

**DETERMINACIÓN DE UN MODELO DE RED NEURONAL ARTIFICIAL PARA
LA PREDICCIÓN DE LA PENETRACIÓN DEL ASFALTO PRODUCIDO EN LA
T-2103 DE LA GERENCIA REFINERÍA BARRANCABERMEJA**

WILMAN DAVID POLO



**UNIVERSIDAD INDUSTRIAL DE SANTANDER
ESCUELA DE INGENIERÍA QUÍMICA
FACULTAD DE INGENIERÍA FÍSICO-QUÍMICAS
BUCARAMANGA**

2018

**DETERMINACIÓN DE UN MODELO DE RED NEURONAL ARTIFICIAL PARA
LA PREDICCIÓN DE LA PENETRACIÓN DEL ASFALTO PRODUCIDO EN LA
T-2103 DE LA GERENCIA REFINERÍA BARRANCABERMEJA**

WILMAN DAVID POLO

**Trabajo de grado para optar al título de
MAESTRIA EN INGENIERIA QUIMICA**

Director

RAMIRO MARTINEZ REY. Ph.D

Dr. En Ingeniería Química



**UNIVERSIDAD INDUSTRIAL DE SANTANDER
ESCUELA DE INGENIERÍA QUÍMICA
FACULTAD DE INGENIERÍA FÍSICO-QUÍMICAS
BUCARAMANGA**

2018

CONTENIDO

	Pág.
INTRODUCCION	13
1. PLANTEAMIENTO DEL PROBLEMA	14
2. MARCO TEÓRICO	20
2.1 LA UNIDAD DE DESTILACIÓN PRIMARIA U-2100	20
2.2 REDES NEURONALES ARTIFICIALES (RNA)	22
2.3 APLICACIONES DE LAS RNA EN PREDICCIONES REFERENTES AL ASFALTO	25
3. OBJETIVOS	29
3.1 OBJETIVO GENERAL	29
3.2 OBJETIVOS ESPECIFICOS	29
4. METODOLOGÍA	30
5. RESULTADOS	37
5.1 DATOS HISTÓRICOS	37
5.2 IDENTIFICACIÓN DE MODOS O ESQUEMAS OPERACIONALES	39
5.3 REGRESIÓN LINEAL MÚLTIPLE	45
5.4 REDES NEURONALES ARTIFICIALES	50
5.5 IMPACTO DE LOS MODELOS DE RNA	70
6. CONCLUSIONES	71
7. RECOMENDACIONES	73
REFERENCIAS	74
ANEXOS	79

LISTA DE FIGURAS

	Pág.
Figura 1. Tanques de asfalto con reproceso.	17
Figura 2. Tendencia tanques a reproceso vs tanques preparados, año 2016.	18
Figura 3. Ejemplo de una estructura de RNA multicapa.	23
Figura 4. Funciones de activación habitualmente utilizadas en RNA.	24
Figura 5. Etapas de la metodología.	30
Figura 6. Detección valores atípicos por medio del gráfico <i>boxplot</i> .	31
Figura 7. Tendencia para algunas variables en las muestras.	38
Figura 8. Matriz de correlación de Pearson. En círculos rojos se indican dependencias importantes entre pares de variables (correlaciones con valor mayor a 0.65).	41
Figura 9. Variación de la distancia interna con el aumento en el número de clústers.	42
Figura 10. Coeficiente silueta para diferentes agrupaciones definidas en el método <i>kmean</i> .	43
Figura 11. Clústeres formados y su relación con las variables temperatura de precaliente, T_{pre_A} , y reflujos de cima en la torre de vacío, $RefCim_V$. Color celeste para el esquema 1 y fucsia para el esquema 2.	47
Figura 12. Pruebas F y <i>one-way</i> ANOVA para la comparación de las varianzas y las medias, respectivamente, de las penetraciones reportadas por los esquemas operacionales según el método <i>kmeans</i> .	48
Figura 13. Visualización de las agrupaciones en coordenadas de la penetración.	49
Figura 14. Resultados reportados por R para el modelo lineal de predicción de la penetración (PEN) con todos los datos históricos, indicando un R^2 de 0.5715.	49
Figura 15. Resultados reportados por R para el modelo lineal de predicción de la penetración (PEN) lineal en cada esquema operacional. Esquema operativo 1, R^2 de 0.5992; esquema operativo 2, R^2 de 0.5646.	50

Figura 16. Gráficas de influencia resultantes de las regresiones RLM para la penetración.	51
Figura 17. Ilustración RNA <i>feedforward</i> 26:2:1 Utilizada para la regresión de la penetración (abreviaciones definidas en el anexo 1); dos neuronas capa interna.	52
Figura 18. Desempeño obtenido con la red 26:x:1 – LogSig con el número de neuronas internas x, con todos los datos históricos (20 repeticiones).	53
Figura 19. Desempeño obtenido con la red 26:x:1 – TanH con el número de neuronas internas x, con todos los datos históricos (20 repeticiones).	54
Figura 20. Desempeños obtenidos con las redes utilizando dos capas internas con todos los datos históricos (20 repeticiones).	55
Figura 21. Resultados para <i>kmeans</i> =3. Celeste agrupación 1, verde agrupación 2 y fucsia agrupación 3.	59
Figura 22. Desempeño de las redes 26:x:1 – LogSig con los históricos sin datos de la agrupación 3 y PEN hasta 100 mm/10 (20 repeticiones).	60
Figura 23. Desempeño de las redes 26:x:1 – TanH con los históricos sin datos de la agrupación 3 y PEN hasta 100 mm/10 (20 repeticiones).	61
Figura 24. Desempeño de las redes 26:32:x:1 – LogSig y 26:35:x:1 – TanH con los históricos sin datos de la agrupación 3 y PEN hasta 100 mm/10 (20 repeticiones).	62
Figura 25. Desempeño de la red 26:35:4:1 – TanH con los históricos sin datos de la agrupación 3 y PEN hasta 100 mm/10 (20 repeticiones).	63
Figura 26. Desempeño de las redes 26:x:1 – LogSig con los históricos sin datos de la agrupación 3 y PEN hasta 60 mm/10 (20 repeticiones).	63
Figura 27. Desempeño de las redes 26:x:1 – TanH con los históricos sin datos de la agrupación 3 y PEN hasta 60 mm/10 (20 repeticiones).	64
Figura 28. Desempeño red 26:12:1 – TanH con los históricos sin la agrupación 3 y hasta 60 mm/10 (20 repeticiones).	64
Figura 29. Desempeño de las redes 26:21:x:1 – LogSig con los históricos sin datos de la agrupación 3 y PEN hasta 60 mm/10 (20 repeticiones).	65

- Figura 30. Desempeño de las redes 26:12:x:1 – TanH con los históricos sin datos de la agrupación 3 y PEN hasta 60 mm/10 (20 repeticiones). 67
- Figura 31. Resultados prueba de sensibilidad red 26:12:1 – TanH con los históricos sin datos de la agrupación 3 y PEN hasta 60 mm/10 (20 repeticiones). 69
- Figura 32. Resultados prueba de sensibilidad red 26:35:4:1 – TanH con los históricos sin datos de la agrupación 3 y PEN hasta 100 mm/10 (20 repeticiones). 69

LISTA DE TABLAS

	Pág.
Tabla 1. Parámetros de calidad del asfalto producto 60/70 mm/10.	15
Tabla 2. Resultados del método <i>kmean</i> . STC: Suma total de cuadrados.	40
Tabla 3. Resultados ajustes de las arquitecturas de red propuestas en el presente trabajo	68

LISTA DE ANEXOS

	Pág.
ANEXO 1. VARIABLES DE LA UNIDAD U-2100 UTILIZADAS EN EL TRABAJO Y SUS CORRESPONDIENTES ABREVIATURAS	80
ANEXO 2. CÓDIGO EN R PARA EL ENTRENAMIENTO Y LA VALIDACIÓN DE LAS RNA UTILIZANDO LA FUNCIÓN <i>NEURALNET</i>	82

RESUMEN

TITULO: * DETERMINACIÓN DE UN MODELO DE RED NEURONAL ARTIFICIAL PARA LA PREDICCIÓN DE LA PENETRACIÓN DEL ASFALTO PRODUCIDO EN LA T-2103 DE LA GERENCIA REFINERÍA BARRANCABERMEJA

AUTOR WILMAN DAVID POLO**

PALABRAS CLAVE: Asfaltos, penetración, destilación, redes neuronales artificiales, neuralnet.

DESCRIPCION

En la Gerencia de la Refinería Barrancabermeja, el fondo de vacío de la unidad U-2100 es utilizado en la preparación de cemento asfáltico utilizado en pavimentación. Entre las especificaciones comerciales para el asfalto se tiene que su penetración debe oscilar entre 60 y 70 mm/10. La penetración es determinada experimentalmente con un tiempo de respuesta de 5 horas, con lo cual la corrección del asfalto preparado fuera de especificaciones impacta en la logística y en las entregas del producto final. En este trabajo se expone una metodología dirigida al entrenamiento de redes neuronales artificiales (RNA) para la predicción de la penetración del fondo de vacío de la U-2100 utilizando datos históricos de las variables de proceso. Según los resultados, un análisis estadístico de los históricos condujo a la definición de 26 variables para el desarrollo de modelos de predicción. Asimismo, un análisis de agrupaciones revela dos esquemas operativos para U-2100; el esquema 1 contiene muestras anteriores a la reparación de la unidad y se caracteriza por valores de penetración altos; el esquema 2 posee muestras posteriores a la reparación de la unidad, con menores valores de penetración. Las arquitecturas de RNA mostraron diferentes desempeños según la función de activación, el número de neuronas y el número de capas internas. El mejor desempeño en la predicción de la penetración hasta 100 mm/10 fue obtenido con la estructura 26:35:4:1 – TanH. Por otra parte, el mejor desempeño en la predicción de la penetración hasta 60 mm/10 fue obtenido con la estructura 26:12:1 – TanH. Las desviaciones estándar de las mejores redes correspondieron a 6.5 mm/10, las cuales resultaron comparables con la repetibilidad promedio de 4 mm/10 de la norma ASTM-D5-06. Asimismo, los valores de las desviaciones estándar de las mejores RNA resultaron comparables con trabajos previos de la literatura abierta.

* Proyecto de Grado

** Escuela de Ingeniería Química, Facultad de Ingeniería Físico-Químicas, Director

RAMIRO MARTINEZ REY. Ph.DDr. En Ingeniería Química

ABSTRACT

TITLE: * DETERMINATION OF AN ARTIFICIAL NEURONAL NETWORK MODEL FOR THE PREDICTION OF THE PENETRATION OF THE ASPHALT PRODUCED IN THE T-2103 OF THE BARRANCABERMEJA REFINERY MANAGEMENT

AUTHOR ** WILMAN DAVID POLO

KEYWORDS: Asphalts, penetration, distillation, artificial neural networks, neuralnet.

DESCRIPTION

In "Gerencia de la Refinería Barrancabermeja", vacuum bottom of U-2100 unit is used to prepare road asphalt. One of the most important commercial requirements for road asphalt corresponds to penetration, which have to be in the range 60-70 mm/10. This property is measured in lab with a reporting time of 5 hours for the result, which affects logistic for delivering commercial products due to asphalt out of specifications. Herein a methodology for training artificial neural network (ANN) for prediction of penetration of U-2100 unit is developed based on process historic data. According to results, an statistics univariate analysis showed 26 important variables for the prediction models. Likewise, a clustering analysis revealed two operational modes for the U-2100. First operational mode contained samples characteristics for high values of penetration, whereas the second operational mode possessed samples with low values of penetration. ANN showed different performance according to activation function, number of neurons and number of hidden layers. The best performance was obtained with the ANN 26:35:4:1 – TanH for penetration up to 100 mm/10. On the other hand, ANN 26:12:1 – TanH showed the best performance for penetration up to 60 mm/10. Deviations of trained ANN were in excellent agreement with the ASTM-D5-06 standard, as well as, models reported in literature.

* Graduation project

** School of Chemical Engineering, Faculty of Physical-Chemical Engineering, Director RAMIRO MARTINEZ REY. Ph.DDr. In Chemical Engineering

INTRODUCCION

En la Gerencia de la Refinería Barrancabermeja, el fondo de vacío de la unidad U-2100 es utilizado en la preparación de cemento asfáltico utilizado en pavimentación. Entre las especificaciones comerciales para el asfalto se tiene que su penetración debe oscilar entre 60 y 70 mm/10. La penetración es determinada experimentalmente con un tiempo de respuesta de 5 horas, con lo cual la corrección del asfalto preparado fuera de especificaciones impacta en la logística y en las entregas del producto final. En este trabajo se expone una metodología dirigida al entrenamiento de redes neuronales artificiales (RNA) para la predicción de la penetración del fondo de vacío de la U-2100 utilizando datos históricos de las variables de proceso. Según los resultados, un análisis estadístico de los históricos condujo a la definición de 26 variables para el desarrollo de modelos de predicción. Asimismo, un análisis de agrupaciones revela dos esquemas operativos para U-2100; el esquema 1 contiene muestras anteriores a la reparación de la unidad y se caracteriza por valores de penetración altos; el esquema 2 posee muestras posteriores a la reparación de la unidad, con menores valores de penetración. Las arquitecturas de RNA mostraron diferentes desempeños según la función de activación, el número de neuronas y el número de capas internas. El mejor desempeño en la predicción de la penetración hasta 100 mm/10 fue obtenido con la estructura 26:35:4:1 – TanH. Por otra parte, el mejor desempeño en la predicción de la penetración hasta 60 mm/10 fue obtenido con la estructura 26:12:1 – TanH. Las desviaciones estándar de las mejores redes correspondieron a 6.5 mm/10, las cuales resultaron comparables con la repetibilidad promedio de 4 mm/10 de la norma ASTM-D5-06. Asimismo, los valores de las desviaciones estándar de las mejores RNA resultaron comparables con trabajos previos de la literatura abierta.

1. PLANTEAMIENTO DEL PROBLEMA

Los inventarios de fondos de vacío se han convertido en los últimos años en un problema importante en la refinería de Barrancabermeja, debido a la disminución en la producción de crudos livianos en los campos nacionales y al descubrimiento y la reactivación de pozos de crudos pesados. Con los crudos pesados se obtienen rendimientos de fondos de vacío de aproximadamente 26% por cada barril que carga la refinería, equivalente a 60000 barriles diarios (BD). Uno de los destinos para estos fondos es la planta Demex, la cual procesa cerca de 45000 BD, el flujo restante de fondos es utilizado en la producción de crudo reconstituido y en la producción de cemento asfáltico, materia prima fundamental en la fabricación de pavimentos asfálticos.

La producción de cemento asfáltico es un proceso estratégico para la refinería; ya que permite mantener en control los fondos de vacío dando como resultado el sostenimiento de los niveles de carga. También se destaca su beneficio económico (47 US\$ por barril) debido a su demanda cada vez más creciente en el mercado nacional. Por esta razón, es de gran importancia controlar los parámetros de calidad que se observan en la Tabla 1 para asfalto producto. Los asfaltos a ventas, deben cumplir con especificaciones de calidad para su comercialización; entre ellas la penetración, la cual, es el parámetro más importante; de hecho, da el nombre comercial al producto (Asfalto 60/70). La penetración estima la dureza o consistencia relativa de un cemento asfáltico por medio de la distancia vertical que una aguja normalizada recorre en una muestra, en condiciones específicas de temperatura, carga y tiempo (ASTM D5-06, 2011). El valor requerido de penetración del cemento asfáltico debe oscilar entre 60 y 70 mm/10.

La Gerencia de la Refinería de Barrancabermeja (GRB) dispone de cinco unidades de destilación primaria U-200, U-150, U-250, U-2000 y U-2100, las cuales están compuestas de dos secciones principales: la sección atmosférica y la

sección al vacío, siendo esta última donde se produce el cemento asfáltico a partir del crudo reducido proveniente de la sección atmosférica. La U-2100 es la unidad mediante la cual la GRB lleva a cabo la preparación del cemento asfáltico, dado que las otras unidades han sido destinadas a la elaboración de otros productos o han tenido adaptaciones físicas para incrementar la severidad del proceso y así aumentar la recuperación de gasóleos.

Tabla 1. Parámetros de calidad del asfalto producto 60/70 mm/10.

Asfaltos 60/70				
PROPIEDADES /	UNIDADES	MÉTODO	ESPECIFICACIONES	
		ASTM	MÍNIMO	MÁXIMO
Ductilidad al 25°C, 5 cm/min	cm	D113	100	
Penetración a 25°C, 100g, 5s	mm/10	D5	60	70
Punto de ablandamiento	°C	D36	45	55
Punto de inflamación	°C	D92	232	
Solubilidad en tricloroetileno	g/100g	D2042	99	
Pérdida de masa	g/100g	D2872		1,0
Viscosidad a 60, 80, 100, 135 y 150 °C	cP	ASTM D	Reportar	
		4402		

Fuente: Ecopetrol S.A.

Por otra parte, la U-150 corresponde a una unidad de destilación combinada, por lo tanto, su sección al vacío provee de destilados a la planta de parafinas. En la U-250 se realizó una adaptación en la metalurgia de los sistemas y en equipos, lo cual permite aumentar la temperatura de salida del horno H-253 obteniendo de esta manera mayor recuperación de gasóleos en la torre al vacío y asfaltos con menores valores de penetración; dando como resultado, la no conveniencia en la preparación el asfalto en esta unidad. Por otra parte, la U-2000 se direccionó para la preparación de crudo reconstituido y la U-200 tiene la flexibilidad de operar como viscorreductora y unidad de destilación primaria, por lo que no cuenta con horno para la preparación del cemento asfáltico.

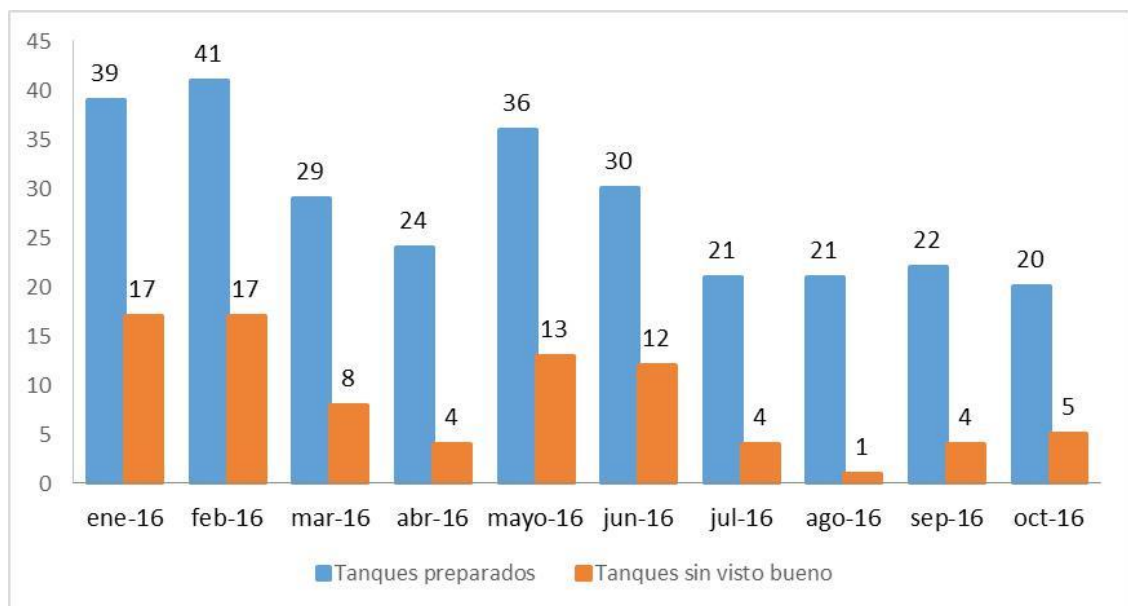
Por lo tanto, la unidad U-2100 y su torre de vacío (T-2103) serán objeto de estudio en el presente trabajo de aplicación. La T-2103 tiene una recuperación aproximadamente del 50% de su carga como gasóleos, el otro 50% corresponde a la producción de cemento asfáltico. Para producir el cemento asfáltico dentro de especificaciones expuestas en la Tabla 1, en la T-2103 se deben manipular las variables como temperatura de salida del horno H-2102, el flujo de vapor al fondo de la torre, la presión de vacío de la torre y los reflujos, entre otras, dependiendo de las características con las que llega el crudo de alimentación a la unidad y el desempeño de la torre atmosférica.

La variación de las dietas de crudo disponibles en la GRB hace que el control automático de la torre deje de ser predictivo, situación que obliga a migrar la operación a control regulatorio; lo que retarda los tiempos de respuesta a perturbaciones en el proceso. Por otra parte, los parámetros de calidad del asfalto son determinados mediante pruebas de laboratorio, donde el lapso que transcurre desde la toma de la muestra hasta el reporte y la disponibilidad de los datos entregados es en promedio de 5 horas, dificultando el control efectivo de las variables de proceso y afectando las especificaciones de calidad del asfalto producido. Cuando se obtiene asfalto con un grado de penetración por fuera de 60/70 se debe realizar el reproceso del producto y corregir la penetración. Si el asfalto reprocesado presenta una penetración inferior a 60 mm/10 se debe corregir mezclándolo con un volumen conveniente de un asfalto con alta penetración; por otra parte, si el asfalto reprocesado presenta una penetración mayor a 70 mm/10 se debe mezclar con un asfalto de baja penetración buscando obtener grados de penetración entre 60-70 mm/10; es importante mencionar que los tanques deben contar con capacidad disponible para realizar este procedimiento.

En la Figura 1 se exhibe el número de tanques de asfalto preparados a ventas y los tanques sin visto bueno durante el periodo enero-octubre de 2016. Según esta

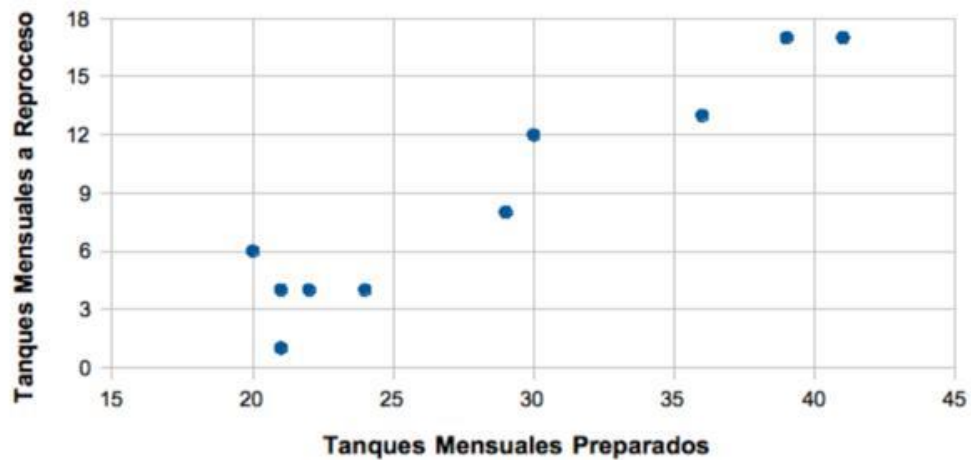
figura, el reproceso en la preparación de tanques de cemento asfáltico para ventas alcanzó aproximadamente el 30% durante el 2016. Asimismo, en la Figura 2 se presenta la tendencia entre la cantidad de tanques preparados por mes y el número de tanques a reproceso durante el año 2016; de esta figura es posible deducir que a mayor número de tanques preparados mayor cantidad de tanques sin visto bueno enviados a reproceso. El costo por reproceso involucra el número de horas hombre, retrasos en la operación y penalidades comerciales relacionadas con entregas de productos a destiempo. Con lo anterior, la disminución en el porcentaje de tanques a reproceso aumentaría positivamente la operación de la U-2100, liberando capacidad de almacenamiento y beneficiando el margen de ganancia grueso de refinación, así como la imagen corporativa de Ecopetrol.

Figura 1. Tanques de asfalto con reproceso.



Fuente: Ecopetrol S.A.

Figura 2. Tendencia tanques a reproceso vs tanques preparados, año 2016.



Fuente: Autor.

La GRB actualmente cuenta con una capacidad de almacenamiento máxima de 32000 barriles de cemento asfáltico, distribuidos en tres tanques: El K-200 con una capacidad máxima de almacenamiento de 10000 barriles, el K-205 y el K-206 con una capacidad máxima de 11000 bls. La demanda actual de asfaltos en el país es de aproximadamente 15000 bls/día. Ante la preparación de un asfalto fuera de especificaciones comerciales, este debe ser sometido al procedimiento de reproceso. Este reproceso envuelve trasegar aproximadamente 2000 bls del asfalto preparado (dependiendo del valor de penetración) a otro tanque que puede ser al K-205 o K-206, según disponibilidad. Simultáneamente, el operador de consola ajusta las condiciones en la torre de vacío para obtener un asfalto corrector con la penetración adecuada de modificación de la preparación; las desiciones de reprocesamiento se apoyan en hojas de cálculos de Excel en donde se estiman las penetraciones de mezcla. Una vez se termina de adicionar el volumen de asfalto corrector se realiza una recirculación en el tanque por una hora, para homogenizar la mezcla; después se toman nuevamente muestras del asfalto para verificar el visto bueno. Todo este proceso puede tardar entre 6 y 8 horas en total. Con lo anterior, un tanque sin visto bueno trae como consecuencia

un problema logístico y económico que causa un impacto negativo en los objetivos diarios de la GRB.

En consecuencia, predecir de forma adecuada la calidad y las características del cemento asfáltico a ventas es una prerrogativa que ofrece una posible solución a los problemas anteriormente expuestos. Por lo tanto, el alcance de este trabajo de aplicación será desarrollar, diseñar y validar un modelo de predicción tipo red neuronal que permita relacionar las variables de control del proceso y la dieta de la carga del crudo, con las especificaciones técnicas del cemento asfáltico.

2. MARCO TEÓRICO

2.1 LA UNIDAD DE DESTILACIÓN PRIMARIA U-2100

Esta unidad fue diseñada por Arthur G. McKee en el año de 1968 y entró en funcionamiento en el año 1971. Actualmente tiene una capacidad para cargar 38000 barriles por día de crudo pesado (21.4°API). La unidad 2100 está compuesta por dos secciones principales. En la primera sección, la atmosférica, se efectúa la destilación fraccionada, la cual extrae los productos livianos de los crudos recibidos. En la segunda sección, el crudo reducido generado en la primera sección es sometido a una destilación al vacío en la torre T-2103 para extraer el producto liviano residual, denominado gasóleo, y generar un fondo o cemento asfáltico.

La torre de vacío T-2103 tiene las siguientes secciones: Una zona flash, una zona de despojo, zona de recicló, zona de gasóleo pesado de vacío (GPV) y zona de gasóleo liviano de vacío (GLV); también cuenta con eyectores para la generación de vacío en la T-2103. El crudo reducido parcialmente vaporizado a 725° F sale del Horno H-2102 y por la línea de transferencia entra a la torre de vacío T-2103 a la altura de la zona de vaporización instantánea (zona flash). La fracción líquida de la carga fluye hacia el fondo de la torre, donde existen 5 platos de despojo y la fracción vaporizada asciende hacia la zona de fraccionamiento, donde se dispone de tres zonas empacadas y tres platos colectores. La condensación de los vapores de hidrocarburo ocurre a medida que la temperatura en la torre se reduce, mediante los reflujos externos e internos, produciéndose corrientes laterales líquidas que se recuperan en los colectores de salida de la torre.

Los circuitos principales que componen la T 2103 son:

El circuito de cima: en este circuito los vapores salen de la torre a 180°F y contienen vapor de agua, gases condensables y no condensables. Los hidrocarburos y el vapor de agua condensado se acumulan y se separan en el tambor de sello D-2104 donde se obtiene como producto gasóleo liviano en proporciones no mayores al 1% respecto a la carga, los no condensables son enviados a la atmósfera o al sistema de tea.

Circuito de Gasóleo Liviano: la corriente de GLV se extrae de la zona superior de la torre de vacío T-2103 y va hacia la succión de las bombas de reflujo de cima P-2109C/D. El gasóleo sale de la bomba, para dividirse en dos circuitos: uno va como reflujo de cima para retornar a la cima de la T-2103 y el segundo circuito va como producto hacia tanques de almacenamiento.

Circuito de Gasóleo Pesado: La corriente de GPV se extrae de la zona media de la torre de vacío T-2103 hacia las bombas de reflujo medio. El líquido descargado por estas bombas se divide en dos circuitos, el primero pasa por el controlador de reflujo caliente FRC2144 y regresa a la torre. La corriente principal de GPV pasa por el lado tubos de los enfriadores de reflujo medio E-2110A/B con crudo, para dividirse en dos corrientes: una corriente va hacia la torre para entrar por encima de la malla de la sección media como reflujo, El resto de GPV que va como subproducto, a tanques de almacenamiento.

Circuito de Reciclo: Debajo de la malla Glitsh se encuentra un colector de reciclo que recoge todo el líquido que baja de la sección de lavado. Este líquido contiene parte del gasóleo que se usó como lavado caliente y contiene además la fracción de fondos que se condensó en la sección de lavado. Se extrae por encima de la zona de vaporización instantánea y se acumula en el tambor de reciclo D-2106. Los vapores o livianos no condensados regresan por la parte superior del D-2106 a la torre de vacío T-2103. El reciclo se extrae con las bombas de reciclo P-2111A/B, la descarga se divide en dos corrientes. La primera se recircula a la torre

de vacío T-2103 como reciclo a la malla. La segunda corriente mantiene el nivel de líquido en el tambor de reciclo D-2106, la cual se envía al tambor de crudo reducido.

Circuito de Fondos: La fracción líquida que llega a la zona de vaporización de la torre de vacío T-2103 cae a través de la sección de despojo de la misma, formada por cinco platos perforados que se numeran de abajo hacia arriba. Debajo del plato 1 se inyecta vapor sobrecalentado del horno H-2102 a 720°F y 50 psig; los vapores ascienden en contracorriente con el producto, despojándolo de las fracciones livianas que arrastró. La cantidad de vapor se controla con el FIC2140.

El residuo de vacío sale por el fondo de la torre por medio de las bombas de fondo SP 2112 C/D; el flujo descargado por estas bombas se divide en tres circuitos. El primer circuito dirige el asfalto caliente a la unidad viscorreductora (cuando se encuentra en servicio). El segundo es utilizado en arrancadas de la unidad como circuito de recirculación al H-2102. El tercero circuito (flujo normal) dirige el fondo a producción de asfalto, después de intercambiar calor con el crudo fresco en el segundo tren de precalentamiento.

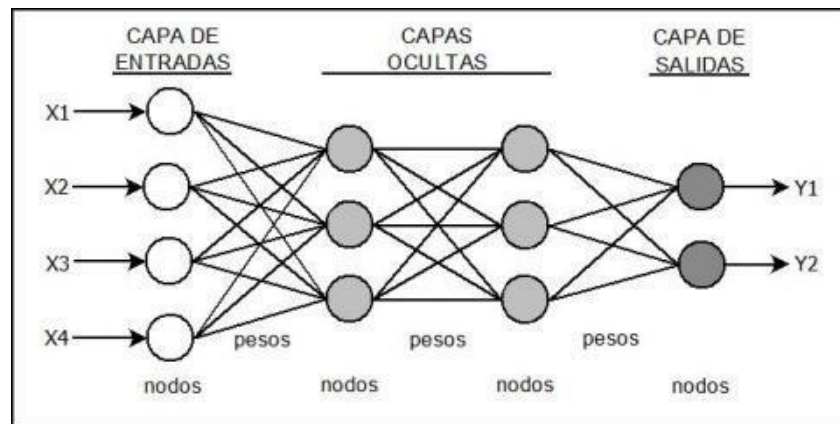
2.2 REDES NEURONALES ARTIFICIALES (RNA)

Una red neuronal artificial es una interconexión masiva de elementos simples en paralelo usualmente adaptativos y con una organización jerárquica definida, que logra emular de manera esquemática la estructura neuronal del cerebro, logrando reproducir sus capacidades en la solución de problemas. Una RNA interactúa con objetos o elementos del entorno o el mundo real de forma similar que lo hace el sistema nervioso biológico, lo cual le permite proporcionar un modelado preciso de un conjunto de datos no lineales útil para la toma decisiones y realización de clasificaciones, con el fin de dar solución a un determinado problema. Los modelos de predicción mediante RNA son útiles para aplicaciones en las cuales

los problemas requieren soluciones con algoritmos complejos y extensos o en aquellos sistemas que no poseen una solución computacional precisa^{1 2}

Un modelo de red neuronal artificial está compuesto por un conjunto de elementos llamados nodos o neuronas los cuales se encuentran unidos entre sí, a través de señales denominadas pesos, siendo estos, la intensidad entre los enlaces sinápticos de las neuronas. Cada neurona posee una función interior que le permite cambiar su nivel de activación, siempre que la cantidad de información entrante logre superar el umbral asignado por la función de activación^{3 4}. Los nodos y pesos están organizados en un conjunto de capas tal como se observa en la Figura 3, las neuronas pertenecientes a la capa de entrada y salida son las encargadas de recibir la información proveniente del exterior y de entregar información arrojada por el sistema, respectivamente^{5 6}

Figura 3. Ejemplo de una estructura de RNA multicapa.



Fuente: Ref. 4.

¹ Hilera, J.R., & Martínez, V.J. (2000). Redes Neuronales Artificiales Fundamentos, Modelos y Aplicaciones. Madrid España: *Alfaomega*.

² Del Brío, B., & Sanz, M.A. (2002). Redes Neuronales y Sistemas Difusos. 2da Ed. México D.F, México: *Alfaomega*.

³ Ponce, C.P. (2010). Inteligencia artificial con aplicaciones a la ingeniería. 1ra Ed. México: *Alfaomega*.

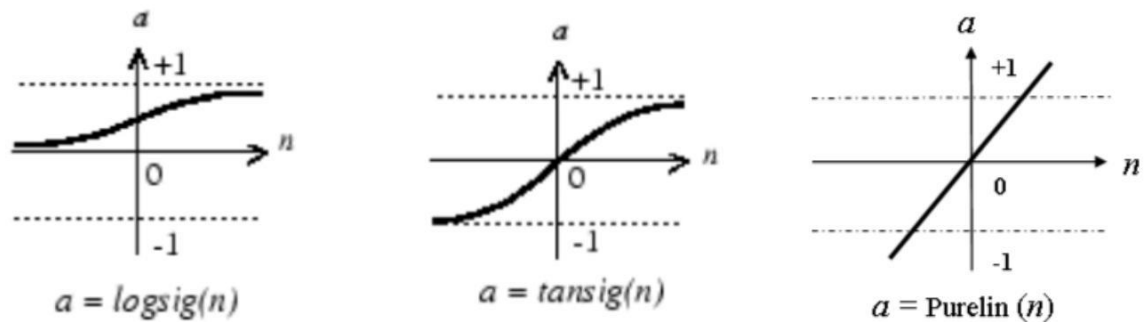
⁴ Isasi, P., Galvan, I. (2004). Redes neuronales artificiales enfoque práctico. Madrid: *Pearson*.

⁵ Hilera, J.R., & Martínez, V.J. (2000). Redes Neuronales Artificiales Fundamentos, Modelos y Aplicaciones. Madrid España: *Alfaomega*.

⁶ Isasi, P., Galvan, I. (2004). Redes neuronales artificiales enfoque práctico. Madrid: *Pearson*.

Por otra parte, las neuronas presentes en la capa oculta pertenecen al conjunto de niveles formados por neuronas que no tienen contacto directo con el entorno exterior y son las encargadas del procesamiento de la información. El número de estas capas puede variar entre cero y un número considerable determinando así las distintas tipologías de las redes neuronales^{7 8}. Cuando determinada neurona recibe una cantidad de información de entrada, cada señal de entrada se multiplica por el peso asociado a cada una de las conexiones. Posteriormente se suma las entradas ponderadas con los pesos y se aplica la función de activación para generar la señal de salida. Estas funciones de activación deben ser continuas y derivables en todos los puntos; en la Figura 4 se encuentran las funciones más utilizadas^{9 10}.

Figura 4. Funciones de activación habitualmente utilizadas en RNA.



Fuente: Del Brío & Sanz, 2002.

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

⁷ Ponce, C.P. (2010). Inteligencia artificial con aplicaciones a la ingeniería. 1ra Ed. México: Alfaomega.

⁸ Isasi, P., Galvan, I. (2004). Redes neuronales artificiales enfoque práctico. Madrid: Pearson.

⁹ Del Brío, B., & Sanz, M.A. (2002). Redes Neuronales y Sistemas Difusos. 2da Ed. México D.F, México: Alfaomega.

¹⁰ Isasi, P., Galvan, I. (2004). Redes neuronales artificiales enfoque práctico. Madrid: Pearson.

más utilizado, ya que presenta información sobre la sensibilidad de las predicciones de la red respecto a los pesos de las señales; el algoritmo actualiza los pesos con base en el gradiente de una función de costo –que indica que tan eficiente es la predicción, por ejemplo, el error cuadrático medio–. Las ecuaciones derivadas del algoritmo backpropagation corresponden a: una ecuación de error en la capa de entrada, una ecuación de error en términos de los errores de la capa anterior y dos ecuaciones diferenciales para el cambio de la función de costo (respecto a los pesos y a los sesgos)^{11 12}. Las funciones son resueltas enmarcadas en la minimización de la función de costo, por medio de un procedimiento como el método de Levenberg-Marquardt.

2.3 APLICACIONES DE LAS RNA EN PREDICCIONES REFERENTES AL ASFALTO

En las últimas décadas, la comunidad científica ha desarrollado modelos de redes neuronales artificiales para predecir las especificaciones de las emulsiones asfálticas. Tarefder *et al.*¹³ realizaron una investigación donde propusieron una red neuronal artificial de cuatro capas ocultas para construir un mapa de diseño de mezcla; el entrenamiento de la red se realizó con más de 50 diferentes tipos de mezclas; la red creada mostró una excelente correlación con los datos de laboratorio y funcionó como una herramienta útil para la construcción y mantenimiento de pavimentos de asfalto. Por otra parte, en el campo de la emulsificación de asfalto Ozsahin y Oruc¹⁴, mostraron el uso potencial de una red neuronal *feedforward* de tres capas con el fin de predecir con suficiente proximidad la relación entre el módulo elástico de los asfaltos y nivel de adición de cemento,

¹¹ Ponce, C.P. (2010). Inteligencia artificial con aplicaciones a la ingeniería. 1ra Ed. México: Alfaomega.

¹² Isasi, P., Galvan, I. (2004). Redes neuronales artificiales enfoque práctico. Madrid: Pearson.

¹³ Tarefder, R.A.; Ahsan, S. (2014). Neural network modelling of asphalt adhesion determined by AFM. *Journal of microscopy*, 254, p. 31-41.

¹⁴ Ozsahin, T.S.; Oruc, S. (2008). Neural network model for resilient modulus of emulsified asphalt mixtures. *Construction and Building Materials*, 22, p. 1436-1445.

tiempo de curación y contenido residual de asfalto. Por otra parte, Ozgan¹⁵ presentó el ajuste a través de inteligencia artificial de la estabilidad de Marshall del asfalto bajo diferentes condiciones de temperatura, tiempo de exposición y propiedades físicas; en este trabajo se encontraron las óptimas condiciones que conllevan a la estabilidad de Marshall. Oztukur *et al.*¹⁶, publicaron un modelo de RNA de tres capas “*backpropagation*” para el diseño volumétrico a través del método “*Superpave*” utilizando como datos de entrada la graduación de la mezcla, gravedad específica de los agregados, grados de desempeño de los agregados y contenido de agregados en la mezcla. En el entrenamiento de la red se utilizaron 1800 diseños de mezcla. Con el fin de determinar el daño por humedad del asfalto, Tarefder *et al.*¹⁷ Construyeron una red neuronal para cuantificar datos de adhesión de fuerzas atómicas microscópicas del asfalto, donde tuvo en cuenta la variación y porcentaje de polímero utilizado para el entrenamiento de la misma. La arquitectura de la red consta de una capa oculta con nueve nodos de entrada y cinco nodos de salida; los resultados del modelo aplicado mostraron una excelente correlación con los datos experimentales.

Nejad *et al.*¹⁸, modelaron la relación de la resistencia de Marshall en función de las características físicas del asfalto a través de inteligencia artificial (RNA). Las características físicas utilizadas fueron el porcentaje de bitumen, porcentaje de agregados y la dirección de los agregados. El modelo investigado por los autores provee una solución eficiente para la predicción de la resistencia de Marshall. Así mismo, Shafabakhsh *et al.*¹⁹, desarrolló un modelo matemático RNA para predecir

¹⁵ Ozgan, E. (2011). Artificial neural network based modelling of the Marshall Stability of asphalt concrete. *Expert Systems with Applications*, 38, p. 6025-6030.

¹⁶ Ozturk, H.I.; Kutay, M.E. (2014). An artificial neural network model for virtual Superpave asphalt mixture design. *International Journal of Pavement Engineering*, 15, p. 151-162.

¹⁷ Tarefder, R.A.; Ahsan, S. (2014). Neural network modelling of asphalt adhesion determined by AFM. *Journal of microscopy*, 254, p. 31-41.

¹⁸ Nejad, F.M.; Mehrabi, A.; Zakeri, H. (2015). Prediction of Asphalt Mixture Resistance Using Neural Network via Laboratorial X-ray Images. *Journal of Industrial and Intelligent Information*, 3, 1.

¹⁹ Shafabakhsh, G.H.; Ani, O.J.; Talebsafa, M. (2015). Artificial neural network modeling (ANN) for predicting rutting performance of nano-modified hot-mix asphalt mixtures containing steel slag aggregates. *Construction and Building Materials*, 85, p. 136-143.

la calidad final de una mezcla de asfalto en función de los tipos de agregados, tipos y porcentajes de aditivos, temperatura de preparación y esfuerzo. Es importante mencionar que, en las diferentes publicaciones científicas consultadas, no se encontraron aplicaciones de redes neuronales artificiales para la elaboración de cementos asfálticos a partir de datos históricos de proceso. En este sentido, los sensores virtuales (*soft sensors*) o “estimadores de estado”, aplican modelos matemáticos creados a partir de datos experimentales y datos de planta, para predecir variables de proceso que no cuentan con su respectivo medidor físico; una de las ventajas en su funcionamiento es el de requerir múltiples entradas, con lo cual la probabilidad de error en sus predicciones es en ciertos casos menor que la de un sensor real^{20 21}. Los modelos más utilizados para calibrar los *soft sensors* corresponden a los obtenidos a partir de redes neuronales artificiales (RNA), debido a su alto desempeño y exactitud en diferentes áreas como el control de procesos. El uso de los sensores virtuales en la predicción de la precipitación de asfáltenos, estimación continua de la calidad de productos, procesamiento de datos en línea y optimización del uso de recursos, hacen que sean atractivos para aplicaciones industriales²².

Un ejemplo de esto, es el proceso de craqueo catalítico (FCC). Este presenta un comportamiento dinámico fuertemente no lineal y altamente interactivo, así como una cinética de reacción poco comprendida con muchas variables operativas no medibles que desempeñan papeles importantes en la determinación del rendimiento y dan lugar a dificultades en el control. Los *soft sensors*, que utilizan mediciones de temperatura, presión y caudal fácilmente medidos y continuamente disponibles de los servidores, junto con modelos de sistemas que utilizan ecuaciones de conservación para asimilar parámetros de aprendizaje en línea,

²⁰ Ponce, C.P. (2010). Inteligencia artificial con aplicaciones a la ingeniería. 1ra Ed. México: Alfaomega.

²¹ Rovert, R. (2002). Neural virtual sensor for the inferencial prediction of product quality from process variables. *Universitat Rovira i Virgili*.

²² Fortuna, L., et al. (2007). Soft sensors for monitoring and control of industrial processes. *Springer Science & Business Media*.

son esenciales para el control avanzado. Yang *et al.*²³ Publicaron la aplicación de una estrategia de control predictivo utilizando dos *soft sensors* en un proceso industrial FCC existente. Por su parte, Bolf *et al.*²⁴, desarrollaron sensores para la predicción del punto final del queroseno basados en mediciones de temperatura y flujo. Los autores de manera general, muestran en sus conclusiones la posibilidad de aplicar *soft sensors* para la estimación de la calidad de los productos de refinería y el control inferencial como una alternativa para analizadores de procesos y ensayos de laboratorio.

Según la revisión anterior, las RNA han sido aplicadas en la predicción de diversas propiedades de flujos a partir de las mediciones de las variables históricas en diferentes procesos de la industria del petróleo, demostrando la utilidad del análisis de los desempeños de las unidades de proceso; estas RNA han sido capitalizadas en el desarrollo de lo que se ha denominado como *soft sensors* o sensores virtuales en línea. Con esto, el desarrollo de un protocolo basado en el análisis de diferentes arquitecturas RNA puede conducir a la predicción de la penetración del asfalto generado en la U-2100, basada en datos históricos de la unidad.

²³ Yang, S.H., et al. (1998). Soft sensor based predictive control of industrial fluid catalytic cracking processes. *Chemical Engineering Research and Design*, 76, p. 499-508.

²⁴ Bolf, N.; Ivandic, M.; Galinec, G. (2008). Soft sensors for crude distillation unit product properties estimation and control. *Control and Automation*, 16th Mediterranean Conference IEEE, p. 1804-1809.

3. OBJETIVOS

3.1 OBJETIVO GENERAL

Desarrollar un modelo de red neuronal artificial válido para la predicción de la penetración del asfalto producido en la T-2103 a partir de datos históricos de proceso de la unidad U-2100 de la Gerencia Refinería Barrancabermeja.

3.2 OBJETIVOS ESPECIFICOS

Generar una base de datos históricos depurada con las condiciones operativas de la U-2100 y las especificaciones de calidad del asfalto producto.

Determinar diferentes esquemas operativos de la unidad U-2100 según su relación con los desempeños de los equipos y el valor de la penetración del asfalto producido.

Evaluar los desempeños en entrenamiento y validación de diferentes arquitecturas de RNA tipo *feedforward backpropagation* en la predicción de la penetración a partir de los datos históricos.

4. METODOLOGÍA

La siguiente figura resume las etapas seguidas para el desarrollo del presente documento.

Figura 5. Etapas de la metodología.

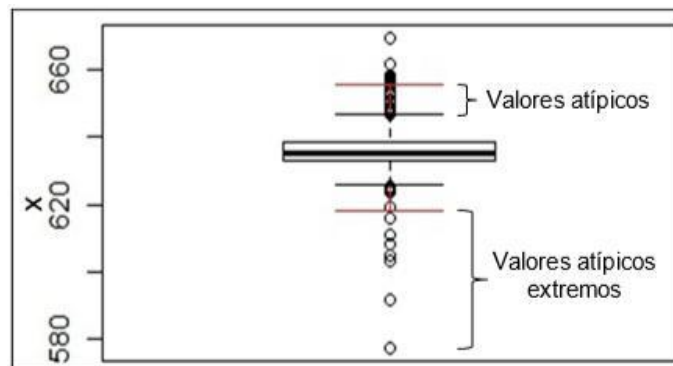


Etapa I. Recolección y depuración datos de proceso: Los datos históricos de las variables de operación en la producción de cemento asfáltico en la unidad U-2100 fueron recolectados desde enero de 2015 hasta enero de 2018. Las variables recolectadas fueron API del crudo de carga a la U-2100, la temperatura de salida del horno H-2102, flujo de vapor en el fondo de la torre T-2103, presión de vacío de la torre T-2103, reflujos, producción total de gasóleos y el grado de penetración del cemento asfáltico, entre otras. Los datos fueron tomados del programa PI Data Link, mientras que los valores de la calidad del asfalto (grado de penetración) fueron obtenidos de los servidores del software de gestión de datos de laboratorio, CDLAB. Los históricos recolectados fueron depurados, con

lo cual, muestras con valores vacíos o con errores fueron descartadas de la matriz de datos.

Etapa II. Análisis estadístico univariado y bivariado: En esta etapa, la estadística univariada, según códigos del software libre R, fue aplicada para determinar los valores máximos y mínimos, valores promedio y sus desviaciones estándar; estos estadísticos ayudaron en la identificación de muestras con respuestas de sensores descalibrados o fuera de línea en la matriz de datos. Asimismo, los datos atípicos desde un punto de vista univariado fueron eliminados de los históricos. Para esto, el rango intercuartil (RI) fue utilizado para la identificación de los atípicos; la Figura 6 ilustra el gráfico *boxplot* derivado del rango intercuartil. Los datos atípicos reportan valores inferiores a $Q1 - 1.5*RI$, o valores superiores a $Q3 + 1.5*RI$, donde $Q1$ y $Q3$ corresponden al primer cuartil y al tercer cuartil, respectivamente²⁵.

Figura 6. Detección valores atípicos por medio del gráfico *boxplot*.



Posteriormente, un análisis de estadística bivariada, según R, fue aplicado para la determinación de los coeficientes de correlación por pares de variables y, con

²⁵ Hair, J.J., Black, W., Babin, B., Anderson, R. (2009). Multivariate Data Analysis. 7th edition. USA: Prentice Hall.

estos, las variables dependientes fueron detectadas; estos coeficientes de correlación fueron ordenados para formar la matriz de correlación²⁶

Etapa III: Identificación de esquemas operacionales: En esta etapa, el método de detección de agrupamientos denominado *kmean* fue aplicado en la identificación de posibles esquemas operativos subyacentes en los datos históricos de la unidad. El método *kmean* divide los datos en *k* agrupaciones de tal forma que la suma de las distancias entre los datos de las agrupaciones sea máxima, mientras las distancias internas (de los datos a su centro en una agrupación) sean mínimas²⁷. Una de las distancias (*d*) más utilizadas corresponde a la euclidiana, la cual se encuentra definida según,

$$d = \left(\sum_{j=1}^k \sum_{i \in C_j} (x_i - z_j)^2 \right)^{1/2}$$

Donde, *c_j* es la *j*-ésima agrupación, *z_j* es el centroide de la *j*-ésima agrupación, *x_j* es el dato de la *i*-ésima muestra. Las *k* agrupaciones son obtenidas en un algoritmo iterativo con un punto de *k* centroides como partida. A partir de este punto, las distancias entre las agrupaciones y las distancias internas de las muestras a los centroides son determinadas; los nuevos centroides son determinados como el punto central de las agrupaciones. El proceso es repetido hasta un número determinado de iteraciones o una tolerancia definida. El número de las agrupaciones adecuadas para los históricos de la U-2100 fue definido por medio del coeficiente silueta, *Si (silhouette coefficient)*^{28 29}; los *k* agrupamientos

²⁶ Hair, J.J., Black, W., Babin, B., Anderson, R. (2009). *Multivariate Data Analysis*. 7th edition. USA: *Prentice Hall*.

²⁷ Thakare, Y.S. & Bagal, S.B. 2015. Performance Evaluation of K-means Clustering Algorithm with Various Distance Metrics. *Int. J. Compt. App.*, 110, 12-16.

²⁸ Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. comp. App. Math.*, 20, 53-65.

con mayor valor en el *Si* promedio fueron propuestos como esquemas operacionales para la U-2100. La verificación de los esquemas operativos fue efectuada por medio de la prueba F y el análisis de varianza *one-way ANOVA*³⁰.

Etapa IV. Análisis de la regresión lineal múltiple: Seguidamente, la regresión lineal múltiple (RLM) fue aplicada como proceso exploratorio previo al desarrollo de las RNA; un rendimiento bajo de la RLM en el ajuste de la penetración del asfalto de la unidad justifica el desarrollo de una RNA; diferentes reportes de literatura verifican la utilidad del desarrollo de RNAs por medio de los desempeños bajos de los modelos RLM en el sistema en análisis^{31 32}. La RLM fue aplicada según los códigos del programa de uso libre R (<https://www.r-project.org>)³³. Para la identificación del mejor modelo, diferentes regresiones RLM fueron aplicadas en el marco del método *stepwise* utilizando los datos históricos para la predicción de la penetración del asfalto generado; la función de R utilizada fue *stepwise* con la dirección de búsqueda 'backward/forward' y el criterio de información a Akaike (AIC); el modelo RLM con el menor valor de AIC fue propuesto como la mejor regresión y la participación de las muestras en este modelo fue analizada por

²⁹ Rao, S.G. & Govardhan. (2015). Performance Validation of the Modified K-means Clustering Algorithm cluster Data. *International Journal of Scientific and Engineering Research*, 6, 726-730.

³⁰ Hair, J.J., Black, W., Babin, B., Anderson, R. (2009). *Multivariate Data Analysis*. 7th edition. USA: *Prentice Hall*.

³¹ Piloto-Rodríguez, R., Sánchez-Borroto, Y., Lapuerta, M., Goyos-Pérez, L. 2013. Prediction of the cetane number of biodiesel using artificial neural networks and multiple linear regression. *Energy Conversion and Management*, 65, 255–261.

³² Rogina, A., Šiško, I., Mohler, I., Ujević, Ž., Bolf, N. 2011. Soft sensor for continuous product quality estimation (in crude distillation unit). *Chem. Eng. Res. Des.*, 89, 2070-2077.

³³ El programa R es un lenguaje y un ambiente para cálculos y gráficas estadísticas, que se encuentra diseñado sobre el sistema operativo GNU (<http://www.gnu.org>). R fue desarrollado en los laboratorios Bell (anteriormente AT&T), bajo la filosofía de desarrollo de librerías y funciones que puedan ser compartidas por usuarios a nivel mundial. Diferentes procedimientos de análisis y procesamiento de datos pueden ser ejecutados en R; entre estos procedimientos se tienen, análisis estadísticos univariados, análisis estadísticos multivariados, regresiones múltiples lineales, regresiones logísticas, regresiones PLS, regresiones PCR, ajustes por RNA y análisis de agrupaciones por kmean, entre otros. Existen extensiones o paquetes desarrollados por programadores de R a nivel mundial, los cuales pueden ser descargados e instalados bajo la filosofía del software libre. Asimismo, los desarrollos y las aplicaciones realizadas con base en R pueden ser publicadas en la web denominada como The R Journal. También, los usuarios de R cuentan con libros y sitios web en donde pueden detallar rutinas, desarrollos y aplicaciones de este programa; los interesados en profundizar y aplicar el programa R cuentan con cursos libres en diferentes páginas de universidades e institutos de educación.

medio de los gráficos de influencia; estos gráficos presentan los estadísticos: residuales estudentizados, valores de la diagonal principal de la matriz Hat, y la distancia de Cook, indicando las posibles muestras atípicas e influyentes^{34 35}.

Etapa V: Desarrollo código de uso función *neuralnet* en R: En esta etapa se codificaron en el programa de uso libre R (<https://www.r-project.org>) las líneas de código requeridas para la aplicación de la función *neuralnet* del paquete *Neuralnet*^{36 37}. En las opciones de esta función se establecieron el algoritmo de entrenamiento, la función de error, las funciones de activación (LogSig o TanH), el número de neuronas internas y el número de capas ocultas. En el entrenamiento se utilizó el algoritmo de aprendizaje *Backpropagation* con el método de optimización de *Levenberg-Marquardt*, ya que es el más rápido para este algoritmo de aprendizaje^{38 39}. Como función de errores se utilizó la suma del error cuadrático medio (MSE: *mean squared error*), según la expresión

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2$$

Donde, y_i , \tilde{y}_i , n corresponden al valor experimental, al valor predicho y al número de datos, respectivamente. Asimismo, la función *neuralnet* utiliza un método de regularización que aplica y detiene el entrenamiento de las RNA en el punto

³⁴ Piloto-Rodríguez, R., Sánchez-Borroto, Y., Lapuerta, M., Goyos-Pérez, L. 2013. Prediction of the cetane number of biodiesel using artificial neural networks and multiple linear regression. *Energy Conversion and Management*, 65, 255–261.

³⁵ Rogina, A., Šiško, I., Mohler, I., Ujević, Ž., Bolf, N. 2011. Soft sensor for continuous product quality estimation (in crude distillation unit). *Chem. Eng. Res. Des.*, 89, 2070-2077.

³⁶ Zhang, Z. (2016). Neural Networks: further insights into error function, generalized weights and others. *Ann Transl Med*, 4, 300.

³⁷ Blanco, M., Maspoch, S., Villarroya, I., Peralta, X., González, J.M., Torres, J. (2000). Determination of the penetration value of bitumens by near infrared spectroscopy. *Analyst*, 125, 1823–1828.

³⁸ Swingler, K. (1996). Applying Neural Networks, A Practical Guide. *Press Limited Oval Road London NW1*.

³⁹ Flórez, R., Fernández, J. (2008). Las Redes Neuronales Artificiales. Oleiros, La Coruña: *Netbiblo*.

mínimo de la curva de aprendizaje para evitar el sobreajuste (*early stopping*)^{40 41}. En el anexo 2 se detalla el código de R desarrollado, en el marco de la aplicación de la función *neuralnet* con los datos históricos de la U-2100.

Etapa VI. Entrenamiento y validación de estructuras RNA: En esta etapa se entrenaron y validaron diferentes estructuras de red. El entrenamiento y la validación de las estructuras de RNA envolvió la variación del número de capas, el número de neuronas por capa y las funciones de activación, según lo recomendado en la literatura^{42 43}. En la mayoría de los casos una única capa oculta resulta suficiente para obtener errores bajos^{44 45}, sin embargo, estructuras RNA con dos capas ocultas fueron exploradas en busca de mejores desempeños. El número de neuronas de la capa oculta fue modificado desde 1 hasta el doble del número de variables de entrada^{46 47}.

Para el entrenamiento y la validación de las redes, la base de datos fue dividida en dos conjuntos, el 70% de los datos se destinará al entrenamiento y el 30% restante a la validación; el conjunto de variables de entrada descartó las variables con coeficiente de correlación elevados (Etapa II). El entrenamiento de las RNA ejecutó simultáneamente un procedimiento de validación de la estructura de red utilizando el criterio de evaluación MSE. Los desempeños de las estructuras de

⁴⁰ Zhang, Z. (2016). Neural Networks: further insights into error function, generalized weights and others. *Ann Transl Med*, 4, 300.

⁴¹ Berthold, M. & Hand, D.J. (2007). Intelligent Data Analysis: An Introduction. *Springer-Verlag*, 2nd edition, Berlin.

⁴² Rao, S.G. & Govardhan. (2015). Performance Validation of the Modified K-means Clustering Algorithm cluster Data. *International Journal of Scientific and Engineering Research*, 6, 726-730.

⁴³ Piloto-Rodríguez, R., Sánchez-Borroto, Y., Lapuerta, M., Goyos-Pérez, L. 2013. Prediction of the cetane number of biodiesel using artificial neural networks and multiple linear regression. *Energy Conversion and Management*, 65, 255–261.

⁴⁴Ibíd.

⁴⁵ Liu, G., Wang, L., Qu, H., Shen, H., Zhang, X., Zhang, S., Mi, Z. 2007. Artificial neural network approaches on composition–property relationships of jet fuels based on GC–MS. *Fuel*, 86, 2551-2559.

⁴⁶ Rogina, A., Šiško, I., Mohler, I., Ujević, Ž., Bolf, N. 2011. Soft sensor for continuous product quality estimation (in crude distillation unit). *Chem. Eng. Res. Des.*, 89, 2070-2077.

⁴⁷ Berry, M.J.A., Linoff, G. (1987). Data mining techniques. *John Wiley & Sons*. Nueva York.

red fueron determinados con los estadísticos MSE y R2; estos estadísticos fueron promediados evaluando las etapas de entrenamiento y validación 20 veces por cada red, como es recomendado en la literatura⁴⁸; esto dada la característica aleatoria de la selección de los conjuntos y el algoritmo de aprendizaje utilizado.

Etapa VII. Análisis de sensibilidad de las redes validadas: En esta etapa final, las estructuras de red con los mejores desempeños según los estadísticos fueron sometidas a un análisis de sensibilidad para determinar la influencia de las variables de entrada en la predicción de la penetración. Este análisis aplicó el procedimiento *input perturbation*⁴⁹, el cual aplica perturbaciones del 50% del valor de cada entrada manteniendo las demás con los valores iniciales; ante la perturbación de cada variable, los estadísticos son recalculados por medio de la ejecución de las respectivas redes. Las entradas más influyentes corresponden a aquellas que alteran en mayor grado los valores de los estadísticos comparados con los desempeños de las redes originales.

⁴⁸ Swingler, K. (1996). *Applying Neural Networks, A Practical Guide*. Press Limited Oval Road London NW1.

⁴⁹ Gevrey, M.; Dimopoulos, I.; Lek, S. (2003). Review and comparison of methods to study to contribution of variables in artificial neural network models. *Ecological Modelling*, 160, 349-264

5. RESULTADOS

5.1 DATOS HISTÓRICOS

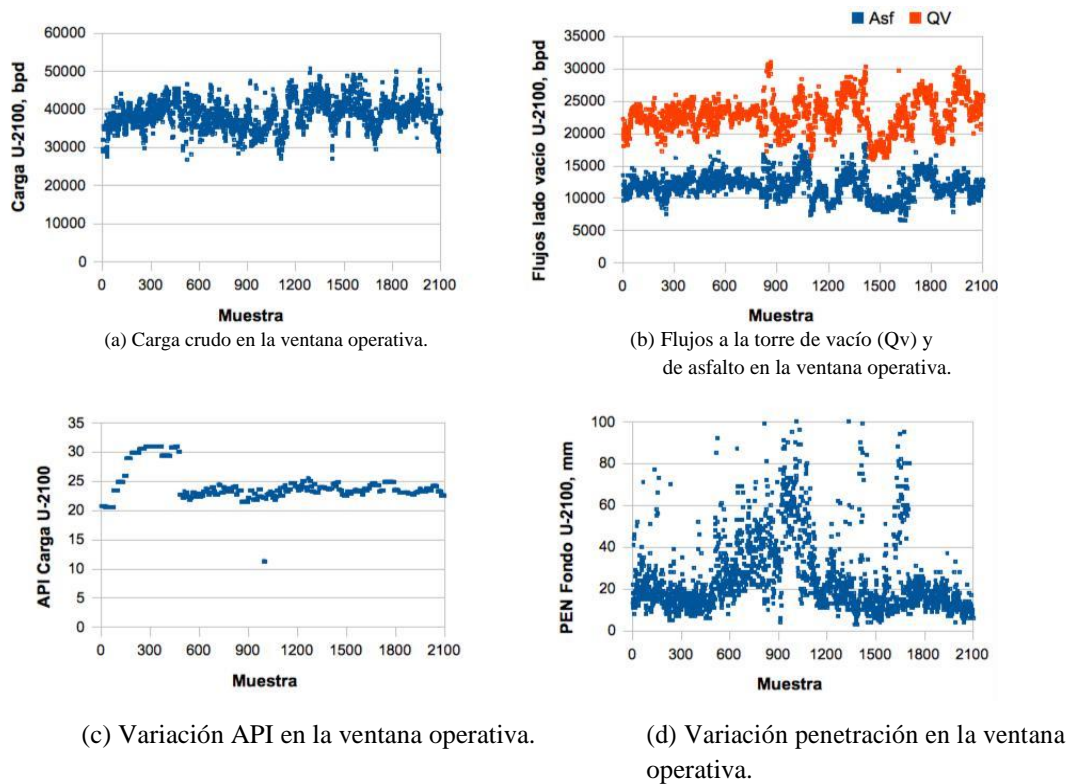
La ventana de operación recolectada corresponde al periodo entre enero 2015 y enero 2018, con lo cual se tienen 2910 muestras, 35 variables operacionales y 2 propiedades (API crudo carga y PEN asfalto de salida de la unidad). Muestras con valores “No Good Data For Calculation”, con valores vacíos o con 0 y muestras con incumplimiento del balance de masa fueron excluidas del conjunto de muestras recolectadas; los valores anteriores reportados por los históricos representan fallas y falta de calibración de los sensores respectivos. Con lo anterior, la base de datos fue reducida a 2175 muestras. Seguidamente, los valores atípicos exhibidos por cada variable según el rango intercuartíl fueron removidos para acercarse a un estado pseudoestacionario univariado (Figura 7). El estado pseudoestacionario para la variable flujo de alimentación a la unidad fue verificado por medio de los valores para la curtosis y el sesgo de -0.18 y 0.05, respectivamente, lo cual presenta una distribución cercana a la gaussiana (curtosis 0 y sesgo 0); asimismo, las muestras con valores de penetración por encima de 100 mm/10 fueron considerados como atípicas y por lo tanto descartadas. Con lo anterior, el número de muestras fue reducido a 2119.

Por la parte bivariada, la matriz de correlación mostrada en la Figura 8 presenta dependencias entre diferentes pares de variables que conforman la base de datos históricos (en el anexo 1 se especifican las abreviaciones para cada variable); estas dependencias (coeficiente de Pearson con valores mayores a 0.65, $r > 0.65$), son indicadas con círculos rojos. Las dependencias se presentan entre: la temperatura de crudo de precaliente (Tpre_A) y otras seis variables, el refluo de cima de la torre de vacío (RefCim_V) y la temperatura de la zona de flash de la torre atmosférica (Tflash_A), la temperatura de la zona de flash de la torre de vacío (Tflash_V) y la temperatura del flujo de crudo reducido (Tcred). También, el

reflujo inferior de la torre atmosférica (RI_A), la temperatura del asfalto (Tasf), la temperatura del acpm (Tacpm_A), la temperatura del gasóleo pesado de vacío (TGP_V) y la temperatura de salida del crudo del horno atmosférico (Tsal_HA) exhiben correlación con otras tres variables (Tcred, Tacmp_A, Tflash_A).

Las dependencias importantes encontradas en la matriz de correlación conducen a la reducción en el número de variables que describen al sistema; esta reducción en el número de variables es tomada con base en valores de [25, 26, 28]. Las variables Pcima_A, Pflash_A, Tflash_A, Tcred y Tsal_HA exhiben dependencias importantes con Tflash_A, Pcima_A, Tpre_A, Tsal_HA, Tcred, RefCim_V, Tflash_V, TGP_V, Tasf y Asf, por lo cual estas últimas variables deben ser omitidas en los modelos de redes neuronales */r/ 0.80* artificiales con la base de datos de la U-2100.

Figura 7. Tendencia para algunas variables en las muestras.



5.2 IDENTIFICACIÓN DE MODOS O ESQUEMAS OPERACIONALES

La técnica de agrupaciones denominada *kmeans* fue aplicada considerando la distancia euclidiana (raíz de la suma de las distancias al cuadrado)⁵⁰. La técnica de agrupaciones fue aplicada para definir posibles esquemas operacionales de la U-2100, subyacentes en los datos históricos. La Tabla 2 muestra los resultados de la aplicación de *kmeans* a los datos históricos autoescalados considerando números de clústeres entre 2 y 6. Según los resultados de la Tabla 1, el aumento en el número de clústeres disminuye la suma total de cuadrados (STC) dentro de los clústeres, mientras que aumenta la STC entre los clústeres definidos; la Figura 9 presenta la variación de la STC dentro de los clústeres para diferente número de agrupaciones. El número de clústeres sugerido para los datos históricos es aquel que presenta la menor STC interna (compactación) y la mayor STC entre los clústeres (separación)⁵¹. La obtención del número de agrupaciones adecuadas se efectuó por medio del método del coeficiente silueta, *Si* (*silhouette coefficient*), el cual determina si las muestras son agrupadas en una forma acertada^{52 53}; una muestra es agrupada acertadamente si *Si* es cercano a la unidad; por otra parte, un valor de *Si* cercano a 0 indica una muestra perteneciente a varios clústeres; un valor de *Si* negativo indica una agrupación errónea para la muestra respectiva⁵⁴. La Tabla 1 presenta en su última columna el valor de los *Si* promedio para los diferentes clústeres, mientras que la Figura 10 ilustra los diferentes valores de los *Si* para las muestras según los clústeres definidos; el eje horizontal en las Figuras 10a-c hace referencia al valor del *Si*, mientras que en el eje vertical se ubican, con diferente coloración, las diferentes agrupaciones obtenidas de la aplicación del método *kmean*. Estas figuras corresponden a un gráfico de frecuencia según el

⁵⁰ Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. comp. App. Math.*, 20, 53-65.

⁵¹ *Ibíd.*

⁵² *Ibíd.*

⁵³ Rao, S.G. & Govardhan. (2015). Performance Validation of the Modified K-means Clustering Algorithm cluster Data. *International Journal of Scientific and Engineering Research*, 6, 726-730.

⁵⁴ Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. comp. App. Math.*, 20, 53-65.

valor de S para cada muestra en cada agrupación. En la parte derecha de las Figuras 10a-c se indica el número de muestras constituyentes de cada agrupación, así como el correspondiente valor del coeficiente silueta promedio. Los valores de los S_i promedio (Figura 10d) indican que el número de agrupaciones adecuado para definir los esquemas operacionales de la U-2100 corresponde a 2. Por otra parte, los valores bajos para los S_i pueden deberse a muestras que representan operación en estado dinámico y a las fluctuaciones respecto a los crudos cargados a la unidad (las propiedades de los crudos se actualizan semanalmente).

Tabla 2. Resultados del método *kmean*. STC: Suma total de cuadrados.

Clústeres	STC dentro de los clústeres	STC entre los clústeres	S_i promedio*
2	57128	19119	0.40
3	51723	24524	0.34
4	46606	29641	0.35
5	42504	33743	0.37
6	39583	36664	0.30

*Coeficiente silueta

Figura 8. Matriz de correlación de Pearson. En círculos rojos se indican dependencias importantes entre pares de variables (correlaciones con valor mayor a 0.65). Las abreviaciones de las variables son definidas en el anexo 1.

	ACPM_A	API	Asf	Cred	GL_V	GO_A	GP_V	Jet_A	Nafta_A	Pcima_A	Pcima_V	PEN	Pflash_A	QA	QV	RC_A	RefCal_V	RefCim_V	RefMed_V	RI_A	RM_A
ACPM_A	1.00	0.10	-0.02	-0.05	0.32	-0.36	0.05	0.12	0.11	-0.35	0.20	-0.27	-0.23	0.38	0.14	0.57	0.03	0.36	-0.36	0.36	-0.26
API	0.10	1.00	-0.01	-0.06	-0.08	0.11	0.09	0.04	-0.23	0.06	0.35	-0.27	0.00	0.14	-0.04	-0.17	-0.22	-0.20	0.11	0.08	0.36
Asf	-0.02	-0.01	1.00	0.23	0.08	0.33	0.16	0.05	-0.12	0.20	0.00	0.30	0.22	0.09	0.71	-0.05	0.11	-0.30	0.38	-0.15	0.20
Cred	-0.05	-0.06	0.23	1.00	0.45	-0.12	-0.32	0.15	0.28	0.15	0.15	0.05	0.17	0.07	0.22	0.13	0.27	0.18	-0.05	0.11	-0.01
GL_V	0.32	-0.08	0.08	0.45	1.00	-0.35	-0.24	0.37	0.29	-0.27	0.09	-0.25	-0.22	0.30	0.35	0.43	0.57	0.46	-0.17	0.50	-0.34
GO_A	-0.36	0.11	0.33	-0.12	-0.35	1.00	0.09	-0.11	-0.26	0.55	0.06	0.26	0.52	-0.10	0.05	-0.38	-0.44	-0.63	0.30	-0.36	0.59
GP_V	0.05	0.09	0.16	-0.32	-0.24	0.09	1.00	-0.04	-0.23	0.07	-0.26	-0.14	0.03	0.07	0.46	-0.06	-0.04	-0.24	0.52	-0.01	0.12
Jet_A	0.12	0.04	0.05	0.15	0.37	-0.11	-0.04	1.00	0.08	0.01	0.04	-0.15	0.04	0.18	0.17	0.31	0.24	0.06	0.08	0.16	0.00
Nafta_A	0.11	-0.23	-0.12	0.28	0.29	-0.26	-0.23	0.08	1.00	-0.09	-0.01	-0.05	0.00	0.33	-0.05	0.34	0.22	0.45	-0.35	0.33	-0.27
Pcima_A	-0.35	0.06	0.20	0.15	-0.27	0.55	0.07	0.01	-0.09	1.00	-0.03	0.30	0.96	-0.12	0.00	-0.20	-0.42	-0.57	0.29	-0.46	0.54
Pcima_V	0.20	0.35	0.00	0.15	0.09	0.06	-0.26	0.04	-0.01	-0.03	1.00	-0.13	-0.02	0.20	-0.03	0.12	-0.12	0.23	-0.39	0.24	0.00
PEN	-0.27	-0.27	0.30	0.05	-0.25	0.26	-0.14	-0.15	-0.05	0.30	-0.13	1.00	0.31	-0.35	0.08	-0.20	-0.12	-0.25	0.03	-0.33	0.18
Pflash_A	-0.23	0.00	0.22	0.17	-0.22	0.52	0.03	0.04	0.00	0.96	-0.02	0.31	1.00	-0.04	0.03	-0.05	-0.41	-0.52	0.21	-0.43	0.48
QA	0.38	0.14	0.09	0.07	0.30	-0.10	0.07	0.18	0.33	-0.12	0.20	-0.35	-0.04	1.00	0.16	0.32	0.05	0.28	-0.17	0.41	-0.05
QV	0.14	-0.04	0.71	0.22	0.35	0.05	0.46	0.17	-0.05	0.00	-0.03	0.08	0.03	0.16	1.00	0.13	0.31	0.00	0.34	0.17	-0.03
RC_A	0.57	-0.17	-0.05	0.13	0.43	-0.38	-0.06	0.31	0.34	-0.20	0.12	-0.20	-0.05	0.32	0.13	1.00	0.33	0.42	-0.25	0.30	-0.41
RefCal_V	0.03	-0.22	0.11	0.27	0.57	-0.44	-0.04	0.24	0.22	-0.42	-0.12	-0.12	-0.41	0.05	0.31	0.33	1.00	0.48	0.10	0.43	-0.48
RefCim_V	0.36	-0.20	-0.30	0.18	0.46	-0.63	-0.24	0.06	0.45	-0.57	0.23	-0.25	-0.52	0.28	0.00	0.42	0.48	1.00	-0.56	0.69	-0.64
RefMed_V	-0.36	0.11	0.38	-0.05	-0.17	0.30	0.52	0.08	-0.35	0.29	-0.39	0.03	0.21	-0.17	0.34	-0.25	0.10	-0.56	1.00	-0.29	0.36
RI_A	0.36	0.08	-0.15	0.11	0.50	-0.36	-0.01	0.16	0.33	-0.46	0.24	-0.33	-0.43	0.41	0.17	0.30	0.43	0.69	-0.29	1.00	-0.43
RM_A	-0.26	0.36	0.20	-0.01	-0.34	0.59	0.12	0.00	-0.27	0.54	0.00	0.18	0.48	-0.05	-0.03	-0.41	-0.48	-0.64	0.36	-0.43	1.00
Tacpm_A	0.39	0.10	-0.07	-0.23	0.14	-0.28	0.01	0.42	0.04	-0.29	0.12	-0.30	-0.25	0.31	0.05	0.18	-0.07	0.23	-0.25	0.19	-0.18
Tasf	0.21	-0.06	-0.22	0.11	0.41	-0.49	-0.09	0.08	0.36	-0.53	0.19	-0.26	-0.51	0.27	0.02	0.35	0.53	0.75	-0.38	0.64	-0.58
Tcima_A	-0.06	-0.01	0.03	0.02	-0.14	0.28	-0.18	-0.10	0.00	0.34	-0.01	0.14	0.35	-0.07	-0.19	-0.10	-0.34	-0.33	-0.07	-0.34	0.32
Tcima_V	-0.23	-0.23	0.24	0.08	-0.04	0.31	0.24	-0.01	-0.04	0.46	-0.05	0.35	0.45	-0.11	0.29	-0.07	-0.14	-0.31	0.14	-0.26	0.26
Tcred	0.40	0.10	-0.04	-0.39	0.02	-0.10	-0.01	-0.02	-0.18	-0.35	0.24	-0.26	-0.31	0.22	0.04	0.09	-0.11	0.25	-0.30	0.31	-0.24
Tflash_A	0.48	0.02	-0.06	-0.36	0.11	-0.14	-0.04	0.03	-0.05	-0.42	0.24	-0.24	-0.37	0.28	0.05	0.17	-0.04	0.37	-0.38	0.39	-0.33
Tflash_V	0.44	0.11	-0.32	-0.01	0.46	-0.54	0.10	0.18	0.16	-0.58	0.20	-0.56	-0.56	0.41	0.04	0.41	0.40	0.67	-0.32	0.65	0.50
TG_A	0.44	0.04	-0.04	-0.22	0.17	-0.08	-0.09	0.14	0.09	-0.36	0.22	-0.24	-0.31	0.30	0.05	0.22	0.04	0.37	-0.32	0.58	-0.25
TGL_V	0.05	0.03	0.56	0.38	0.43	0.28	0.06	0.19	-0.10	0.35	-0.02	0.14	0.38	0.02	0.56	0.10	0.08	-0.36	0.22	-0.18	0.26
TGP_V	0.49	-0.18	-0.14	0.13	0.70	-0.39	-0.22	0.25	0.32	-0.40	0.21	-0.26	-0.31	0.39	0.14	0.52	0.33	0.63	-0.60	0.53	-0.49
TGR_V	-0.01	-0.37	0.02	0.00	0.17	-0.04	-0.06	0.17	0.13	0.01	0.03	0.07	0.03	-0.01	0.09	0.21	0.33	0.17	-0.03	0.08	-0.26
TH_V	0.27	-0.38	-0.34	0.08	0.39	-0.44	-0.20	0.09	0.46	-0.40	0.02	-0.10	-0.34	0.24	-0.10	0.36	0.32	0.70	-0.55	0.48	-0.53
Tjet_A	-0.13	0.11	0.14	0.49	0.03	0.11	-0.51	0.09	-0.08	0.29	0.17	0.03	0.31	-0.21	-0.10	0.00	-0.16	-0.25	-0.02	-0.30	0.14
Tpre_A	0.45	-0.12	-0.06	-0.19	0.46	-0.46	0.09	0.11	0.19	-0.61	0.11	-0.30	-0.58	0.37	0.24	0.37	0.43	0.67	-0.33	0.67	0.59
Tsal_HA	0.40	0.11	-0.06	-0.44	0.00	-0.11	0.02	-0.04	-0.21	-0.37	0.22	-0.25	-0.34	0.19	0.02	0.07	-0.13	0.24	-0.29	0.31	-0.25
VaporF_A	0.16	0.21	-0.12	-0.13	0.06	-0.24	0.18	0.08	0.17	-0.28	0.00	-0.19	-0.28	0.29	0.02	0.03	0.12	0.26	-0.03	0.38	-0.14

	Tacpm_A	Tasf	Tcima_A	Tcima_V	Tcred	Tflash_A	Tflash_V	TG_A	TGL_V	TGP_V	TGR_V	TH_V	Tjet_A	Tpre_A	Tsal_HA	VaporF_A
ACPM_A	0.39	0.21	-0.06	-0.23	0.40	0.48	0.44	0.44	0.05	0.49	-0.01	0.27	-0.13	0.45	0.40	0.16
API	0.10	-0.06	-0.01	-0.23	0.10	0.02	0.11	0.04	0.03	-0.18	-0.37	-0.38	0.11	-0.12	0.11	0.21
Asf	-0.07	-0.22	0.03	0.24	-0.04	-0.06	-0.32	-0.04	0.56	-0.14	0.02	-0.34	0.14	-0.06	-0.06	-0.12
Cred	-0.23	0.11	0.02	0.08	-0.39	-0.36	-0.01	-0.22	0.38	0.13	0.00	0.08	0.49	-0.19	-0.44	-0.13
GL_V	0.14	0.41	-0.14	-0.04	0.02	0.11	0.46	0.17	0.43	0.70	0.17	0.39	0.03	0.46	0.00	0.06
GO_A	-0.28	-0.49	0.28	0.31	-0.10	-0.14	-0.54	-0.08	0.28	-0.39	-0.04	-0.44	0.11	-0.46	-0.11	-0.24
GP_V	0.01	-0.09	-0.18	0.24	-0.01	-0.04	0.10	-0.09	0.06	-0.22	-0.06	-0.20	-0.51	0.09	0.02	0.18
Jet_A	0.42	0.08	-0.10	-0.01	-0.02	0.03	0.18	0.14	0.19	0.25	0.17	0.09	0.09	0.11	-0.04	0.08
Nafta_A	0.04	0.36	0.00	-0.04	-0.18	-0.05	0.16	0.09	-0.10	0.32	0.13	0.46	-0.09	0.19	-0.21	0.17
Pcima_A	-0.29	-0.53	0.34	0.46	-0.35	-0.42	-0.58	-0.36	0.35	-0.40	0.01	-0.40	0.29	-0.61	-0.37	-0.28
Pcima_V	0.12	0.19	-0.01	-0.05	0.24	0.24	0.20	0.22	-0.02	0.21	0.03	0.02	0.17	0.11	0.22	0.00
PEN	-0.30	-0.26	0.14	0.35	-0.26	-0.24	-0.56	-0.24	0.14	-0.26	0.07	-0.10	0.03	-0.30	-0.25	-0.19
Pflash_A	-0.25	-0.51	0.35	0.45	-0.31	-0.37	-0.56	-0.31	0.38	-0.31	0.03	-0.34	0.31	-0.58	-0.34	-0.28
QA	0.31	0.27	-0.07	-0.11	0.22	0.28	0.41	0.30	0.02	0.39	-0.01	0.24	-0.21	0.37	0.19	0.29
QV	0.05	0.02	-0.19	0.29	0.04	0.05	0.04	0.05	0.56	0.14	0.09	-0.10	-0.10	0.24	0.02	0.02
RC_A	0.18	0.35	-0.10	-0.07	0.09	0.17	0.41	0.22	0.10	0.52	0.21	0.36	0.00	0.37	0.07	0.03
RefCal_V	-0.07	0.53	-0.34	-0.14	-0.11	-0.04	0.40	0.04	0.08	0.33	0.33	0.32	-0.16	0.43	-0.13	0.12
RefCim_V	0.23	0.75	-0.33	-0.31	0.25	0.37	0.67	0.37	-0.36	0.63	0.17	0.70	-0.25	0.67	0.24	0.26
RefMed_V	-0.25	-0.38	-0.07	0.14	-0.30	-0.38	-0.32	-0.32	0.22	-0.60	-0.03	-0.55	-0.02	-0.33	-0.29	-0.03
RI_A	0.19	0.64	-0.34	-0.26	0.31	0.39	0.65	0.38	-0.18	0.53	0.08	0.48	-0.30	0.67	0.31	0.38
RM_A	-0.18	-0.58														

Figura 9. Variación de la distancia interna con el aumento en el número de clústers.

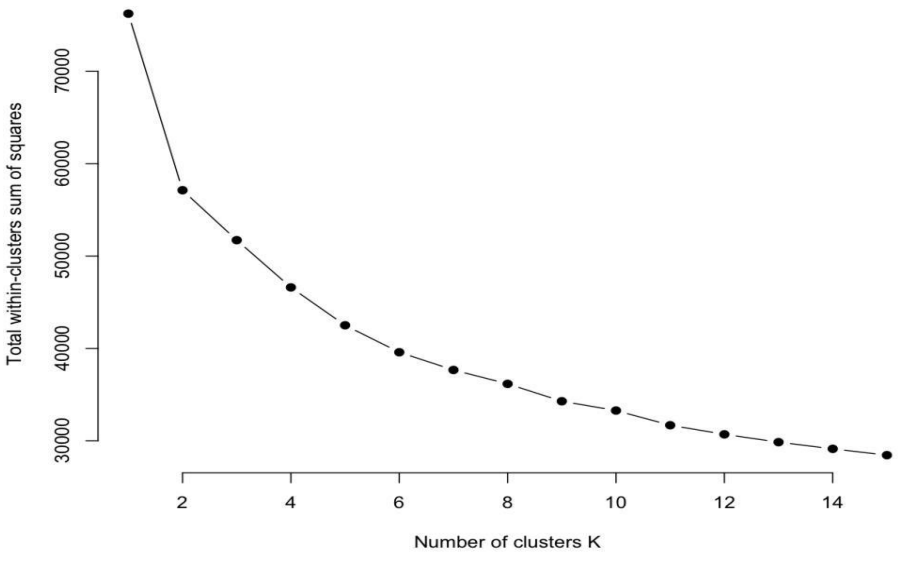
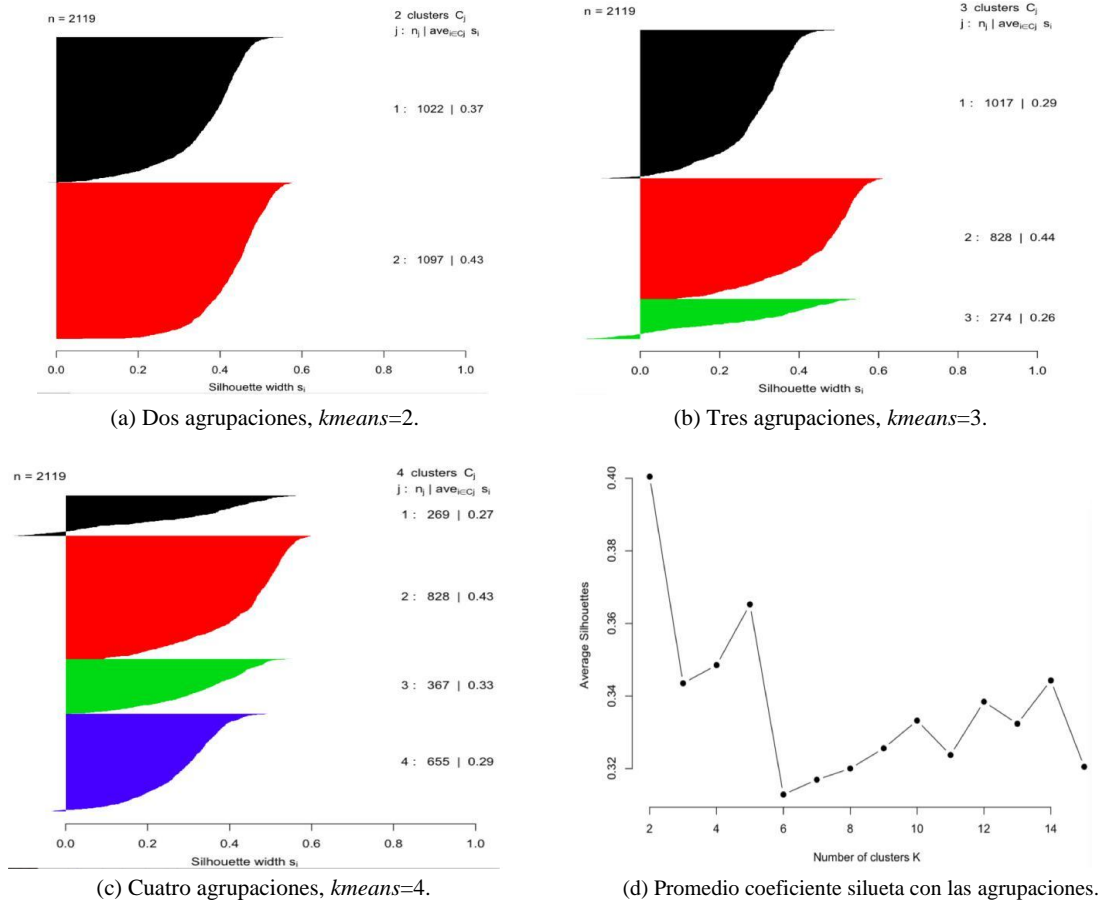


Figura 10. Coeficiente silueta para diferentes agrupaciones definidas en el método *kmean*.



En la parte derecha de las siluetas se indica el número de muestras constituyentes de cada agrupación, así como el correspondiente valor del coeficiente silueta promedio.

Las gráficas de dispersión de las variables de la U-2100 en los 2 agrupamientos pueden indicar las diferencias importantes entre los dos esquemas operacionales. Las Figuras 11a-c ilustran dispersiones en tres dimensiones de diferentes variables operacionales de la U-2100. En estas figuras se presentan los valores de las variables para cada agrupación o esquema operacional utilizando un color diferente, celeste para el esquema 1 y fucsia para el esquema 2; asimismo, los

elipsoides encierran el 50% de los datos para cada agrupación. De acuerdo a las tendencias en los elipsoides de las dispersiones, las muestras del esquema operacional 1 se diferencian de las muestras del esquema operacional 2 en los ejes relacionados con las variables reflujo de cima torre de vacío (RefCim_V) (Figuras 11a-c), temperatura de precaliente atmosférica (Tpre_A) (Figuras 11b-c) y penetración (PEN) (Figuras 11b-c). Asimismo, la Figura 5.5d muestra las contribuciones de cada variable para la formación de cada agrupación (1 o 2); según esta figura, las variables Tpre_A y RefCim_V son las mayores contribuyentes para la formación de las agrupaciones.

El esquema operativo 1 comprende la ventana de operación entre enero de 2015 y junio de 2016, mientras que desde diciembre de 2016 a enero de 2018 la U-2100 envuelve el esquema operativo 2; los datos entre julio y noviembre de 2016 fueron removidos de los históricos por valores incompletos en las variables. Los dos esquemas operacionales encontrados se diferencian debido a las condiciones de limpieza de los equipos. El esquema operacional 1 reporta una condición subestándar (equipos sin mantenimiento), especialmente en muestras pertenecientes al año 2016, en donde se consume una mayor cantidad de servicios debido a un ensuciamiento constante de los equipos; la zona empacada de la torre atmosférica en este esquema exhibe adelgazamiento por desgaste mecánico con los fluidos; el material del empaque de esta torre pierde consistencia por un periodo de uso de alrededor de 5 años. Asimismo, los hornos y los intercambiadores de calor presentan disminución en sus respectivas eficiencias; lo anterior conduce menores temperaturas de precaliente (Figuras 11b-c), menores valores del reflujo de cima en la torre de vacío (Figuras 11a-c) y por lo tanto mayores valores de penetración por contenidos importantes de gasóleo en los asfaltos generados (Figuras 11b-c).

De otro lado, el esquema operacional 2 está compuesto por muestras indicando una disminución en los consumos de servicios industriales debido a los beneficios

obtenidos de la aplicación previa de una reparación y un mantenimiento realizados a los equipos de la unidad. La reparación de la unidad fue efectuada entre los meses de octubre y noviembre de 2016, con premisas de eliminación de ensuciamientos en los equipos de transferencia de calor, cambio en los empaques de las torres y mantenimientos estructurales en equipos. La operación de la U-2100 después de la reparación exhibe mayor transferencia de calor en los hornos y en los intercambiadores (mayor T_{pre_A} , Figuras 11b-c), mayor valor en el reflujó de cima en la torre de vacío (Figuras 11a-c) y menor valor en la penetración del asfalto generado (Figuras 11b-c) por mayor despojo de su contenido de gasóleo.

A nivel estadístico, las pruebas F y *one-way* ANOVA fueron aplicadas para verificar las diferencias entre los dos esquemas operacionales considerando a la penetración como el factor de respuesta. La Figura 5.6a presenta la salida reportada por Rcommander para la prueba F, la cual rechaza la hipótesis nula de relación de varianzas igual a la unidad para los esquemas operacionales; con esto, los esquemas operacionales conducen a diferentes valores de penetración para el asfalto generado. También, la Figura 11b muestra el resultado de la prueba *one-way* ANOVA para la comparación de las medias entre los valores de las penetraciones para los dos esquemas operacionales de la U-2100; el resultado de la ANOVA muestra que las medias son diferentes para las penetraciones de los dos esquemas.

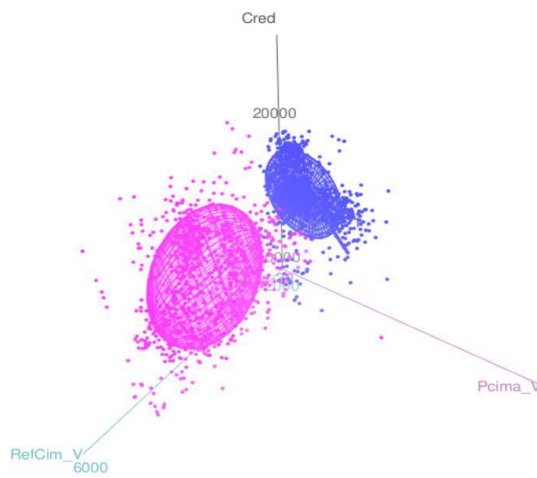
5.3 REGRESIÓN LINEAL MÚLTIPLE

La regresión *stepwise* fue aplicada a los datos históricos para identificar las variables estadísticamente significativas para la predicción de la penetración del asfalto generado; la función de R *stepwise* fue aplicada según lo definido en la metodología. La Figura 14 presenta los resultados finales para la predicción de la penetración (PEN) a partir de los resultados de la regresión *stepwise*; según los resultados, la regresión con todos los datos históricos presenta un desempeño

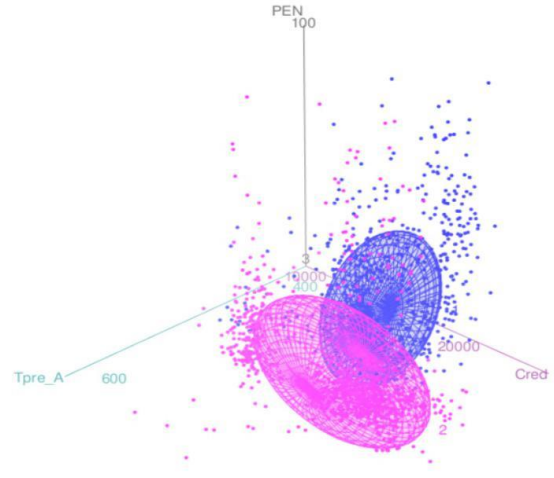
bajo al reproducir sólo el 57% de la varianza de la PEN a partir de 23 variables. Asimismo, la regresión *stepwise* aplicada a cada uno de los esquemas operativos, no reporta mejoras significativas para el ajuste de PEN; la Figura 15 muestra que para el primer esquema operativo, la varianza explicada corresponde al 61% utilizando 21 variables, mientras que para el segundo esquema la varianza explicada es del 57% utilizando 24 variables.

Las gráficas de influencia resultantes de las regresiones lineales son mostradas en la Figura 16. Las gráficas de influencia presentan tres estadísticos de diagnóstico (los residuales estudentizados, los valores de la diagonal principal de la matriz Hat, y la distancia de Cook) de las muestras en la regresión; las posibles muestras atípicas e influyentes en las regresiones se localizan en las regiones con residuales estudentizados fuera del intervalo entre -2 y 2 y, valores Hat superiores a 0.05; respecto a la distancia de Cook, las posibles muestras atípicas e influyentes se diferencian por un mayor tamaño de circunferencia. Con lo anterior, las regresiones para la penetración (PEN) exhiben diferentes puntos atípicos e influyentes según los correspondientes datos históricos. Estos puntos atípicos e influyentes precisamente corresponden a los puntos con bajos valores de coeficiente silueta (Figura 9), lo cual hace referencia a las muestras alrededor del periodo de mantenimiento de la unidad; los desempeños de las regresiones son disminuidos debido a estos puntos.

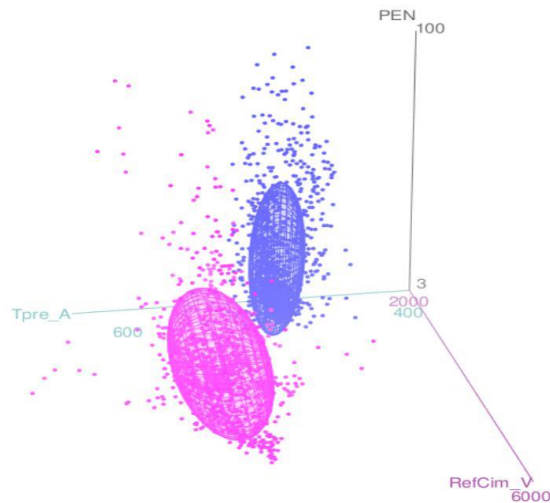
Figura 11. Clústeres formados y su relación con las variables temperatura de precaliente, Tpre_A, y reflujo de cima en la torre de vacío, RefCim_V. Color celeste para el esquema 1 y fucsia para el esquema 2.



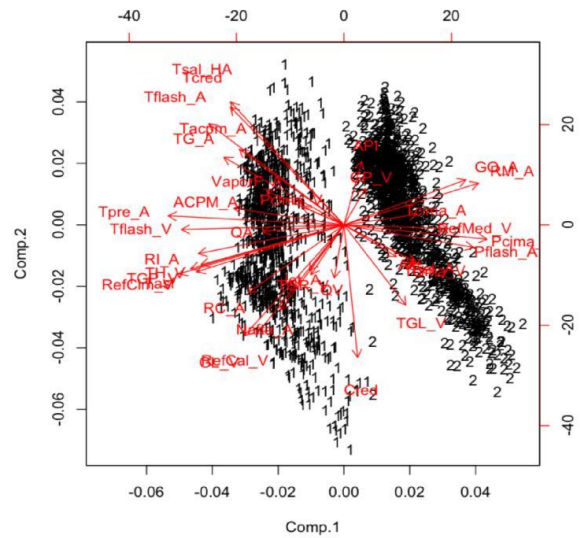
(a) Crudo reducido vs Reflujo de cima de vacío vs Presión cima atmosférica.



(b) Penetración vs T de precaliente vs Crudo reducido.



(c) Penetración vs T de precaliente vs Reflujo de cima de vacío.



(d) Componentes resultantes con *kmeans=2*.

Figura 12. Pruebas F y *one-way* ANOVA para las comparaciones de las varianzas y las medias, respectivamente, de las penetraciones reportadas por los esquemas operacionales según el método *kmeans*.

```
> var.test(PEN ~ KMeans, alternative='two.sided', conf.level=.95, data=Ini)

      F test to compare two variances

data:  PEN by KMeans
F = 1.5966, num df = 1093, denom df = 1024, p-value = 4.308e-14
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 1.415049 1.801090
sample estimates:
ratio of variances
 1.596635
```

a. Prueba F. Hipótesis nula: relación entre las varianzas de los esquemas operacionales igual a 1.

```
> summary(AnovaModel.3)
      Df Sum Sq Mean Sq F value Pr(>F)
KMeans    1  58495   58495   199.5 <2e-16 ***
Residuals 2117 620646    293
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> with(Ini, numSummary(PEN, groups=KMeans, statistics=c("mean", "sd")))
      mean      sd data:n
1 30.19561 18.91707  1094
2 19.68195 14.97101  1025

> oneway.test(PEN ~ KMeans, data=Ini) # Welch test

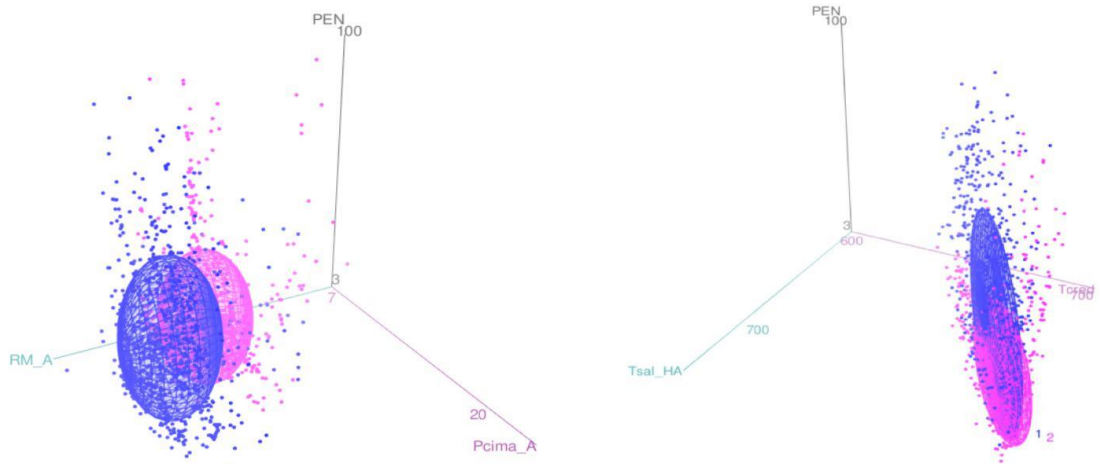
      One-way analysis of means (not assuming equal variances)

data:  PEN and KMeans
F = 202.53, num df = 1.0, denom df = 2060.1, p-value < 2.2e-16
```

b. *One-way* ANOVA. Hipótesis nula: las medias de los esquemas operacionales son iguales.

Según estas pruebas, las respectivas hipótesis nulas son rechazadas, con lo cual, las varianzas y las medias de las penetraciones generadas en cada esquema operacional resultan ser diferentes.

Figura 13. Visualización de las agrupaciones en coordenadas de la penetración.



(a) Penetración vs Reflujo medio atmosférico vs Presión cima atmosférica.

(b) Penetración vs T de Horno atmosférico vs T crudo reducido.

Figura 14. Resultados reportados por R para el modelo lineal de predicción de la penetración (PEN) con todos los datos históricos, indicando un R2 de 0.5715.

```
Call:
lm(formula = PEN ~ Asf + Cred + GL_V + GO_A + GP_V + Nafta_A +
    Pcima_V + Pflash_A + QA + QV + RC_A + RefCim_V + RefMed_V +
    RI_A + Tacpm_A + Tasf + Tcima_V + Tcred + Tflash_A + Tflash_V +
    TGP_V + TH_V + Tjet_A, data = Autoescalados)

Residuals:
    Min       1Q   Median       3Q      Max
-4.0516 -0.3656 -0.0792  0.2515  3.6384

Coefficients:
(Intercept)  0.34758    0.06153    5.649 1.83e-08 ***
Asf           0.24471    0.02875    8.511 < 2e-16 ***
Cred         -0.06108    0.02605    2.345 0.019129 *
GL_V        -0.24853    0.03718   -6.684 2.97e-11 ***
GO_A        -0.09753    0.02331   -4.184 2.98e-05 ***
GP_V        -0.46916    0.03299  -14.221 < 2e-16 ***
Nafta_A     -0.10734    0.02067   -5.193 2.27e-07 ***
Pcima_V     -0.07213    0.01879   -3.839 0.000127 ***
Pflash_A    0.11631    0.02278    5.106 3.58e-07 ***
QA          -0.22418    0.01937  -11.573 < 2e-16 ***
QV          0.24312    0.03491    6.964 4.42e-12 ***
RC_A        0.06839    0.01965    3.481 0.000510 ***
RefCim_V   -0.10505    0.03513   -2.990 0.002820 **
RefMed_V   -0.19574    0.03082   -6.351 2.61e-10 ***
RI_A        0.06579    0.02486    2.647 0.008188 **
Tacpm_A    -0.08180    0.02270   -3.603 0.000322 ***
Tasf       -0.05241    0.02521   -2.079 0.037731 *
Tcima_V    0.16648    0.02019    8.248 2.82e-16 ***
Tcred     -0.17902    0.04177   -4.286 1.90e-05 ***
Tflash_A   0.12745    0.04402    2.896 0.003824 **
Tflash_V  -0.25252    0.03190   -7.916 3.93e-15 ***
TGP_V     -0.15648    0.04906   -3.190 0.001444 **
TH_V       0.20270    0.03059    6.626 4.37e-11 ***
Tjet_A    -0.38434    0.02597  -14.799 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6546 on 2095 degrees of freedom
Multiple R-squared:  0.5762, Adjusted R-squared: 0.5715
F-statistic: 123.8 on 23 and 2095 DF, p-value: < 2.2e-16
```

Figura 15. Resultados reportados por R para el modelo lineal de predicción de la penetración (PEN) lineal en cada esquema operacional. Esquema operativo 1, R2 de 0.5992; esquema operativo 2, R2 de 0.5646.

```

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.032e-15  1.914e-02  0.000 1.000000
API          -9.768e-02  3.019e-02  -3.235 0.001252 **
Cred         2.971e-01  3.680e-02  8.073 1.83e-15 ***
GL_V        -1.441e-01  3.404e-02  -4.233 2.50e-05 ***
GP_V        -4.552e-01  3.567e-02  -12.763 < 2e-16 ***
Jet_A       5.554e-02  2.644e-02  2.100 0.035920 *
Nafta_A     -6.594e-02  2.361e-02  -2.793 0.005310 **
Pcima_A     6.997e-02  3.099e-02  2.258 0.024158 *
Pcima_V     -9.485e-02  2.620e-02  -3.620 0.000308 ***
QA          -1.933e-01  2.372e-02  -8.150 1.01e-15 ***
QV          2.903e-01  3.087e-02  9.406 < 2e-16 ***
RC_A        1.040e-01  2.211e-02  4.705 2.87e-06 ***
RefCim_V    5.843e-02  2.320e-02  2.518 0.011943 *
RefMed_V    -1.425e-01  2.562e-02  -5.561 3.38e-08 ***
Tacpm_A     -1.341e-01  4.552e-02  -2.946 0.003290 **
Tcima_V     1.743e-01  3.119e-02  5.587 2.93e-08 ***
Tcred       -1.976e-01  4.915e-02  -4.020 6.22e-05 ***
Tflash_V    -8.053e-02  3.030e-02  -2.657 0.007990 **
TGL_V       -2.483e-01  5.240e-02  -4.739 2.44e-06 ***
TGP_V       2.199e-01  6.135e-02  3.585 0.000352 ***
Tjet_A      -8.207e-02  3.339e-02  -2.458 0.014141 *
Tpre_A      1.694e-01  5.081e-02  3.335 0.000883 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6331 on 1072 degrees of freedom
Multiple R-squared:  0.6069, Adjusted R-squared:  0.5992
F-statistic: 78.8 on 21 and 1072 DF,  p-value: < 2.2e-16

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -7.211e-14  2.061e-02  0.000 1.00000
ACPM_A      -1.544e-01  3.586e-02  -4.305 1.84e-05 ***
API          8.667e-02  2.659e-02  3.259 0.00116 **
Asf          2.795e-01  5.312e-02  5.262 1.75e-07 ***
GO_A        -9.570e-02  3.009e-02  -3.181 0.00152 **
GP_V        -3.484e-01  5.910e-02  -5.896 5.10e-09 ***
Nafta_A     -2.911e-01  3.776e-02  -7.711 3.01e-14 ***
Pcima_A     -2.844e-01  7.131e-02  -3.988 7.15e-05 ***
Pcima_V     -1.010e-01  3.124e-02  -3.232 0.00127 ***
Pflash_A    3.178e-01  8.064e-02  3.941 8.66e-05 ***
QV          1.727e-01  5.307e-02  3.254 0.00117 **
RC_A        9.778e-02  3.700e-02  2.643 0.00836 **
RefCal_V    1.091e-01  4.189e-02  2.605 0.00933 **
RefMed_V    -3.214e-01  5.269e-02  -6.100 1.51e-09 ***
RI_A        7.238e-02  2.624e-02  2.758 0.00593 **
RM_A        -1.571e-01  2.716e-02  -5.785 9.71e-09 ***
Tcima_A     1.122e-01  2.766e-02  4.056 5.37e-05 ***
Tcima_V     1.204e-01  2.811e-02  4.284 2.01e-05 ***
Tflash_A    2.545e-01  6.222e-02  4.089 4.67e-05 ***
Tflash_V    -4.980e-01  4.027e-02  -12.368 < 2e-16 ***
TGP_V       -2.026e-01  5.556e-02  -3.646 0.00028 ***
TH_V        2.538e-01  3.816e-02  6.652 4.77e-11 ***
Tjet_A      -4.820e-01  4.212e-02  -11.444 < 2e-16 ***
Tpre_A      -9.901e-02  3.991e-02  -2.481 0.01328 *
Tsal_HA     -1.378e-01  6.161e-02  -2.237 0.02551 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6599 on 1000 degrees of freedom
Multiple R-squared:  0.5748, Adjusted R-squared:  0.5646
F-statistic: 56.32 on 24 and 1000 DF,  p-value: < 2.2e-16

```

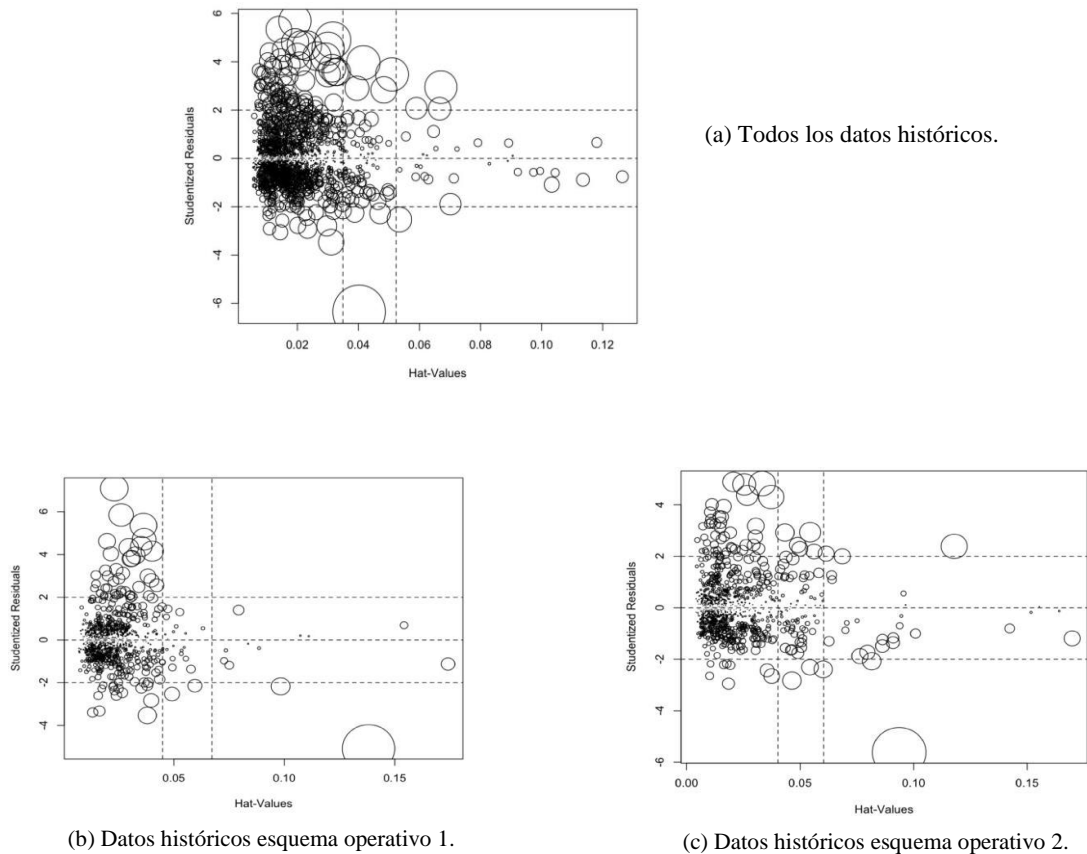
(a) Esquema operativo 1

(b) Esquema operativo 2

5.4 REDES NEURONALES ARTIFICIALES

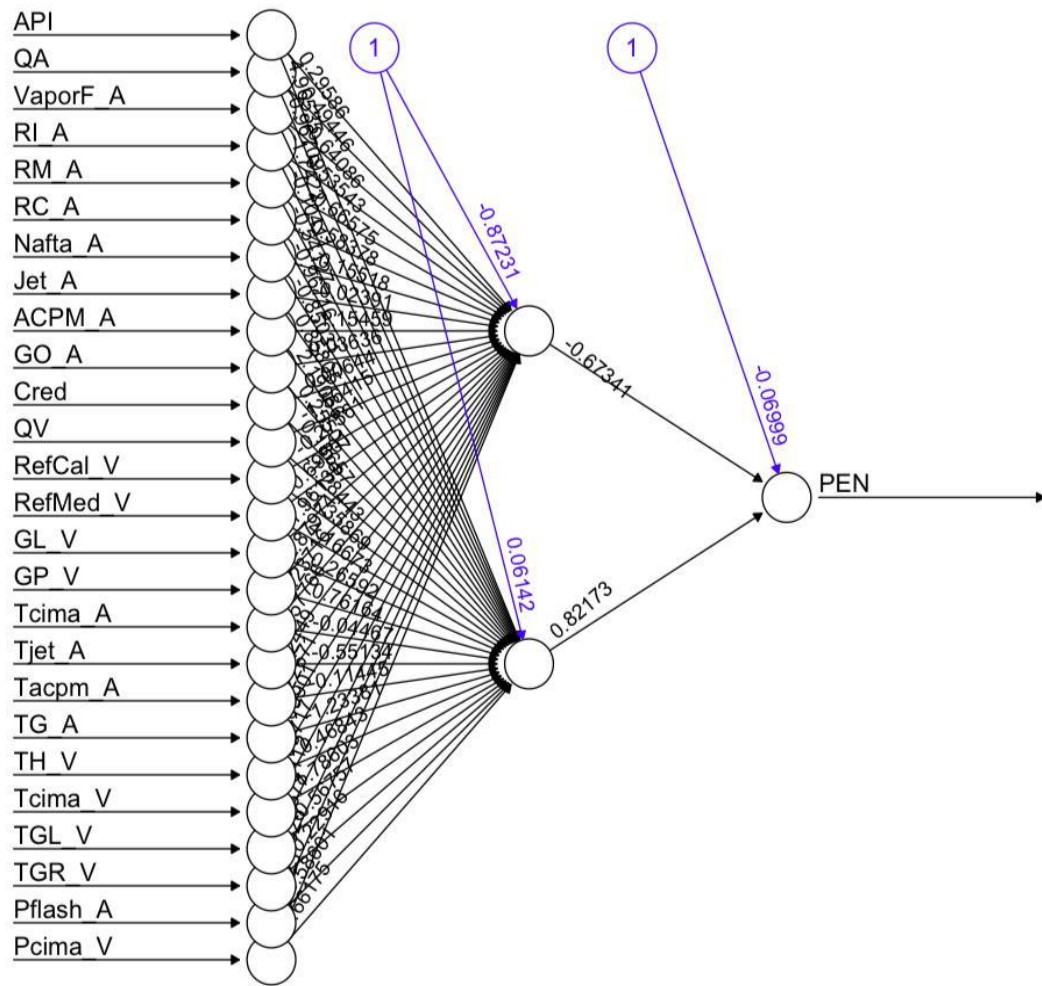
Para la red, las variables Tflash_A, Pcima_A, Tpre_A, Tsal_HA, Tcred, RefCim_V, Tflash_V, TGP_V, Tasf y Asf (ver anexo 1) fueron eliminadas de los datos históricos debido a su correlación elevada con otras variables (indicadas con círculos en la Figura 8). Con esto, un total de 26 variables de entrada y una variable de salida (PEN) fueron definidas para el entrenamiento de las RNA, conduciendo a arquitecturas tipo *feedforward*, es decir, 26 neuronas de entrada, x neuronas en la capa interna y 1 neurona de salida (Figura 17); la cantidad de neuronas de la capa interna fue evaluada hasta un máximo de 60, según lo 26: $x : 1$ definido en la metodología.

Figura 16. Gráficas de influencia resultantes de las regresiones RLM para la penetración.



La Figura 17 representa una arquitectura de RNA definida para la función *neuralnet* del paquete *Neuralnet* en el programa R. La arquitectura definida en esta figura corresponde a la 26:2:1, la cual indica 26 variables de entrada (abreviaciones en el anexo 1), 2 neuronas en la capa intermedia y la neurona de salida que corresponde a la penetración (PEN); el código mostrado en el anexo 2 fue desarrollado para el uso de *neuralnet*. En las Figuras 18 – 23 se presentan los resultados obtenidos en el entrenamiento y la validación de las redes neuronales utilizando todos los datos históricos.

Figura 17. Ilustración RNA *feedforward* 26:2:1 Utilizada para la regresión de la penetración (abreviaciones definidas en el anexo 1); dos neuronas capa interna.

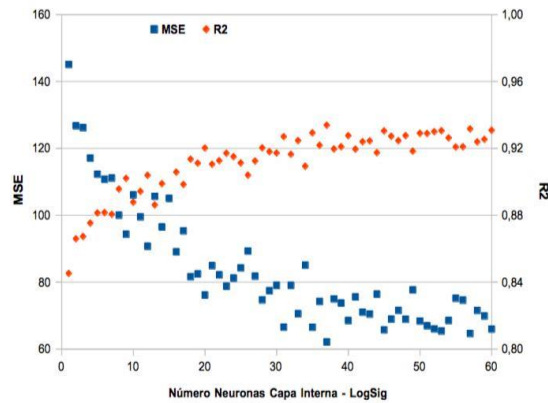


Los valores en las líneas de flujo entre las neuronas corresponden a los pesos de las señales, mientras que los valores en las líneas provenientes de los círculos 1 corresponden a los términos independientes (*bias*) que ingresan a las neuronas.

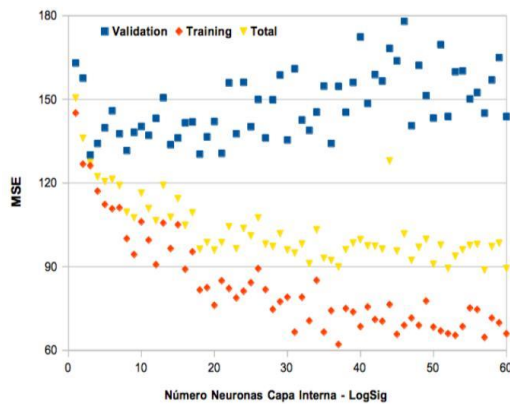
Las tendencias mostradas en las Figuras 18 y 19 indican que un aumento en el número de neuronas en la capa interna mejora el desempeño de las redes en entrenamiento, mientras que disminuye el desempeño de las mismas en validación; el desempeño total (validación y entrenamiento) mejora ligeramente con las neuronas de la capa interna. Los mejores desempeños (totales) se presentan con las estructuras 26:37:1 – LogSig (MSEall=89.9 mm²/100 y

$R^2_{all}=0.905$, Figura 18) y 26:32:1 – TanH ($MSE_{all}=85.6 \text{ mm}^2/100$ y $R^2_{all}=0.909$, Figura 19). Estas redes presentan una predicción con una desviación estándar de aproximadamente 9.5 mm/10; es decir, las predicciones reportar un intervalo de confianza (95%) de $\pm 19 \text{ mm}/10$. El tamaño del intervalo de confianza (38 mm/10) resulta superior comparado con el valor promedio de los históricos, 24.4 mm/10, con lo cual, las RNA no presentan una predicción aceptable para ayudar en la disminución de la cantidad de reprocesos por preparaciones de asfalto con PEN fuera de especificaciones.

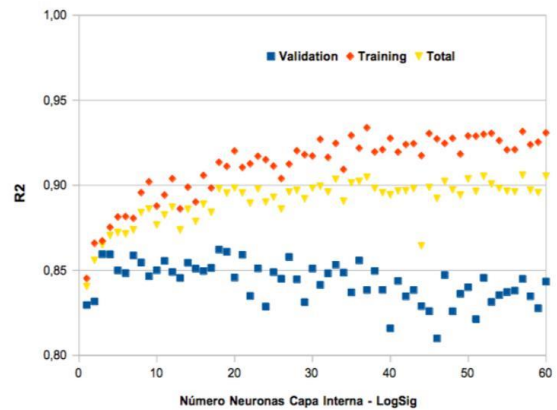
Figura 18. Desempeño obtenido con la red 26:x:1 – LogSig con el número de neuronas internas x, con todos los datos históricos (20 repeticiones).



(a) MSE y R^2 entrenamiento vs Neuronas capa interna.

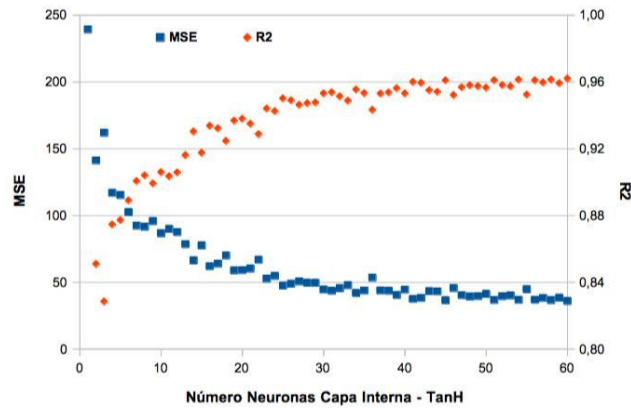


(b) MSE vs Neuronas capa interna.

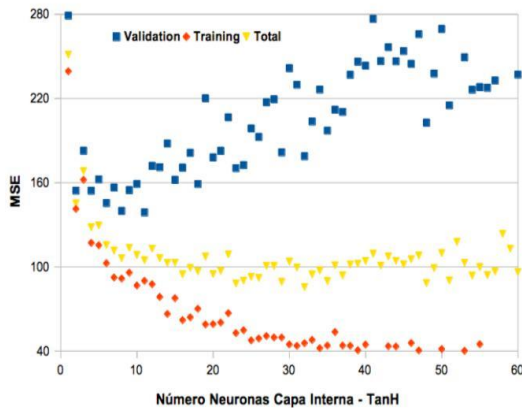


(c) R^2 vs Neuronas capa interna.

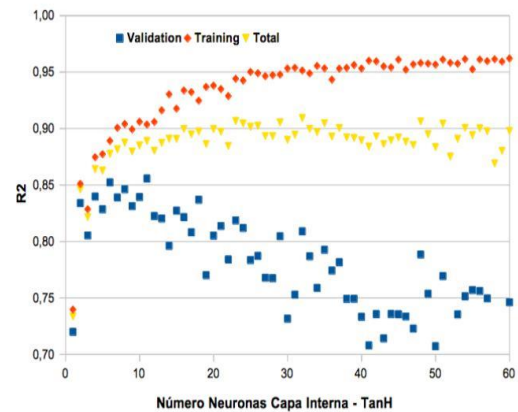
Figura 19. Desempeño obtenido con la red 26:x:1 – TanH con el número de neuronas internas x, con todos los datos históricos (20 repeticiones).



(a) MSE y R^2 entrenamiento vs Neuronas capa interna



(b) MSE vs Neuronas capa interna

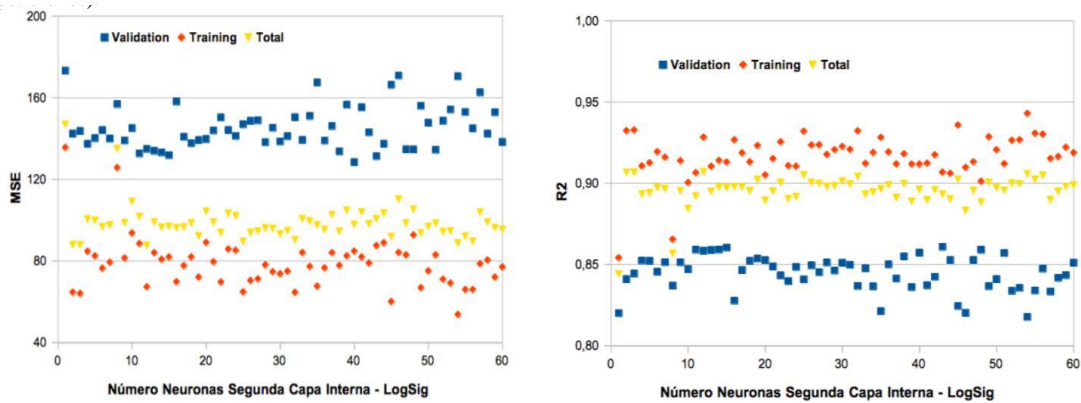


(c) R^2 vs Neuronas capa interna

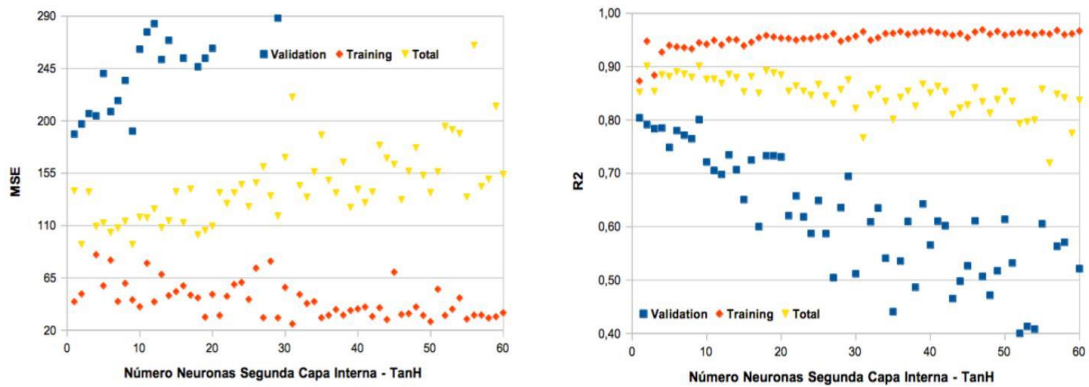
Debido a los bajos desempeños en validación y en el ajuste total, una nueva capa oculta fue incluida en la mejor arquitectura de red para cada función (26:37:1 – LogSig y 26:32:1 – TanH). Según los nuevos entrenamientos, tanto para LogSig y como para TanH, la segunda capa no conlleva a mejoras en los desempeños de las redes (Figura 20). El mejor desempeño en total para estas redes corresponde a $R^2_{all}=0.907$ y $MSE_{all}=88 \text{ mm}^2/100$, obtenido con la arquitectura 26:37:3:1 – LogSig (Figura 19a). La desviación estándar de esta red es 9.4 mm/10, lo cual supera los valores máximos de repetibilidad (4 mm/10) y reproducibilidad (11 mm/10) especificados en la norma ASTM D5-06 para los valores de PEN

considerados. Los anteriores desempeños de las RNA sugieren la presencia de datos atípicos en los históricos.

Figura 20. Desempeños obtenidos con las redes utilizando dos capas internas con todos los datos históricos (20 repeticiones).



(a) Arquitectura 26:37:x:1 - LogSig.



(b) Arquitectura 26:32:x:1 - TanH.

Según el análisis de agrupaciones por *kmeans*, la reparación de la unidad divide a los históricos en dos conjuntos (Figura 10d). Sin embargo, la división en tres agrupaciones puede identificar muestras atípicas presentes en los históricos. La Figura 21 ilustra los componentes obtenidos con el método *kmeans* definiendo tres agrupaciones (Figura 21a, en color celeste la agrupación 1, en color verde la agrupación 2 y en fucsia la agrupación 3). El análisis de las tendencias en las Figuras 21 muestra que los datos en la agrupación 3 exhiben bajas temperaturas de crudo reducido obtenido de la torre de destilación atmosférica (Figura 21b-e).

Esta operación se debe a la baja eficiencia del horno atmosférico (Figura 21d), el cual disminuye la cantidad de ACPM (Figura 21e), disminuye la temperatura del crudo reducido y aumenta la penetración del asfalto generado (Figura 21f).

La disminución en el desempeño de la unidad exhibida por las muestras de la agrupación 3 afecta la predicción de la penetración con las RNA, por lo cual, estas muestras fueron eliminada de la base de históricos. En total, 274 puntos fueron eliminados (Figura 10b, silueta color verde), lo cual corresponde al 13% de los datos y a 3 meses de operación antes de la parada para la reparación de la unidad; la operación con las características de la agrupación 3 es mostrada en diversas muestras en el periodo de febrero a junio de 2016; la parada de planta se efectuó de octubre 20 a noviembre 21 de 2016. Con esto, los históricos fueron reducidos a 1845 muestras (87% de los históricos tratados en las primeras estructuras de red), con un valor de penetración máximo de 100, mínimo de 3, promedio de 22 y desviación estándar de 15 mm/10.

Los históricos sin los datos de la agrupación 3 fueron escalados entre los valores 0 y 1, y utilizados para la predicción de los respectivos valores de penetración con las arquitecturas de red; los desempeños fueron ponderados sobre 20 ejecuciones de cada estructura de red. Las Figura 22 y 23 presentan las variaciones para los desempeños de las estructuras 26:x:1 – LogSig y 26:x:1 – TanH en la predicción de la penetración sin la agrupación 3; según estas figuras, la omisión de los datos cercanos al periodo de la reparación (agrupación 3) aumenta los desempeños y disminuye el ruido en los ajustes, especialmente en lo referente a las etapas de validación. Los mejores desempeños se reportan con las redes 26:32:1 – LogSig y 26:35:1 – TanH, con valores de MSE_{all} de 74.3 y 65.2 mm²/100, respectivamente, y valores de R²_{all} de 0.894 y 0.907, respectivamente. Con los valores obtenidos de MSE_{all}, las redes presentan una desviación de 8.5 mm/10, lo cual conduce al intervalo de ± 17 mm/10 para la predicción (95% de confianza). Este intervalo de confianza resulta mayor en 12 unidades que el promedio de los datos reducidos; a

pesar de que hay una mejora respecto a las predicciones con todos los históricos, el desempeño de las redes actuales presenta una posibilidad de mejora, asignando una nueva capa oculta a las mejores estructuras de red. La Figura 24 presenta la variación en el desempeño con la asignación de la segunda capa oculta; según esta figura, las estructuras 26:32:8:1 – LogSig y 26:35:4:1 – TanH exhiben los mejores desempeños totales con MSE_{all} de 72.2 y 43.5 mm²/100, respectivamente, y R²_{all} de 0.897 y 0.917, respectivamente. De estas redes, la estructura 26:35:4:1 – TanH presenta la menor desviación estándar (6.6 mm/10) y el menor intervalo de confianza hasta ahora obtenidos (± 13 mm/10, 95% confianza); la desviación estándar de la anterior red se encuentra cercana a la repetibilidad definida en la norma (4 mm/10) y es comparable con trabajos previos reportados en la literatura basados en el espectro NIR (6.6 mm/10)^{55 56}. De la Figura 25a, que muestra la dispersión entre los valores experimentales y los ajustados por la RNA, es posible enunciar que una gran cantidad de muestras se concentran en las inmediaciones de la línea de pendiente uno (roja); también, los errores exhiben una distribución normal con media en 0 y desviación estándar de 6.6 mm/10 (Figura 25b).

Por otra parte, la Figura 21 muestra que existen diversos valores por fuera de la elipse de Hotelling (95% de confianza) para las agrupaciones 1 y 3; estos datos atípicos pueden provenir de estados dinámicos o condiciones operacionales anómalas que conducen a valores de penetración elevados, mayores de 60 mm/10. Los valores atípicos fueron eliminados conduciendo a una reducción adicional de los históricos a 1785 muestras; es decir, 60 muestras fueron eliminadas, con lo cual los históricos disponibles para la predicción reportaron un valor de penetración máximo de 60, mínimo de 3, promedio de 20.2 y desviación

⁵⁵ Blanco, M., MasPOCH, S., Villarroya, I., Peralta, X., González, J.M., Torres, J. (2000). Determination of the penetration value of bitumens by near infrared spectroscopy. *Analyst*, 125, 1823–1828.

⁵⁶ Blanco, M., MasPOCH, S., Villarroya, I., Peralta, X., González, J.M., Torres, J. (2001). Determination of physical properties of bitumens by use of near-infrared spectroscopy with neural networks. Joint modelling of linear and non-linear parameters. *Analyst*, 126, 378–382.

estándar de 11.4 mm/10. Estos valores atípicos coincidieron con valores negativos del coeficiente silueta para las respectivas muestras del agrupamiento 1 (Figura 10b).

Como en el caso anterior, los históricos sin los datos de la agrupación 3 y con la restricción de penetración inferior o igual a 60 mm/10 fueron normalizados a valores entre 0 y 1, y utilizados para el entrenamiento de las redes con funciones de activación LogSig y TanH; los desempeños fueron ponderados sobre 20 ejecuciones de cada red. Las Figura 26 y 27 exhiben las tendencias para los desempeños de las estructuras 26:x:1 – LogSig y 26:x:1 – TanH en la predicción de la penetración; en estas figuras es posible apreciar que la restricción de penetración hasta 60 mm/10 mejora los desempeños de las redes, especialmente en validación. Los mejores desempeños se reportan para las redes 26:21:1 – LogSig y 26:12:1 – TanH, con valores de MSEall de 42.8 y 41.7 mm²/100, respectivamente, y valores de R2all de 0.918 y 0.921, respectivamente. Con esto, las redes presentan una desviación estándar de 6.5 mm/10, lo cual conduce al intervalo de ± 13 mm/10 para la predicción (95% de confianza). La desviación estándar de estas redes es comparable con el valor de repetibilidad (4 mm/10) especificado según la norma y trabajos previos reportados en la literatura basados en el espectro NIR (6.6 mm/10)^{57 58}.

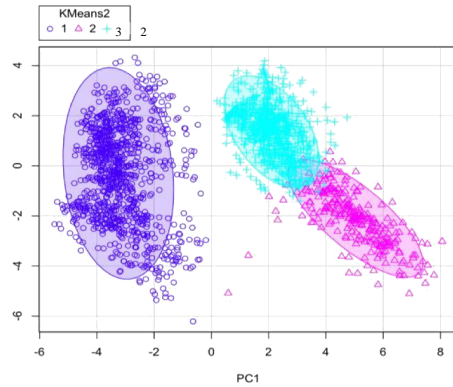
La Figura 28a muestra la dispersión de los históricos vs la predicción para la PEN obtenida con la estructura 26:12:1 – TanH; según esta figura, la dispersión de las predicciones se efectúa alrededor de la pendiente de valor 1. Asimismo, la Figura 28b presenta una distribución normal de los residuales con valor medio de 0 y desviación estándar de 10 mm/10. Las arquitecturas fueron modificadas para adicionar una nueva capa oculta; los resultados de los entrenamientos de

⁵⁷ Blanco, M., MasPOCH, S., Villarroja, I., Peralta, X., González, J.M., Torres, J. (2000). Determination of the penetration value of bitumens by near infrared spectroscopy. *Analyst*, 125, 1823–1828.

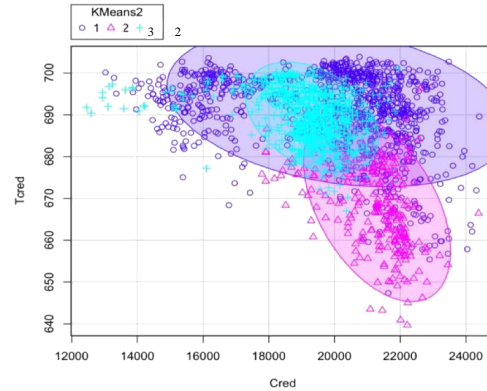
⁵⁸ *Ibíd.*

26:21:x:1 – LogSig y 26:12:x:1 – TanH pueden ser consultados en las Figuras 29 y 30, respectivamente; según estas figuras, los desempeños disminuyen con la nueva capa oculta en cada caso.

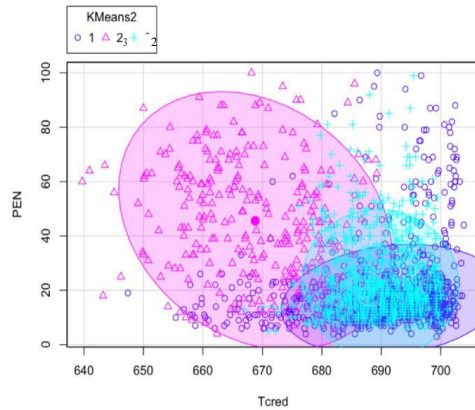
Figura 21. Resultados para $kmeans=3$. Celeste agrupación 1, verde agrupación 2 y fucsia agrupación 3.



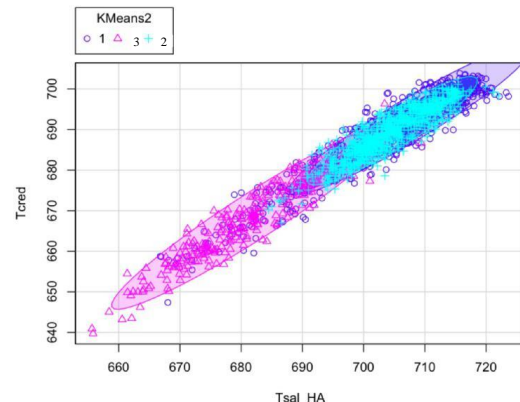
(a) Gráfica componentes 1 y 2



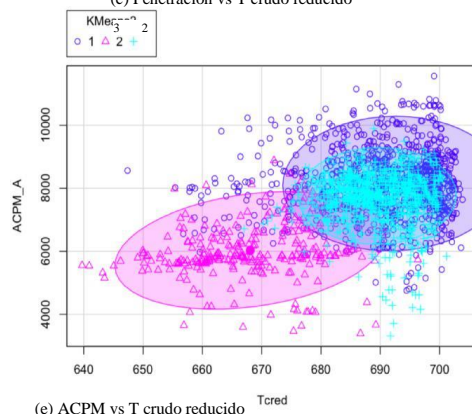
(b) T crudo reducido vs flujo crudo reducido



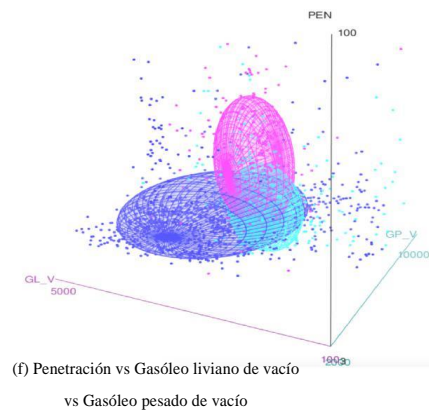
(c) Penetración vs T crudo reducido



(d) T crudo reducido vs T horno atmosférico

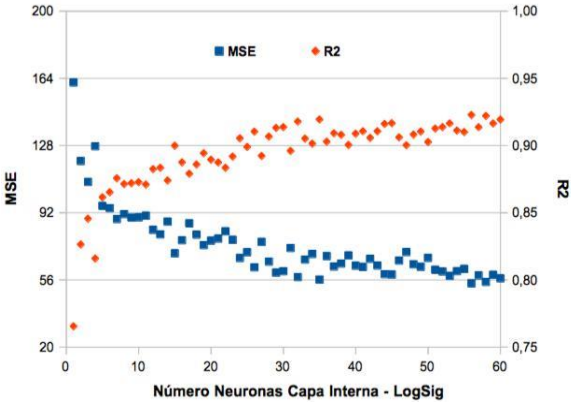


(e) ACPM vs T crudo reducido

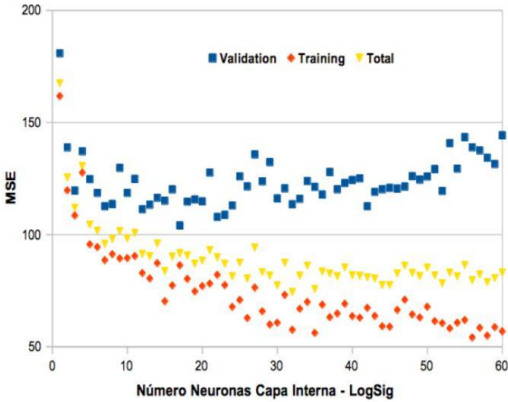


(f) Penetración vs Gasóleo liviano de vacío vs Gasóleo pesado de vacío

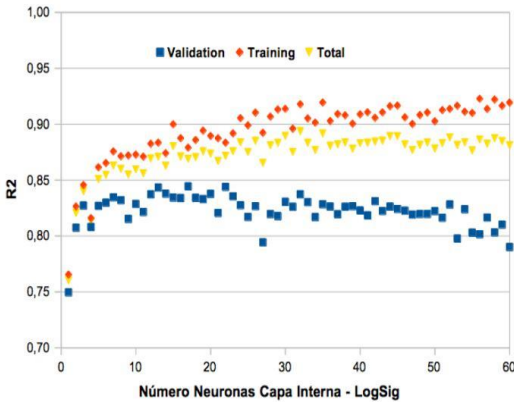
Figura 22. Desempeño de las redes 26:x:1 – LogSig con los históricos sin datos de la agrupación 3 y PEN hasta 100 mm/10 (20 repeticiones).



(a) MSE y R^2 entrenamiento vs Neuronas capa interna

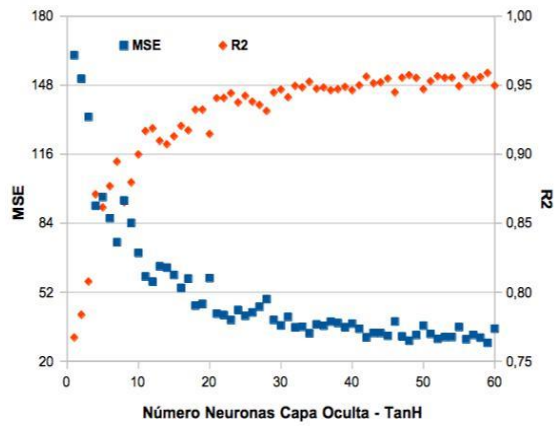


(b) MSE vs Neuronas capa interna

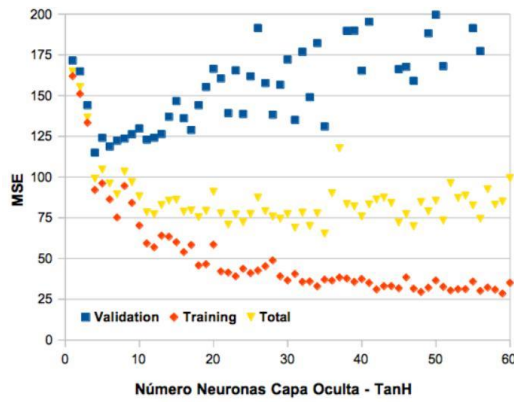


(c) R^2 vs Neuronas capa interna

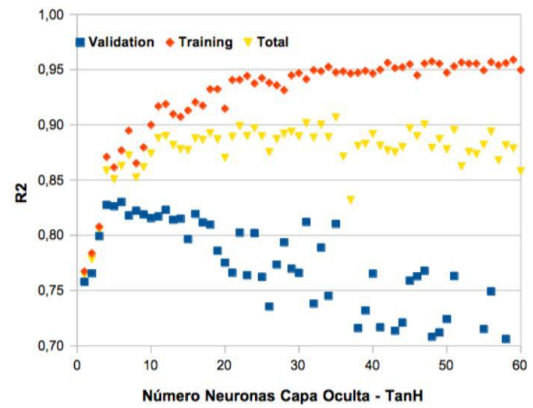
Figura 23. Desempeño de las redes 26:x:1 – TanH con los históricos sin datos de la agrupación 3 y PEN hasta 100 mm/10 (20 repeticiones).



(a) MSE y R² entrenamiento vs Neuronas capa interna

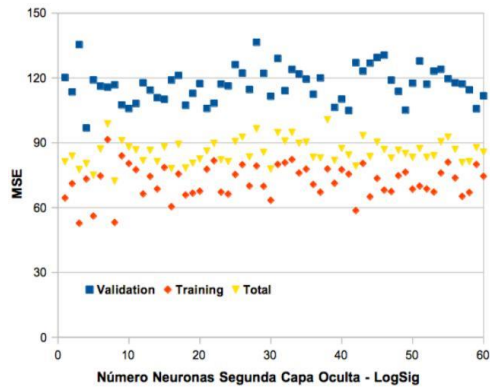


(b) MSE vs Neuronas capa interna

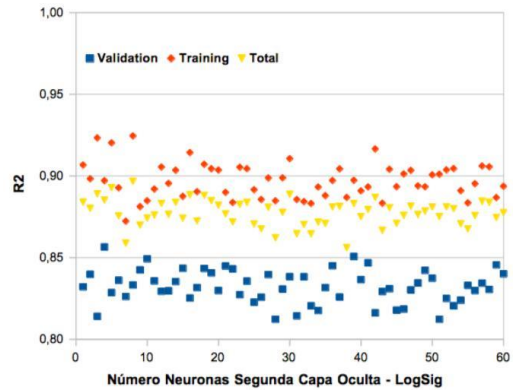


(c) R² vs Neuronas capa interna

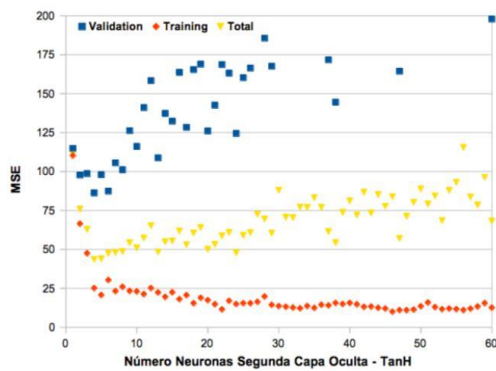
Figura 24. Desempeño de las redes 26:32:x:1 – LogSig y 26:35:x:1 – TanH con los históricos sin datos de la agrupación 3 y PEN hasta 100 mm/10 (20 repeticiones).



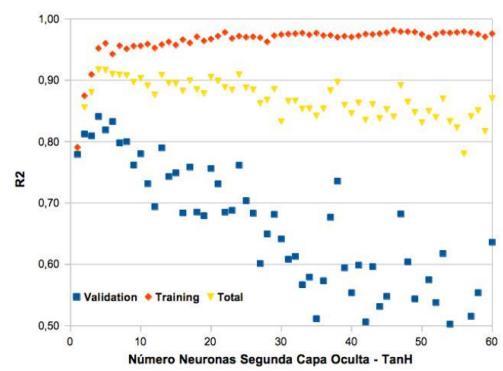
(a) MSE vs Neuronas segunda capa interna para LogSig.



(b) R^2 vs Neuronas segunda capa interna para LogSig.

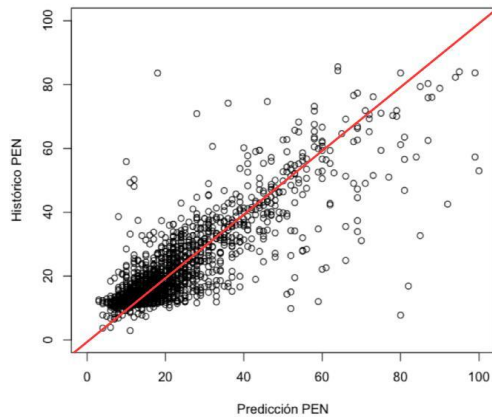


(c) MSE vs Neuronas segunda capa interna para TanH.

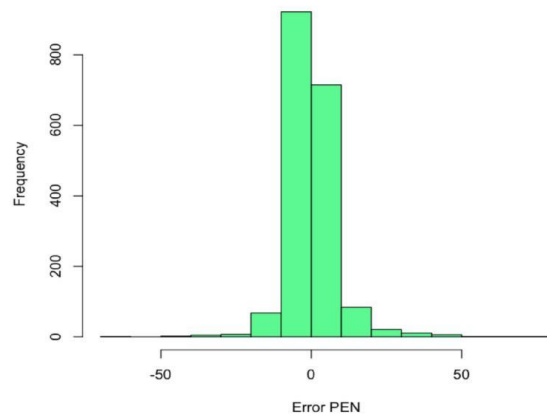


(d) R^2 vs Neuronas segunda capa interna para TanH.

Figura 25. Desempeño de la red 26:35:4:1 – TanH con los históricos sin datos de la agrupación 3 y PEN hasta 100 mm/10 (20 repeticiones). La línea roja indica predicción exacta del valor de la penetración.

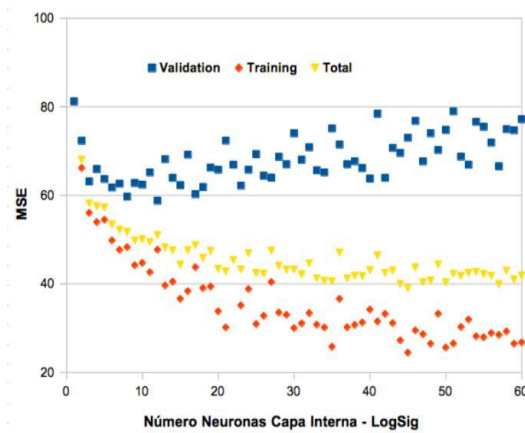


(a) Predicción PEN vs Histórico PEN red 26:35:4:1 – TanH.

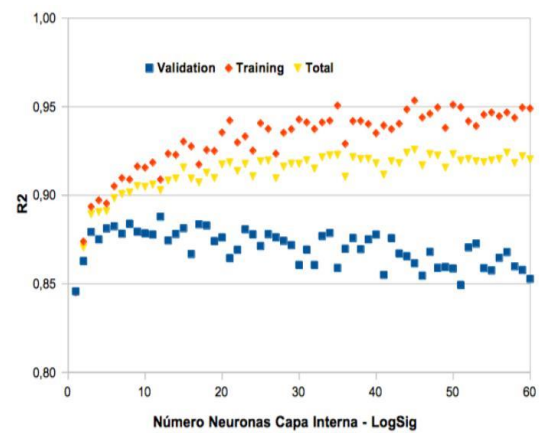


(b) Errores obtenidos con 26:35:4:1 – TanH.

Figura 26. Desempeño de las redes 26:x:1 – LogSig con los históricos sin datos de la agrupación 3 y PEN hasta 60 mm/10 (20 repeticiones).



(a) MSE vs Neuronas capa interna



(b) R^2 vs Neuronas capa interna

Figura 27. Desempeño de las redes 26:x:1 – TanH con los históricos sin datos de la agrupación 3 y PEN hasta 60 mm/10 (20 repeticiones).

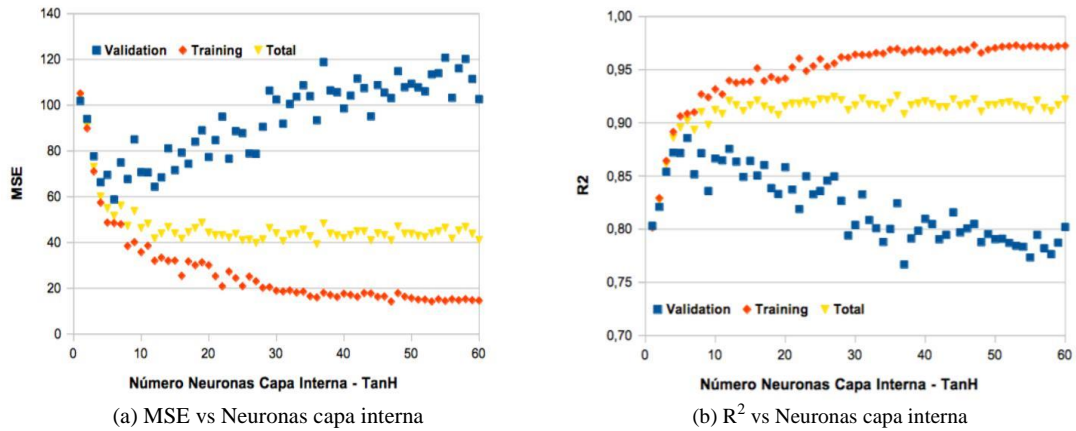
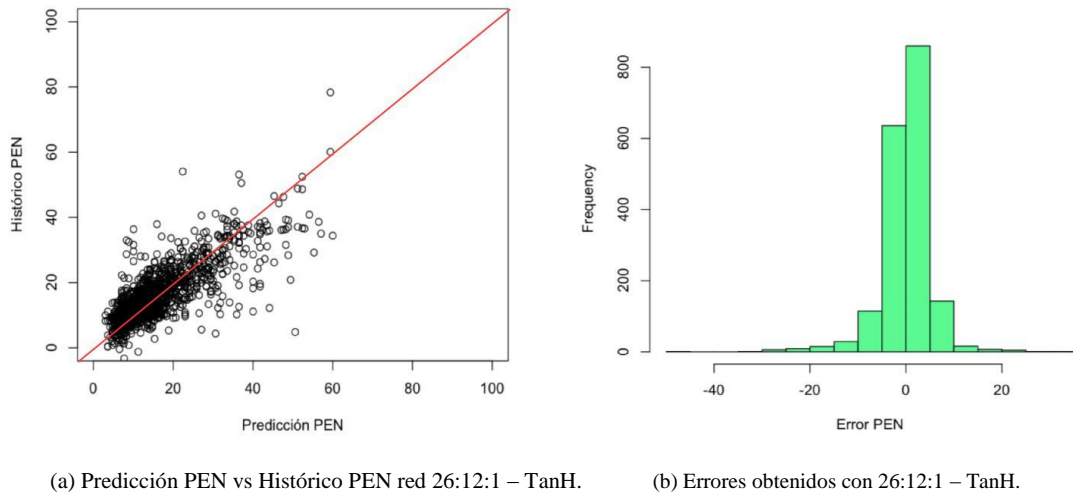


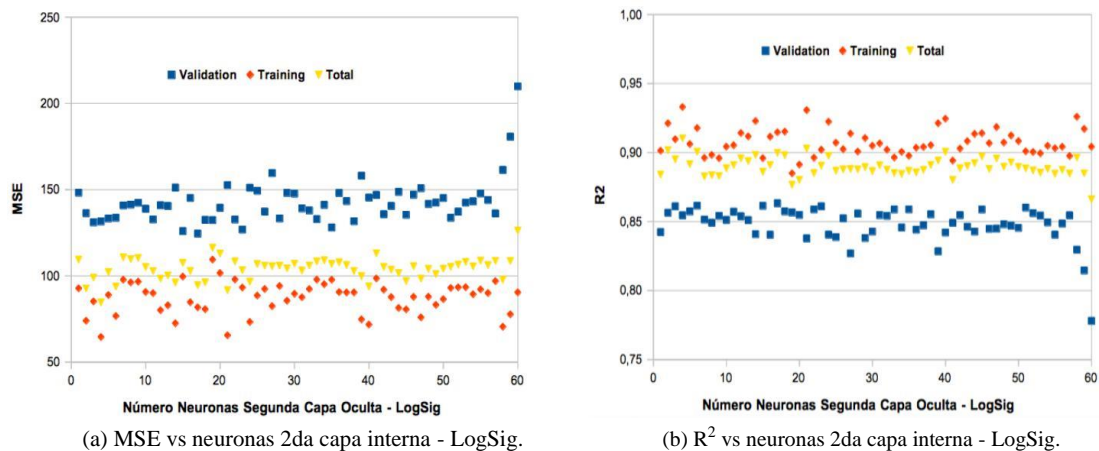
Figura 28. Desempeño red 26:12:1 – TanH con los históricos sin la agrupación 3 y hasta 60 mm/10 (20 repeticiones). La línea roja indica predicción exacta del valor de la penetración.



La estructura 26:12:1 – TanH aplicada a los históricos sin la agrupación 3 y con la restricción de $PEN \leq 60$ mm/10, presenta un desempeño similar con la estructura 26:35:4:1 – TanH aplicada a los históricos sin la agrupación 3 (26:35:4:1, $MSE_{all}=43.5$ mm²/100 y $R^2_{all}=0.917$; 26:12:1 – TanH, $MSE_{all}=41.7$ mm²/100 y $R^2_{all}=0.921$). La red 26:12:1 presenta una ligera ventaja ante la red 26:35:4:1 por

su distribución de los residuales, aunque pierde generalidad para tratar muestras con mayor valor de PEN. Sin embargo, una importante anotación se debe realizar en este aspecto, las muestras con valores elevados de PEN hacen referencia a operaciones atípicas de la unidad o estados dinámicos indeseados, durante la ventana de operación analizada. Con lo anterior, la red 26:12:1 – TanH entrenada con los históricos sin la agrupación 3 (muestras 3 meses antes de la reparación de la unidad) y con restricción de valores de $PEN \leq 60$ mm/10, corresponde a la estructura con mejor desempeño en la predicción de la penetración del asfalto generado por la U-2100 en la ventana de operación (enero 2015 a enero 2018). Por otra parte, la política de planeación actual (año 2018) para la planta U-2100 dirige la producción de fondos a asfalto 60/70 (con penetraciones entre 60 mm/10 y 70 mm/10, según norma comercial, Tabla 1). Con esto, la red más indicada para la predicción de las penetraciones en operación actual corresponde a la estructura 26:35:4:1 – TanH aplicada a los históricos sin la agrupación 3.

Figura 29. Desempeño de las redes 26:21:x:1 – LogSig con los históricos sin datos de la agrupación 3 y PEN hasta 60 mm/10 (20 repeticiones).



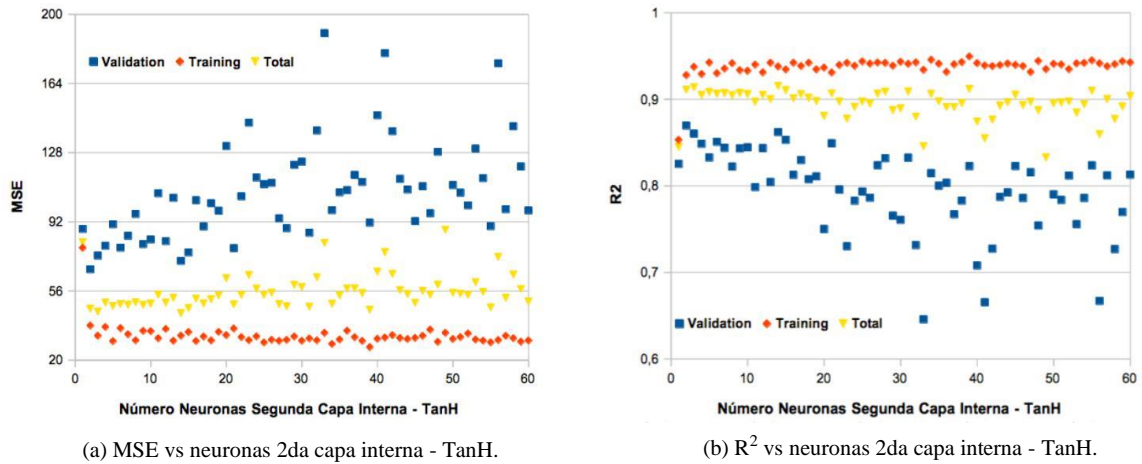
La Tabla 2, resume los desempeños obtenidos en el presente trabajo, con las diferentes arquitecturas y los diferentes tamaños de la base de históricos en la predicción de la penetración del asfalto generado en la U-2100. Según esta tabla,

los desempeños reportados por el uso de la función de activación TanH supera a los desempeños obtenidos con la función LogSig. Asimismo, la adición de una segunda capa oculta mejora los desempeños sólo para las redes entrenadas con los históricos sin la agrupación 3. También, la disminución del 16% de los datos incluyendo la restricción en PEN conduce a los mejores desempeños en las estructuras de red. La disminución adicional en los históricos aplicando un mayor número de agrupaciones (*kmeans*) no se vislumbra como una opción que pueda aportar un mejor ajuste; un mayor número de agrupaciones definiría escenarios de crudos con propiedades diferentes, sin embargo, podría carecer de una aplicación operativa ya que no se cuenta con propiedades adicionales al API para el crudo alimentado a la planta. Asimismo, el desempeño obtenido con las RNA es suficiente para una aplicación cuantitativa en la predicción de la penetración y una posible mejora del esquema de generación de asfalto actual en la GRB; las desviaciones estándar de las mejores redes (6.5 mm/10) resultaron comparables con el valor de repetibilidad (4 mm/10) especificado según la norma y trabajos previos reportados en la literatura basados en el espectro NIR (6.6 mm/10)^{59 60}.

⁵⁹ Blanco, M., MasPOCH, S., Villarroja, I., Peralta, X., González, J.M., Torres, J. (2000). Determination of the penetration value of bitumens by near infrared spectroscopy. *Analyst*, 125, 1823–1828.

⁶⁰ *Ibíd.*

Figura 30. Desempeño de las redes 26:12:x:1 – TanH con los históricos sin datos de la agrupación 3 y PEN hasta 60 mm/10 (20 repeticiones).



La estructura 26:12:1 – TanH fue sometida a un análisis de sensibilidad para determinar la influencia de las variables de entrada en la predicción de la PEN. Este análisis fue efectuado por medio de perturbaciones del 50% del valor de cada entrada manteniendo las demás con los valores iniciales, lo cual es denominado como procedimiento *input perturbation*⁶¹. La Figura 31 exhibe los valores obtenidos por la prueba de sensibilidad para la mejor red (Tabla 3). Esta figura presenta los valores obtenidos de MSEall y de R2all para cada una de las 26 variables, así como la referencia de comparación (MSEall=41.7 y el valor de R2all=0.92, línea negra punteada). Según la Figura 31, la variable que posee mayor influencia en la predicción en la arquitectura 26:12:1 – TanH corresponde a la presión de cima de la torre de vacío, Pcima_V, (círculo rojo), mientras que la temperatura del flujo de jet A, Tjet_A, corresponde a la variable menos influyente en la predicción. Otras variables influyentes, según la Figura 31 corresponden a la temperatura de gasóleo liviano de vacío, TGL_V, la carga a la torre de vacío, QV, y el reflujo medio de vacío, RefMed_V; lo anterior indica que la predicción de la

⁶¹ Gevrey, M.; Dimopoulos, I.; Lek, S. (2003). Review and comparison of methods to study to contribution of variables in artificial neural network models. *Ecological Modelling*, 160, 349-264.

PEN depende fuertemente de las condiciones de operación de la torre de vacío, según lo esperado.

Tabla 3. Resultados ajustes de las arquitecturas de red propuestas en el presente trabajo

RNA	Entrenamiento		Validación		Total		Datos
	MSE	R2	MSE	R2	MSEAll	R ² All	
26:37:1 – LS	62.1	0.934	154.0	0.833	89.9	0.905	
26:32:1 – TH	45.6	0.951	178.8	0.809	85.6	0.909	2119
26:37:3:1 – LS	64.0	0.933	143.7	0.844	88.0	0.907	3≤PEN≤100
26:32:2:1 – TH	49.1	0.947	197.3	0.791	93.6	0.901	
26:32:1 – LS	57.5	0.918	113.5	0.837	74.3	0.894	
26:35:1 – TH	37.0	0.947	131.0	0.810	65.2	0.907	1845
26:32:8:1 – LS	53.1	0.833	115.7	0.924	72.2	0.897	3≤PEN≤100
26:35:4:1 – TH	25.1	0.952	86.3	0.840	43.5	0.917	
26:21:1 – LS	30.1	0.942	72.3	0.864	42.8	0.918	1785
							3≤PEN≤100
26:12:1 – TH	32.0	0.940	64.3	0.875	41.7	0.921	
26:21:4:1 – LS	64.5	0.933	131.5	0.854	84.6	0.910	
26:12:3:1 – TH	32.7	0.937	74.4	0.860	45.2	0.914	

Complementariamente, la red 26:35:4:1 – TanH también fue analizada por el procedimiento *input perturbation* para determinar la influencia de las variables en la predicción de la PEN. La Figura 32 presenta los valores de MSEall y R²all para cada una de las 26 variables según la prueba de sensibilidad (referencia de comparación MSEall=43.5 mm²/100 y R²all=0.917, línea negra punteada). Según esta figura, las variables con mayor influencia en la predicción de PEN con la red 26:35:4:1 – TanH corresponden a flujo de gasóleo pesado de vacío, GP_V, la temperatura del gasóleo de vacío de reflujo, TGR_V, el flujo de gasóleo atmosférico de la torre atmosférica, GO_A, y el reflujo medio atmosférico, RM_A. Con lo anterior, el contenido de gasóleo que posee el crudo reducido procesado en la torre de vacío presenta la mayor influencia en la predicción, lo cual coincide

con valores elevados de PEN utilizados en el entrenamiento y la validación de esta red.

Figura 31. Resultados prueba de sensibilidad red 26:12:1 – TanH con los históricos sin datos de la agrupación 3 y PEN hasta 60 mm/10 (20 repeticiones).

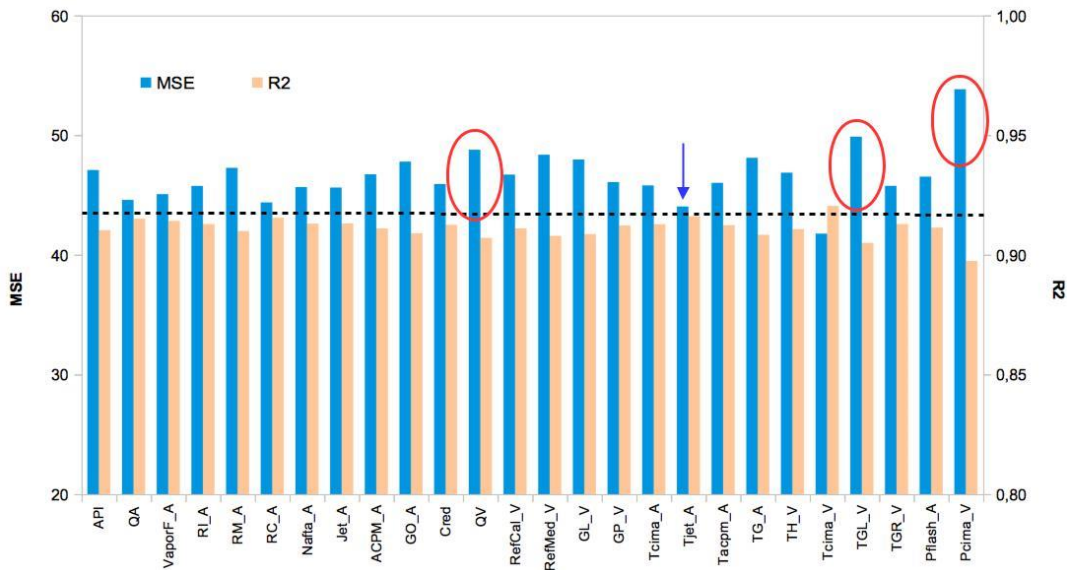
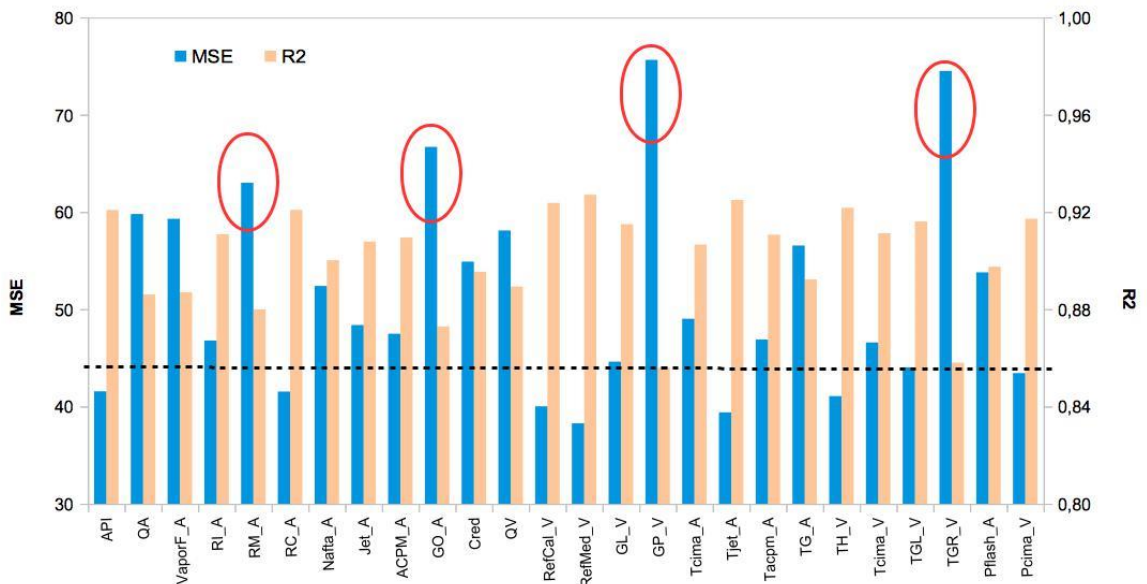


Figura 32. Resultados prueba de sensibilidad red 26:35:4:1 – TanH con los históricos sin datos de la agrupación 3 y PEN hasta 100 mm/10 (20 repeticiones).



5.5 IMPACTO DE LOS MODELOS DE RNA

La cantidad de reprocesos en la generación de cemento asfáltico en la refinería ha llegado a valores importantes (Figuras 1, 2), con lo cual el 30% de las preparaciones deben ser modificadas para alcanzar la restricción comercial 60/70 mm/10 en penetración. Según los desempeños mostrados por las RNA entrenadas y validadas con los históricos de la U-2100, la aplicación en un *soft sensor* conllevará a una disminución en la frecuencia de reprocesos. Esto puede ser estimado a partir de la desviación estándar de las redes validadas y por la consideración de una desviación mayor del método utilizado convencionalmente en la preparación del asfalto; el valor de desviación del método convencional puede ser reconsiderado a partir de los históricos de las preparaciones incluyendo las predicciones realizadas.

Comparando la desviación de 6.6 mm/10 mostrada por las redes con la reproducibilidad de 4 mm/10 de la norma ASTM D5-06, es posible aproximar que la aplicación de un *soft sensor* disminuirá la cantidad de reprocesos en aproximadamente la mitad de los casos reportados actualmente; la desviación considerada del método convencional fue de 13 mm/10. Con esto la eficiencia de la preparación de asfalto con visto bueno podrá aumentar, disminuyendo la cantidad de reprocesos y con el correspondiente aumento en el margen de refinación. Una estimación más elaborada sobre los impactos económico y operacional de la aplicación de la red neuronal enmarcada en un *soft sensor* es recomendada a partir de los históricos de las preparaciones dirigidas a asfalto de penetración 60/70.

6. CONCLUSIONES

El análisis de los procedimientos aplicados durante el desarrollo del presente trabajo de aplicación, conduce a las siguientes conclusiones:

La depuración y el análisis estadístico de los datos históricos de proceso de la unidad de destilación primaria U-2100 conducen a la definición de 26 variables para el desarrollo de modelos de predicción; estas 26 variables se caracterizan por exhibir baja dependencia entre ellas, basado en un valor para el coeficiente de correlación de Pearson inferior a 0.6.

La aplicación del método *kmeans* para la identificación de agrupaciones genera dos esquemas operativos para la operación de la unidad de destilación de crudos U-2100; el análisis de varianza entre los valores de penetración de los dos esquemas operativos muestra que los asfaltos producidos en estos esquemas poseen diferentes valores de penetración. El esquema operativo 1 se caracteriza por presentarse en las fechas anteriores a la reparación de la unidad; las gráficas de dispersiones entre las variables de operación muestra una disminución en la eficiencia de calentamiento en la unidad. Asimismo, el esquema operativo 2 se caracteriza por presentar muestras con fechas posteriores a la reparación de la unidad, con el respectivo aumento en la eficiencia de calentamiento.

Las arquitecturas de RNA entrenadas y validadas mostraron diferentes desempeños de predicción según la presencia de datos atípicos. Las arquitecturas con mejor desempeño corresponden a 26:35:4:1 – TanH, la cual reporta valores de $MSE_{all}=43.5 \text{ mm}^2/100$ y de $R^2_{all}=0.917$, con los datos históricos sin muestras cercanas a la fecha de la reparación de la unidad (agrupación 3), y la red 26:12:1 – TanH, la cual reporta valores de $MSE_{all}=41.7 \text{ mm}^2/100$ y $R^2_{all}=0.92$, con los datos históricos sin la agrupación 3 y con valores de penetración por debajo de 60 mm/10. Las desviaciones estándar de las mejores redes (6.5 mm/10) resultaron

comparables con el valor de repetibilidad (4 mm/10) especificado según la norma y trabajos previos reportados en la literatura basados en el espectro NIR (6.6 mm/10)

El análisis de sensibilidad conduce a la identificación de la presión de la cima de la torre de vacío de la U-2100 como la variable con mayor influencia en la predicción de la penetración realizada por la estructura 26:12:1 – TanH. De igual forma, el análisis de sensibilidad para la estructura 26:35:4:1 – TanH muestra que el contenido de gasóleo en el crudo reducido procesado en la torre de vacío afecta mayoritariamente la predicción de la penetración del asfalto; lo anterior resulta como consecuencia de la presencia de muestras con valores elevados de penetración.

7. RECOMENDACIONES

Continuar con el desarrollo de RNA para la predicción de otras propiedades de los asfaltos generados en la GRB.

Reentrenar las redes propuestas en el documento con cada cambio de esquema operativo o con cada cambio en la política para la especificación de penetración del asfalto generado en la unidad.

REFERENCIAS

Alaei, J. & Nariman, E. (2015). Simulation of hydrodesulfurization unit for natural gas condensate with high sulfur content. *Applied Petrochemical Research*, 107.

Berry, M.J.A. & Linoff, G. (1997). Data Mining Techniques. *John Wiley & Sons*. Nueva York. Berrocal, J.L.A., Paniagua, C.G.F. & Rodríguez, A.F.Z. (2006). Mejoras en la recuperación de información en la Web mediante el tratamiento de la información de los enlaces. *Scire: Representación y organización del conocimiento*, 2006, 197-209.

Berry, M.J.A., Linoff, G. (1987). Data mining techniques. *John Wiley & Sons*. Nueva York.

Berthold, M. & Hand, D.J. (2007). Intelligent Data Analysis: An Introduction. *Springer-Verlag*, 2nd edition, Berlin.

Blanco, M., MasPOCH, S., Villarroya, I., Peralta, X., González, J.M., Torres, J. (2000). Determination of the penetration value of bitumens by near infrared spectroscopy. *Analyst*, 125, 1823–1828.

Blanco, M., MasPOCH, S., Villarroya, I., Peralta, X., González, J.M., Torres, J. (2001). Determination of physical properties of bitumens by use of near-infrared spectroscopy with neural networks. Joint modelling of linear and non-linear parameters. *Analyst*, 126, 378–382.

Bolf, N., Ivandic, M., Galinec, G. (2008). Soft sensors for crude distillation unit product properties estimation and control. *Control and Automation*, 16th Mediterranean Conference on. IEEE. p. 1804-1809.

Bolf, N.; Ivandic, M.; Galinec, G. (2008). Soft sensors for crude distillation unit product properties estimation and control. *Control and Automation*, 16th Mediterranean Conference IEEE, p. 1804-1809.

Del Brío, B. & Sanz, M.A. (2002). Redes Neuronales y Sistemas Difusos. 2da Ed. México D.F, México. *Alfaomega*.

Devore, J.L. (2008). Probabilidad y Estadística para Ingeniería y Ciencias. 7ma edición.

Flórez, R. & Fernández, J. (2008). Las Redes Neuronales Artificiales. Oleiros, La Coruña: *Netbiblo*.

Fortuna, L. et al. (2007). Soft sensors for monitoring and control of industrial processes. *Springer Science & Business Media*.

Freeman, J.A.; Skapura, D.M. (1991). Neural Networks: Algorithms, Applications and Programming Techniques. *Addison-Wesley*.

Gevrey, M.; Dimopoulos, I.; Lek, S. (2003). Review and comparison of methods to study to contribution of variables in artificial neural network models. *Ecological Modelling*, 160, 349-264.

Hair J.J., Black, W., Babin, B., Anderson, R. (2009). Multivariate Data Analysis. 7th edition. USA: *Prentice Hall*.

Hilera, J.R., & Martínez, V.J. (2000). Redes Neuronales Artificiales Fundamentos, Modelos y Aplicaciones. Madrid España: *Alfaomega*.

Isasi, P., Galvan, I. (2004). *Redes neuronales artificiales enfoque práctico*. Madrid: Pearson.

Journal of microscopy, 254, 1, p. 31-41.

Lippmann, R. (1987). An introduction to computing with neural nets. *ASSPMagazine*, IEEE.

Liu, G., Wang, L., Qu, H., Shen, H., Zhang, X., Zhang, S., Mi, Z. 2007. Artificial neural network approaches on composition–property relationships of jet fuels based on GC–MS. *Fuel*, 86, 2551-2559.

Mendenhall, W. (1987). *Introducción a la probabilidad y la estadística*. p. 551-652.

Mongomery, D.C. & Runger, G.C. (1996). *Probabilidad y estadística aplicadas a la ingeniería*, México, *Mc Graw-Hill*.

Nejad, F.M., Mehrabi, A., Zakeri, H. (2015). Prediction of Asphalt Mixture Resistance Using Neural Network via Laboratorial X-ray Images. *Journal of Industrial and Intelligent Information*, 3, no 1.

Nejad, F.M.; Mehrabi, A.; Zakeri, H. (2015). Prediction of Asphalt Mixture Resistance Using Neural Network via Laboratorial X-ray Images. *Journal of Industrial and Intelligent Information*, 3, 1.

Ozgan, E. (2011). Artificial neural network based modelling of the Marshall Stability of asphalt concrete. *Expert Systems with Applications*, 38, p. 6025-6030.

Ozsahin, T.S., Oruc, S. (2008). Neural network model for resilient modulus of emulsified asphalt mixtures. *Construction and Building Materials*, 22, p. 1436-1445.

Ozturk, H.I., Kutay, M.E. (2014). An artificial neural network model for virtual Superpave asphalt mixture design. *International Journal of Pavement Engineering*, 15, 2, p. 151-162.

Piloto-Rodríguez, R., Sánchez-Borroto, Y., Lapuerta, M., Goyos-Pérez, L. 2013. Prediction of the cetane number of biodiesel using artificial neural networks and multiple linear regression. *Energy Conversion and Management*, 65, 255–261.

Ponce, C.P. (2010). Inteligencia artificial con aplicaciones a la ingeniería. 1ra Ed. México: Alfaomega.

Robert, R. (2002). Neural virtual sensor for the inferencial prediction of product quality from process variables. *Universitat Rovira i Virgili*.

Rao, S.G. & Govardhan. (2015). Performance Validation of the Modified K-means Clustering Algorithm cluster Data. *International Journal of Scientific and Engineering Research*, 6, 726-730.

Rogina, A., Šiško, I., Mohler, I., Ujević, Ž., Bolf, N. 2011. Soft sensor for continuous product quality estimation (in crude distillation unit). *Chem. Eng. Res. Des.*, 89, 2070-2077.

Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. comp. App. Math.*, 20, 53-65.

Rovert, R. (2002). Neural virtual sensor for the inferencial prediction of product quality from process variables. *Universitat Rovira i Virgili*.

Shafabakhsh, G.H., Ani, O.J., Talebsafa, M. (2015). Artificial neural network modeling (ANN) for predicting rutting performance of nano-modified hot-mix asphalt mixtures containing steel slag aggregates. *Construction and Building Materials*, 85, p. 136-143.

Swingler, K. (1996). Applying Neural Networks, A Practical Guide. *Press Limited Oval Road London NW1*.

Tarefer, R.A.; Ahsan, S. (2014). Neural network modelling of asphalt adhesion determined by AFM.

Thakare, Y.S. & Bagal, S.B. 2015. Performance Evaluation of K-means Clustering Algorithm with Various Distance Metrics. *Int. J. Compt. App.*, 110, 12-16.

Watkins, R. (1979). Petroleum Refinery Distillation (Segunda edición ed.). Gulf publishing Co. Wauquier, J. (2004). El Refino Del Petróleo: Petróleo crudo, Productos Petrolíferos, Esquemas de Fabricación. España: Ediciones Díaz de Santos.

Yang, S.H., et al. (1998). Soft sensor based predictive control of industrial fluid catalytic cracking processes. *Chemical Engineering Research and Design*, 76, p. 499-508.

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

ANEXOS

ANEXO 1. VARIABLES DE LA UNIDAD U-2100 UTILIZADAS EN EL TRABAJO Y SUS CORRESPONDIENTES ABREVIATURAS

Abreviatura	Variable Correspondiente
A P I	Gravedad API crudo de alimentación a la unidad
ACPM_A	Flujo de acpm torre atmosférica T-2101, bbpd.
A s f	Flujo de asfalto torre de vacío T-2103, bbpd.
Cred	Flujo de crudo reducido torre atmosférica T-2101, bbpd.
GL_V	Flujo de gasóleo liviano torre de vacío T-2103, bbpd.
GO_A	Flujo de gasóleo torre atmosférica T-2101, bbpd.
GP_V	Flujo de gasóleo pesado torre de vacío T-2103, bbpd.
Jet_A	Flujo de jet torre atmosférica T -2101, bbpd.
Nafta_A	Flujo de nafta torre atmosférica T-2101 , bbpd.
Pcima_A	Presión de cima torre atmosférica T-2101, °F. PSI
Pcima_V	Presión de cima torre de vacío T-2103, °F. IN HG
PEN	Penetración asfalto generado, mm/10
Pflash_A	Presión zona flash torre atmosférica T-2101, °F. PSI
QA	Carga de crudo de alimentación a la unidad, bbpd
QV	Carga de crudo reducido a la torre de vacío T-2103, bbpd.
RC_A	Reflujo de cima torre atmosférica T-2101 , bbpd.
RefCal_V	Reflujo caliente torre de vacío T-2103, bbpd.
RefCim_V	Reflujo de cima torre de vacío T-2103, bbpd.
RefMed_V	Reflujo medio torre de vacío T-2103, bbpd.
RI_A	Reflujo inferior torre atmosférica T-2101 , bbpd.
RM_A	Reflujo medio torre atmosférica T-2101 , bbpd.
Tacpm_A	Temperatura producto acpm torre atmosférica T-2101, °F .
Tasf	Temperatura producto asfalto torre de vacío T-2103, °F.
Tcima_A	Temperatura de cima torre atmosférica T-2101, °F .
Tcima_V	Temperatura de cima torre de vacío T-2103, °F.
Tcred	Temperatura producto crudo reducido torre atmosférica T-2101, °F .
Tflash_A	Temperatura zona flash torre atmosférica T-2101, °F .
Tflash_V	Temperatura zona flash torre de vacío T-2103, °F.
TG_A	Temperatura producto gasóleo torre atmosférica T-2101, °F.
TGL_V	Temperatura producto gasóleo liviano torre de vacío T-2103, °F.

Abreviatura	Variable Correspondiente
TGP_V	Temperatura producto gasóleo pesado torre de vacío T-2103, °F.
TGR_V	Temperatura gasóleo de reciclaje torre de vacío T-2103, °F.
TH_V	Temperatura salida horno de vacío, °F.
Tjet_A	Temperatura producto jet torre atmosférica T-2101, °F .
Tpre_A	Flujo de gasóleo liviano torre de vacío T-2103, bbpd.
Tsal_HA	Temperatura de precaliente flujo de alimentación torre atmosférica T-2101, °F .
VaporF_A	Flujo de vapor de agua a la torre atmosférica T-2101, lb/h.

ANEXO 2. CÓDIGO EN R PARA EL ENTRENAMIENTO Y LA VALIDACIÓN DE LAS RNA UTILIZANDO LA FUNCIÓN *NEURALNET*

```
data<-Kmean13
n<-names(data)
maxs <- apply(data[,1:27], 2, max)
mins <- apply(data[,1:27], 2, min)
scaled01=normalizeData(data[,1:27],type="0_1")
samplesize = as.integer(0.70 * nrow(data))
set.seed(nrow(data))
i<-0
MSEnn<-matrix(nrow=60,ncol=5)
MSEnnt<-matrix(nrow=60,ncol=5)
MSEnntot<-matrix(nrow=60,ncol=5)
R2_<-matrix(nrow=60,ncol=5)
R2_t<-matrix(nrow=60,ncol=5)
R2_tot<-matrix(nrow=60,ncol=5)
for (k in 1:60) {
  i=i+1
  for (j in 1:5) {
    index = sample(seq_len(nrow(scaled01)), size = samplesize)
    Dtrain<- scaled01[index,]
    Dtest <- scaled01[-index,]
    colnames(Dtrain) <- n
    colnames(Dtest) <- n
    f <- as.formula(paste("PEN ~", paste(n[!n %in% "PEN"], collapse = " +
    ""))) model0 <- neuralnet(f,data=Dtrain,hidden=c(32, k),linear.output=T,
    err.fct="sse",act.fct="logistic",stepmax=100000,threshold=0.2,
    algorithm="rprop+",rep=3) pr.nn <- compute(model0,Dtest[,1:26])
    pr.nn_ <- pr.nn$net.result*(max(data$PEN)-min(data$PEN))+min(data$PEN)
```

```

test.r <- Dtest[,27]*(max(data$PEN)-min(data$PEN))+min(data$PEN)
ym_<-mean(Dtest[,27])
MSEnn[i,j]<-sum((test.r-pr.nn_)^2)/nrow(Dtest)
R2_[i,j]<-1-sum((test.r-pr.nn_)^2)/sum((test.r-ym_)^2)
pr.nn <- compute(model0,Dtrain[,1:26])
pr.nn_t <- pr.nn$net.result*(max(data$PEN)-min(data$PEN))+min(data$PEN)
test.rt<- Dtrain[,27]*(max(data$PEN)-min(data$PEN))+min(data$PEN)
ym_t<-mean(Dtrain[,27])
MSEnnt[i,j]<-sum((test.rt-pr.nn_t)^2)/nrow(Dtrain)
R2_t[i,j]<-1-sum((test.rt-pr.nn_t)^2)/sum((test.rt-ym_t)^2)
MSEnnot[i,j]=(sum((test.r-pr.nn_)^2)+sum((test.rt-pr.nn_t)^2))/(nrow(data))
R2_tot[i,j]<-1-(sum((test.r-pr.nn_)^2)+sum((test.rt-pr.nn_t)^2))/(sum((test.r-
ym_)^2)+sum((test.rt-ym_t)^2))
}
}

```