscopy. *Fuel* [Internet]. Elsevier Ltd; 2011;90(5):2007–15.
25. MELÉNDEZ L V., LACHE A, ORREGO-RUIZ JA, PACHON Z, MEJÍA-OSPINO E. Prediction of the SARA analysis of Colombian crude oils using ATR–FTIR spectroscopy and chemometric methods. *J Pet Sci Eng* [Internet]. 2012;90-91:56–60.
26. FEI F, QIONGSHUI W, LIBO Z. Rapid analysis of diesel fuel properties by near infrared reflectance spectra. *Spectrochim Acta Part A Mol Biomol Spectrosc* [Internet]. Elsevier B.V.; 2015;149:271–8.
27. MENDENHALL W. *Introducción a la probabilidad y la estadística*. 1987. p. 626.
28. KANDEL E, JESSELL T, SHWARTZ J. *Neurociencia y Conducta*. 1996. p. -.
29. HILERA J, MARTÍNEZ V. *Redes Neuronales Artificiales, fundamentos, modelos y aplicaciones*.
30. CAICEDO EF, LÓPEZ JA. *Una aproximación práctica a las Redes Neuronales Artificiales*. Universidad del Valle; 2009. 217 p.
31. PORCEL M. *Aplicación de técnicas quimiométricas para el desarrollo de nuevos métodos cinético-espectrofotométricos de análisis*. Universidad Autónoma de Barcelona; 2001.
32. QIUWEN C, TIESHENG G, LIU Y, RUONAN L, FRIEDRICH R. Online forecasting chlorophyll a concentrations by an auto-regressive integrated moving average model : Feasibilities and potentials. *Harmful Algae* [Internet]. Elsevier B.V.; 2015;43:58–65.
33. ABYANEH HZ. Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. *J Environ Heal Sci Eng* [Internet]. *Journal of Environmental Health Science and Engineering*; 2014;12(1):1–8.

34. RAMAADHAS AS, JAYARAJ S, MURALEEDHARAN C, PADMAKUMARI K. Artificial neural networks used for the prediction of the cetane number of biodiesel. *Renew Energy* [Internet]. 2006;31(15):2524–33.
35. SÁNCHEZ B. Y, PILOTO R. R, ERRASTI M, SIERENS R, VERHELST S. Prediction of Cetane Number and Ignition Delay of Biodiesel Using Artificial Neural Networks. *Energy Procedia* [Internet]. Elsevier B.V.; 2014;57:877–85.
36. ROCABRUNO V. CI, RAMÍREZ V. LF, HERNÁNDEZ JA. Artificial neural network models to predict density, dynamic viscosity, and cetane number of biodiesel. *Fuel* [Internet]. 2015;147:9–17.
37. BOLANCA T, MARINOVIC S, UKIC S, JUKIC A, RUKAVINA V. Development of Artificial Neural Network Model for Diesel Fuel Properties Prediction using Vibrational Spectroscopy. *Acta Chim, Slov.* 2012;59:249–57.
38. YANG H, RING Z, BRIKER Y, MCLEAN N, FRIESEN W, FAIRBRIGDE C. Neural network prediction of cetane number and density of diesel fuel from its chemical composition determined by LC and GC-MS. *Fuel.* 2002;81(1):65–74.
39. PASADAKIS N, SOURLIGAS S, FOTEINOPOULOS C. Prediction of the distillation profile and cold properties of diesel fuels using mid-IR spectroscopy and neural networks. *Fuel* [Internet]. 2006;85:1131–7.
40. STRATIEV D, MARINOV I, DINKOV R, SHISHKOVA I, VELKOV I, Sharafutdinov I, et al. Opportunity to Improve Diesel-Fuel Cetane-Number Prediction from Easily Available Physical Properties and Application of the Least-Squares Method and Artificial Neural Networks. *Energy & Fuels* [Internet]. 2015;29(3):1520–33.

BIBLIOGRAFIA

ABYANEH, Hamid Zare. Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. Journal of Environmental Health Science and Engineering. Acta [online] 2014. Vol. 12 N° 1.

AMERICAN SOCIETY FOR TESTING MATERIALS. Standard Test Method for Distillation of Petroleum Products at Atmospheric Pressure. ASTM D86. Pennsylvania, EU. ASTM, 2015.

AMERICAN SOCIETY FOR TESTING MATERIALS. Standard Terminology Relating to Petroleum, Petroleum Products, and Lubricants. ASTM D4175. Pennsylvania, EU. ASTM, October 2015.

AMERICAN SOCIETY FOR TESTING MATERIALS. Standard Test Method for Cetane Number of Diesel Fuel Oil. ASTM D613. Pennsylvania, EU. ASTM, 2012.

ARDILA ANTOLINES, Jessica Viviana y PINILLA TORRES, Ana María. Optimización de un proceso limpio para la desulfurización de una muestra de diésel colombiano enriquecida utilizando el líquido iónico tetrafluoroborato de 1-Butil-3-Metilimidazolio [BMIM][BF₄]. Tesis de grado Ingeniera Química. Bucaramanga: Universidad Industrial de Santander, Facultad de Ingenierías Fisico-Químicas, Escuela de Ingeniería Química, 2014

BALABIN, Roman M., LOMAKINA, Ekaterina I. y SAFIEVA, Ravilya Z.. Neural network (ANN) approach to biodiesel analysis: Analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (NIR) spectroscopy. *Fuel. Acta* [online] 2011. Vol. 90 N° 5. p. 2007–2015. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

BOLANCA, T., *et al.* Development of Artificial Neural Network Model for Diesel Fuel Properties Prediction using Vibrational Spectroscopy. *Acta Chim. Slov.* Acta [online] 2012. Vol. 59. p. 249-257. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

BOMBERGER, John D. y SEBORG, Dale E.. Determination of model order for NARX models directly from input-output data. *Journal of Process Control. Acta* [online] 1998. Vol. 8 N° 5.-6. p. 459-468. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

CAICEDO BRAVO, Eduardo Francisco and LÓPEZ SOTELO, Jesus Alfonso. *Una aproximación práctica a las Redes Neuronales Artificiales..* 1ed. Universidad del Valle, 2009. 217p.

DRAPER, Norman Richard and SMITH, Harry. *Applied Regression Analysis.* 3ed. New York. John Wiley, 1998. 706p.

DUEZ, Pierre y VICENTE, Kim J.. Ecological interface design and computer network management: The effects of network size and fault frequency. *International Journal of Human-Computer Studies* .Acta [online] 2005. Vol. 63 N° 6, p. 565–586. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

ERDEM, S. y KOCAKOC, I. D.. A new ANN training approach for efficiency evaluation. *Hacettepe Journal of Mathematics and Statistics* .Acta [online] 2010. Vol. 39, p. 439–447. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

FALLA, F. S., *et al.* A comparative study of diesel analysis by FTIR, FTNIR and FT-Raman spectroscopy using PLS and artificial neural network analysis. *Journal of Petroleum Science and Engineering*. Acta [online] 2006. Vol. 51 N° 1-2, p. 127–137. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

FEI, Feng, QIONGSHUI, Wu y LIBO, Zeng. Rapid analysis of diesel fuel properties by near infrared reflectance spectra. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. Acta [online] 2015. Vol. 149, p. 271–278. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

GONZAGA, Fabiano Barbieri y PASQUINI, Celio. A low cost short wave near infrared spectrophotometer: Application for determination of quality parameters of diesel fuel. *Analytica Chimica Acta*. Acta [online] 2010. Vol. 670 N° 1-2, p. 92–97. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

HAIR, Joseph, BLACK, William, BABIN, Barry and ANDERSON, Rold. *Applied Regression Analysis*. 7ed. E.U. Pearson Prentice Hall. 2009. 752p.

HILERA, J., y MARTÍNEZ, V.. *Análisis estadístico multivariante: un enfoque teórico y práctico*. McGraw-Hill. 2000.

JAMIESON, G. y VICENTE, K. J.. Ecological interface design for petrochemical applications: supporting operator adaptation, continuous learning, and distributed, collaborative work. *Computers & Chemical Engineering*. Acta [online] 2001. Vol. 25, p. 1055–1074. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

KANDEL, Eric R., JESSELL, Thomas M., and SHWARTZ, James H. *Neurociencia y Conducta*. PRENTICE-HALL. 1997. 832p.

KARONIS, D., *et al.* A Neural Network Approach for the Correlation of Exhaust Emissions from a Diesel Engine with Diesel Fuel Properties. *Energy & Fuels*. Acta [online] 2003. Vol. 17 N° 9, p. 1259–1265. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

MELÉNDEZ, L. V., *et al.* Prediction of the SARA analysis of Colombian crude oils using ATR–FTIR spectroscopy and chemometric methods. *Journal of Petroleum Science and Engineering*. Acta [online] 2012. Vol. 90-91, p. 56–60. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

MENDENHALL, William, BEAVER, Robert J. y BEAVER, Barbara M.. Introducción a la probabilidad y la estadística. Cengage Learning. 13ed. 2005. 780p.

PASADAKIS, N., SOURLIGAS, S. and FOTEINOPOULOS, C.. Prediction of the distillation profile and cold properties of diesel fuels using mid-IR spectroscopy and neural networks. *Fuel*. Acta [online] 2006. Vol. 85, p. 1131–1137. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

QINQIN, C., *et al.* Environmental risk source management system for the petrochemical industry. *J Process Safety and Environmental Protection*. Acta [online] 2014. Vol. 92, p. 251–260. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

QIUWEN, C., *et al.* Online forecasting chlorophyll a concentrations by an autoregressive integrated moving average model: Feasibilities and potentials. *Harmful Algae*. Acta [online] 2014. Vol. 43, p. 58–65. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

RAGHAVACHARI, Ramesh. *Near-Infrared Applications in Biotechnology*. Nueva York: Marcel Decker, Inc. 2000. 1ed.392p.
https://books.google.com.co/books?hl=es&lr=&id=R4xb4NPg8_kC&oi=fnd&pg=PR3&dq=Near-Infrared+Applications+in+Biotechnology.&ots=H-qsRVyzFV&sig=HjeWwYNWm9afrlbtr8x7oRahnOM#v=onepage&q=Near-Infrared%20Applications%20in%20Biotechnology.&f=false

RAMAADHAS, A. S., *et al.* Artificial neural networks used for the prediction of the cetane number of biodiesel. *Renewable Energy*. Acta [online] 2006. Vol. 31 N° 15, p. 2524–2533. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

RIAZI, M. R. *Characterization and properties of petroleum fractions*. 1ed. Philadelphia. Pearson Prentice Hall. 2005. 427p.

ROCABRUNO V., C. I., RAMÍREZ V., L. F. and HERNÁNDEZ, J. A.. Artificial neural network models to predict density, dynamic viscosity, and cetane number of biodiesel. *Fuel*. Acta [online] 2015. Vol. 147, p. 9–17. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

SÁNCHEZ B., *et al.* Prediction of Cetane Number and Ignition Delay of Biodiesel Using Artificial Neural Networks. *Energy Procedia*. Acta [online] 2014. Vol. 57 N, p. 877–885. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

SANTOS, V. O., *et al.* A comparative study of diesel analysis by FTIR, FTNIR and FT-Raman spectroscopy using PLS and artificial neural network analysis. *Analytica Chimica Acta* [online] 2005. Vol 547 N° 2 [citado 4 agosto 2015] p. 188–196. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

SMILDE, Age, BRO, Rasmus and GELADI, Paul. *Multi-Way Analysis with Applications in the Chemical Sciences* . 2005. 3ed. England: John Wiley & Sons. <http://doi.org/10.1002/0470012110>

SPEIGHT, James G.. *The Chemistry and Technology of Petroleum*. 4ed. Taylor & Francis Group. 2014. 984p.

STRATIEV, D., *et al.* Opportunity to Improve Diesel-Fuel Cetane-Number Prediction from Easily Available Physical Properties and Application of the Least-Squares Method and Artificial Neural Networks. *Energy & Fuels*. Acta [online] 2015. Vol. 29 N° 3, p. 1520–1533. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

XINSHUAI, L., and ZHENYI, L.. Research progress on flash point prediction. *Journal of Chemical and Engineering. Acta* [online] 2010. Vol. 55 N°9, p. 2943–2950. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

YANG, H., *et al.* Neural network prediction of cetane number and density of diesel fuel from its chemical composition determined by LC and GC-MS. *Fuel. Acta* [online] 2002. Vol. 81 N° 1, p. 65–74. Available from: ELSEVIER-SCIENCE DIRECT. [Biblioteca Universidad Industrial de Santander].

ANEXOS

ANEXO A: BASE DE DATOS

Tabla 22. Propiedades fisicoquímicas de las muestras del diésel de estudio

N°	CI ASTM D-4737	Densidad 15°C [g/cm ³]	ASTM 10% [v/v]	D-86 50% [v/v]	D. [°C] 90% [v/v]	Viscosidad 40°C [mm ² /s]	Peso molecular [g/mol]	Índice de refracción 20°C	Punto de anilina [°C]	Contenido H2 [% m/m]
1	45,0	0,841	200	245	311	2,32	176	1,4708	60,7	12,91
2	51,0	0,821	196	239	315	2,03	175	1,4574	67,3	13,58
3	51,0	0,842	209	275	337	3,22	200	1,4699	71,2	13,21
4	47,0	0,833	193	242	324	2,19	175	1,4657	63,0	13,15
5	50,0	0,836	202	259	332	2,64	188	1,4667	68,0	13,25
6	43,0	0,877	252	290	336	4,41	205	1,4931	61,6	12,14
7	45,6	0,801	179	200	229	1,32	150	1,4463	61,6	13,83
8	54,4	0,844	243	282	335	3,52	206	1,4708	72,9	13,22
9	56,0	0,842	250	284	324	3,56	208	1,4690	74,7	13,33
10	52,5	0,844	228	279	333	3,39	203	1,4707	72,0	13,2
11	53,7	0,838	229	272	323	3,07	199	1,4672	71,9	13,33
12	52,0	0,841	234	270	316	3,08	196	1,4692	69,9	13,2
13	57,0	0,856	272	308	347	5,00	227	1,4775	77,1	13,06
14	48,6	0,834	196	250	337	2,39	181	1,4658	65,5	13,22
15	55,0	0,855	268	300	341	4,56	219	1,4775	74,5	13
16	45,0	0,804	174	204	230	1,29	153	1,4481	61,7	13,76
17	51,0	0,845	231	274	320	3,06	199	1,4718	69,6	13,1
18	47,8	0,859	235	285	338	3,88	205	1,4811	67,3	12,7
19	49,3	0,849	230	275	324	3,32	199	1,4747	68,1	12,96
20	55,0	0,862	275	310	350	5,13	227	1,4814	75,3	12,87
21	39,7	0,800	182	185,5	198	1,41	141	1,4466	56,4	13,64

Tabla 23. Propiedades fisicoquímicas de las muestras del diésel de estudio (Continuación)

N°	CI ASTM D-4737	Densidad 15°C [g/cm ³]	ASTM 10% [v/v]	D-86 50% [v/v]	D. [°C] 90% [v/v]	Viscosidad 40°C [mm ² /s]	Peso molecular [g/mol]	Índice de refracción 20°C	Punto de anilina [°C]	Contenido H2 [%, m/m]
22	47,8	0,837	242	245,5	258	2,58	177	1,4677	62,8	13,07
23	50,6	0,846	262	265,5	278	3,15	191	1,4733	65,8	12,95
24	56,6	0,863	302	305,5	318	6,50	222	1,4826	73,0	12,78
25	64,8	0,873	342	345,5	358	10,01	259	1,4870	83,2	12,83
26	39,0	0,802	182	185,5	198	1,21	140	1,4477	55,8	13,58
27	44,6	0,827	222	225,5	238	1,99	164	1,4621	59,8	13,19
28	48,5	0,835	242	245,5	258	2,53	178	1,4666	63,5	13,14
29	50,3	0,847	262	265,5	278	3,08	191	1,4739	65,5	12,92
30	52,5	0,857	282	285,5	292	3,53	206	1,4792	68,7	12,81
31	56,2	0,864	302	305,5	318	6,89	222	1,4831	72,7	12,76
32	59,5	0,871	322	325,5	338	8,95	239	1,4867	77,0	12,71
33	41,3	0,797	182	185,5	198	1,17	141	1,4443	58,0	13,78
34	46,4	0,822	222	225,5	238	1,95	165	1,4593	61,6	13,35
35	50,2	0,831	242	245,5	258	2,48	178	1,4641	65,1	13,27
36	61,8	0,866	322	325,5	338	7,43	241	1,4833	79,1	12,88
37	41,5	0,796	182	185,5	198	1,12	141	1,4440	58,2	13,8
38	44,6	0,809	202	205,5	218	1,39	153	1,4510	60,3	13,61
39	46,8	0,822	222	225,5	238	1,76	165	1,4587	62,0	13,38
40	48,7	0,834	242	245,5	258	2,27	178	1,4662	63,8	13,15
41	48,9	0,850	262	265,5	278	2,93	190	1,4759	64,2	12,81
42	55,2	0,851	282	285,5	292	4,05	207	1,4753	71,2	13,01
43	40,7	0,798	182	185,5	198	1,17	141	1,4452	57,4	13,73

Tabla 24. Propiedades fisicoquímicas de las muestras del diésel de estudio (Continuación)

N°	CI ASTM D-4737	Densidad 15°C [g/cm ³]	ASTM			Viscosidad 40°C [mm ² /s]	Peso molecular [g/mol]	Índice de refracción 20°C	Punto de anilina [°C]	Contenido H2 [% m/m]
			10% [v/v]	D-86 50% [v/v]	D. [°C] 90% [v/v]					
44	49,2	0,833	242	245,5	258	2,51	178	1,4656	64,2	13,19
45	50,1	0,848	262	265,5	278	2,90	191	1,4742	65,3	12,9
46	54,9	0,852	282	285,5	292	4,16	207	1,4758	70,9	12,98
47	59,1	0,858	302	305,5	318	5,78	224	1,4790	75,4	12,97
48	58,8	0,872	322	325,5	338	7,93	239	1,4877	76,4	12,66
49	46,0	0,804	175	206	232	1,31	154	1,4476	62,8	13,81
50	46,0	0,806	180	209	234	1,37	156	1,4488	63,0	13,78
51	53,0	0,847	245	280	329	3,48	203	1,4729	70,9	13,09
52	52,0	0,839	240	267	291	2,67	194	1,4682	69,6	13,24
53	54,0	0,858	261	303	345	4,76	221	1,4790	74,5	12,95
54	51,0	0,890	314	333	360	8,48	242	1,4997	71,5	12,13
55	48,8	0,886	297	320	351	6,80	230	1,4971	68,8	12,17
56	20,0	0,911	208	240	265	2,45	159	1,5195	30,8	10,43
57	15,0	0,943	220	246	273	2,55	157	1,5414	21,6	9,51
58	13,0	1,013	300	322	348	9,32	203	1,5859	23,9	8,39
59	43,3	0,872	210	308	359	5,42	223	1,4879	70,5	12,53
60	37,5	0,790	128	164	246	0,95	128	1,4410	52,8	13,71
61	49,8	0,833	196	261,2	328	2,46	191	1,4644	70,2	13,39
62	27,6	0,931	258,1	307,5	342,1	5,15	208	1,5289	47,1	10,61
63	35,6	0,881	220,7	292,9	338,2	3,80	206	1,4956	61,1	12,04
64	30,4	0,891	196,7	276,3	332,1	2,92	190	1,5035	51,2	11,52
65	38,3	0,864	203,4	275,9	331,5	2,98	196	1,4847	62,1	12,45

Tabla 25. Propiedades fisicoquímicas de las muestras del diésel de estudio (Continuación)

N°	CI ASTM D-4737	Densidad 15°C [g/cm ³]	ASTM 10% [v/v]	D-86 50% [v/v]	D. [°C] 90% [v/v]	Viscosidad 40°C [mm ² /s]	Peso molecular [g/mol]	Índice de refracción 20°C	Punto de anilina [°C]	Contenido H2 [%, m/m]
66	52,5	0,844	230	279	329	3,40	203	1,4708	71,9	13,19
67	66,7	0,817	253	274	308	2,87	206	1,4528	82,2	14,14
68	51,4	0,856	245	292	346	4,15	212	1,4782	71,4	12,91
69	40,0	0,892	257	299	345	5,23	209	1,5025	59,0	11,75
70	45,3	0,872	257	286	330	4,12	203	1,4899	62,3	12,27
71	55,7	0,823	219	255	322	2,41	188	1,4580	72,4	13,69
72	57,7	0,865	288,5	329	369	6,87	245	1,4823	80,9	12,96
73	35,9	0,888	239	275	315	3,77	190	1,5017	51,8	11,6
74	40,4	0,876	241,5	276,5	317	3,71	194	1,4933	57,2	12,02
75	41,1	0,833	166	220	303	1,75	159	1,4669	54,8	12,87
76	41,0	0,853	178	250	328	2,54	177	1,4789	57,3	12,52
77	46,0	0,838	177	250	328	2,42	180	1,4685	63,8	13,07
78	33,3	0,893	201	285	340	4,35	197	1,5043	53,5	11,56
79	37,5	0,883	211	285	342	4,21	199	1,4970	57,7	11,91
80	44,2	0,864	209	291	341	4,25	209	1,4840	67,5	12,6
81	55,2	0,829	200	274	356	3,02	203	1,4610	76,7	13,68
82	55,8	0,806	179	242	345	1,98	181	1,4471	75,4	14,19
83	58,3	0,827	217	279	360	3,16	207	1,4593	79,4	13,81
84	59,6	0,820	214	272	359	2,85	203	1,4550	80,1	14
85	62,7	0,811	211	268	351	2,62	202	1,4491	82,8	14,3
86	55,3	0,829	213	269	351	2,86	198	1,4612	74,8	13,63
87	53,4	0,827	210	256	343	2,47	188	1,4606	70,9	13,55

Tabla 26. Propiedades fisicoquímicas de las muestras del diésel de estudio (Continuación)

N°	CI ASTM D-4737	Densidad 15°C [g/cm³]	ASTM 10% [v/v]	D-86 50% [v/v]	D. [°C] 90% [v/v]	Viscosidad 40°C [mm²/s]	Peso molecular [g/mol]	Índice de refracción 20°C	Punto de anilina [°C]	Contenido H2 [%, m/m]
88	48,1	0,841	210	254	346	2,55	183	1,4703	64,0	13,01
89	60,2	0,814	213	260	328	2,45	194	1,4515	78,4	14,1
90	59,1	0,817	213	261	330	2,50	194	1,4535	77,4	13,99
91	60,2	0,826	228	278	323	3,12	207	1,4587	79,5	13,84
92	58,8	0,826	227	272	325	3,08	202	1,4590	77,3	13,77
93	52,6	0,837	219	269	326	2,82	196	1,4667	71,2	13,33
94	59,0	0,834	235	288	345	3,85	214	1,4636	79,6	13,65
95	52,2	0,844	224	280	339	3,11	204	1,4709	72,2	13,2
96	54,2	0,838	220	279	350	3,26	205	1,4668	74,5	13,4
97	53,1	0,842	223	281	344	3,35	205	1,4694	73,4	13,28
98	57,0	0,835	240,2	276,3	324,6	2,84	203	1,4649	74,9	13,49
1	37,8	0,834	185	214	254	1,66	154	1,4680	52,2	12,75
2	46,0	0,804	182	205	231	1,32	153	1,4481	62,1	13,78
3	43,8	0,828	222	225,5	238	2,23	164	1,4633	59,0	13,12
4	53,1	0,855	282	285,5	292	3,33	206	1,4783	69,3	12,85
5	64,8	0,873	342	345,5	358	10,79	259	1,4870	83,2	12,83
6	43,0	0,812	202	205,5	218	1,60	152	1,4534	58,7	13,47
7	51,5	0,844	262	265,5	278	3,07	192	1,4719	66,7	13,02
8	54,1	0,853	282	285,5	292	3,52	207	1,4769	70,1	12,93
9	58,1	0,860	302	305,5	318	5,98	223	1,4805	74,4	12,89
10	59,6	0,857	302	305,5	318	5,54	224	1,4783	75,8	13
11	57,7	0,875	322	325,5	338	7,98	238	1,4892	75,4	12,59

Tabla 27. Propiedades fisicoquímicas de las muestras del diésel de estudio (Continuación)

N°	CI ASTM D-4737	Densidad 15°C [g/cm ³]	ASTM 10% [v/v]	D-86 50% [v/v]	D. [°C] 90% [v/v]	Viscosidad 40°C [mm ² /s]	Peso molecular [g/mol]	Índice de refracción 20°C	Punto de anilina [°C]	Contenido H2 [%, m/m]
12	46,8	0,822	222	225,5	238	1,89	165	1,4587	62,0	13,38
13	46,0	0,809	172	210	252	1,37	156	1,4512	61,7	13,65
14	58,0	0,862	289	317	352	6,06	234	1,4809	77,9	12,95
15	54,0	0,859	260	311	370	4,91	229	1,4797	76,7	12,97
16	19,0	0,967	228	268	328	3,38	169	1,5568	21,0	9,07
17	39,0	0,870	206,5	286,4	336,4	3,34	204	1,4881	63,5	12,36
18	36,0	0,865	194,9	272,3	329,5	2,80	193	1,4857	60,4	12,37
19	34,1	0,877	207,6	281,5	333	3,10	198	1,4936	58,6	12,05
20	41,7	0,854	196,7	269,3	329,3	2,74	193	1,4783	63,9	12,72
21	31,0	0,900	221,3	297,3	339,1	4,70	206	1,5085	55,1	11,46
22	45,6	0,873	248	297	346	4,74	212	1,4896	66,0	12,37
23	34,5	0,914	266	303	349	5,95	208	1,5175	51,8	11,08
24	24,4	0,947	280	310	351	7,24	206	1,5403	41,9	10,13
25	58,9	0,839	243,5	295,5	349	4,02	219	1,4667	80,0	13,53
26	57,6	0,857	270	320,5	366,5	5,92	238	1,4775	81,2	13,14
27	31,9	0,900	236	273	312,5	3,76	186	1,5096	46,8	11,21
28	45,6	0,864	245	279,5	318	3,69	199	1,4846	63,4	12,48
29	36,9	0,849	165	221	304	1,84	156	1,4776	48,6	12,31
30	45,9	0,818	166	220	302	1,67	162	1,4567	61,5	13,44
31	36,6	0,868	176	251	328	2,67	175	1,4893	51,3	12
32	45,4	0,852	221	263	312	2,93	188	1,4775	62,4	12,71
33	55,6	0,836	221	283	365	3,44	209	1,4652	76,8	13,52

Tabla 28. Propiedades fisicoquímicas de las muestras del diésel de estudio (Continuación)

N°	CI ASTM D-4737	Densidad 15°C [g/cm ³]	ASTM			Viscosidad 40°C [mm ² /s]	Peso molecular [g/mol]	Índice de refracción 20°C	Punto de anilina [°C]	Contenido H2 [% m/m]
			10% [v/v]	50% [v/v]	D. [°C] 90% [v/v]					
34	57,7	0,821	212	265	349	2,65	197	1,4560	77,0	13,88
35	57,9	0,814	208	252	324	2,26	187	1,4520	75,4	14
36	54,4	0,818	206	245	319	2,14	181	1,4551	70,9	13,76
37	58,9	0,826	228	272	324	3,09	202	1,4590	77,3	13,77
38	60,9	0,826	229	281	326	2,90	210	1,4585	80,6	13,87
39	58,5	0,826	228	270	326	2,88	200	1,4591	76,6	13,75
40	74,8	0,807	234	296	346	3,90	229	1,4449	95,2	14,75
41	52,5	0,814	213	231	269	1,93	171	1,4532	67,5	13,74
42	53,2	0,839	214	279	342	2,57	204	1,4675	74,0	13,37

ANEXO B: MATRIZ DE CORRELACIÓN

Tabla 29. Correlaciones entre las variables

	ρ	T10	T50	T90	ϑ	PM	IR	PA	H2	TME	FP
ρ	1										
T10	0,4946	1									
T50	0,6467	0,8081	1								
T90	0,5059	0,4082	0,8362	1							
ϑ	0,6696	0,8632	0,8724	0,6053	1						
PM	0,4575	0,8081	0,9723	0,8112	0,8405	1					
IR	0,9970	0,4447	0,5857	0,4523	0,6244	0,3878	1				
PA	-0,4759	0,3290	0,3614	0,3468	0,1890	0,5604	-0,5420	1			
H2	-0,9497	-0,2733	-0,3802	-0,2722	-0,4601	-0,1634	-0,9703	0,7247	1		
TME	0,6163	0,8280	0,9913	0,8461	0,8762	0,9718	0,5545	0,3904	-0,3454	1	
FP	0,4995	0,9944	0,8167	0,4213	0,8332	0,8117	0,4489	0,3335	-0,2742	0,8340	1
IC	-0,5330	0,3565	0,2655	0,2171	0,1471	0,4676	-0,5924	0,9628	0,7571	0,3162	0,3552

ANEXO C: CÓDIGO EN R

Esta estructura de código realizada en R se utiliza para los demás casos.

CASO I, pasos

Paso 1: Transportar los datos

```
read.excel<-  
function(header=TRUE,...){read.table("clipboard",sep="\t",header=header,dec=",",...  
.)}
```

```
Datoscasol=read.excel()
```

```
fix(Datoscasol)
```

Paso 2: Efectuar el modelo lineal

```
RegresionCasolZ=lm(CI~0+Densidad+T10+T50+T90+Viscosidad+Puntoanilina+ConH2+FlashPoint, data=Datoscasol)
```

```
summary(RegresionCasol)
```

```
layout(matrix(c(1,2,3,4),2,2)) # Figuras por página
```

```
plot(RegresionCasol)
```

```
residuos=residuals(RegresionCasol)
```

```
shapiro.test(residuos)
```

Paso 3: Leverage

```
lev = hat(model.matrix(RegresionCasol))
```

```
levmax=2*sum(lev)/98 # Máximo leverage.
```

```
plot(lev)
```

```
abline(levmax,0,col="blue",lwd=1,lty=3)
```

```
abline(levmax,0)
Datoscasol[lev>levmax,]
levmax
lev[lev>levmax]
```

Paso 4: Distancia de cook

```
cook= cooks.distance(RegresionCasol)
cookmax=2*sum(cook)/98 # Máximo cook.
plot(cook,ylab="Distanciade Cook")
abline(cookmax,0,col="red",lwd=1,lty=3)
abline(cook,0)
Datoscasol[cook>cookmax,]
cookmax
cook[cook>cookmax]
```

Paso 5: Análisis de residuos studentizados

```
restud=rstudent(RegresionCasol)
plot(restud)
abline(restud,0)
Datoscasol[abs(restud)>1.9546,] # Puntos fuera del rango (el 1.95 se calculó en
matlab)
abline(1.9546,0,col="red",lty=3)
abline(-1.9546,0,col="Green
",lty=3)
```

Paso :Análisis de residuos

```
plot(Datoscasol$Densidad,residuos)
plot(Datoscasol$T10,residuos)
plot(Datoscasol$T50,residuos)
plot(Datoscasol$T90,residuos)
plot(Datoscasol$Viscosidad,residuos)
plot(Datoscasol$Puntoanilina,residuos)
plot(Datoscasol$ConH2,residuos)
plot(Datoscasol$FlashPoint,residuos)
```

Paso 6: Análisis de residuos standarizados

```
rstand=rstandard(RegresionCasol)
plot(rstand)
abline(2,0,col="blue",lty=3)
abline(-2,0,col="maroon",lty=3)
```

Paso 7: Analizar anova - residuos

```
anova(RegresionCasol) # tabla anova
```

Paso 7: Coeficientes de correlación

```
cor(Datoscasol$ Densidad,residuos)
cor(Datoscasol$T10,residuos)
cor(Datoscasol$T50,residuos)
cor(Datoscasol$T90,residuos)
```

```
cor(Datoscasol$ Viscosidad,residuos)
cor(Datoscasol$ Puntoanilina,residuos)
cor(Datoscasol$ ConH2,residuos)
cor(Datoscasol$ FlashPoint,residuos)
```

```
# Intervalo de confianza
```

```
confint.default(RegresionCasol)
```

ANEXO D: MODELOS DE REGRESIÓN LINEAL MULTIPLE

CASO I

Tabla 30. Resumen ajuste caso I ($R^2=0.9832$, $R^2_{ajust.}=0.9817$).

	Coeficiente	Error estándar	t - student	Pr(> t)
Densidad	0,7661	0,2992	2,5600	1,21E-02
T10	1,3290	0,2652	5,0110	2,68E-06
T50	-2,2461	0,5699	-3,9410	1,60E-04
T90	0,3333	0,0480	6,9510	5,57E-10
Viscosidad	-0,2528	0,0639	-3,9580	1,51E-04
Punto anilina	2,7304	0,7332	3,7240	3,41E-04
Contenido H2	-1,1980	0,7245	-1,6540	0,101709
P. inflamación	-0,6838	0,2501	-2,7340	7,52E-03

Figura 11. Leverage para el caso I

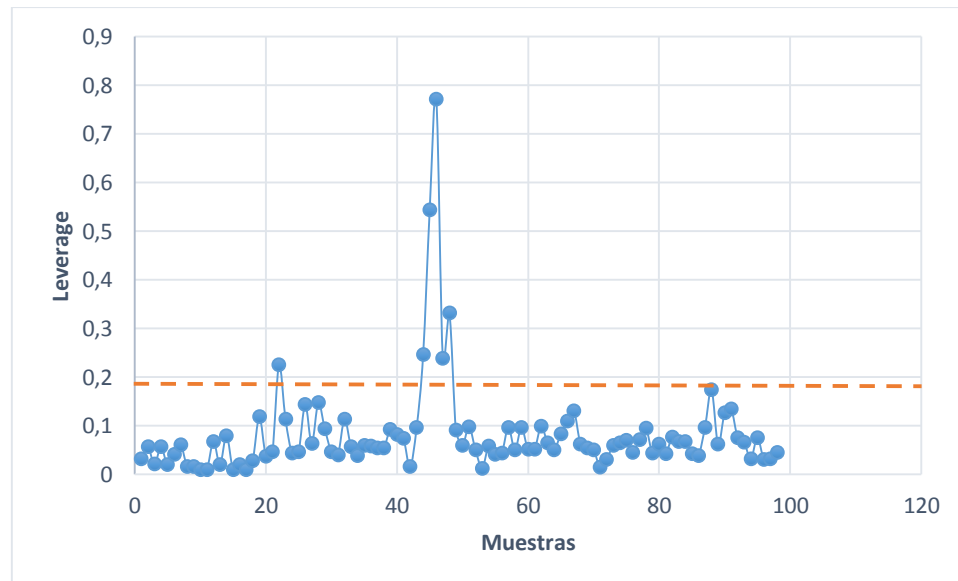
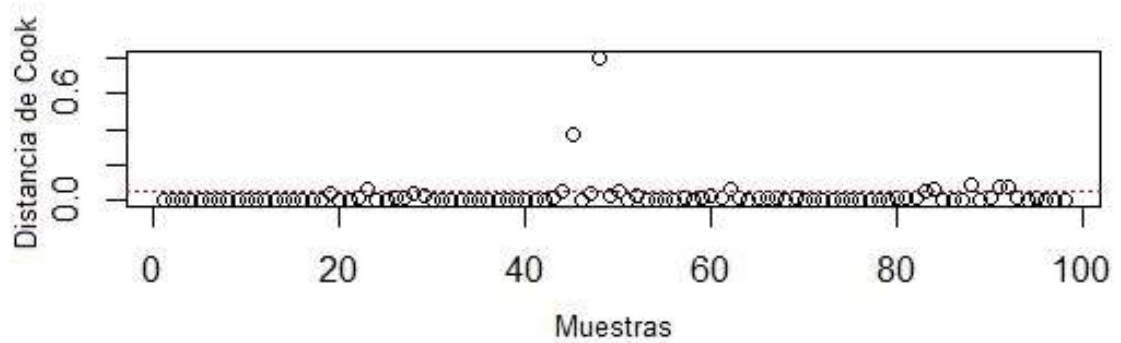


Figura 12. Distancia de Cook caso I



CASO II

Tabla 31. Resumen ajuste caso II ($R^2=0.9884$, $R^2_{ajust.}=0.9874$).

	Coeficiente	Error estándar	t - student	Pr(> t)
Densidad	-9,0698	2,1148	-4,2890	4,51E-05
T10	1,4097	0,1710	8,2450	1,27E-12
T90	0,2150	0,0425	5,0540	2,25E-06
Peso Molecular	-3,4478	0,3814	-9,0410	2,83E-14
Índice Refracción	7,4830	1,8557	4,0320	1,15E-04
Punto anilina	6,5136	0,7221	9,0210	3,12E-14
Contenido H2	-5,6823	0,7931	-7,1650	2,07E-10
P. inflamación	-0,8978	0,1748	-5,1360	1,62E-06

Figura 13. Leverage para el caso II

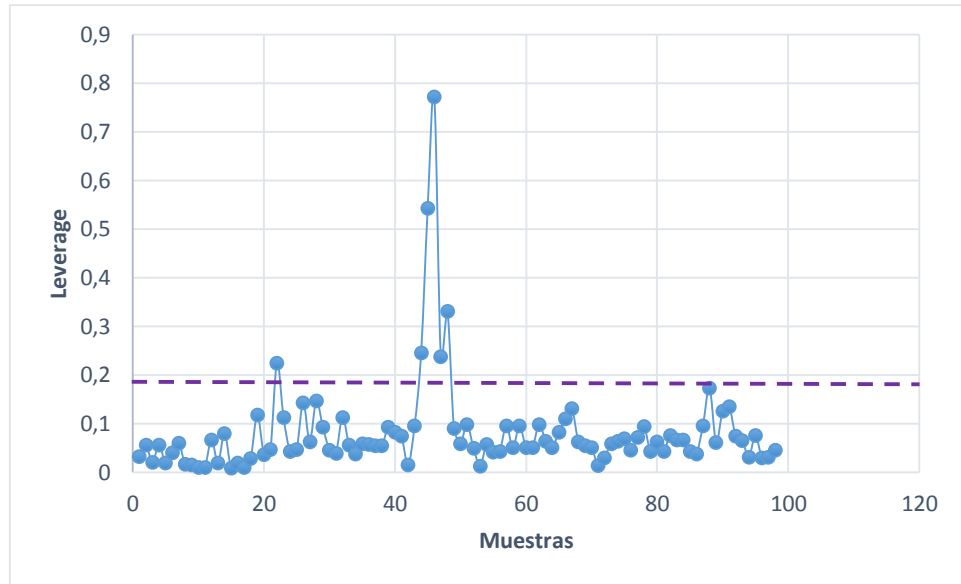
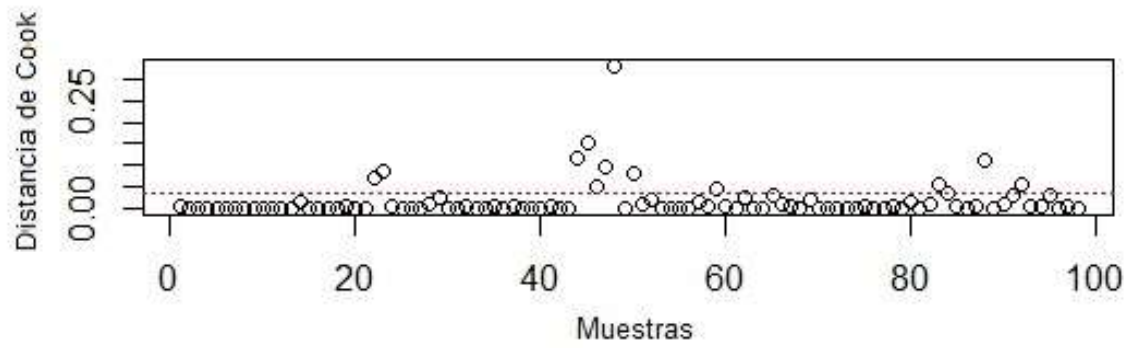


Figura 14. Distancia de Cook caso II



CASO III

Tabla 32. Resumen ajuste caso III ($R^2=0.9892$, $R^2_{ajust.}=0.9884$).

	Coefficiente	Error estándar	t - student	Pr(> t)
T10	1,4130	0,1598	8,8430	6,72E-14
T50	-2,3592	0,4220	-5,5900	2,35E-07
T90	0,2191	0,0406	5,3910	5,48E-07
Peso Molecular	-3,0907	0,3492	-8,8520	6,45E-14
Punto anilina	7,6422	0,8013	9,5370	2,37E-15
Contenido H2	-5,9835	0,7493	-7,9850	4,11E-12
P. inflamación	-0,8966	0,1652	-5,4260	4,73E-07

Figura 15. Leverage para el caso III

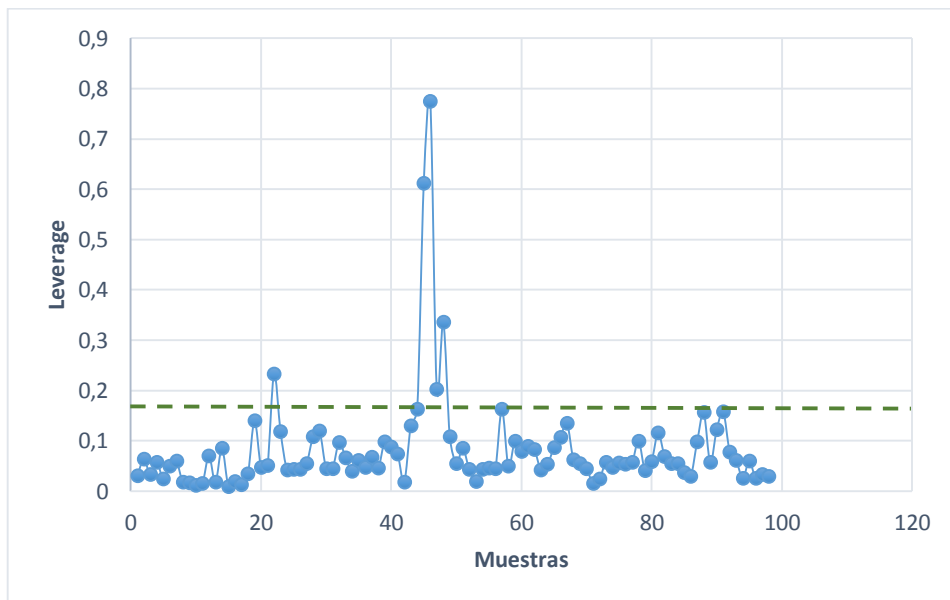
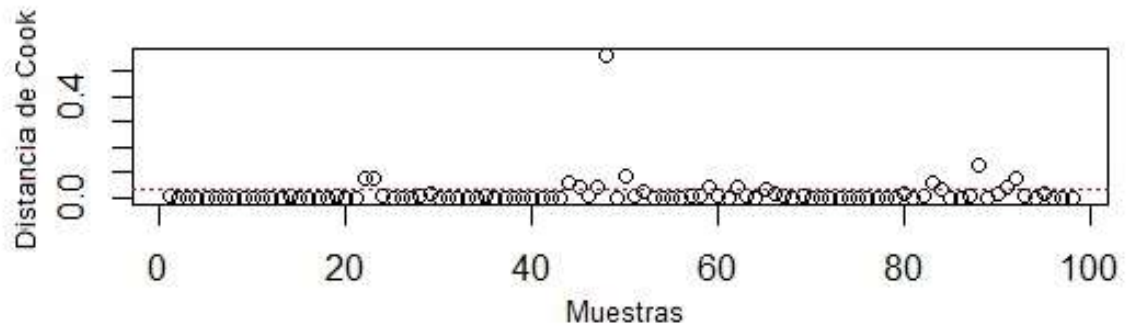


Figura 16. Distancia de Cook caso III



CASO IV

Tabla 33. Resumen ajuste caso IV ($R^2=0.9827$, $R^2_{ajust.}=0.9813$).

	Coeficiente	Error estándar	t - student	Pr(> t)
Densidad	0,8223	0,3001	2,7400	7,39E-03
T10	1,4790	0,2516	5,8790	6,74E-08
T50	-1,4088	0,2641	-5,3350	6,93E-07
T90	0,3340	0,0484	6,8980	6,80E-10
Viscosidad	-0,2678	0,0638	-4,1960	6,30E-05
Punto anilina	1,5858	0,2439	6,5030	4,16E-09
P. inflamación	-0,8222	0,2379	-3,4570	8,33E-04

Figura 17. Leverage para el caso IV

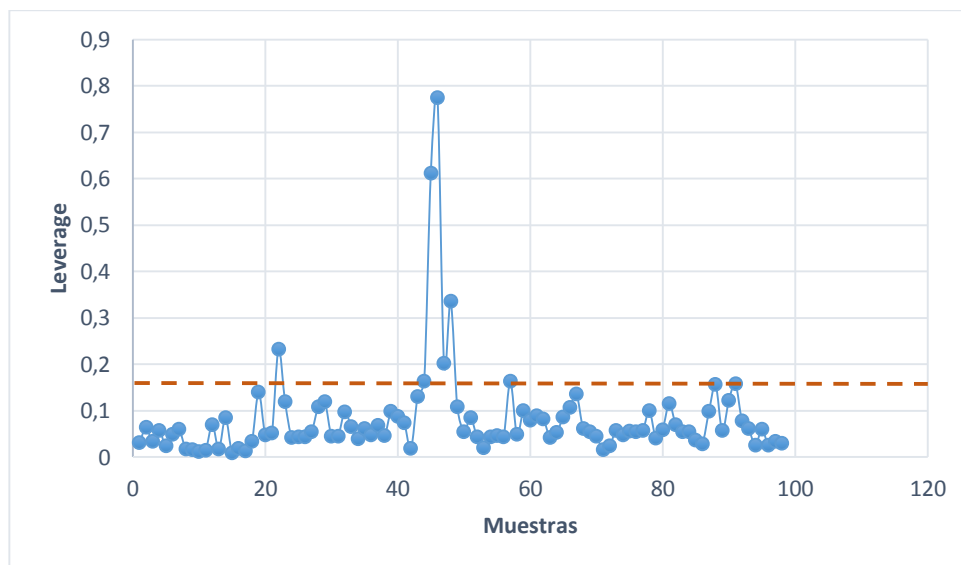
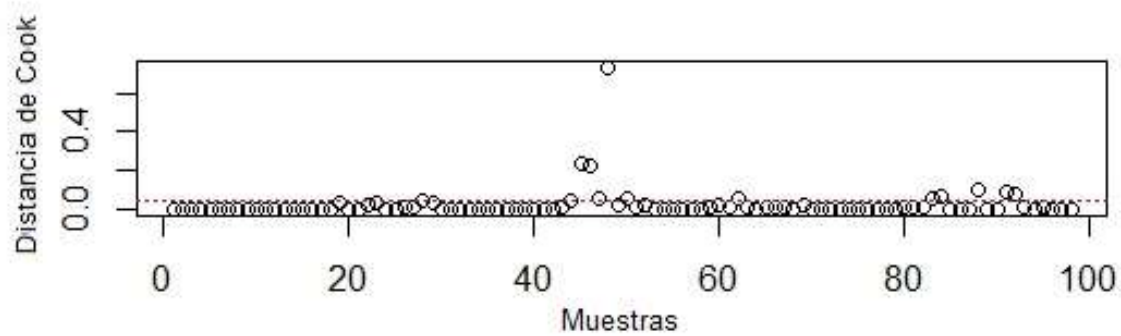


Figura 18. Distancia de Cook caso IV



CASO V

Tabla 34. Resumen ajuste caso V ($R^2=0.9783$, $R^2_{ajust.}=0.9771$).

	Coefficiente	Error estándar	t - student	Pr(> t)
T10	0,3413	0,0277	12,3080	< 2e-16
T50	-2,9751	0,5763	-5,1630	1,38E-06
Peso Molecular	-2,1772	0,3308	-6,5810	2,72E-09
Punto anilina	7,5983	1,0330	7,3560	7,35E-11
Contenido H2	-6,1432	0,9867	-6,2260	1,37E-08

Figura 19. Leverage para el caso V

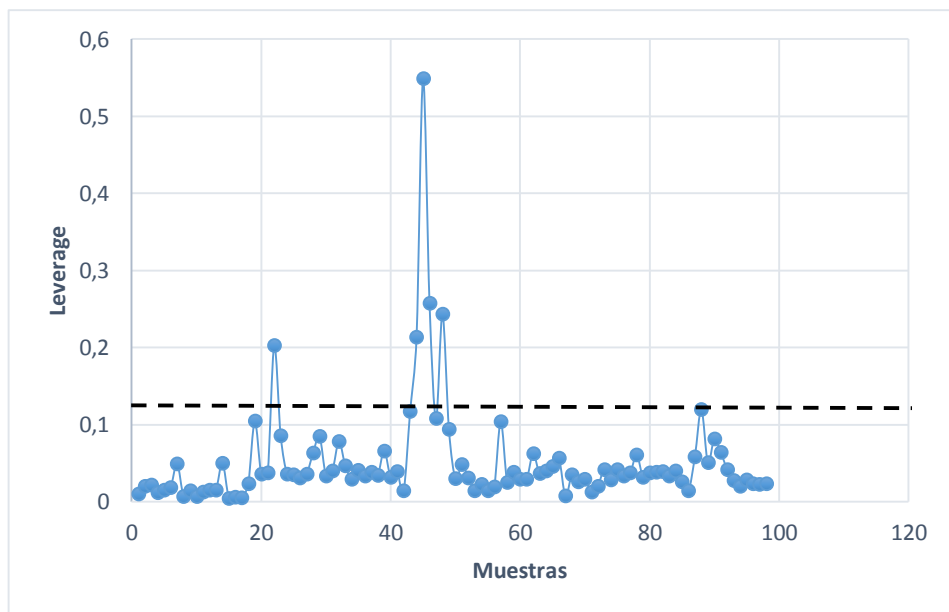
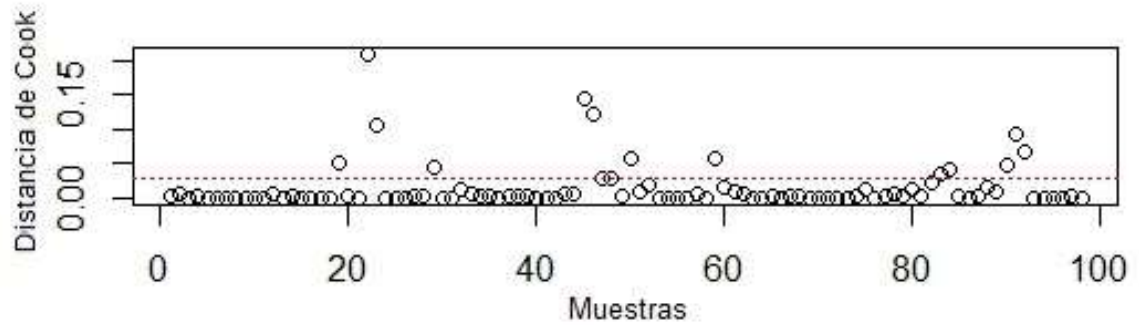


Figura 20. Distancia de Cook caso V



CASO VI

Tabla 35. Resumen ajuste caso VI ($R^2=0.943$, $R^2_{ajust.}=0.9405$).

	Coefficiente	Error estándar	t - student	Pr(> t)
T50	-3,4321	0,9274	-3,7010	3,62E-04
Peso Molecular	-1,0334	0,5120	-2,0180	4,64E-02
Punto anilina	6,9548	1,6638	4,1800	6,53E-05
Contenido H2	-5,7570	1,5904	-3,6200	4,77E-04

Figura 21. Leverage para el caso VI

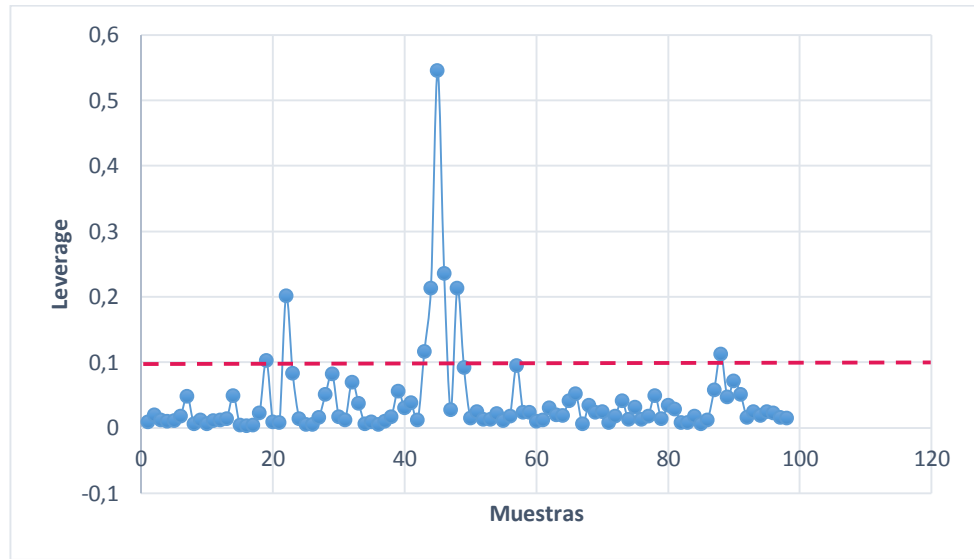
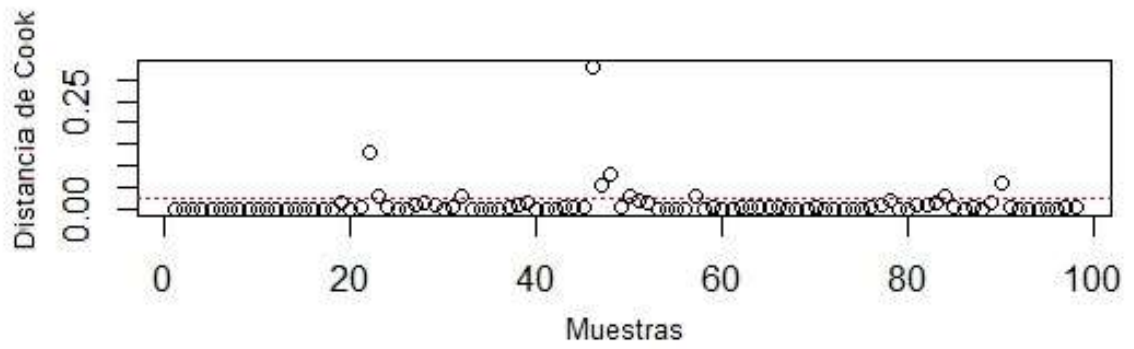


Figura 22. Distancia de Cook caso VI



CASO VII

Tabla 36. Resumen ajuste caso VI ($R^2=0.9434$, $R^2_{ajust.}=0.941$).

	Coeficiente	Error estándar	t - student	Pr(> t)
Densidad	1,0870	0,4910	2,2140	2,93E-02
T50	-3,6740	0,9564	-3,8420	2,22E-04
Punto anilina	5,3156	1,2021	4,4220	2,63E-05
Contenido H2	-3,4600	1,1598	-2,9830	3,63E-03

Figura 23. Leverage para el caso VII

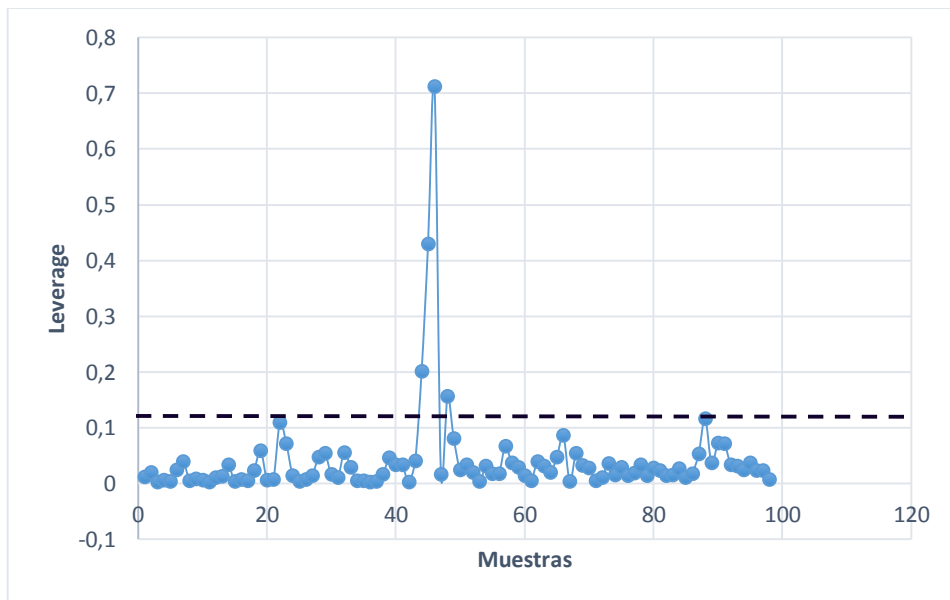
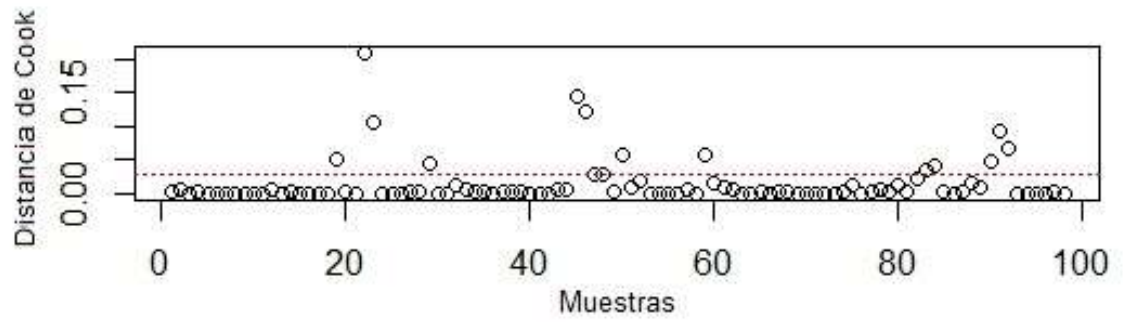


Figura 24. Distancia de Cook caso VII



ANEXO E: CORRELACIONES PARA EL PUNTO DE ANILINA Y CONTENIDO DE HIDROGENO

Las ecuaciones 1,2,3 y 4 corresponden al índice de refracción (IR), el punto de anilina (PA) y el contenido de hidrogeno (H2) fueron tomadas de Stratiev y colaboradores [40], las cuales se presentan a continuación:

$$PA = -183,3 + 0,27 * API * \sqrt[3]{T_{50}} + 0,317 * T_{50} \quad \text{Ecuación 1}$$

Donde: T_{50} es la temperatura de ebullición al 50% (v/v).

$$IR_{20^{\circ}C} = 1 + 0,8447 * SG^{1,2056} * T_{50}^{-0,0557} * PM^{-0,0044} \quad \text{Ecuación 2}$$

Donde: $IR_{20^{\circ}C}$ corresponde al índice de refracción a 20 °C, SG es la gravedad específica.

$$H2 = 30,346 + \frac{82,952 - 65,341 * IR_{20}}{\rho_4^{20}} - \frac{306}{PM} \quad \text{Ecuación 3}$$

Donde: ρ_4^{20} es la densidad relativa, y se representa:

$$\rho_4^{20} = \frac{\text{densidad a } 20^{\circ}C (g/cm^3)}{\text{densidad del agua a } 4^{\circ}C (g/cm^3)} \quad \text{Ecuación 4}$$

Y la densidad del agua a 4°C es: 1000 g/cm³

Las ecuaciones 5 y 6 que corresponden al peso molecular (PM) y a la gravedad específica (SG) fueron tomadas de Riazi y colaboradores [16], las cuales se presentan a continuación:

$$PM = 42,965 * \left[e^{\frac{(2,097x 10^{-4} T_b - 7,78712 * SG + 2,08476x 10^{-3} T_b * SG)}{}} \right] * T_b^{1,26007} * SG^{4,98308} \quad \text{Ecuación 5}$$

Donde: T_b es la temperatura de ebullición.

$$\rho_{20} = SG - 4,5 \times 10^{-3} * (2,34 - 1,9 * SG) \quad \text{Ecuación 6}$$

Donde: ρ_{20} corresponde a densidad a 20°C

Temperatura de ebullición al 50%

$$ASTM D 86(50 v/v\%) = 255.4 + 079424[SD(50 p/p\%) - 255.4]^{1.0395} \quad \text{Ecuación 7}$$

Dónde: SD (50% wt%) y ASTM D 86(50 vol%) son las temperaturas en 50% destiladas en Kelvin. La diferencia entre los puntos de corte adyacentes se calcula de la siguiente ecuación.

$$U_i = ET_i^F \quad \text{Ecuación 8}$$

Donde: U_i corresponde a la diferencia en la norma ASTM D86 entre la temperatura en dos puntos de corte [K].

T_i corresponde a la diferencia observada en las temperaturas SD entre dos puntos de corte [K].

E, F son constantes que varían en cada punto de corte, ver tabla 35

Tabla 37. Correlaciones para las constantes E y F [16].

i	Intervalo de punto de corte [%]	E	F	Rango permitido T_i [°C]
1	100-90	2,13092	0,6596	55
2	90-70	0,35326	1,2341	55
3	70-50	0,19121	1,4287	55
4	50-30	0,10949	1,5386	55
5	30-oct	0,08227	1,5176	85
6	10-0	0,3281	1,125	85

Para determinar la temperatura a partir de la norma ASTM D86 en cualquier porcentaje destilado, pueden emplearse las ecuaciones del 9 al 14, las cuales se presentan a continuación:

$$ASTM D86 (100\%) = ASTM D86(50\%) - U_4 - U_5 - U_6 \quad \text{Ecuación 9}$$

$$ASTM D86 (10\%) = ASTM D86(50\%) - U_4 - U_5 \quad \text{Ecuación 10}$$

$$ASTM D86 (30\%) = ASTM D86(50\%) - U_4 \quad \text{Ecuación 11}$$

$$ASTM D86 (70\%) = ASTM D86(50\%) - U_3 \quad \text{Ecuación 12}$$

$$ASTM D86 (90\%) = ASTM D86(50\%) + U_3 + U_2 \quad \text{Ecuación 13}$$

$$ASTM D86 (100\%) = ASTM D86(50\%) + U_3 + U_2 + U_1 \quad \text{Ecuación 14}$$

ANEXO F: CORRELACIÓN DEL ÍNDICE DE CETANO Y EL NÚMERO DE CETANO

Tabla 38. Datos CI ASTM976 y ASTM4737

Nº	CI ASTM D-976	CI ASTM D-4737
1	44,8426	45
2	50,2246	51
3	51,5083	51
4	46,6589	47
5	50,0645	50
6	43,3676	43
7	44,0411	45,6
8	52,1806	54,4
9	53,3832	56
10	51,752	52,5
11	52,3269	53,7
12	50,9211	52
13	52,45	57
14	48,5178	48,6
15	51,489	55
16	44,4703	45
17	50,4045	51
18	47,8022	47,8
19	49,1575	49,3
20	50,8477	55
21	37,7873	39,7
22	46,4837	47,8
23	48,0759	50,6
24	49,8231	56,6
25	51,1014	64,8
26	37,187	39
27	43,9737	44,6
28	47,0729	48,5
29	47,8124	50,3
30	48,7643	52,5
31	49,6109	56,2
32	49,9694	59,5

Tabla 39. Datos CI ASTM976 y ASTM4737

Nº	CI ASTM D-976	CI ASTM D-4737
33	39,2016	41,3
34	45,4659	46,4
35	48,3354	50,2
36	51,4194	61,8
37	39,3644	41,5
38	43,3532	44,6
39	45,7970	46,8
40	47,2470	48,7
41	46,8333	48,9
42	50,5800	55,2
43	38,6336	40,7
44	47,5965	49,2
45	47,6483	50,1
46	50,3543	54,9
47	51,4860	59,1
48	49,5329	58,8
49	45,5157	46
50	45,9101	46
51	50,8474	53
52	50,8896	52
53	51,2000	54
54	45,2843	51
55	45,1499	48,8
56	23,4478	20
57	18,6894	15
58	23,7282	13
59	47,7212	43,3
60	30,3730	37,5
61	51,6959	49,8
62	33,1800	27,6
63	42,7836	35,6
64	37,1066	30,4

Tabla 40. Datos CI ASTM976 y ASTM4737

N°	CI ASTM D-976	CI ASTM D-4737
65	44,7173	38,3
66	51,7186	52,5
67	60,2444	66,7
68	50,1520	51,4
69	40,8910	40
70	44,1404	45,3
71	53,7583	55,7
72	52,1261	57,7
73	37,5846	35,9
74	41,1712	40,4
75	39,7486	41,1
76	42,1347	41
77	47,1685	46
78	38,2420	33,3
79	41,0594	37,5
80	47,3920	44,2
81	55,8501	55,2
82	56,7626	55,8
83	57,5479	58,3
84	58,7167	59,6
85	61,2731	62,7
86	54,8124	55,3
87	52,5523	53,4
88	47,1688	48,1
89	58,3235	60,2
90	57,4239	59,1
91	57,7169	60,2
92	56,5196	58,8
93	52,0068	52,6
94	56,6945	59
95	51,8414	52,2
96	53,6817	54,2

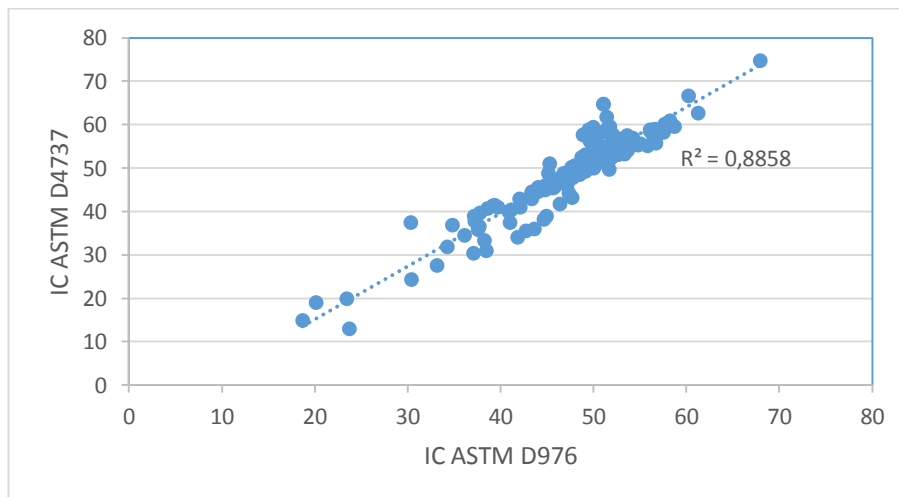
Tabla 41. Datos CI ASTM976 y ASTM4737

N°	CI ASTM D-976	CI ASTM D-4737
97	52,6992	53,1
98	54,2295	57
99	37,2345	37,8
100	44,8761	46
101	43,3269	43,8
102	49,1739	53,1
103	51,1014	64,8
104	42,0413	43
105	48,7389	51,5
106	49,8413	54,1
107	50,8030	58,1
108	51,7675	59,6
109	48,8991	57,7
110	45,7970	46,8
111	44,8920	46
112	51,7660	58
113	51,7451	54
114	20,1433	19
115	44,9402	39
116	43,6595	36
117	41,8700	34,1
118	46,4381	41,7
119	38,4824	31
120	45,8032	45,6
121	36,1337	34,5
122	30,4072	24,4
123	56,1282	58,9
124	53,6186	57,6
125	34,3118	31,9
126	45,3837	45,6
127	34,8501	36,9
128	45,1810	45,9

Tabla 42. Datos CI ASTMD976 y ASTMD4737

N°	CI ASTM D-976	CI ASTM D-4737
129	37,7479	36,6
130	45,6268	45,4
131	55,1159	55,6
132	56,8450	57,7
133	56,3620	57,9
134	52,9996	54,4
135	56,5196	58,9
136	58,2856	60,9
137	56,1022	58,5
138	67,9925	74,8
139	50,3853	52,5
140	53,3395	53,2

Figura 25. Correlación IC ASTM 4737 vs IC ASTM D976



ANEXO G: MÉTODO DE AUTOESCALADO- MEAD AND CENTER

Para el escalamiento de los datos iniciales se utilizó el método de Mead and Center, la expresión del modelo de regresión múltiple que se obtuvo fue:

$$CI = a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_5X_5 + \dots + a_nX_n$$

Donde:

a_1, a_2, \dots, a_n corresponde a cada uno de los coeficientes del modelo

X_1, X_2, \dots, X_n corresponde a cada uno de las propiedades autoescaladas, puede calcularse con la siguiente expresión:

$$X_1 = \frac{x - \bar{x}}{\sigma_x}$$

x : Corresponde al valor de la variable independiente (propiedad macroscópica)

\bar{x} : Corresponde al promedio de la variable independiente (propiedad macroscópica)

σ_x : Corresponde a la desviación estándar de la variable independiente (propiedad macroscópica).

Para regresar a las variables originales:

Para obtener los valores correspondientes a las variables originales, se presenta a continuación la siguiente expresión, la cual debe usarse con los valores de la desviación estándar y los promedios de cada una de las propiedades que se encuentran en la tabla 41.

$$x = (X_1 * \sigma_x) + \bar{x}$$

Tabla 43. Desviación estándar y promedio de las propiedades autoescaladas

Propiedades Autoescaladas	Promedio	Desviación Estándar
Densidad	47,6837	10,2258
T10	849,6469	35,0753
T50	504,5765	42,5642
T90	540,9041	37,7536
Viscosidad	583,3459	44,1269
Peso Molecular	0,0000	0,0000
Índice Refracción	193,7755	26,6955
Punto Anilina	274,6255	0,0228
Contenido H2	338,7133	11,4621
P. inflamación	12,8722	0,9650