

**PROCESO DE HIDROTRATAMIENTO: PREDICCIONES MEDIANTE
REGRESION LINEAL MULTIVARIADA Y REGRESIÓN PLS UTILIZANDO
DATOS HISTORICOS DE OPERACION DE UNA UNIDAD HDT**

GUSTAVO ADOLFO CARVAJAL CIPAMOCHA

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

2017

**PROCESO DE HIDROTRATAMIENTO: PREDICCIONES MEDIANTE
REGRESION LINEAL MULTIVARIADA Y REGRESION PLS UTILIZANDO
DATOS HISTORICOS DE OPERACION DE UNA UNIDAD HDT**

GUSTAVO ADOLFO CARVAJAL CIPAMOCHA

**Trabajo de grado presentado como requisito para optar por el título de
Ingeniero Químico**

**Director:
GIOVANNI MORALES MEDINA, I.Q., Dr.**

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

2017

DEDICATORIA Y AGRADECIMIENTOS

En primer lugar a la divina providencia, ya que desde el inicio de mis estudios universitarios, no me desamparo en ningún instante. Me brindo fortaleza en los momentos difíciles y me otorgo sabiduría para continuar y culminar este trabajo de la mejor manera.

A mis padres Audias y en especial a mi Madre Maria Estela, por su acompañamiento, orientación por el buen camino, comprensión y apoyo incondicional desde el inicio de mis estudios universitarios. Jamás hubiese logrado esto sin ustedes. Me enseñaron que las cosas que valen la pena en la vida son difíciles pero vale la pena emprender sin desfallecer hasta lograr las metas propuestas.

Al profe PhD. Giovanni Morales Medina, Le agradezco grandemente su acompañamiento, colaboración, paciencia y asesoría en este proyecto.

CONTENIDO

| | Pág. |
|---|-------------|
| INTRODUCCION | 12 |
| 1. OBJETIVOS..... | 14 |
| 1.1. OBJETIVO GENERAL..... | 14 |
| 1.2. OBJETIVOS ESPECIFICOS | 14 |
| 2. MARCO TEORICO | 15 |
| 2.1. PROCESO DE HIDROTRATAMIENTO | 15 |
| 2.1.1 Variables que influyen en el proceso de HDT. | 16 |
| 2.2 ESTADO DEL ARTE | 17 |
| 2.3. REGRESION LINEAL MULTIVARIADA (RML) | 19 |
| 3. METODOLOGIA | 22 |
| 3.1 DESCRIPCION GENERAL SOBRE LOS PROCEDIMIENTOS LLEVADOS A CABO EN EL PRESENTE TRABAJO | 23 |
| 4. ANALISIS Y DISCUSION DE RESULTADOS | 24 |
| 4.1 ANÁLISIS CLÚSTER NO JERÁRQUICO <i>K-MEANS</i> : | 26 |
| 4.2 PLANTEAMIENTO Y CALIBRACIÓN DE LOS MODELOS DE RML: | 29 |
| 4.3 PLANTEAMIENTO Y CALIBRACIÓN MODELOS DE PREDICCIÓN MEDIANTE REGRESIÓN PLS..... | 33 |
| BIBLIOGRAFIA..... | 42 |
| ANEXOS..... | 46 |

LISTA DE TABLAS

| | Pág. |
|--|-------------|
| Tabla 1. Análisis univariado sobre la matriz de datos empleada en el presente trabajo..... | 24 |
| Tabla 2. Matriz de correlación entre las variables consideradas..... | 25 |
| Tabla 3. Distancias inmtragrupales y suma de cuadrados caso de estudio IV | 28 |
| Tabla 4. Coordenadas centroides de las variables consideradas caso de estudio IV | 28 |
| Tabla 5. Matriz de distancias intergrupales caso de estudio IV | 28 |
| Tabla 6. Resultados casos de predicción mediante regresión multivariada..... | 30 |

LISTA DE FIGURAS

| | Pág. |
|--|-------------|
| Figura 1. Diagrama de bloques metodología..... | 22 |
| Figura 2. Dendograma análisis cluster jerarquico realizado sobre la totalidad de las variables consideradas en este trabajo..... | 25 |

LISTA DE ANEXOS

| | Pág. |
|--|-------------|
| ANEXO A. MÉTODO DE ESTANDARIZACION MEAD AND CENTER..... | 46 |
| ANEXO B. ANÁLISIS CLÚSTER NO JERÁRQUICO K-MEANS..... | 47 |
| ANEXO C. ANALISIS CLUSTER JERARQUICO | 50 |
| ANEXO D. REGRESION POR MINIMOS CUADRADOS (PLS)..... | 50 |
| ANEXO E. GRAFICAS VARIABLES CONSIDERADAS CONTRA EL NÚMERO DE OBSERVACION SOBRE LA MATRIZ FINAL DE DATA SELECCIONADA . | 54 |
| ANEXO F. ANALISIS CLUSTER K- MEDIAS SOBRE LAS OBSERVACIONES DE LAS VARIABLES CONSIDERADAS EN ESTE TRABAJO | 59 |
| ANEXO G. WORKSHEETS MINITAB 17 CASOS DE PREDICCION MEDIANTE RML..... | 64 |
| ANEXO H. CASOS DE PREDICCION I, IV Y V MEDIANTE RML UTILIZANDO EL METODO DE ELIMINACION HACIA ATRÁS | 68 |
| ANEXO I. GRAFICAS DE SELECCIÓN MODELO DE REGRESION, DE RESPUESTA DE REGRESION PLS Y DE COEFICIENTES. CASO DE ESTUDIO II. | 74 |
| ANEXO J. GRAFICAS DE SELECCIÓN MODELO DE REGRESION, DE RESPUESTA DE REGRESION PLS Y DE COEFICIENTES. CASO DE ESTUDIO III. | 75 |

RESUMEN

TITULO: PROCESO DE HIDROTRATAMIENTO: PREDICCIONES MEDIANTE REGRESION LINEAL MULTIVARIADA Y REGRESIÓN PLS UTILIZANDO DATOS HISTORICOS DE OPERACION DE UNA UNIDAD HDT.*

AUTOR: GUSTAVO ADOLFO CARVAJAL CIPAMOCHA**

PALABRAS CLAVE: HIDROTRATAMIENTO, PREDICCIONES, ESTADISTICA UNIVARIADA Y MULTIVARIADA, REGRESION LINEAL MULTIPLE, REGRESION PLS, CORRELACION LINEAL.

En el presente trabajo, utilizando estadística univariada y multivariada; se analizó una matriz de data histórica de operación tomada de una unidad HDT, con el fin de analizar el comportamiento de las observaciones y realizar el planteamiento de modelos matemáticos para la predicción de variables de salida en función de variables de operación y de entrada comúnmente consideradas por la comunidad científica en el análisis del proceso HDT. Se tuvieron en cuenta doce variables, como se presenta más adelante en este trabajo.

Inicialmente, se llevó a cabo un análisis estadístico univariado, donde se calcularon parámetros estadísticos clave como el valor máximo, el mínimo, la media y la desviación estándar de cada variable respectivamente. En seguida, se realizó la matriz de correlación entre las variables consideradas, un análisis estadístico no jerárquico Clúster k-medias y análisis Cluster jerárquico (dendograma). A partir de estos tres análisis, se determinó que las variables consideradas, casi en su totalidad, presentaron una muy baja correlación y una alta variabilidad entre si. Finalmente, se establecieron cinco casos de predicción mediante regresión lineal multivariada; se obtuvieron buenos resultados (R^2 ajustado > 0,9) en dos de los cinco casos de predicción planteados. Los buenos resultados obtenidos; se debieron a la alta correlación lineal evidenciada entre las variables flujo de hidrogeno a hidrotreatar en la alimentación a la unidad HDT, relación hidrocarburo/hidrogeno alimentados a la zona de reacción, y el flujo de diésel hidrotreatado a la salida de la unidad HDT. Estos dos casos de estudio, adicionalmente, se validaron mediante regresión PLS. Se encontró que la variación del grado de ajuste en la estimación pasando de tres a los nueve predictores considerados inicialmente apenas incrementaba de manera insignificante. Esto se justifica, en la alta correlación lineal entre las tres variables mencionadas anteriormente, en la baja influencia de las demás variables predictoras en la predicción y a la baja correlación existente entre casi la totalidad de variables consideradas.

* Trabajo de grado

** Facultad De Ingenierías Físicoquímicas. Escuela De Ingeniería Química. Director: Giovanni Morales Medina,

ABSTRACT

TITLE: HYDROTREATMENT PROCESS: PREDICTIONS USING MULTIPLE LINEAR REGRESSION AND PLS REGRESSION TAKING HISTORIC DATA FROM AN HDT TREATMENT UNITY.*

AUTHOR: GUSTAVO ADOLFO CARVAJAL CIPAMOCHA**

KEY WORDS: HIDROTREATMENT, PREDICTIONS, UNIVARIATED AND BIVARIATED STATISTICS, MULTIPLE LINEAR REGRESSION, PLS REGRESSION, LINEAR CORRELATION.

In this paperwork, it was analyzed, using univariate and multivariate statistics, a matrix of historic operation HDT process data in order to study the data behaviour and also in order to propose mathematical models to predict outlet variables in function of inlet and operation variables commonly considered within the HDT process analysis by scientific community all over the world. Twelve variables were taken in account as it is shown forward into this work.

At first, it was run an univariate statistics analysis and were calculated key statistic parameters like minimum and maximum values, the average and standard deviation of each variable taken in account, respectively. After, it was made the variables correlation matrix, among all the variables that were taken in account, an statistic cluster k-means analysis and a no jerarquic analysys (dendogram). It was found from these three last procedures that the analyzed data showed a high range of variability and mainly low correlation coefficients among the variables considered. Finally, were established five prediction cases using multiple linear regression and were obtained great results in two of them (R^2 **adjusted** > 0,9). It was because it was found a strong linear correlation among variables: inlet Diesel flow to the HDT unity, relationship between hydrocarbon/hydrogen flow fed to reaction zone and outlet Diesel flow. These two cases were also estimated using PLS regression. It was also found that the adjust square coeficient variation was very low with the estimation taking in account three predictors in comparison to take in account nine predictors. It's justified on the linear strong relationship among the three variables that were mentioned before, the low variables correlation coefficients among them and It's also justified on the low influence of the additional predictory variables that were taken in account for prediction in these two cases.

* Degree work

** Faculty of Physicochemical Engineering. School Of Chemical Engineering. Director: Giovanni Morales Medina

INTRODUCCION

Uno de los procesos más importantes en la industria de los hidrocarburos, particularmente, en la refinación del petróleo es el de hidrotratamiento (HDT) debido a sus implicaciones técnicas, económicas y ambientales [1] [2]. Por medio del proceso HDT, el azufre y el nitrógeno contenidos en diferentes cortes del petróleo (desde nafta ligera hasta residuo de vacío), son removidos por hidrogenación catalítica a temperaturas y presiones elevadas. Adicionalmente, diferentes tipos de compuestos de alto peso molecular y compuestos insaturados son fraccionados e hidrogenados para reducir el contenido de aromáticos y olefinas [1] [2] [4] [5].

Los contenidos de azufre elevados en combustibles fósiles como el diésel, generan emisiones de gases de efecto invernadero contribuyendo a los fenómenos de calentamiento global, lluvia ácida, muerte de fauna y flora, e impactos nocivos en la calidad de vida de las personas [1] [3] [4]. En los motores, los contenidos elevados de azufre generan la aparición de depósitos que taponan los inyectores, generan pérdidas del combustible y corroen los anillos, los pistones y los cilindros. Por todo lo anterior, la comunidad internacional impone normativas ambientales cada vez más estrictas en cuanto al contenido de aromáticos, mercaptanos, y principalmente de azufre, en los combustibles de origen fósil [3] [4].

Para cumplir con la normatividad ambiental, las refinerías ahondan esfuerzos en el análisis de sus procesos para definir las condiciones operativas que conlleven a la consecución de los respectivos estándares. Uno de los procedimientos utilizados para definir las condiciones operativas apropiadas consiste en analizar los datos históricos de proceso por medio de algoritmos de estadística multivariada [2][5]. Algoritmos como el análisis por componentes principales (PCA) y el método de agrupación *k-mean* permiten dilucidar la relación entre las diferentes variables independientes de un proceso y generar casos operacionales representativos según los valores de estas variables. De otro lado, métodos de ajuste como la regresión

lineal múltiple (RLM) y la regresión por mínimos cuadrados parciales (PLS) permiten la generación de ecuaciones de predicción de variables de salida de los procesos. Con el soporte de los resultados de la estadística multivariada es posible definir regiones operativas que conlleven a la consecución de las metas en las unidades HDT.

En el presente trabajo se realizó un análisis clasificatorio y se realizaron predicciones, de variables de operación y de salida, de interés para el proceso HDT tales como flujo de diésel hidrotratado y su contenido de azufre, relación hidrocarburo/hidrogeno alimentado a la zona de reacción, hidrogeno total alimentado, y cambio de temperatura en la zona de reacción. La predicción se efectúa utilizando datos históricos de planta, correspondientes a la operación de la unidad de HDT de la refinería de Barrancabermeja de Ecopetrol S.A.

1. OBJETIVOS

1.1. OBJETIVO GENERAL

Calibrar y validar modelos de predicción mediante estadística multivariada utilizando regresión lineal multivariada (RML) y regresión por mínimos cuadrados parciales (PLS) para el proceso de hidrotatamiento de la refinería de Barrancabermeja.

1.2. OBJETIVOS ESPECIFICOS

- Desarrollar análisis estadísticos univariado y bivariado, a una base de datos histórica para el proceso HDT de la GRB, con el fin de identificar puntos atípicos, determinar la influencia de variables de proceso y la correlación entre las mismas.
- Aplicar análisis estadísticos clúster jerárquico *k-means* y no jerárquico, para determinar diferentes agrupaciones y variables influyentes en la reproducción de la varianza de las variables independientes de la base de datos.
- Calibrar y validar diferentes modelos de regresión lineal multivariada (RLM) y regresión por mínimos cuadrados parciales (PLS) a la base de datos históricos.

2. MARCO TEORICO

2.1. PROCESO DE HIDROTRATAMIENTO

En el proceso HDT se lleva a cabo la hidrogenólisis de compuestos nitrogenados, sulfurados, oxidados y la hidrogenación de aromáticos y olefinas, con el objetivo de mejorar la calidad del petróleo crudo y/o de sus fracciones. En la industria, el proceso se realiza en reactores de lecho fijo tubulares, donde se alimenta hidrogeno, con el fin de ponerlo en contacto con las fracciones de petróleo, en presencia de un catalizador metálico (tipo sulfuro). La reacción se lleva a cabo bajo condiciones de alta presión y temperatura, con el fin de obtener productos con concentraciones muy bajas de los contaminantes. Se usan catalizadores compuestos por metales de transición, como el cobalto, molibdeno, wolframio, hierro o níquel, dado que estos metales evidencian una alta actividad frente a otros compuestos [4] [5].

El proceso HDT consiste de reacciones de hidrogenación (HDO), hidrogenación (HDN), hidrogenación (HDA), hidrogenación (HDS) e hidrogenación (HDM). Dependiendo de los compuestos contenidos en el crudo a hidrotatar. En el presente trabajo, la reacción principal a tener en cuenta es la hidrogenación (HDS), en la cual los compuestos con contenido de azufre tales como sulfuros, bisulfuros, tiofenos, mercaptanos, benzotiofenos y dibenzotiofenos; reaccionan con el hidrogeno. La reacción es exotérmica e irreversible a condiciones normales de proceso; en consecuencia, se rompen los enlaces del carbono-azufre y se adiciona el hidrogeno en el lugar vacío dejado por el azufre. Por esto se obtienen fracciones de petróleo de mejor calidad mientras el azufre se remueve como sulfuro de hidrogeno (H_2S) [4]. Los compuestos aromáticos, mezclados con azufre como lo son el tiofeno, benzotiofeno y

dibenzotiofeno son más difíciles de eliminar dado que poseen puntos de ebullición altos, peso molecular alto, lo que causa cierta resistencia a la remoción [6].

El contenido de azufre es uno de los parámetros de control más importantes de calidad del diésel comercial, debido a los impactos ambientales generados por la emisión de óxidos de azufre (SO_x), que son generados durante la combustión del diesel, los cuales se convierten en ácidos al ponerse en contacto con el vapor de agua presente en la troposfera. Por ejemplo, el ácido sulfúrico (H_2SO_4), ya que es producido en presencia del dióxido de azufre (SO_2) [9]. Por esto, la legislación colombiana en la más reciente resolución 90963 de 2014, en correspondencia con los criterios de calidad de los biocombustibles para su uso en motores que operan con diésel como componente de la mezcla; establece que el diésel comercial, debe contener un 2% de biodiesel y un 98% de diésel de origen fósil. Adicionalmente, establece que el contenido de azufre permitido en el diésel comercial debe estar por debajo de las 50 ppm [10].

2.1.1 Variables que influyen en el proceso de HDT. Las variables que tienen influencia en el proceso HDT son la velocidad espacial del líquido, la relación hidrogeno/hidrocarburo alimentados a la unidad HDT, la presión parcial del hidrogeno alimentado y la temperatura. Según reportes de la literatura, si el hidrocarburo a hidrotreatar presenta bajos valores de propiedades tales como el punto final de ebullición y la densidad, la actividad de la reacción de HDS es baja, por lo que esta condición demanda tiempos de retención menor y condiciones de operación extremas de presión y temperatura. Solamente así, es posible obtener el hidrocarburo hidrotreatado con bajos contenidos de azufre bajo estas condiciones [11] [12].

Con respecto a las variables, relación hidrocarburo/hidrogeno, temperatura y presión parcial del hidrogeno. Con el aumento de sus valores, se favorece ampliamente las reacciones de HDS, debido a que se facilita la disolución del H_2S ,

ya que este compuesto inhibe la reacción. De esta manera, se obtienen productos hidrotratados con muy alta remoción de azufre [13] [14]. Por otra parte, si se aumenta excesivamente la temperatura se pueden producir daños en el catalizador [11] [15]. Además, la velocidad espacial del líquido corresponde al cociente entre el flujo de hidrocarburo en la alimentación sobre el volumen de catalizador en la zona de reacción, esta relación es inversamente proporcional al tiempo de residencia. Si por ejemplo, se tiene una unidad HDT donde el volumen de catalizador en la zona de reacción es constante, entre menor sea el flujo de hidrocarburo en la alimentación el tiempo de retención será considerablemente mayor, lo que garantiza un mayor tiempo de contacto entre las fases y el catalizador, favoreciendo así la reacción de HDS [11] [14].

Dado que se estandarizaron todas las observaciones de las diferentes variables consideradas en el desarrollo del presente trabajo, con el fin de disminuir la variabilidad causada por la dependencia de cada variable con respecto a sus unidades de medida. En el anexo A, se describe el método de estandarización *mead and center*.

Adicionalmente, en el presente trabajo se realiza un análisis estadístico no jerárquico *k-means*, cuyos casos de estudio establecidos, variables consideradas y resultados se presentan en la sección cuatro del presente trabajo. Una breve introducción de este procedimiento se describe en el anexo B.

2.2 ESTADO DEL ARTE

En el año 2015 Narimani et al, tomando una planta de hidrotratamiento, correspondiente a la refinería de *South pars* llevaron a cabo una simulación de gas con elevado contenido de azufre con el fin de obtener gas libre de impurezas. La simulación se realizó en el software ASPEN HYSIS. Se obtuvo como resultado que

el gas objeto de análisis tenía contenidos menores a 10 ppm de azufre, dos de las corrientes de salida reportaron valores de 11 y 27 ppm respectivamente, y tres de las corrientes de salida no reportaron contenido de azufre. Se concluyó que era factible retirar el aceite de disulfuro [7].

En 2015 también, Shokri et al, llevaron a cabo la predicción del contenido de azufre en el diésel de una unidad HDT, utilizando el software MATLAB, mediante un modelo de *soft sensor*. Se llevaron a cabo alrededor de 300 experimentos y se obtuvo como resultado; que el modelo permitió tener una mayor predicción para el contenido de azufre, ya que este ofreció un algoritmo de optimización. El valor del error medio absoluto obtenido fue de 0,0745 y el valor de R^2 fue de 0,997 [5].

En el año 2012, en el ámbito local, Álvarez *et al.* llevaron a cabo numerosas pruebas con diferentes mezclas de diésel, aceite liviano de ciclo y Jet. Desarrollaron, para predecir la cinética de reacción en la unidad HDT de la refinería, un modelo experimental de predicción. También, lograron predecir las concentraciones de azufre, nitrógeno y aromáticos [6]. Este reporte fue el primero realizado en dicha refinería para el estudio del proceso HDT. Presenta la desventaja de que depende de los resultados de planta piloto que pueden divergir de los de planta industrial por cuestiones de escalado.

En 2016 Liscano, tomando una muestra histórica de data correspondiente a la operación de la unidad de HDT de diésel U-4700 de la GRB, planteó y calibró casos de predicción mediante modelos RML y redes neuronales artificiales (RNA). Lo anterior, con el fin de predecir, en función de variables de entrada y operación y propiedades del hidrocarburo, variables de operación y salida del proceso HDT tales como el flujo de diésel hidrotratado a la salida de la unidad, su respectivo contenido de azufre, la relación de flujo hidrocarburo/hidrogeno alimentado a la zona de reacción, el hidrogeno total alimentado, entre otras variables. Los resultados evidenciaron que casi la totalidad de los modelos calibrados mediante RML

presentaron un desempeño insuficiente para eventual aplicación en planta debido a los bajos grados de ajuste obtenidos. Por otra parte, en la validación de los modelos de RNA reportó valores bajos para el error cuadrático medio (MSE). Dichos valores se obtuvieron utilizando la topología 09:12:02 con las funciones de activación *logsig-purelin* [27].

2.3. REGRESION LINEAL MULTIVARIADA (RML)

Es un modelo matemático reducido mediante el cual es posible explicar las interacciones existentes entre un grupo determinado de predictores o variables, en en cual se define una variable independiente y se incluye de forma simultanea o individual la interacción de las variables dependientes con ésta. El método se basa en un análisis de regresión lineal clásico en el cual se minimizan las diferencias de la suma de los cuadrados entre una variable de respuesta denominada variable dependiente y una combinación ponderada de las variables predictoras, llamadas variables independientes [8]. Donde se busca obtener una ecuación como la presentada a continuación:

$$z = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \dots + \alpha_{j-1} x_{j-1} + \alpha_j x_j + \delta \quad (1)$$

En la ecuación (1), z es la variable dependiente, α_i son los coeficientes de regresión. Estos reflejan la influencia de los cambios de los x_i , los cuales, son las variables predictoras o independientes y δ es el error presente entre la predicción y el valor real. Por otra parte, en la calibración de un modelo de regresión lineal, se parte de la premisa de que el error (δ) es independiente, tiene una media cero, posee una distribución normal y una varianza en común para cualquier valor de x_i [16]. La ecuación (1) sin el término δ representa la ecuación de predicción de un modelo de regresión lineal multiple donde la variable a predecir se denota como \hat{z} . Utilizando

arreglos matriciales y el procedimiento de mínimos cuadrados es posible calcular los valores de los coeficientes de regresión [17]:

$$\alpha = (X'X)^{-1}X'z \quad (2)$$

Donde, X' es la matriz transpuesta de X , X y z son la matriz de datos iniciales y α es el vector de coeficientes de regresión. A continuación, realizando un análisis de varianza (ANOVA) se debe valorar la significancia del modelo y sus coeficientes, con el fin de determinar el modo en el que cada una de las variables independientes x_i afectan el promedio de la variable de salida \hat{z} [16].

El grado de ajuste del modelo de regresión lo evidencia el coeficiente de determinación R cuadrado (R^2), el cual establece la fuerza de la relación lineal entre los predictores y la variable de respuesta, se calcula con la expresión que se presenta a continuación:

$$R^2 = 1 - \frac{SSR}{SSF} \quad (3)$$

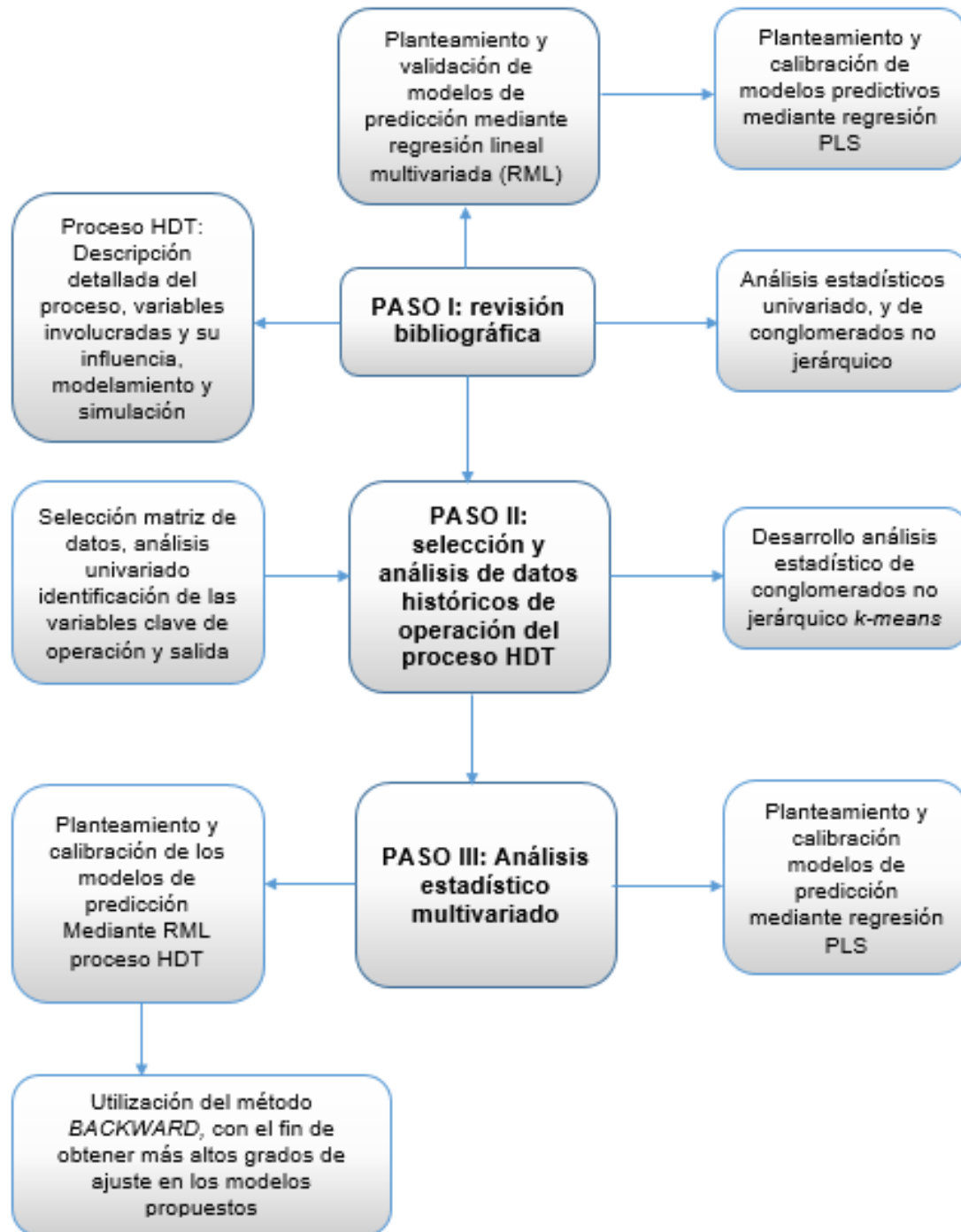
Donde SSF es la varianza total de los datos y SSR es la suma de los cuadrados del error según el conjunto externo. Se puede deducir adicionalmente, a partir de la ecuación 3, que el coeficiente de determinación no es función de la cantidad de variables presentes en el modelo; por esto, es posible calcular un valor incorrecto debido a la inclusión de más variables independientes y se debe plantear un ajuste al R^2 como el que se presenta en la ecuación 4, donde n es igual al número de observaciones y j corresponde al número de variables independientes [17][18].

$$R^2_{ajustado} = 1 - \frac{SSR/(n-j-1)}{SSF/(n-1)} \quad (4)$$

Cabe resaltar, que también se utiliza el método de eliminación hacia atrás en el presente trabajo, con el fin de obtener mejores grados de ajuste sobre los modelos de RLM a plantear. También, se utiliza la regresión PLS, como se describirá en la sección 3. Por efectos de distribución de este trabajo la descripción de este método se presenta en el anexo C.

3. METODOLOGIA

Figura 1. Diagrama de bloques metodología.



3.1 DESCRIPCION GENERAL SOBRE LOS PROCEDIMIENTOS LLEVADOS A CABO EN EL PRESENTE TRABAJO

En principio, se realiza una revisión bibliográfica con el objetivo de recopilar y analizar información del proceso HDT. También, se consulta literatura referente a análisis estadísticos univariados y análisis de conglomerados jerárquico y no jerárquico (cluster k-means), respectivamente, y al planteamiento y calibración de modelos de predicción mediante RML y regresión PLS. En seguida, para la realización del presente trabajo, se empleó una muestra de datos históricos con dimensiones de 12 variables por 25000 observaciones, correspondiente a registros minuto a minuto de la operación de la unidad de HDT U-4700, de la refinería de Ecopetrol S.A. de Barrancabermeja, Colombia. Esta matriz se obtuvo producto de la selección aleatoria, a partir de la matriz original con dimensiones de 14 variables por 172420 observaciones. Las observaciones de las 12 variables tenidas en cuenta, posterior a la selección de la matriz empleada en este trabajo, fueron graficadas con respecto al número de observación. Las gráficas obtenidas se presentan en el anexo D de este trabajo. Posterior a esto, se realizaron análisis estadísticos univariado, y análisis bivariados, secuencialmente, de la forma como se describe en la figura 1. Los resultados y demás detalles de los diferentes procedimientos realizados se exhiben en la siguiente sección. Finalmente, es muy importante resaltar (no aparece en la figura 1) que se realizó también un análisis clúster jerárquico con el fin de complementar el análisis cluster *k-means* realizado previamente y lograr una explicación gráfica más detallada sobre la similaridad existente entre pares, grupos y la totalidad de las observaciones de las variables para verificar la existencia de esquemas operativos en la unidad. Por último, se aplican las regresiones RLM y PLS sobre cuatro casos de predicción utilizando la totalidad de los datos iniciales considerados.

4. ANALISIS Y DISCUSION DE RESULTADOS

A continuación, en la tabla 1 se presentan los parámetros más relevantes calculados en la realización del análisis estadístico univariado, sobre todas las variables consideradas para el desarrollo del presente trabajo. Adicionalmente, se exhiben sus unidades de medición y la nomenclatura empleada con el fin de facilitar su manipulación, en el desarrollo y presentación de los diferentes análisis univariados y bivariados llevados a cabo.

Tabla 1. Análisis univariado sobre la matriz de datos empleada en el presente trabajo.

| VARIABLE | UNIDADES | SÍMBOLO | MÍNIMO | MÁXIMO | PROMEDIO | DESVIACIÓN ESTÁNDAR |
|---|---------------------|---------|----------|----------|----------|---------------------|
| Flujo de diésel a hidrotratar | $\frac{bbl}{min}$ | F1 | 31851.20 | 66973.10 | 52597.70 | 5608.20 |
| Densidad relativa diésel | °API | API | 28.40 | 30.90 | 29.70 | 0.39 |
| Concentración de azufre en F1 | ppm | Sin | 3985 | 5603 | 4789.50 | 350.10 |
| Temperatura inicial de ebullición | °F | IBP | 333.50 | 414.90 | 390.04 | 19.99 |
| Temperatura final de ebullición | °F | FBP | 719.10 | 759.70 | 737.20 | 8.38 |
| Flujo de hidrogeno alimentado a la zona de reacción | $\frac{mol}{min}$ | FH | 105.75 | 142.70 | 126.40 | 5.86 |
| Relación hidrocarburo/hidrogeno alimentados a la zona de reacción | $\frac{bbl}{mol}$ | RH | 232.32 | 557.51 | 417.96 | 55.67 |
| Presión de operación zona de reacción | $\frac{lb}{pulg^2}$ | P | 662.27 | 694.74 | 684.91 | 0.90 |
| Temperatura de entrada a la zona de reacción | °F | Tin | 585.65 | 667.94 | 636.07 | 3.52 |
| Cambio de temperatura en la zona de reacción | °F | ΔT | 1.43 | 92.72 | 41.93 | 1.96 |
| Flujo de diésel hidrotratado | $\frac{bbl}{min}$ | F2 | 24.51 | 60178.92 | 47565.25 | 5574.84 |
| Concentración de azufre en F2 | ppm | Sout | 5 | 50 | 26.95 | 10.65 |

A pesar de que anteriormente se explicó la influencia de la velocidad espacial del hidrocarburo (VELM) en el proceso HDT. Se tomó la decisión de no tener en cuenta esta variable, debido a que corresponde al cociente entre el flujo de diésel a hidrotratar sobre el volumen de catalizador en la zona de reacción. El volumen de catalizador en la zona de reacción es constante para la unidad que es objeto de estudio en este trabajo, y la variable F1 ya es tenida en cuenta y esta representa completamente a la variable VELM, como se muestra en la ecuación (5) por lo que no se justifica su inclusión en la estimación mediante RML Y regresión PLS para los diferentes casos de predicción planteados.

$$VELM = \frac{F1}{Vol. catalizador} = \frac{F1}{cte} \quad (5)$$

A continuación se presenta la matriz de correlación entre las variables consideradas. Los coeficientes fueron calculados tomando la muestra de datos que fue estandarizada, utilizando el método *mead and center*.

Tabla 2. Matriz de correlación entre las variables consideradas.

| | F1 | API | S in | IBP | FBP | FH | RH | P | T in | ΔT | F2 | S out |
|-------|--------------|--------|--------|--------|--------|--------|--------------|--------|--------------|-------|-------|-------|
| F1 | 1 | | | | | | | | | | | |
| API | -0.01 | 1 | | | | | | | | | | |
| S in | -0.174 | -0.233 | 1 | | | | | | | | | |
| IBP | 0.07 | -0.298 | -0.166 | 1 | | | | | | | | |
| FBP | 0.268 | -0.39 | 0.087 | 0.085 | 1 | | | | | | | |
| FH | -0.452 | -0.327 | 0.088 | 0.285 | -0.019 | 1 | | | | | | |
| RH | 0.948 | 0.103 | -0.172 | -0.045 | 0.208 | -0.709 | 1 | | | | | |
| P | -0.031 | 0.031 | 0.012 | -0.078 | -0.013 | 0.016 | -0.031 | 1 | | | | |
| T in | 0.517 | -0.339 | 0.056 | 0.111 | 0.491 | -0.066 | 0.435 | -0.044 | 1 | | | |
| ΔT | 0.377 | -0.03 | -0.095 | 0.111 | -0.018 | -0.052 | 0.313 | 0.099 | -0.052 | 1 | | |
| F2 | 0.962 | -0.062 | -0.163 | 0.107 | 0.258 | -0.442 | 0.916 | -0.032 | 0.541 | 0.343 | 1 | |
| S out | 0.221 | 0.063 | 0.271 | -0.294 | 0.185 | -0.183 | 0.224 | -0.02 | 0.224 | 0.009 | 0.219 | 1 |

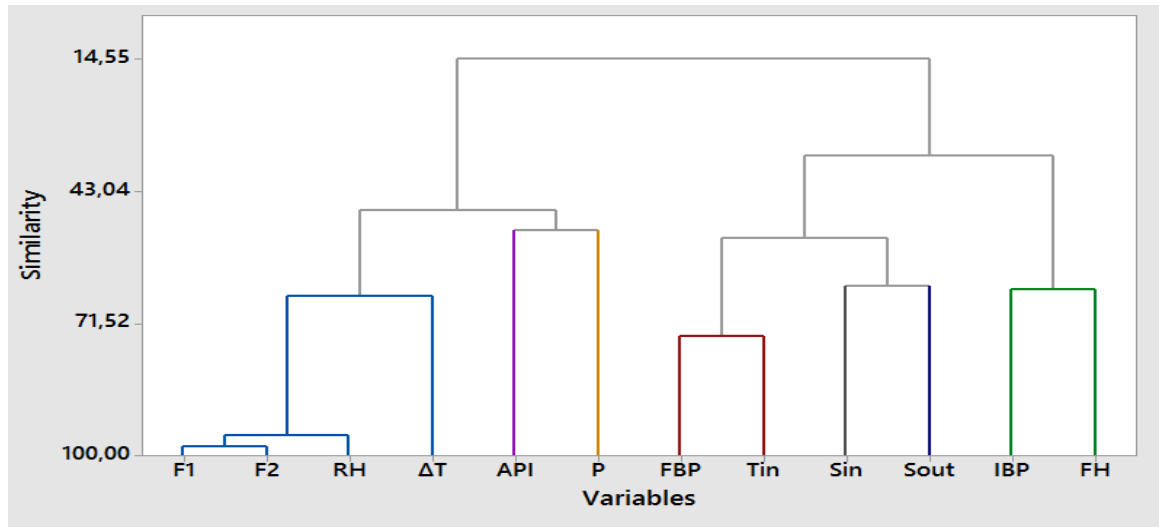
4.1 ANÁLISIS CLÚSTER JERARQUICO Y NO JERÁRQUICO *K-MEANS*:

Se tomó la matriz seleccionada previamente de 12 variables por 25000 observaciones estandarizadas de cada variable, correspondiente a las variables F1, FBP, IBP, API, Sin, FH, RH, T in, ΔT , P y F2. Se llevó a cabo análisis de conglomerados no jerárquico estableciendo cuatro casos de análisis, como se exhibe a continuación:

- **Caso I:** clasificación categórica en dos grupos o clúster.
- **Caso II:** clasificación categórica en tres grupos o clúster.
- **Caso III:** clasificación categórica en cinco grupos o clúster.
- **Caso IV:** clasificación categórica en siete grupos o clúster.

A continuación, se presentan en las tablas 3, 4 y 5, los resultados proporcionados por el software minitab 17 para el caso de estudio IV, ya que después de realizar los análisis para los cuatro casos planteados, se determinó que esta fue la clasificación más apropiada, en función de la variabilidad de los datos, entre los cuatro casos propuestos; los siete grupos obtenidos sugieren siete esquemas operativos diferentes para la operación de la unidad HDT de la GRB. Como se observa en las tablas 3 y 5 las distancias intragrupal promedio son levemente menores a las distancias intergrupales, por lo que la clasificación es pertinente; los detalles, sobre los demás casos de análisis desarrollados se presentan en el anexo E. La figura 2 presenta el dendograma correspondiente al análisis clúster jerárquico que fue realizado sobre las mismas 12 variables mencionadas al inicio de esta sesión. La descripción de este análisis se presenta en el anexo.

Figura 2. Dendograma analisis cluster jerarquico sobre las variables consideradas en este trabajo.



Según la figura 2, la similitud que se presenta entre las variables F1 y F2 es casi del 100%, debido a que estas variables corresponden a los flujos de alimentación y salida de diésel de la unidad HDT U-4700, respectivamente. En concordancia con la totalidad de las observaciones objeto de análisis en este trabajo, se presentan valores del contenido de azufre en la corriente de alimentación del combustible considerablemente inferiores al 1% p/p por lo que al final del paso del diésel por la unidad HDT se reportan valores del flujo de salida numéricamente muy similares a los del flujo en la alimentación. Por otra parte, como se observa en el dendograma, la variable RH también presenta un porcentaje de similitud mayor al 90 % en relación con las variables F1 y F2, ya que esta se encuentra relacionada con F1; la relación entre el flujo de hidrógeno alimentado a la zona de reacción y F1 es aproximadamente constante en las observaciones proporcionadas para este trabajo. Las demás variables (figura 12) presentan similitud moderada, y esta disminuye altamente al considerar grupos cada vez más grandes de variables, llegando a un valor de apenas el 15% de similitud en las observaciones para las variables consideradas en este trabajo. Los resultados también sugieren diferentes esquemas operativos para la unidad HDT.

Tabla 3. Distancias intragrupales y suma de cuadrados caso de estudio IV

| Conglomerado | Numero de observaciones en el conglomerado | Suma de cuadrados dentro del conglomerado | Distancia promedio desde el centroide del conglomerado | Distancia Máxima desde el centroide del conglomerado |
|------------------|--|---|--|--|
| Conglomerado I | 276 | 10385.321 | 4.926 | 22.056 |
| Conglomerado II | 6471 | 33478.671 | 2.154 | 8.321 |
| Conglomerado III | 1634 | 6596.444 | 1.954 | 8.586 |
| Conglomerado IV | 3126 | 18827.695 | 2.254 | 16.201 |
| Conglomerado V | 4350 | 25271.346 | 2.200 | 30.219 |
| Conglomerado VI | 5975 | 21629.980 | 1.795 | 7.479 |
| Conglomerado VII | 3165 | 13778.976 | 1.983 | 6.450 |

Tabla 4. Coordenadas centroides de las variables consideradas caso de estudio IV

| variable | Centroides conglomerado 1 | Centroides conglomerado 2 | Centroides conglomerado 3 | Centroides conglomerado 4 | Centroides conglomerado 5 | Centroides conglomerado 6 | Centroides conglomerado 7 | Centroide principal |
|----------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------|
| F1 | -0.0519 | 0.4427 | -0.7450 | -2.0253 | -0.2509 | 0.7354 | 0.4409 | -0.000 |
| FBP | 0.0341 | -0.1394 | -0.5504 | -0.3178 | -0.9171 | 0.2717 | 1.6275 | 0.000 |
| IBP | 0.7611 | -1.1943 | 0.2278 | -0.1703 | 0.4297 | 0.7206 | 0.4752 | -0.000 |
| API | -0.1949 | 0.6466 | -1.1873 | 0.0824 | 0.4820 | -0.0394 | -1.3615 | -0.000 |
| Sin | -0.1220 | 0.2476 | 1.4841 | 0.2896 | -0.7163 | -0.4634 | 0.3115 | 0.000 |
| FH | -0.0334 | -0.7060 | 1.5453 | 0.6648 | 0.2963 | -0.4184 | 0.3745 | -0.000 |
| RH | -0.0262 | 0.5978 | -1.0842 | -1.8202 | -0.3137 | 0.7259 | 0.1984 | 0.000 |
| T1 | 0.5847 | 0.0014 | -0.1888 | -0.9431 | -0.6078 | 0.1506 | 1.5261 | 0.000 |
| ΔT | -1.3216 | 0.0155 | 0.3854 | -1.2255 | 0.5145 | 0.0705 | 0.2549 | 0.000 |
| P | -6.6010 | 0.0865 | 0.0931 | 0.1273 | 0.0828 | 0.0340 | 0.0471 | 0.000 |
| F2 | 0.0358 | 0.3687 | -0.6381 | -2.0060 | -0.2525 | 0.7607 | 0.4646 | 0.000 |

Tabla 5. Matriz de distancias intergrupales caso de estudio IV

| | Centroides conglomerado 1 | Centroides conglomerado 2 | Centroides conglomerado 3 | Centroides conglomerado 4 | Centroides conglomerado 5 | Centroides conglomerado 6 | Centroides conglomerado 7 |
|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Centroides conglomerado 1 | 0.0000 | 7.2626 | 7.5538 | 7.7859 | 7.1838 | 6.9432 | 7.2375 |
| Centroides conglomerado 2 | 7.2626 | 0.0000 | 4.1926 | 4.8260 | 2.7425 | 2.2756 | 3.6967 |
| Centroides conglomerado 3 | 7.5538 | 4.1926 | 0.0000 | 3.3532 | 3.2469 | 4.1866 | 3.8481 |
| Centroides conglomerado 4 | 7.7859 | 4.8260 | 3.3532 | 0.0000 | 3.6961 | 5.2478 | 5.5611 |
| Centroides conglomerado 5 | 7.1838 | 2.7425 | 3.2469 | 3.6961 | 0.0000 | 2.4898 | 4.1006 |
| Centroides conglomerado 6 | 6.9432 | 2.2756 | 4.1866 | 5.2478 | 2.4898 | 0.0000 | 2.6934 |
| Centroides conglomerado 7 | 7.2375 | 3.6967 | 3.8481 | 5.5611 | 4.1006 | 2.6934 | 0.0000 |

4.2 PLANTEAMIENTO Y CALIBRACIÓN DE LOS MODELOS DE RML:

En este trabajo se establecieron cinco casos de predicción, los cuales se describen a continuación. Para su planteamiento, se usaron 17500 observaciones de proceso, los cuales corresponden al 70% de la matriz de datos seleccionada, para su calibración se emplearon las 7500 observaciones restantes de esta misma matriz correspondientes al restante 30%:

- **Caso I:** Predicción del contenido de azufre en el ACPM hidrotratado (Sout).
- **Caso II:** Predicción del flujo de ACPM hidrotratado (F2).
- **Caso III:** Predicción de la relación hidrocarburo/hidrógeno alimentados a la zona de reacción (RH).
- **Caso IV:** Predicción del flujo de hidrógeno alimentado a la zona de reacción (FH).
- **Caso V:** Predicción del cambio de temperatura en la zona de reacción (ΔT).

Para los casos de predicción I y II las variables consideradas en condición de variables independientes fueron, las variables F1, FBP, IBP, API, Sin, RH, ΔT , P y Tin. Por otra parte, para los casos de predicción III y IV se tomaron como variables predictoras las variables F1, FBP, IBP, API, Sin, Sout, ΔT , P y Tin. Para el caso de estudio V, se consideraron en condición de variables predictoras, las variables F1, FBP, IBP, API, Sin, Sout, FH, P, Y Tin.

En primer lugar se realizaron estimaciones mediante regresión lineal multivariada para los cinco casos de estudio listados anteriormente y su respectiva calibración. En la Tabla 3 se presentan los valores de R^2 , R^2_{Ajustado} y para el proceso de calibración el valor del error cuadrático medio (MSE), que se obtuvieron para cada caso. El procedimiento fue realizado en una versión de prueba del software Minitab 17. Se fijó un intervalo de confianza del 95%. Los detalles de este procedimiento se presentan en el Anexo F.

Tabla 6. Resultados casos de predicción mediante regresión multivariada.

| Caso de predicción | R^2 validación | R^2 ajustado | MSE |
|--------------------|------------------|----------------|---------|
| I | 0.24690 | 0.24660 | 0.76069 |
| II | 0.93290 | 0.93260 | 0.06503 |
| III | 0.92120 | 0.92120 | 0.07870 |
| IV | 0.39420 | 0.39400 | 0.60365 |
| V | 0.28750 | 0.28720 | 0.70377 |

Como se observa en la Tabla 6, el grado de ajuste en las predicciones para los casos I, IV, V, es insignificante (correlación entre las variables según tabla 2). Teniendo en cuenta esto último, es poco probable explicar la relación existente entre las variables consideradas; esto a pesar de que los datos fueron estandarizados utilizando la metodología descrita en el anexo A. A excepción de los valores de R^2 y R^2_{Ajustado} calculados en los casos de estudio II y III, estos presentan altos valores debido al alto grado de correlación entre el flujo de diésel a hidrotratar, la relación hidrocarburo/hidrogeno que ingresan a la unidad HDT y con el flujo de diésel hidrotratado a la salida del proceso. Lo anterior se inferido al comparar los valores de los coeficientes de regresión obtenidos para cada variable independiente en las ecuaciones de ajuste obtenidas, ecuaciones (6) y (7), respectivamente; también puede verificarse en la matriz de correlación.

Las ecuaciones de ajuste obtenidas para los casos de predicción II y III, se presentan a continuación:

- **Caso II.** Predicción del flujo de ACPM hidrotratado (F2):

$$\begin{aligned}
 \mathbf{F2} = & 0,79795*\mathbf{F1} - 0,04199*\mathbf{FBP} + 0,03949*\mathbf{IBP} - 0,05519*\mathbf{API} - 0,00495*\mathbf{Sin} \\
 & + 0,15146*\mathbf{RH} - 0,01031*\mathbf{\Delta T} + 0,00557*\mathbf{P} + 0,06076*\mathbf{Tin}
 \end{aligned}
 \tag{6}$$

Como se observa en la ecuación (6) las variables F1 y RH explican en un muy alto porcentaje la variabilidad de la variable F2 (se evidencia en los valores de los coeficientes de regresión calculados). Entonces, se procede tomando el valor absoluto de los coeficientes de regresión obtenidos para cada una de las variables predictoras tenidas en cuenta para realizar el planteamiento del modelo de predicción de la variable F2. A continuación en las ecuaciones (7) y (8) se calcula el porcentaje de variabilidad de F2 explicado por cada una de estas dos variables respectivamente:

- Variable F1:

$$\frac{0,79795}{0,79795+0,04199+0,03949+0,05519+0,00495+0,15146+0,01031+0,00557+0,06076} * 100 = 68,34\% \quad (7)$$

- Variable RH:

$$\frac{0,15146}{0,79795+0,04199+0,03949+0,05519+0,00495+0,15146+0,01031+0,00557+0,06076} * 100 = 12,97\% \quad (8)$$

Como se mostró anteriormente los predictores F1 y RH, principalmente el predictor F1, explican ambos en conjunto el 81,31% de la varianza total de la variable de respuesta F2. Se fijó un intervalo de confianza del 95%. Por esto, basta con tener en cuenta tres predictores más entre los siete adicionales, tomando como criterio el valor, de manera descendente, de los coeficientes de regresión calculados y presentados en la ecuación (6), para cumplir con el intervalo de confianza establecido; por lo que es posible reducir el modelo a la siguiente expresión:

$$F2 = 0,79795 * F1 - 0,04199 * FBP - 0,05519 * API + 0,15146 * RH + 0,06076 * Tin \quad (9)$$

- **Caso III.** Predicción de la relación hidrocarburo / hidrógeno alimentados a la zona de reacción (RH):

$$\text{RH} = 1,01043 \text{ F1} + 0,00073 \text{ FBP} - 0,08742 \text{ IBP} + 0,07074 \text{ API} + 0,00969 \text{ Sin} - 0,02244 \text{ Sout} - 0,05789 \Delta\text{T} - 0,00610 \text{ P} - 0,05353 \text{ Tin} \quad (10)$$

Del mismo modo que se llevó a cabo para el caso de predicción II, como se observa en la ecuación 10 (valor de los coeficientes de regresión); la variable F1 explica en un alto porcentaje la variabilidad de RH. A continuación se calcula este porcentaje:

$$\frac{1,01043}{1,01043+0,00073+0,08742+0,07074+0,00969+0,02244+0,05789+0,00610+0,05353} * 100 = 76,61\% \quad (11)$$

Dado que la variable F1 explica un 76,61% de la variabilidad total de la variable RH, como se muestra en la ecuación (11). Para este caso también se fijó un intervalo de confianza del 95%, por lo que se tomaron cuatro predictores de los restantes ocho para incluirlos en un modelo final simplificado. Tomando como criterio de selección, de manera descendente, los valores de los coeficientes de regresión que se muestran en la ecuación (10) el modelo se simplifica a la siguiente expresión:

$$\text{RH} = 1,01043 \text{ F1} - 0,08742 \text{ IBP} + 0,07074 \text{ API} - 0,05789 \Delta\text{T} - 0,05353 \text{ Tin} \quad (12)$$

Por otra parte, dado a los grados de ajuste bajos en tres de los cinco casos de predicción mediante RML, se procedió a utilizar el método de eliminación hacia atrás; con este procedimiento es posible determinar las posibles interacciones entre las variables predictoras y de esta forma buscar obtener modelos con valores de R^2 , R^2_{Ajustado} más altos. Se desarrollaron cada uno de los modelos propuestos. Los detalles del análisis se incluyen en el Anexo G y en la tabla 7 se exponen los resultados.

Tabla 7. Resultados casos de predicción mediante RML utilizando el método de eliminación hacia atrás.

| Caso de predicción | R^2 | R^2 ajustado |
|--------------------|--------|----------------|
| I | 0.5817 | 0.5810 |
| IV | 0.5082 | 0.5075 |
| V | 0.4757 | 0.4749 |

Como se observa en la Tabla 7 se obtienen valores más altos de R^2 y $R^2_{Ajustado}$ comparados con los modelos lineales efectuados previamente para los mismos casos de estudio; esto es porque se tienen en cuenta todas las posibles combinaciones lineales entre las variables iniciales y causa que su influencia estadística sea más significativa en la predicción. A pesar de esto, no se obtienen valores de R^2 , $R^2_{Ajustado}$ que permitan otorgar fiabilidad a los modelos para su aplicación. Una mejora significativa en las regresiones para estos casos de predicción puede ser obtenida considerando las agrupaciones o esquemas operativos reportados en la sección 4.1.

4.3 PLANTEAMIENTO Y CALIBRACIÓN MODELOS DE PREDICCIÓN MEDIANTE REGRESIÓN PLS

Como se mostró anteriormente, en la predicción mediante RML se plantearon cinco casos de estudio, pero solamente se obtuvieron altos valores de R^2 y $R^2_{Ajustado}$ en los casos de predicción II y III. Por esto, es pertinente realizar un análisis más detallado de estos dos casos de estudio, ya que los altos grados de ajuste obtenidos se deben, como se explicó en la sección anterior, a la alta colinealidad que se presenta entre la variable F2, con las variables F1 y RH.

Conocidos los objetivos del planteamiento de modelos de regresión PLS, descritos en la sección 2, se realizó la predicción agrupando las nueve variables, inicialmente consideradas en condición de variables independientes, tomando desde uno hasta

nueve variables latentes, para los dos casos de predicción, respectivamente. A continuación se exponen los resultados. Las gráficas de selección de modelo, respuesta de regresión y de coeficientes se exhiben en los anexos H y I para los casos de estudio II y III respectivamente.

- **Caso II:** estimación de la variable F2 en función de las variables F1, FBP, IBP, API, Sin, RH, ΔT , P y Tin.

Tabla 8. Selección y validación del modelo de estimación mediante regresión PLS caso de predicción II.

| Variabes latentes | Varianza de X | Error | R-cuadrado validación |
|-------------------|---------------|---------|-----------------------|
| 1 | 0.288851 | 1976.73 | 0.886542 |
| 2 | 0.447948 | 1525.36 | 0.912449 |
| 3 | 0.547033 | 1327.11 | 0.923828 |
| 4 | 0.623320 | 1287.84 | 0.926082 |
| 5 | 0.710150 | 1265.79 | 0.927347 |
| 6 | 0.771469 | 1221.46 | 0.929892 |
| 7 | 0.831022 | 1189.98 | 0.931699 |
| 8 | 0.935143 | 1189.08 | 0.931750 |

- **Caso III:** estimación de la variable RH en función de las variables F1, FBP, IBP, API, Sin, S out, ΔT , P y T in

Tabla 9. Selección y validación del modelo de estimación mediante regresión PLS caso de predicción III.

| Variabes latentes | Varianza de X | Error | R-cuadrado validación |
|-------------------|---------------|---------|-----------------------|
| 1 | 0.206317 | 4076.75 | 0.766596 |
| 2 | 0.378809 | 2566.80 | 0.853094 |
| 3 | 0.448574 | 1548.91 | 0.911321 |
| 4 | 0.533201 | 1404.14 | 0.919609 |

| | | | |
|----------|----------|---------|----------|
| 5 | 0.631095 | 1380.45 | 0.920966 |
| 6 | 0.783232 | 1379.42 | 0.921025 |
| 7 | 0.837719 | 1379.35 | 0.921029 |
| 8 | 0.926960 | 1379.35 | 0.921029 |

Según los resultados, el desempeño mostrado por PLS es equivalente al obtenido por RLM. La ventaja de utilizar PLS consiste en que las variables latentes obtenidas (7 para el caso II y 6 para el caso III, debido a que con un mayor número no mejora el error reportado) no presentan correlación; estas variables latentes se encuentran en términos de las variables originales del sistema.

De otro lado, el algoritmo PLS reportó también bajo desempeño en los otros casos. Debido a que RLM y PLS no conllevan a regresiones con buena explicación de la varianza para los casos I, IV y V, es recomendable aplicar estos métodos sobre cada una de las agrupaciones o esquemas operativos obtenidos en la sección 4.1.

5. CONCLUSIONES

- Se encontró un alto grado de correlación lineal entre las variables F1, con las variables F2 y RH respectivamente. La correlación entre las variables F1 y F2 se da porque en este caso, los contaminantes que son eliminados presentan bajas concentraciones en el combustible (no superan el 1% p/p de la corriente de diésel en la alimentación en la unidad HDT), comparadas con el orden de magnitud del flujo de diésel en la alimentación a hidrotrotar; esto se debe a que este flujo a lo largo de su paso por la totalidad de la unidad HDT es aproximadamente constante. Por otra parte, a partir de la alta correlación lineal existente entre las variables F1 y RH (por ende también entre F2 y RH) se dedujo que el hidrogeno alimentado a la zona de reacción fue aproximadamente constante en el rango de tiempo en que se realizó el registro de las diferentes observaciones contenidas en la matriz de data que fue analizada en este trabajo.

6. RECOMENDACIONES

- La aplicación de regresiones en cada una de las agrupaciones es recomendada para obtener un mejor desempeño en la predicción de las condiciones operacionales de la unidad HDT.
- Se recomienda utilizar herramientas de predicción basadas en algoritmos de aprendizaje. Estas herramientas sin importar la naturaleza de dispersión de los datos, permiten realizar entrenamiento y validación; por ende, aprenden a partir del comportamiento de los datos, sin importar que no sigan patrones matemáticos definidos y que exista muy baja correlación entre las variables a considerar. Si se realiza un entrenamiento y validaciones adecuadas se pueden obtener grados de predicción considerablemente mejores que utilizando la estadística multivariada.

REFERENCIAS

- [1] FABIAN SALVADOR MENDEROS NIETO “Desarrollo de un modelo dinámico de una planta piloto de hidrotratamiento”. Tesis para obtener el grado de maestro en ciencias con especialidad en ingeniería química. Escuela superior de ingeniería química e industrias extractivas, instituto politécnico nacional, México 2004
- [2] CHOWDHURY, R, PEDERNERA, E. Y REIMERT, R., “Trickle-bed reactor model for desulfurization and dearomatization of diesel”, AIChE Journal, Vol. 48, No. 1, 2002.
- [3] NICOLAS RODRIGUEZ. “Análisis de la reducción del azufre en el combustible diesel en el Salvador, Guatemala Honduras y Nicaragua”.. “Superando obstáculos en la integración energética en Centroamérica y Sudamérica”. Comisión económica para America Latina y el Caribe, Mexico D.F. 2012.
- [4] MIGUEL ANGEL RODRIGUEZ HUERTA, “Modelación de un reactor de HDT de crudo pesado y su aplicación en integración energética”. Tesis para obtener el grado de doctor en ingeniería energética. Universidad Nacional Autonoma de Mexico 2007.
- [5] SAEID SHOKRI, MOHAMMAD TAGHI SADEGHI, MANDHI AHMADI MARVAST, SHANKAR NARASIMHAN. “Improvement of the prediction performance of a soft sensor model based on support vector regression for production of ultra-low sulfur diesel ”., recibido 22/04/2014, publicado Online 13/01/2014. Department of Chemical Engineering, Iran University of Science and Technology (IUST), Tehran, Iran.

- [6] ÁLVAREZ, MAGDA C.; HOYOS, LUIS J.; ZAMBRANO LILIANA G.; Diesel hydrotreating process modelling. Instituto Colombiano del Petróleo (ICP), Ecopetrol S.A., aceptado 27/11/2012.
- [7] Alaei. J; Nariman E. Simulation of hydrodesulfurization unit for natural gas condensate with high sulfur content. Applied Petrochemical Research (107) 2015. 1-10 8.
- [8] HAIR JJ, BLACK W, BABIN B, ANDERSON R. Multivariate Data Analysis. 7th edition. USA: Prentice Hall; 2009.
- [9] ARCE-MEDINA, E., & PAZ-PAREDES, J. (2009). Artificial neural network modeling techniques applied to the hydrodesulfurization process. Mathematical and Computer Modelling An International Journal, (pp. 207-214).
- [10] Ministerio de Minas y Energía. Resolución 90963 de 2014. Recuperado el 05 de abril 2016. <https://www.minminas.gov.co/documents/10180//23517//26101-90963.pdf>
- [11] VEGA, ANDRÉS F.; DE RIO AMADOR, EDUARDO; BARBOSA, AIDA L.; (2014). Hidrodesulfuración de crudos de petróleo: base para el mejoramiento de combustibles. una revisión. Avances en Ciencias e Ingeniería, Julio-septiembre, 37-60.
- [12] GONZÁLEZ-CORTÉS, S.L., RUGMINI, S., XIAO, T., GREEN, M.L.H., RODULFO-BAECHLER, S.M. & IMBERT, F.E. (2014). Deep hydrotreating of different feedstocks over a highly active Al₂O₃-supported NiMoW sulfide catalyst. Applied Catalysis A: General, 475 (0), 270-281.

- [13] KNUDSEN, K.G., COOPER, B.H. & TOPSØE, H. (1999). Catalyst and process technologies for ultra low sulfurdiesel. *Applied Catalysis A: General*, 189 (2), 205-215.
- [14] JONES, D. & PUJADÓ, P. (2006). *Handbook of petroleum processing*: Springer
- [15] ANDARI, M.K., ABU-SEEDO, F., STANISLAUS, A. & QABAZARD, H.M. (1996). Kinetics of individual sulfur compounds in deep hydrodesulfurization of Kuwait diesel oil. *Fuel*, 75 (14), 1664-1670.
- [16] MENDENHALL W. *Introducción a la probabilidad y la estadística*. 1987. p. 500-632.
- [17] DRAPER, N.R.; SMITH, H. (1980) *Applied Regression Analysis*. Editorial Pueblo y Educación.
- [18] WALPOLE, RONALD E., MYERS, RAYMOND H. (1999). *Probabilidad y estadística para ingenieros*. México: Prentice-hall Hispanoamérica s.a.
- [19] Curso regresión lineal múltiple.
<http://personal.us.es/vararey/adatos2/multiple.pdf>
- [20] Suport minitab 17 <http://support.minitab.com/es-mx/minitab/17/topic-library/modeling-statistics/regression-and-correlation/partial-least-squares-regression/what-is-partial-least-squares-regression/>
- [21] Vega Carmen (2008). "Regresión por Mínimos Cuadrados Parciales con Aplicación en Regresión Logística". Tesis UMSA, Carrera de Estadística.

- [22] Revista Varianza versión impresa ISSN 9876-6789. edición n.7 La Paz nov. 2010. http://www.revistasbolivianas.org.bo/scielo.php?pid=S9876-67892010000100005&script=sci_arttext.
- [23] IBM knowledge center. Analisis cluster no jerarquico. http://www.ibm.com/support/knowledgecenter/es/SSLVMB_22.0.0/com.ibm.spss.statistics.help/spss/base/idh_quic.htm
- [24] Análisis de datos multivariable. El procedimiento de conglomerados de k-medias, capítulo 21. Universidad complutense de Madrid http://pendientedemigracion.ucm.es/info/socivmyt/paginas/D_departamento/materiales/analisis_datosyMultivariable/21conglk_SPSS.pdf
- [25] Bailey, Ken (1994). "Numerical Taxonomy and Cluster Analysis". Typologies and Taxonomies. p. 34
- [26] CÉSAR PEREZ LOPEZ. Técnicas de análisis multivariante de datos. Universidad complutense de Madrid. Pearson educación S.A. 2004.
- [27] JUAN DAVID LIZCANO. Aplicación de la regresión lineal múltiple y las redes neuronales artificiales en el análisis del proceso de hidrotratamiento de ACPM. Universidad industrial de Santander, 2016.
- [28] IBM knowledge center. Analisis cluster jerarquico. http://www.ibm.com/support/knowledgecenter/es/SSLVMB_22.0.0/com.ibm.spss.statistics.help/spss/base/idh_clus.htm

BIBLIOGRAFIA

ALAEI, J; NARIMAN E. Simulation of hydrodesulfurization unit for natural gas condensate with high sulfur content. Applied Petrochemical Research (107) 2015. 1-10 8.

ÁLVAREZ, MAGDA C.; HOYOS, LUIS J.; ZAMBRANO LILIANA G.; Diesel hidrotreating process modelling. Instituto Colombiano del Petróleo (ICP), Ecopetrol S.A., aceptado 27/11/2012.

ANALISIS DE DATOS MULTIVARIABLE. El procedimiento de conglomerados de k-medias, capítulo 21. Universidad complutense de Madrid [en línea] [citado 18 de junio de 2016] Disponible en: http://pendientedemigracion.ucm.es/info/socivmyt/paginas/D_departamento/materiales/analisis_datosyMultivariable/21conglk_SPSS.pdf

ANDARI, M.K., ABU-SEEDO, F., STANISLAUS, A. & QABAZARD, H.M. (1996). Kinetics of individual sulfur compounds in deep hydrodesulfurization of Kuwait diesel oil. Fuel, 75 (14), 1664-1670.

ARCE-MEDINA, E., & PAZ-PAREDES, J. (2009). Artificial neural network modeling techniques applied to the hydrodesulfurization process. Mathematical and Computer Modelling An International Journal, (pp. 207-214).

BAILEY, KEN. "Numerical Taxonomy and Cluster Analysis". Typologies and Taxonomies. p. 34. 1994.

CÉSAR PEREZ LOPEZ. Técnicas de análisis multivariante de datos. Universidad complutense de Madrid. Pearson educación S.A. 2004.

CHOWDHURY, R, PEDERNERA, E. Y REIMERT, R., "Tricle-bed reactor model for desulfurización and dearomatization of diesel", AIChE Journal, Vol. 48, No. 1, 2002.
DRAPER, N.R.; SMITH, H. Applied Regression Analysis. Editorial Pueblo y Educación. 1980.

GONZÁLEZ-CORTÉS, S.L., RUGMINI, S., XIAO, T., GREEN, M.L.H., RODULFO-BAECHLER, S.M. & IMBERT, F.E. (2014). Deep hydrotreating of different feedstocks over a highly active Al₂O₃-supported NiMoW sulfide catalyst. Applied Catalysis A: General, 475 (0), 270-281.

HAIR JJ, BLACK W, BABIN B, ANDERSON R. Multivariate Data Analysis. 7th edition. USA: Prentice Hall; 2009.

IBM knowledge center. Análisis cluster no jerarquico. [en línea] [citado 19 de junio de 2016] Disponible en: http://www.ibm.com/support/knowledgecenter/es/SSLVMB_22.0.0/com.ibm.spss.statistics.help/spss/base/idh_quic.htm

IBM knowledge center. Analisis cluster jerarquico. [en línea] [citado 26 de enero de 2017] Disponible en: http://www.ibm.com/support/knowledgecenter/es/SSLVMB_22.0.0/com.ibm.spss.statistics.help/spss/base/idh_clus.htm

JONES, D. & PUJADÓ, P. Handbook of petroleum processing: Springer.2006

JUAN DAVID LIZCANO. Aplicación de la regresión lineal múltiple y las redes neuronales artificiales en el análisis del proceso de hidrotratamiento de ACPM. Universidad industrial de Santander, 2016.

KNUDSEN, K.G., COOPER, B.H. & TOPSØE, H. Catalyst and process technologies for ultra low sulfur diesel. Applied Catalysis A: General, 189 (2), 205-215. 1999.

MENDENHALL W. Introducción a la probabilidad y la estadística. 1987. p. 500-632.

MENDEROS NIETO Fabian Salvador “Desarrollo de un modelo dinámico de una planta piloto de hidrotratamiento”. Tesis para obtener el grado de maestro en ciencias con especialidad en ingeniería química. Escuela superior de ingeniería química e industrias extractivas, instituto politécnico nacional, Mexico 2004

MIGUEL ANGEL RODRIGUEZ HUERTA, “Modelación de un reactor de HDT de crudo pesado y su aplicación en integración energética”. Tesis para obtener el grado de doctor en ingeniería energética. Universidad Nacional Autónoma de México 2007.

MINISTERIO DE MINAS Y ENERGÍA. Resolución 90963 de 2014. [en línea] [citado 05 de abril de 2016] Disponible en: <https://www.minminas.gov.co/documents/10180//23517//26101-90963.pdf>

REGRESIÓN LINEAL MÚLTIPLE Curso regresión lineal múltiple. [en línea] [citado 19 de junio de 2016] Disponible en: <http://personal.us.es/vararey/adatos2/multiple.pdf>

REVISTA VARIANZA VERSIÓN IMPRESA ISSN 9876-6789. edición n.7 La Paz nov. 2010. [en línea] [citado 15 de abril de 2016] Disponible en: http://www.revistasbolivianas.org.bo/scielo.php?pid=S987667892010000100005&script=sci_arttext.

RODRIGUEZ Nicolas. "Análisis de la reducción del azufre en el combustible diesel en el Salvador, Guatemala Honduras y Nicaragua".. "Superando obstáculos en la integración energética en Centroamérica y Sudamérica". Comisión económica para America Latina y el Caribe, Mexico D.F. 2012.

SAEID SHOKRI, MOHAMMAD TAGHI SADEGHI, MANDHI AHMADI MARVAST, SHANKAR NARASIMHAN. "Improvement of the prediction performance of a soft sensor model based on support vector regression for production of ultra-low sulfur diesel "., recibido 22/04/2014, publicado Online 13/01/2014. Department of Chemical Engineering, Iran University of Science and Technology (IUST), Tehran, Iran.

SUPPORT MINITAB 17¿Qué es la regresión de mínimos cuadrados parciales? [en línea] [citado 20 de abril de 2016] Disponible en: <http://support.minitab.com/es-mx/minitab/17/topic-library/modeling-statistics/regression-and-correlation/partial-least-squares-regression/what-is-partial-least-squares-regression/>

VEGA Carmen. "Regresión por Mínimos Cuadrados Parciales con Aplicación en Regresión Logística". Tesis UMSA, Carrera de Estadística. 2008

VEGA, Andrés F.; DE RIO Amador, Eduardo; BARBOSA, Aida L.; Hidrodesulfuración de crudos de petróleo: base para el mejoramiento de combustibles. una revisión. Avances en Ciencias e Ingeniería,

WALPOLE, RONALD E., MYERS, RAYMOND H. Probabilidad y estadística para ingenieros. México: Prentice-hall Hispanoamérica s.a.1999.

ANEXOS

ANEXO A. MÉTODO DE ESTANDARIZACION MEAD AND CENTER

La estandarización *mead and center* es un método que permite realizar el adimensionamiento de datos, de una muestra en particular. Cada observación, de cada variable en un matriz de observaciones, es transformada con respecto a la desviación estándar (σ) y al valor medio de la variable a la que corresponda, de la siguiente manera:

$$y_i^N = \frac{y_i - \bar{y}}{\sigma_y} \quad (1)$$

Donde:

y_i^N : corresponde al valor estandarizado de la observación i de la variable y .

y_i : corresponde al valor real de la observación i de la variable y .

\bar{y} : es el valor medio entre todas las observaciones consideradas de la variable y á contenidos en una determinada matriz.

σ_y : es la desviación estándar de la muestra de observaciones de la variable y .

El objetivo de dicha transformación es eliminar la dependencia de cada variable con respecto a sus unidades de medida y acotar la variabilidad de los datos [26]. Por otra parte, si se desea obtener el valor real de una determinada observación a partir de su valor normalizado, basta con conocer la desviación estándar y la media de la respectiva variable a la que corresponda la observación y despejando el término y_i de la ecuación (1).

ANEXO B. ANÁLISIS CLÚSTER NO JERÁRQUICO K-MEANS

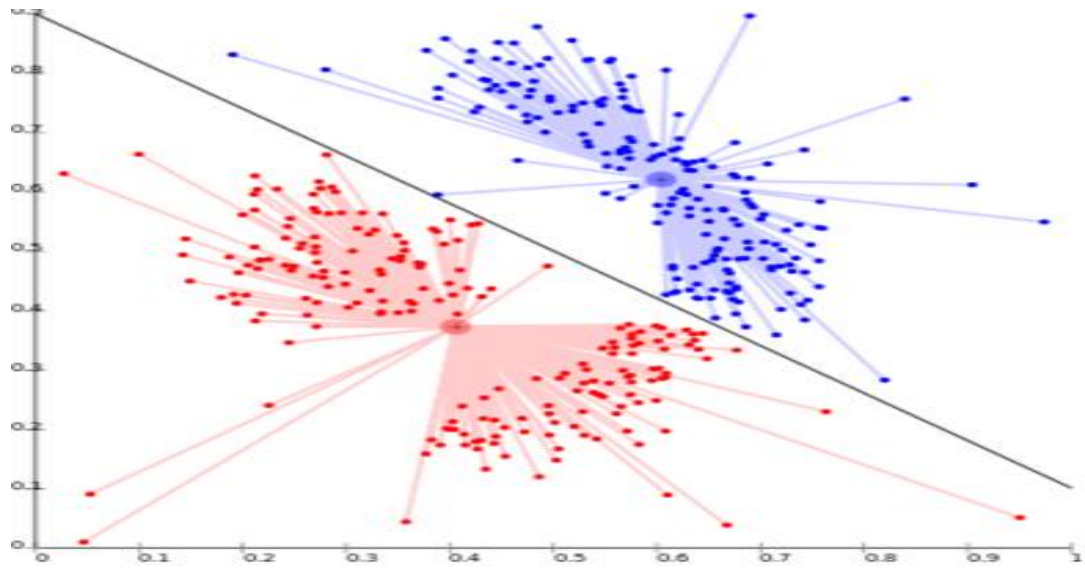
El análisis de conglomerados k-means, es un método de agrupación de muestras de data que se basa en las distancias existentes entre las observaciones en un conjunto de variables. Este algoritmo clasificatorio no permite agrupar variables. Las observaciones, en un determinado caso de estudio, son clasificadas en grupos o clusters, en función de la similaridad existente entre estas, sin importar que correspondan a diferentes variables [24]. Se recomienda estandarizar los datos, previo a la realización del análisis; ya que esta acción permite disminuir la variabilidad y facilita su clasificación en gran parte de los casos, porque se elimina la dependencia de cada variable con respecto a sus respectivas unidades de medida empleadas [23].

Por otra parte, es importante mencionar, que existen dos tipos de análisis cluster; el jerárquico y el no jerárquico. El análisis clúster jerárquico, agrupa las observaciones en un diagrama llamado dendograma, conforme a un algoritmo definido, sin posibilidad de reclasificar las observaciones en los diferentes grupos; mientras que en el análisis no jerárquico, el programador puede fijar inicialmente el número de grupos o clusters. Esto claramente es una ventaja sobre el análisis jerárquico, teniendo en cuenta, que es posible cambiar el número de grupos. Por lo que el observador, puede realizar el análisis tomando diferente número de clusters, y posteriormente elegir la clasificación más apropiada posible, donde se clasifiquen las observaciones con la mejor distribución posible, en función de su similaridad. Para determinar si la clasificación agrupa las observaciones correctamente, en un determinado caso de estudio; existen tres conceptos clave, las distancias máxima y media intragrupal y la distancia intergrupala [23].

Para cada clúster, se realiza el cálculo de un centroide. Este corresponde al punto geométrico medio entre las observaciones clasificadas en cada grupo. Las distancia máxima intragrupal e intragrupal media corresponden a la distancia existente entre

la observación más alejada, clasificada en un determinado clúster con respecto a su centroide, y la distancia promedio entre las observaciones contenidas en cada cluster con respecto al centroide, respectivamente [24]. La distancia intergrupala, es la distancia entre los centroides de los diferentes grupos (si se presentan casos donde se consideren más de dos clusters, por ende se presentará más de una distancia intergrupala).

En este orden de ideas, como se mencionó anteriormente, el observador podrá determinar si la clasificación agrupa de manera adecuada las observaciones en función de su similaridad o no. Para esto, se deben comparar las distancias intragrupalas y las intergrupales. Las distancias máximas intragrupalas, deben ser menores a las intergrupales, el cumplimiento de esta premisa, permitiría afirmar que la variabilidad de las muestras contenidas en cada grupo es menor a la variabilidad existente si se compararan observaciones intergrupos. Sin embargo, es posible que esta condición no se cumpla debido a la naturaleza de los datos que se sometan al análisis o debido al número de clusters que el programador defina. Por esto, es conveniente realizar prueba y error, considerando diferentes números de clúster y así poder elegir la mejor distribución conforme al criterio mencionado anteriormente [23] [24]. Finalmente, en la figura 1 se presenta un ejemplo típico de análisis clúster, considerando dos grupos para una determinada muestra de datos. Se puede observar, que las máximas distancias intragrupalas son mayores a la distancia intergrupala, por lo que en este caso sería conveniente, realizar nuevamente el análisis, tomando más de dos clusters con el fin de buscar una mejor clasificación.



Ejemplo gráfico caso de análisis clúster *k-means* tomando dos grupos [25].

ANEXO C. ANALISIS CLUSTER JERARQUICO

El análisis cluster jerárquico (ACJ) es una herramienta exploratoria diseñada para revelar las agrupaciones naturales (o los conglomerados o clusters) dentro de un conjunto de datos que no sería de otra manera evidente. Es el más útil cuando usted desea agrupar un número pequeño (menos que algunos cientos) de objetos. Los objetos en análisis cluster jerárquico pueden ser casos o variables, dependiendo de si usted desea clasificar casos o examinar relaciones entre las variables.

El Análisis Cluster Jerárquico comienza separando cada objeto en un cluster por sí mismo. En cada etapa del análisis, el criterio por el que los objetos son separados se relaja en orden a enlazar los dos conglomerados más similares hasta que todos los objetos sean agrupados en un árbol de clasificación completo.

El criterio básico para cualquier agrupación es la distancia. Los objetos que estén cerca uno del otro pertenecerían al mismo conglomerado o cluster, y los objetos que estén lejos uno del otro pertenecerán a distintos clusters. Para un conjunto de datos dado, los clusters que se construyen dependen de nuestra propia especificación de los siguientes parámetros:

- El método cluster define las reglas para la formación del cluster. Por ejemplo, cuando calculamos la distancia entre dos clusters, podemos usar el par de objetos más cercado entre clusters o el par de objeto más alejados, o un compromiso entre estos métodos.
- La medida define la formula para el cálculo de la distancia. Por ejemplo, la medida de distancia Euclídea calcula la distancia como una línea recta entre dos clusters. Las medidas de intervalo asumen que las variables están medidas en escala; las medidas de conteo asumen que son números discretos, y las medidas binarias asumen que toman dos valores.

- La estandarización permite igualar el efecto de las variables medidas sobre diferentes escalas [28].

ANEXO D. REGRESION POR MINIMOS CUADRADOS PARCIALES (PLS)

La regresión de mínimos cuadrados parciales (PLS) es una técnica, mediante la cual, es posible reducir las variables independientes definidas en un caso de estudio particular, a un conjunto más pequeño de predictores no correlacionados entre sí. Las variables independientes que estén correlacionadas entre sí, son agrupadas en un mismo predictor. Se obtiene una regresión de mínimos cuadrados con estos predictores, en vez de hacerlo con los datos originales. La regresión PLS es ampliamente útil cuando no se tienen más predictores que observaciones y la regresión de mínimos cuadrados ordinarios produce coeficientes con altos errores estándar o falla por completo [20].

La regresión PLS es empleada mayormente en la industria química, de alimentos, de plásticos y de medicamentos. Una aplicación importante, consiste en modelar la relación entre mediciones espectrales (IR, UV, NIR), que incluye múltiples variables que comúnmente están correlacionadas entre sí, y la composición química u otras propiedades fisicoquímicas. La regresión PLS, hace énfasis en el desarrollo de modelos predictivos. Por lo que no se utiliza a menudo para eliminar variables que no son útiles para explicar la respuesta [21] [22].

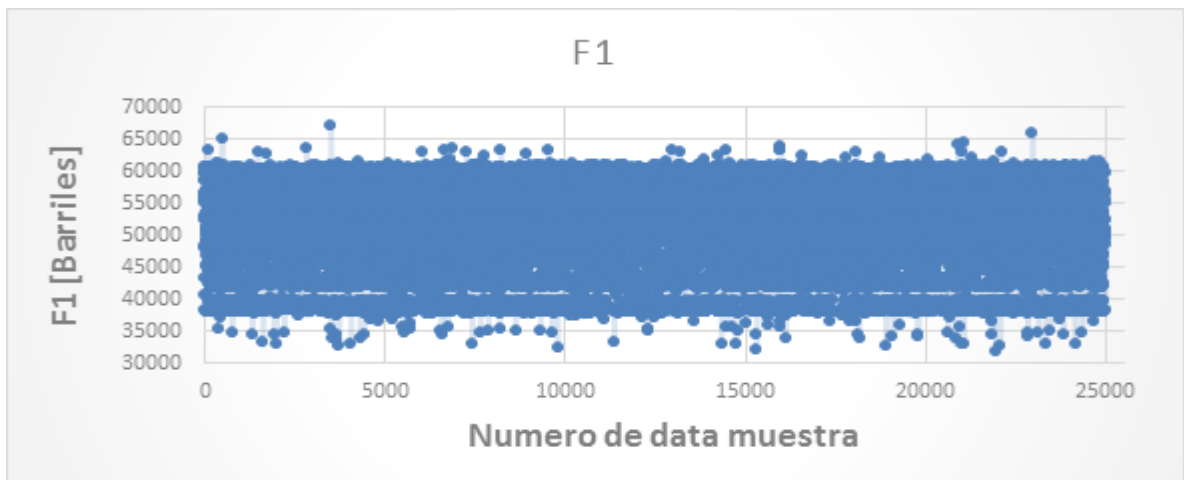
Dado que para el desarrollo del presente método de predicción, en este trabajo se emplea el software Minitab 17; este emplea el algoritmo iterativo no lineal de mínimos cuadrados parciales (NIPALS) creado por Herman Wold. Este algoritmo disminuye el número de predictores, mediante el uso de un método similar al del análisis por componentes principales, para obtener un conjunto de componentes que modele la correlación máxima entre los predictores y las variables de respuesta. Si los predictores están altamente correlacionados, o si un menor número de componentes describe perfectamente la respuesta, entonces el número de componentes en el modelo puede ser mucho menor que el número de predictores. A continuación, Minitab realiza la regresión de mínimos cuadrados en estos

componentes no relacionados. Además, la validación cruzada se utiliza con frecuencia para seleccionar los componentes que aprovechen al máximo la capacidad predictiva del modelo [20].

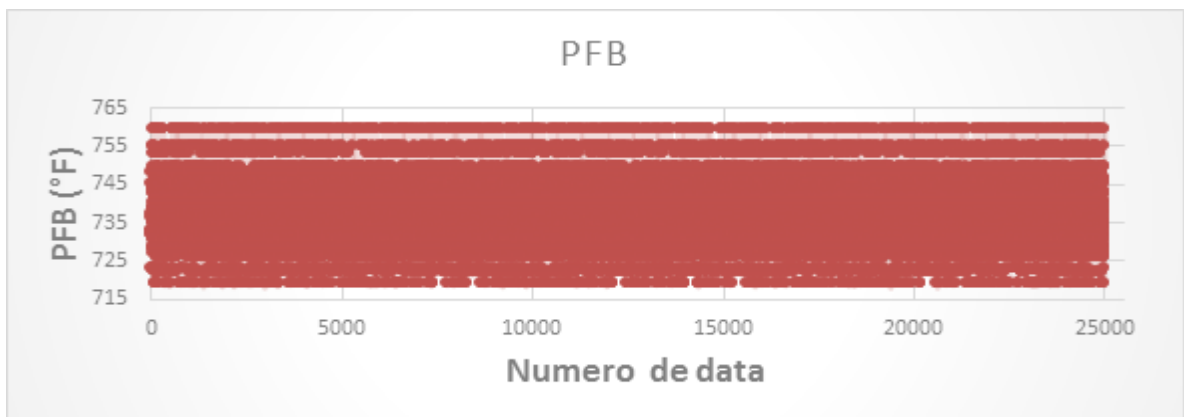
ANEXO E. GRAFICAS VARIABLES CONSIDERADAS CONTRA EL NÚMERO DE OBSERVACION SOBRE LA MATRIZ FINAL DE DATA SELECCIONADA

Después de realizar la selección de la matriz de 12 variables por 25000 observaciones, objeto de análisis en este trabajo; las observaciones tenidas en cuenta de cada variable fueron listadas conforme al orden en que aparecen, con el fin de graficarlas para obtener un mejor panorama del comportamiento univariado de cada variable con respecto al número de observación. A continuación se presentan los gráficos:

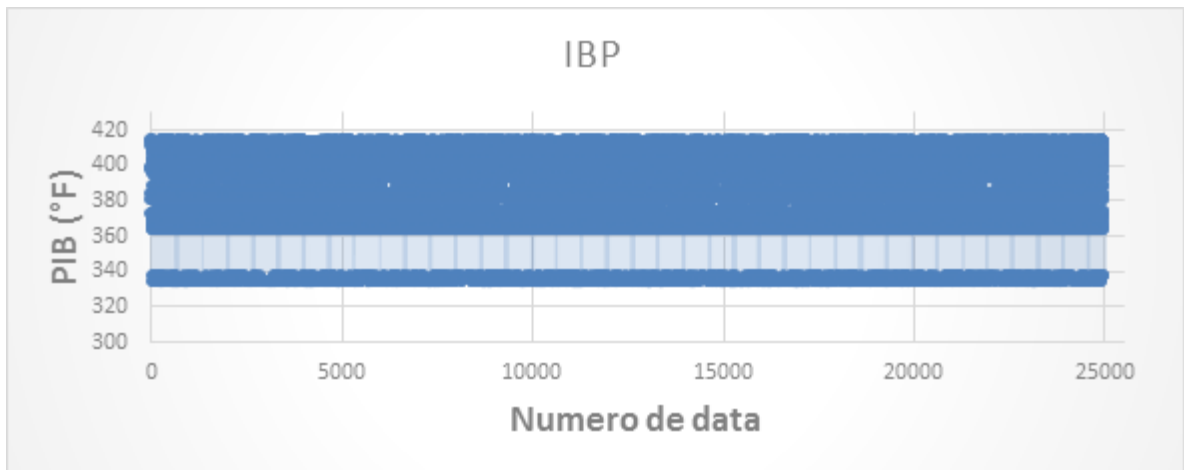
1. Variable F1:



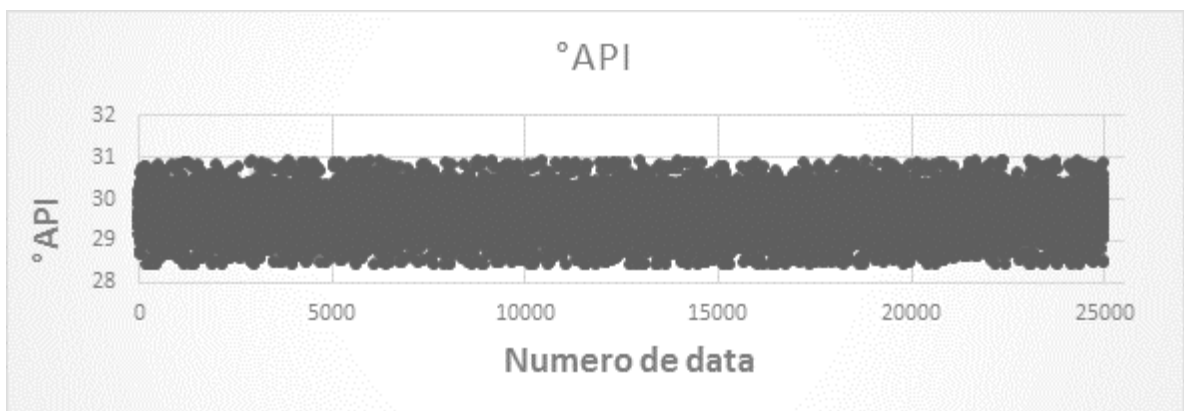
2. Variable PFB:



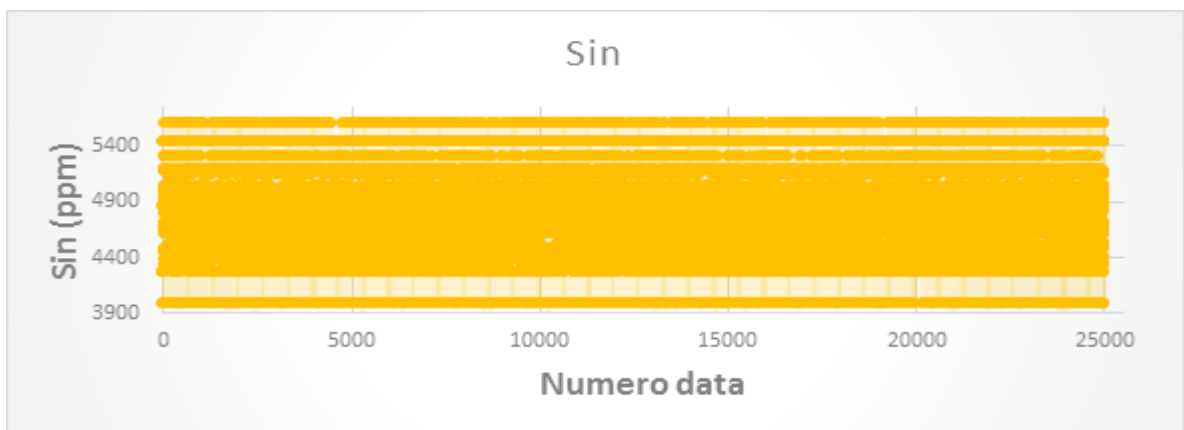
3. Variable IBP:



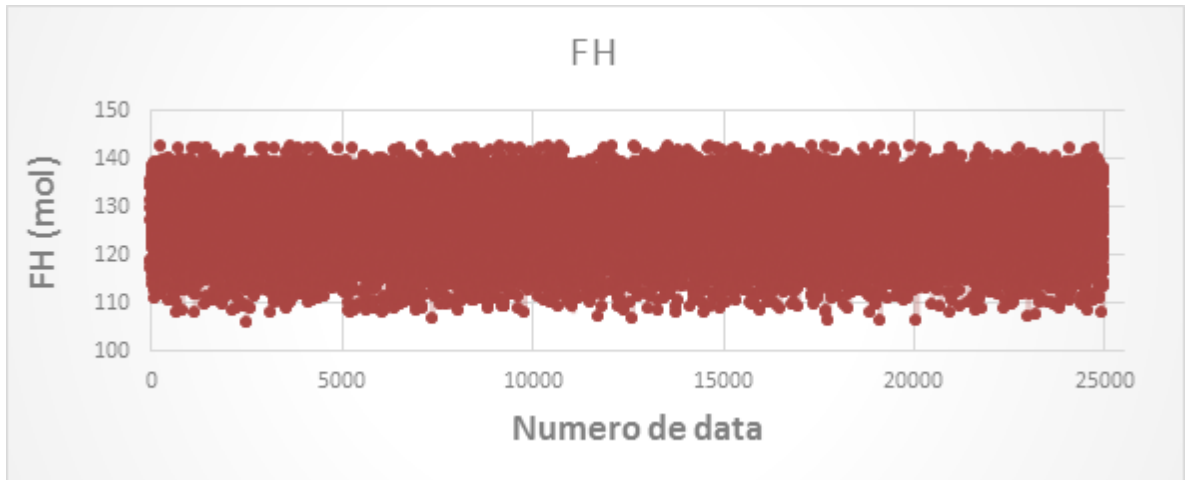
4. Variable °API:



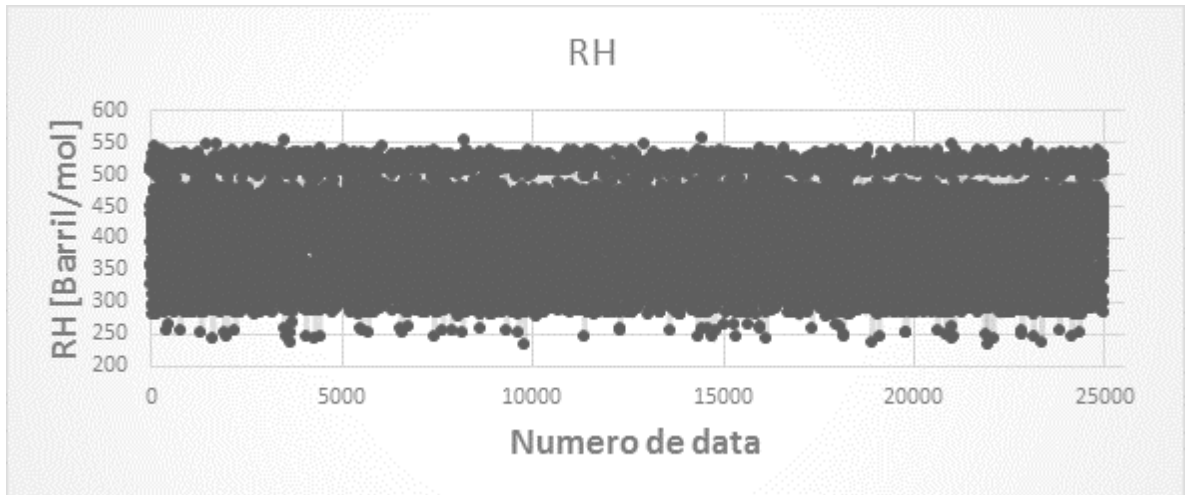
5. Variable Sin:



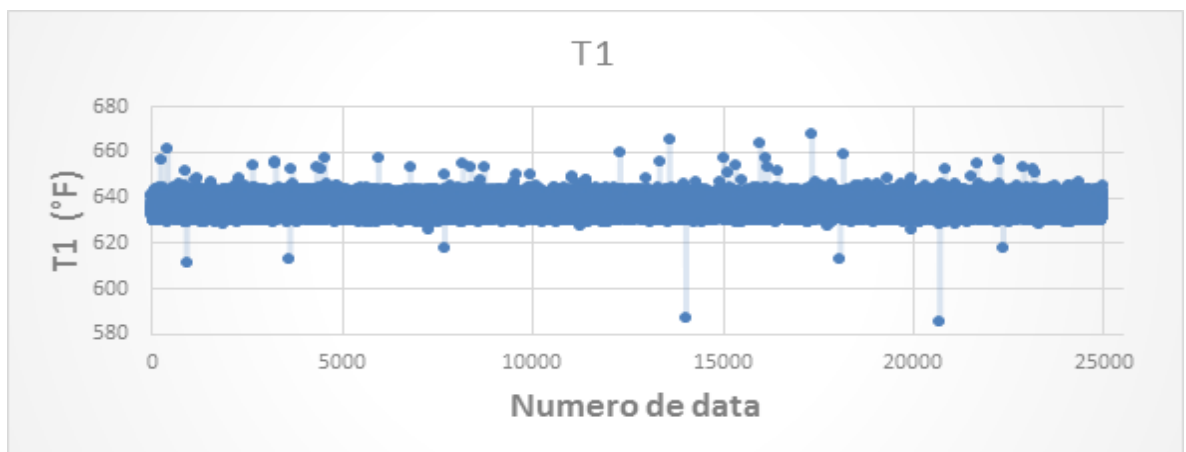
6. Variable FH:



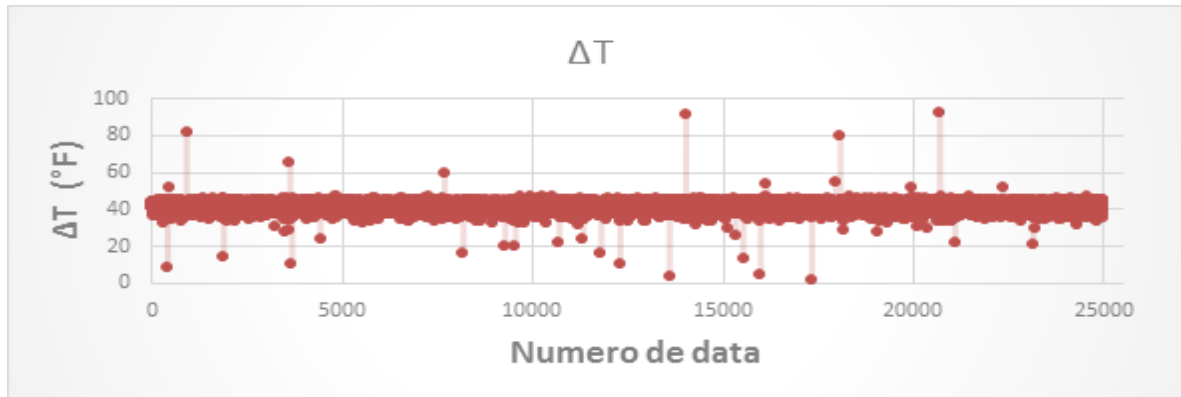
7. Variable RH:



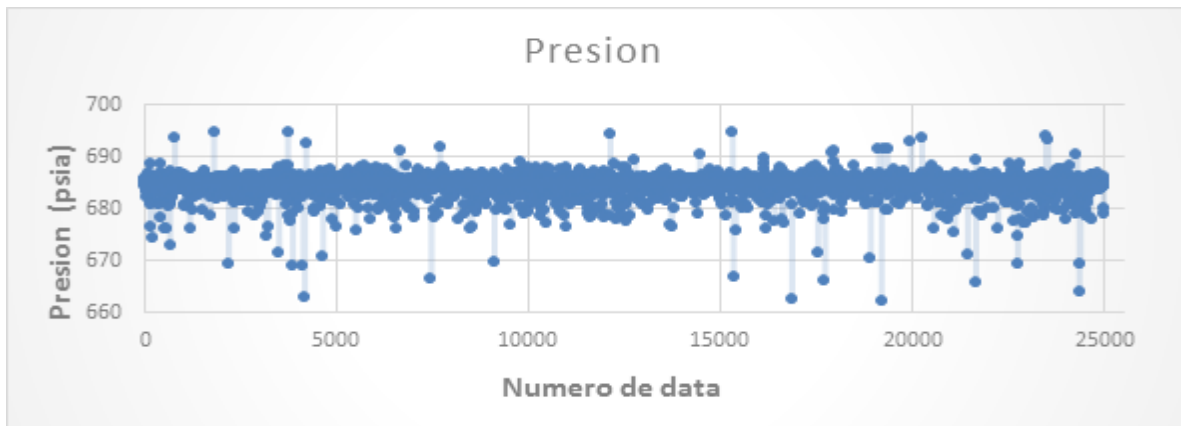
8. Variable T1:



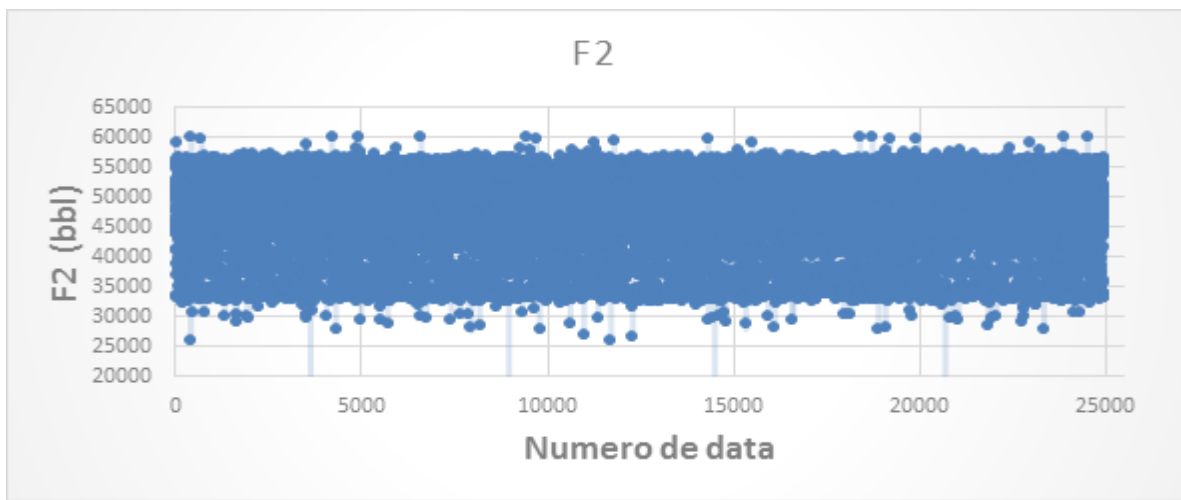
9. Variable ΔT :



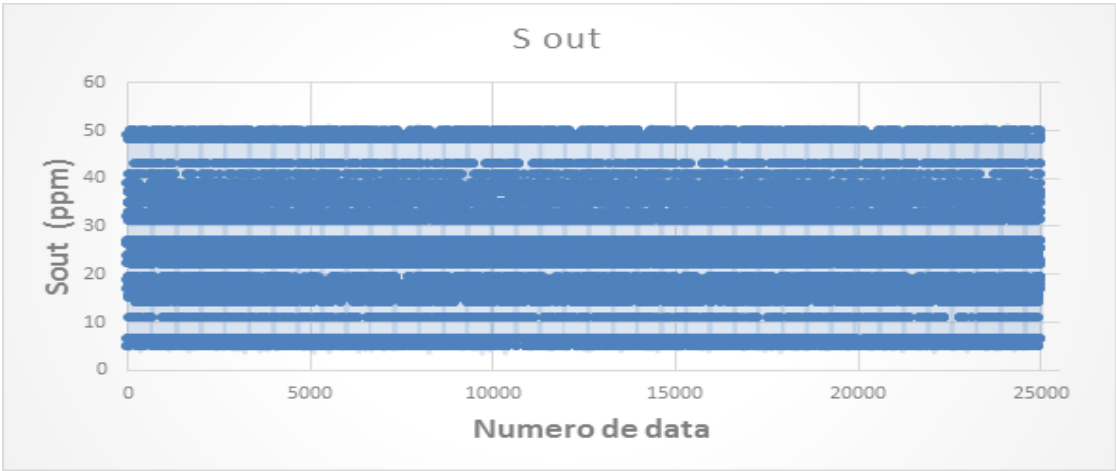
10. Variable P:



11. Variable F2:



12. Variable Sout:



ANEXO F. ANALISIS CLUSTER K- MEDIAS SOBRE LAS OBSERVACIONES DE LAS VARIABLES CONSIDERADAS EN ESTE TRABAJO

A continuación, se presenta las hojas de respuesta proporcionadas por el software minitab 17, dado que el procedimiento se llevó a cabo en dicho software. El análisis fue realizado sobre las variables que se listan a continuación:

- ✓ Carga de alimentación a la zona HDT del diésel a hidrotratar (F1).
- ✓ Punto inicial de ebullición (IBP)
- ✓ Punto final de ebullición carga (FBP)
- ✓ Grados API (API)
- ✓ Contenido de azufre en la carga de alimentación (Sin)
- ✓ Hidrogeno total alimentado zona de HDT (FH)
- ✓ Relación Hidrocarburo/Hidrogeno alimentados a la zona HDT (RH)
- ✓ Temperatura de entrada a la zona de reacción (T1)
- ✓ Cambio de temperatura en la zona de reacción (ΔT)
- ✓ Presión (P)
- ✓ Carga de diésel hidrotratado a la salida de la unidad HDT (F2)

Se tomaron 25000 observaciones, estandarizadas, correspondientes a cada una de las variables listadas anteriormente y se realizó el análisis estableciendo cuatro casos de estudio como se exhibe a continuación:

- ✓ Caso I: clasificación categórica en dos grupos o clusters.
- ✓ Caso II: clasificación categórica en tres grupos o clusters.
- ✓ Caso III: clasificación categórica en cinco grupos o clusters.
- ✓ Caso IV: clasificación categórica en siete grupos o clusters.

En seguida se presentan los resultados obtenidos:

- **CASO I:** clasificación en dos clusters

Análisis de conglomerados de K-medias: F1 FBP IBP API Sin...

Partición final

Número de conglomerados: 2

| Conglomerado | Numero de observaciones en el conglomerado | Suma de cuadrados dentro del conglomerado | Distancia promedio desde el centroide del conglomerado | Distancia Máxima desde el centroide del conglomerado |
|---------------------|---|--|---|---|
| Conglomerado I | 14593 | 105243.201 | 2.443 | 25.233 |
| Conglomerado II | 10404 | 105697.147 | 2.963 | 30.509 |

Centroides de grupo:

| variable | Centroides conglomerado 1 | Centroides conglomerado 2 | Centroide principal |
|-----------------|----------------------------------|----------------------------------|----------------------------|
| F1 | 0.6524 | -0.9150 | -0.0000 |
| FBP | 0.3955 | -0.5548 | 0.0000 |
| IBP | 0.1388 | -0.1947 | -0.0000 |
| API | -0.1390 | 0.1950 | -0.0000 |
| Sin | -0.0758 | 0.1063 | 0.0000 |
| FH | -0.3402 | 0.4771 | -0.0000 |
| RH | 0.6329 | -0.8878 | 0.0000 |
| T1 | 0.5079 | -0.7124 | 0.0000 |
| ΔT | 0.0658 | -0.0923 | 0.0000 |
| P | -0.0404 | 0.0566 | 0.0000 |
| F2 | 0.6474 | -0.9081 | 0.0000 |

Las distancias entre los centroides de conglomerados:

| | Centroides conglomerado 1 | Centroides conglomerado 2 |
|----------------------------------|----------------------------------|----------------------------------|
| Centroides conglomerado 1 | 0.0000 | 3.2465 |
| Centroides conglomerado 2 | 3.2465 | 0.0000 |

- **CASO II:** clasificación en tres clusters

Análisis de conglomerados de K-medias: F1 FBP IBP API Sin,,,

Partición final

Número de conglomerados: 3

| Conglomerado | Numero de observaciones en el conglomerado | Suma de cuadrados dentro del conglomerado | Distancia promedio desde el centroide del conglomerado | Distancia Máxima desde el centroide del conglomerado |
|------------------|--|---|--|--|
| Conglomerado I | 13224 | 80656.672 | 2.298 | 30.860 |
| Conglomerado II | 6238 | 40830.462 | 2.208 | 25.181 |
| Conglomerado III | 5535 | 55973.214 | 2.907 | 22.671 |

Centroides de grupo:

| variable | Centroides conglomerado 1 | Centroides conglomerado 2 | Centroides conglomerado 3 | Centroide principal |
|------------|---------------------------|---------------------------|---------------------------|---------------------|
| F1 | 0.3709 | 0.5477 | -1.5035 | -0.0000 |
| FBP | -0.2929 | 1.0901 | -0.5287 | 0.0000 |
| IBP | -0.3627 | 0.6585 | 0.1245 | -0.0000 |
| API | 0.6039 | -0.9735 | -0.3455 | -0.0000 |
| Sin | -0.1927 | 0.2222 | 0.2099 | 0.0000 |
| FH | -0.5400 | 0.1811 | 1.0860 | -0.0000 |
| RH | 0.4782 | 0.3497 | -1.5366 | 0.0000 |
| T1 | -0.1727 | 0.9858 | -0.6984 | 0.0000 |
| ΔT | 0.1794 | 0.1307 | -0.5758 | 0.0000 |
| P | 0.0494 | -0.1184 | 0.0153 | 0.0000 |
| F2 | 0.3142 | 0.6126 | -1.4412 | 0.0000 |

Las distancias entre los centroides de conglomerados:

| | Centroides conglomerado 1 | Centroides conglomerado 2 | Centroides conglomerado 3 |
|----------------------------------|---------------------------|---------------------------|---------------------------|
| Centroides conglomerado 1 | 0.0000 | 2.7650 | 3.9372 |
| Centroides conglomerado 2 | 2.7650 | 0.0000 | 4.4109 |
| Centroides conglomerado 3 | 3.9372 | 4.4109 | 0.0000 |

- **CASO III:** clasificación en cinco clusters

Análisis de conglomerados de K-medias: F1. FBP. IBP. API. Sin. FH. RH. T1. ΔT.

P. F2

Partición final

Número de conglomerados: 5

| Conglomerado | Numero de observaciones en el conglomerado | Suma de cuadrados dentro del conglomerado | Distancia promedio desde el centroide del conglomerado | Distancia Máxima desde el centroide del conglomerado |
|---------------------|---|--|---|---|
| Conglomerado I | 2731 | 17264.779 | 2.282 | 16.748 |
| Conglomerado II | 6357 | 39824.318 | 2.174 | 25.182 |
| Conglomerado III | 1587 | 8013.538 | 1.970 | 30.528 |
| Conglomerado IV | 3726 | 33132.717 | 2.593 | 22.300 |
| Conglomerado V | 10596 | 52140.539 | 2.081 | 28.509 |

Centroides de grupo:

| variable | Centroides conglomerado 1 | Centroides conglomerado 2 | Centroides conglomerado 3 | Centroides conglomerado 4 | Centroides conglomerado 5 | Centroides principal |
|-----------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|-----------------------------|
| F1 | 1.0257 | 0.5883 | -0.7355 | -1.8885 | 0.1569 | -0.0000 |
| FBP | -0.3436 | 1.1020 | -0.5774 | -0.4038 | -0.3441 | 0.0000 |
| IBP | -0.7904 | 0.5910 | 0.2421 | -0.0220 | -0.1794 | -0.0000 |
| API | 0.2928 | -0.9271 | -1.2018 | 0.0428 | 0.6457 | -0.0000 |
| Sin | -0.1459 | 0.2150 | 1.5441 | -0.0272 | -0.3131 | 0.0000 |
| FH | -1.6868 | 0.1409 | 1.5326 | 0.8473 | -0.1773 | -0.0000 |
| RH | 1.4888 | 0.3950 | -1.0737 | -1.7636 | 0.1602 | 0.0000 |
| T1 | 0.6918 | 0.9905 | -0.2194 | -0.8629 | -0.4363 | 0.0000 |
| ΔT | -0.0763 | 0.1036 | 0.4448 | -1.0709 | 0.2675 | 0.0000 |
| P | -0.0093 | -0.1137 | 0.0671 | -0.0170 | 0.0665 | 0.0000 |
| F2 | 1.0813 | 0.6383 | -0.6296 | -1.8589 | 0.0863 | 0.0000 |

Las distancias entre los centroides de conglomerados:

| | Centroides conglomerado 1 | Centroides conglomerado 2 | Centroides conglomerado 3 | Centroides conglomerado 4 | Centroides conglomerado 5 |
|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
| Centroides conglomerado 1 | 0.0000 | 3.2670 | 5.5026 | 6.1817 | 2.7769 |
| Centroides conglomerado 2 | 3.2670 | 0.0000 | 3.7217 | 5.0985 | 2.8599 |
| Centroides conglomerado 3 | 5.5026 | 3.7217 | 0.0000 | 3.2592 | 3.5956 |
| Centroides conglomerado 4 | 6.1817 | 5.0985 | 3.2592 | 0.0000 | 3.8952 |
| Centroides conglomerado 5 | 2.7769 | 2.8599 | 3.5956 | 3.8952 | 0.0000 |

ANEXO G. WORKSHEETS MINITAB 17 CASOS DE PREDICCIÓN MEDIANTE RML

- ✓ Caso I estimación de la variable S2 en función de las variables: F1, FBP, IBP, API, Sin, RH, ΔT, P y Tin.

Análisis de varianza:

| FUENTE | GL | SC. Ajustado | MC. Ajustado | Valor F | Valor-p |
|-----------|-------|--------------|--------------|---------|---------|
| Regresión | 9 | 6171.5 | 685.72 | 910.20 | 0.000 |
| F1 | 1 | 345.6 | 345.56 | 458.68 | 0.000 |
| FBP | 1 | 401.5 | 401.53 | 532.98 | 0.000 |
| IBP | 1 | 1313.5 | 1313.53 | 1743.53 | 0.000 |
| API | 1 | 337.0 | 336.96 | 447.26 | 0.000 |
| Sin | 1 | 1754.1 | 1754.09 | 2328.32 | 0.000 |
| RH | 1 | 91.0 | 91.0 | 120.79 | 0.000 |
| ΔT | 1 | 12.7 | 12.73 | 16.9 | 0.000 |
| P | 1 | 34.3 | 34.28 | 45.5 | 0.000 |
| Tin | 1 | 4.8 | 4.81 | 6.38 | 0.012 |
| Error | 24987 | 18824.5 | 0.75 | | |
| Total | 24996 | 24996.6 | | | |

Resumen del modelo:

| s | R-cuadrado | R-cuadrado ajustado | R- Cuadrado predicho |
|---------|------------|---------------------|----------------------|
| 0.86797 | 0.2469 | 0.2466 | 0.2462 |

Ecuación de ajuste obtenida:

$$S_{out} = 0,00000 + 0,4512 F1 + 0,15209 FBP - 0,25898 IBP + 0,14568 API + 0,28687 Sin - 0,2144 RH - 0,02676 \Delta T - 0,03749 P + 0,01984 Tin$$

- ✓ Caso II estimación de la variable F2 en función de las variables: F1, FBP, IBP, API, Sin, RH, ΔT, P y Tin.

Análisis de varianza:

| FUENTE | GL | SC. Ajustado | MC. Ajustado | Valor F | Valor-p |
|------------------|-----------|-------------------------|-------------------------|----------------|----------------|
| Regresión | 9 | 23318.4 | 2590.94 | 38591.72 | 0.000 |
| F1 | 1 | 1080.7 | 1080.68 | 16096.64 | 0.000 |
| FBP | 1 | 30.6 | 30.61 | 455.91 | 0.000 |
| IBP | 1 | 30.5 | 30.54 | 454.82 | 0.000 |
| API | 1 | 48.4 | 48.36 | 720.37 | 0.000 |
| Sin | 1 | 0.5 | 0.52 | 7.79 | 0.005 |
| RH | 1 | 45.4 | 45.41 | 676.38 | 0.000 |
| ΔT | 1 | 1.9 | 1.89 | 28.17 | 0.000 |
| P | 1 | 0.8 | 0.76 | 11.29 | 0.001 |
| Tin | 1 | 45.1 | 45.11 | 671.88 | 0.000 |
| Error | 24987 | 1677.6 | 0.07 | | |
| Total | 24996 | 24996.0 | | | |

Resumen del modelo:

| s | R-cuadrado | R-cuadrado ajustado | R- Cuadrado predicho |
|----------|-------------------|--------------------------------|---------------------------------|
| O.259108 | 0.9329 | 0.9326 | 0.9326 |

- ✓ Caso III estimación de la variable RH en función de las variables: F1, FBP, IBP, API, Sin, S out, ΔT, P y T in

Análisis de varianza:

| FUENTE | GL | SC. Ajustado | MC. Ajustado | Valor F | Valor-p |
|------------------|-----------|-------------------------|-------------------------|----------------|----------------|
| Regresión | 9 | 23025.9 | 2558.4 | 32449.06 | 0.000 |
| F1 | 1 | 12486.1 | 12486.1 | 158363.3 | 0.000 |
| FBP | 1 | 0.0 | 0.0 | 0.11 | 0.735 |
| IBP | 1 | 150.6 | 150.6 | 1910.18 | 0.000 |
| API | 1 | 81.3 | 81.3 | 1030.83 | 0.000 |
| Sin | 1 | 1.8 | 1.8 | 23.26 | 0.005 |
| Sout | 1 | 9.5 | 9.5 | 120.79 | 0.000 |
| ΔT | 1 | 61.4 | 61.4 | 778.5 | 0.000 |
| P | 1 | 0.9 | 0.9 | 11.49 | 0.001 |
| Tin | 1 | 35.6 | 35.6 | 451.96 | 0.000 |
| Error | 24987 | 1970.1 | 0.1 | | |

| | | | | | |
|--------------|-------|---------|--|--|--|
| Total | 24996 | 24996.0 | | | |
|--------------|-------|---------|--|--|--|

Resumen del modelo:

| s | R-cuadrado | R-cuadrado ajustado | R- Cuadrado predicho |
|----------|-------------------|----------------------------|-----------------------------|
| 0.280793 | 0.9212 | 0.9212 | 0.9211 |

- ✓ Caso IV estimación de la variable FH en función de las variables: F1, FBP, IBP, API, Sin, S out, ΔT , P y T in

Análisis de varianza:

| FUENTE | GL | SC. Ajustado | MC. Ajustado | Valor F | Valor-p |
|------------------------------|-----------|---------------------|---------------------|----------------|----------------|
| Regresión | 9 | 9853.7 | 1094.85 | 1806.66 | 0.000 |
| F1 | 1 | 4693.3 | 4693.34 | 7744.67 | 0.000 |
| FBP | 1 | 39.5 | 39.47 | 65.13 | 0.735 |
| IBP | 1 | 1038.2 | 1038.19 | 1713.16 | 0.000 |
| API | 1 | 827.9 | 827.87 | 1366.10 | 0.000 |
| Sin | 1 | 18.7 | 18.71 | 30.88 | 0.005 |
| Sout | 1 | 7.4 | 7.38 | 12.18 | 0.000 |
| ΔT | 1 | 435.1 | 435.06 | 717.92 | 0.000 |
| P | 1 | 5.3 | 5.29 | 8.73 | 0.003 |
| Tin | 1 | 416.3 | 416.29 | 686.95 | 0.000 |
| Error | 24987 | 15142.3 | 0.61 | | |
| Total | 24996 | 24996.0 | | | |

Resumen del modelo:

| s | R-cuadrado | R-cuadrado ajustado | R- Cuadrado predicho |
|----------|-------------------|----------------------------|-----------------------------|
| 0.778466 | 0.3942 | 0.3940 | 0.3936 |

Ecuación de ajuste obtenida:

$$FH = -0,00000 - 0,61949 F1 - 0,04819 FBP + 0,22953 IBP - 0,22577 API - 0,03097 Sin + 0,01975 Sout + 0,15412 \Delta T + 0,01474 P + 0,18297 Tin$$

- ✓ Caso V estimación de la variable ΔT en función de las variables: F1, FBP, IBP, API, Sin, FH, P y Tin

Análisis de varianza:

| FUENTE | GL | SC. Ajustado | MC. Ajustado | Valor F | Valor-p |
|-----------|-------|-----------------|-----------------|---------|---------|
| Regresión | 8 | 7185.6 | 898.20 | 1260.17 | 0.000 |
| F1 | 1 | 6045.4 | 6045.39 | 8481.67 | 0.000 |
| FBP | 1 | 32.0 | 31.98 | 44.86 | 0.000 |
| IBP | 1 | 30.2 | 30.22 | 42.41 | 0.000 |
| API | 1 | 181.4 | 181.40 | 254.5 | 0.000 |
| Sin | 1 | 5.1 | 5.08 | 7.13 | 0.008 |
| Sout | 1 | 509.7 | 509.69 | 715.09 | 0.000 |
| P | 1 | 274.5 | 274.48 | 385.1 | 0.000 |
| Tin | 1 | 2289.2 | 2289.17 | 3211.7 | 0.000 |
| Error | 24988 | 17810.4 | 0.71 | | |
| Total | 24996 | 24996.0 | | | |

Resumen del modelo:

| s | R-cuadrado | R-cuadrado ajustado | R- Cuadrado predicho |
|----------|------------|------------------------|-------------------------|
| 0.844251 | 0.2875 | 0.2872 | 0.2850 |

Ecuación de ajuste obtenida:

$$\Delta T = 0,00000 + 0,68441 F1 - 0,04293 FBP + 0,03923 IBP - 0,10716 API + 0,01545 Sin + 0,18085 FH + 0,10527 P - 0,40931 Tin$$

ANEXO H. CASOS DE PREDICCIÓN I, IV Y V MEDIANTE RML UTILIZANDO EL METODO DE ELIMINACIÓN HACIA ATRÁS

A continuación, se presenta un resumen, análisis de varianza, los valores de R cuadrado y R cuadrado ajustado calculados, sobre los regresores que quedan en cada modelo, después de aplicar el método de eliminación hacia atrás en la estimación de las respectivas variables de salida para cada caso. Cabe resaltar, que no se presenta la eliminación de cada predictor paso a paso; se presenta una ficha general para cada caso de estudio. El intervalo de confianza fijado fue del 95%.

- ✓ Caso I estimación de la variable S2 en función de las variables: F1, FBP, IBP, API, Sin, RH, ΔT , P y Tin.

Análisis de varianza:

| FUENTE | GL | SC. AJUSTADO | MC. AJUSTADO | VALOR F | VALOR P |
|-----------------|----|--------------|--------------|---------|---------|
| REGRESION | 38 | 14539.5 | 382.62 | 913.25 | 0.000 |
| F1*F1 | 1 | 228.9 | 228.89 | 546.32 | 0.000 |
| F1*FBP | 1 | 161.2 | 161.20 | 384.76 | 0.000 |
| F1*IBP | 1 | 229.8 | 229.80 | 548.49 | 0.000 |
| F1*API | 1 | 149.4 | 149.44 | 356.69 | 0.000 |
| F1*RH | 1 | 266.7 | 266.70 | 636.58 | 0.000 |
| F1* ΔT | 1 | 414.6 | 414.63 | 989.65 | 0.000 |
| F1*Tin | 1 | 250.8 | 250.78 | 598.57 | 0.000 |
| FBP*FBP | 1 | 546.4 | 546.39 | 1304.15 | 0.000 |
| FBP*IBP | 1 | 462.2 | 462.22 | 1103.24 | 0.000 |
| FBP*API | 1 | 160.5 | 160.46 | 383.00 | 0.000 |
| FBP*Sin | 1 | 81.8 | 81.75 | 195.13 | 0.000 |
| FBP*RH | 1 | 317.3 | 317.29 | 757.32 | 0.000 |
| FBP* ΔT | 1 | 107.9 | 107.92 | 257.59 | 0.000 |
| FBP*P | 1 | 23.4 | 23.39 | 55.82 | 0.000 |
| FBP*Tin | 1 | 427.9 | 427.89 | 1021.30 | 0.000 |
| IBP*IBP | 1 | 1163.4 | 1163.40 | 2776.85 | 0.000 |
| IBP*API | 1 | 147.1 | 147.12 | 351.16 | 0.000 |
| IBP*Sin | 1 | 568.9 | 568.91 | 1357.91 | 0.000 |
| IBP*RH | 1 | 185.1 | 185.08 | 441.76 | 0.000 |
| IBP* ΔT | 1 | 53.1 | 53.08 | 126.70 | 0.000 |

| | | | | | |
|---------|-------|---------|---------|---------|-------|
| IBP*Tin | 1 | 517.4 | 517.45 | 1235.07 | 0.000 |
| API*API | 1 | 51.3 | 51.27 | 122.38 | 0.000 |
| API*Sin | 1 | 1133.9 | 1133.91 | 2706.48 | 0.000 |
| API*RH | 1 | 44.5 | 44.5 | 106.22 | 0.000 |
| API* ΔT | 1 | 19.4 | 19.37 | 46.23 | 0.000 |
| API*P | 1 | 6.2 | 6.18 | 14.75 | 0.000 |
| API*Tin | 1 | 20.1 | 20.08 | 47.92 | 0.000 |
| Sin*Sin | 1 | 3011.4 | 3011.38 | 7187.70 | 0.000 |
| Sin* ΔT | 1 | 604.8 | 604.85 | 1443.68 | 0.000 |
| Sin*P | 1 | 23.6 | 23.55 | 56.21 | 0.000 |
| Sin*Tin | 1 | 5.3 | 5.26 | 12.55 | 0.000 |
| RH*RH | 1 | 320.4 | 320.40 | 764.75 | 0.000 |
| RH* ΔT | 1 | 447.5 | 447.51 | 1068.15 | 0.000 |
| RH*P | 1 | 2.2 | 2.19 | 5.22 | 0.022 |
| RH*Tin | 1 | 321.0 | 321.01 | 766.19 | 0.000 |
| ΔT* ΔT | 1 | 31.2 | 31.19 | 74.44 | 0.000 |
| ΔT*Tin | 1 | 19.1 | 19.06 | 45.49 | 0.000 |
| P*P | 1 | 1.7 | 1.67 | 4.00 | 0.046 |
| Error | 24998 | 10456.5 | 0.42 | | |
| Total | 24996 | 24996.0 | | | |

Resumen del modelo:

| S | R-CUADRADO | R-CUADRADO AJUSTADO | R-CUADRADO PREDICHO |
|----------|------------|---------------------|---------------------|
| 0.647273 | 0.5817 | 0.5810 | 0.5794 |

Ecuación de ajuste obtenida:

$$\begin{aligned}
\text{Sout} = & 0,01203 - 1,1577 F1*F1 - 0,4702 F1*FBP - 0,4733 F1*IBP - 0,5051 F1*API \\
& + 2,2654 F1*RH + 0,6017 F1*\Delta T + 0,6315 F1*Tin \\
& + 0,24242 FBP*FBP + 0,30015 FBP*IBP + 0,13559 FBP*API - 0,08247 FBP*Sin \\
& + 0,6237 FBP*RH - 0,12463 FBP*\Delta T + 0,05480 FBP*P - 0,30610 FBP*Tin \\
& + 0,26398 IBP*IBP - 0,18103 IBP*API - 0,28800 IBP*Sin + 0,4059 IBP*RH \\
& - 0,07975 IBP*\Delta T - 0,29328 IBP*Tin + 0,05150 API*API - 0,34395 API*Sin \\
& + 0,2624 API*RH - 0,05648 API*\Delta T + 0,02538 API*P - 0,0702 API*Tin \\
& - 0,43594 Sin*Sin - 0,21495 Sin*\Delta T + 0,05337 Sin*P - 0,02614 Sin*Tin
\end{aligned}$$

$$- 1,1673 \text{ RH} \cdot \text{RH} - 0,6352 \text{ RH} \cdot \Delta\text{T} + 0,01057 \text{ RH} \cdot \text{P} - 0,7115 \text{ RH} \cdot \text{Tin} + 0,01137 \Delta\text{T} \cdot \Delta\text{T} + 0,02022 \Delta\text{T} \cdot \text{Tin} + 0,000880 \text{ P} \cdot \text{P}$$

- ✓ Caso IV estimación de la variable FH en función de las variables: F1, FBP, IBP, API, Sin, S out, ΔT, P y T in

Análisis de varianza:

| FUENTE | GL | SC AJUSTADO | MC AJUSTADO | VALOR F | VALOR P |
|-----------|----|-------------|-------------|---------|---------|
| REGRESION | 36 | 12704.0 | 352.89 | 716.58 | 0.000 |
| F1*F1 | 1 | 5.9 | 5.86 | 11.90 | 0.001 |
| F1*IBP | 1 | 42.4 | 42.38 | 86.06 | 0.000 |
| F1*Sin | 1 | 27.7 | 27.74 | 56.32 | 0.000 |
| F1*Sout | 1 | 1013.9 | 1013.85 | 2058.72 | 0.000 |
| F1* ΔT | 1 | 64.9 | 64.95 | 131.88 | 0.000 |
| F1*Tin | 1 | 171.0 | 171.05 | 347.33 | 0.000 |
| FBP*IBP | 1 | 632.1 | 632.09 | 1283.52 | 0.000 |
| FBP*API | 1 | 123.0 | 122.97 | 249.71 | 0.000 |
| FBP*Sin | 1 | 3.7 | 3.67 | 7.46 | 0.006 |
| FBP*Sout | 1 | 1184.2 | 1184.21 | 2404.66 | 0.000 |
| FBP* ΔT | 1 | 81.2 | 81.17 | 164.83 | 0.000 |
| FBP*Tin | 1 | 352.1 | 352.06 | 714.89 | 0.000 |
| IBP*IBP | 1 | 1171.3 | 1171.28 | 2378.40 | 0.000 |
| IBP*API | 1 | 23.6 | 23.57 | 47.86 | 0.000 |
| IBP*Sin | 1 | 228.7 | 228.66 | 464.31 | 0.000 |
| IBP*Sout | 1 | 7.2 | 7.15 | 14.53 | 0.000 |
| IBP*P | 1 | 25.1 | 25.10 | 50.97 | 0.000 |
| IBP*Tin | 1 | 244.0 | 243.99 | 495.44 | 0.000 |
| API*API | 1 | 18.4 | 18.42 | 37.41 | 0.000 |
| API*Sin | 1 | 10.1 | 10.12 | 20.56 | 0.000 |
| API*Sout | 1 | 94.3 | 94.33 | 191.54 | 0.000 |
| API* ΔT | 1 | 76.7 | 76.74 | 155.83 | 0.000 |
| API*Tin | 1 | 65.1 | 65.08 | 132.14 | 0.000 |
| Sin*Sin | 1 | 736.7 | 736.70 | 1495.95 | 0.000 |
| Sin*Sout | 1 | 264.9 | 264.92 | 537.95 | 0.000 |
| Sin* ΔT | 1 | 35.8 | 35.77 | 72.63 | 0.000 |
| Sin*P | 1 | 18.2 | 18.23 | 37.02 | 0.000 |
| Sin*Tin | 1 | 252.8 | 252.81 | 513.36 | 0.000 |
| Sout*Sout | 1 | 217.9 | 217.94 | 442.55 | 0.000 |
| Sout* ΔT | 1 | 75.7 | 75.66 | 153.63 | 0.000 |

| FUENTE | GL | SC AJUSTADO | MC AJUSTADO | VALOR F | VALOR P |
|-----------------------|-------|-------------|-------------|---------|---------|
| Sout*Tin | 1 | 71.2 | 71.24 | 144.66 | 0.000 |
| $\Delta T^* \Delta T$ | 1 | 7.1 | 7.09 | 14.41 | 0.000 |
| $\Delta T^* Tin$ | 1 | 44.7 | 44.69 | 90.76 | 0.000 |
| P*P | 1 | 26.3 | 26.32 | 53.45 | 0.000 |
| P*Tin | 1 | 15.5 | 15.50 | 31.47 | 0.000 |
| Tin*Tin | 1 | 85.5 | 85.48 | 173.58 | 0.000 |
| ERROR | 24960 | 12292.0 | 0.49 | | |
| TOTAL | 24996 | 24996.0 | | | |

Resumen del modelo:

| S | R-CUADRADO | R-CUADRADO AJUSTADO | R-CUADRADO PREDICHO |
|----------|------------|---------------------|---------------------|
| 0.701759 | 0.5082 | 0.5075 | 0.5053 |

Ecuación de ajuste obtenida:

$$\begin{aligned}
 FH = & -0,0463 - 0,02539 F1^*F1 + 0,06004 F1^*IBP + 0,07103 F1^*Sin \\
 & + 0,41597 F1^*Sout - 0,06130 F1^*\Delta T - 0,16887 F1^*Tin - 0,3694 FBP^*IBP \\
 & + 0,10160 FBP^*API + 0,02150 FBP^*Sin - 0,37313 FBP^*Sout - 0,10096 FBP^*\Delta T \\
 & + 0,18990 FBP^*Tin - 0,23178 IBP^*IBP - 0,0698 IBP^*API + 0,18654 IBP^*Sin \\
 & - 0,0413 IBP^*Sout - 0,04324 IBP^*P + 0,15839 IBP^*Tin + 0,02977 API^*API \\
 & - 0,03748 API^*Sin + 0,09415 API^*Sout - 0,09531 API^*\Delta T - 0,10331 API^*Tin \\
 & + 0,24882 Sin^*Sin + 0,22300 Sin^*Sout - 0,06204 Sin^*\Delta T - 0,03867 Sin^*P \\
 & - 0,22021 Sin^*Tin - 0,12047 Sout^*Sout + 0,07654 Sout^*\Delta T + 0,1307 Sout^*Tin \\
 & + 0,00567 \Delta T^*\Delta T + 0,04351 \Delta T^*Tin - 0,003533 P^*P - 0,02424 P^*Tin \\
 & + 0,06218 Tin^*Tin
 \end{aligned}$$

- ✓ Caso V estimación de la variable ΔT en función de las variables: F1, FBP, IBP, API, Sin, FH, P y Tin

Análisis de varianza:

| FUENTE | GL | SC AJUSTADO | MC AJUSTADO | VALOR F | VALOR P |
|------------------|-----------|--------------------|--------------------|----------------|----------------|
| REGRESION | 34 | 11889.7 | 349.70 | 666.02 | 0.000 |
| F1*F1 | 1 | 2478.8 | 2478.84 | 4721.15 | 0.000 |
| F1*FBP | 1 | 28.8 | 28.77 | 54.80 | 0.000 |
| F1*IBP | 1 | 62.7 | 62.74 | 119.48 | 0.000 |
| F1*API | 1 | 175.8 | 175.82 | 334.86 | 0.000 |
| F1*Sin | 1 | 14.3 | 14.28 | 27.21 | 0.000 |
| F1*FH | 1 | 340.7 | 340.68 | 648.85 | 0.000 |
| F1*P | 1 | 201.6 | 201.57 | 383.90 | 0.000 |
| F1*Tin | 1 | 624.7 | 624.69 | 1189.77 | 0.000 |
| FBP*IBP | 1 | 49.6 | 49.62 | 94.50 | 0.000 |
| FBP*API | 1 | 8.9 | 8.87 | 16.90 | 0.000 |
| FBP*Sin | 1 | 116.2 | 116.15 | 221.22 | 0.000 |
| FBP*FH | 1 | 118.5 | 118.54 | 225.76 | 0.000 |
| IBP*P | 1 | 19.4 | 19.43 | 37.00 | 0.000 |
| IBP*Tin | 1 | 34.0 | 34.05 | 64.84 | 0.000 |
| API*API | 1 | 21.6 | 21.57 | 41.08 | 0.000 |
| API*Sin | 1 | 148.6 | 148.62 | 283.06 | 0.000 |
| API*Sout | 1 | 27.5 | 27.46 | 52.29 | 0.000 |
| API*FH | 1 | 87.7 | 87.71 | 167.06 | 0.000 |
| API*P | 1 | 142.9 | 142.95 | 272.25 | 0.000 |
| API*Tin | 1 | 125.0 | 125.04 | 238.15 | 0.000 |
| Sin*Sin | 1 | 183.9 | 183.93 | 350.30 | 0.000 |
| Sin*Sout | 1 | 368.5 | 368.47 | 701.79 | 0.000 |
| Sin*FH | 1 | 94.4 | 94.37 | 179.73 | 0.000 |
| Sin*P | 1 | 104.9 | 104.90 | 199.78 | 0.000 |
| Sin*Tin | 1 | 218.2 | 218.21 | 415.60 | 0.000 |
| Sout*Sout | 1 | 297.0 | 297.01 | 565.69 | 0.000 |
| Sout*FH | 1 | 662.3 | 662.34 | 1261.47 | 0.000 |
| Sout*P | 1 | 10.5 | 10.45 | 19.91 | 0.000 |
| Sout*Tin | 1 | 397.1 | 397.05 | 756.22 | 0.000 |
| FH*FH | 1 | 74.8 | 74.85 | 142.55 | 0.000 |
| FH*P | 1 | 49.3 | 49.27 | 93.84 | 0.000 |
| P*P | 1 | 100.7 | 100.68 | 191.75 | 0.000 |
| P*Tin | 1 | 153.7 | 153.67 | 292.67 | 0.000 |
| Tin*Tin | 1 | 19.3 | 19.30 | 36.77 | 0.000 |
| ERROR | 24962 | 13106.3 | 0.53 | | |
| TOTAL | 24996 | 24996.0 | | | |

Resumen del modelo:

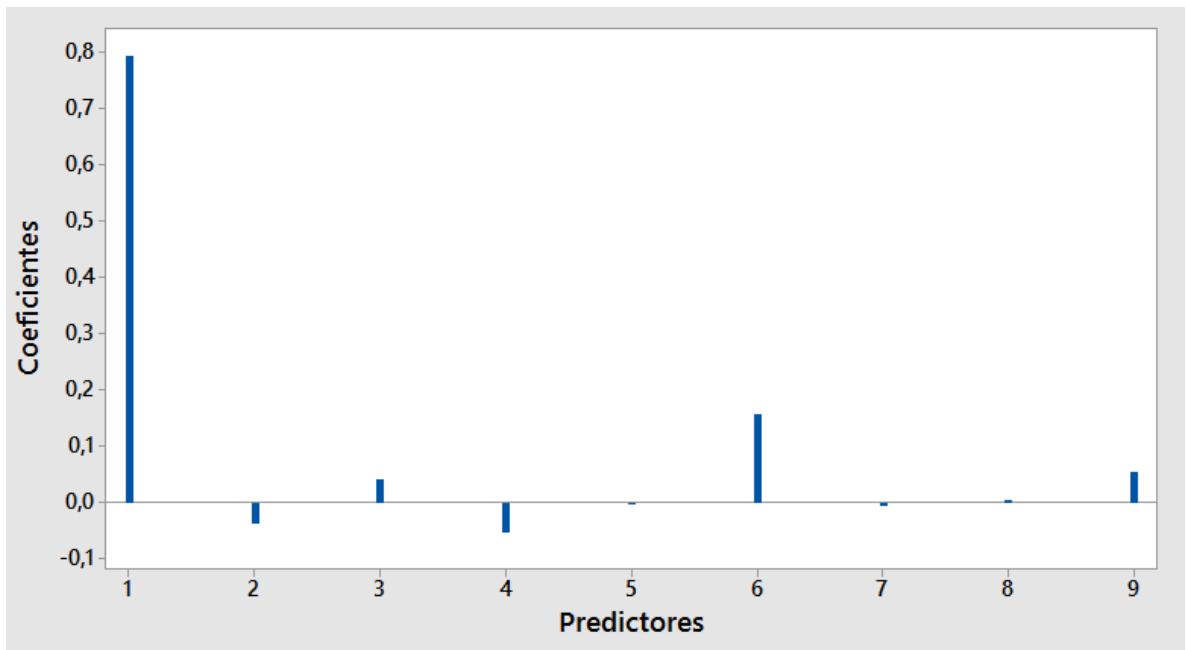
| S | R-CUADRADO | R-CUADRADO AJUSTADO | R-CUADRADO PREDICHO |
|----------|-------------------|----------------------------|----------------------------|
| 0.724604 | 0.4757 | 0.4749 | 0.4495 |

Ecuación de ajuste obtenida:

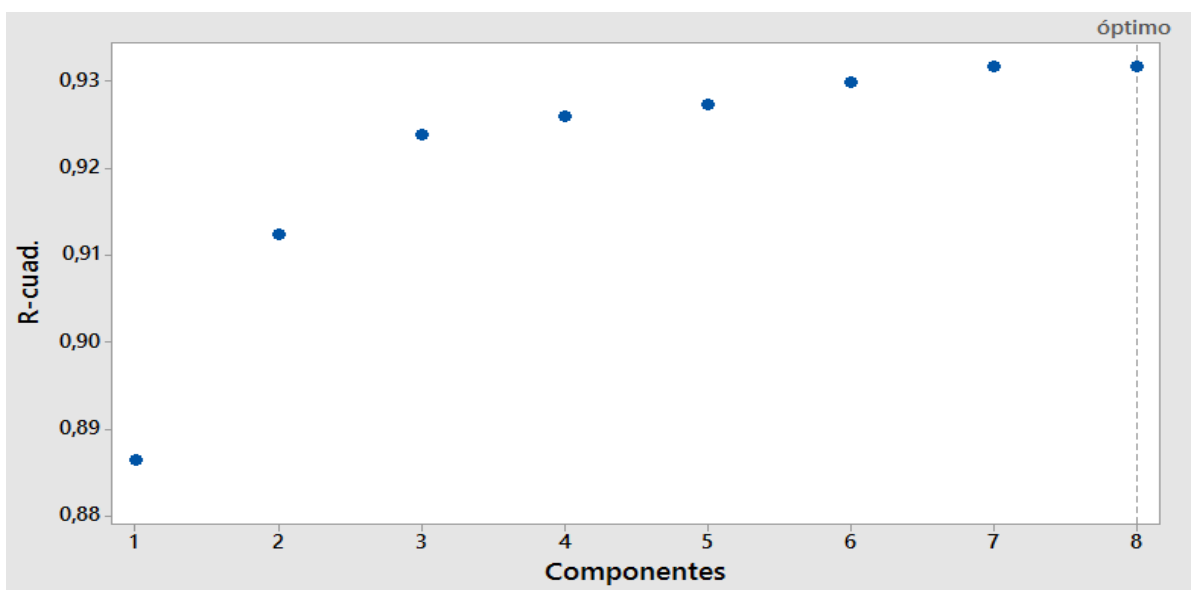
$$\begin{aligned} \Delta T = & 0,11047 - 0,38459 F1 * F1 - 0,05833 F1 * FBP - 0,07541 F1 * IBP \\ & + 0,16416 F1 * API + 0,04516 F1 * Sin - 0,21592 F1 * FH - 0,09694 F1 * P \\ & + 0,24193 F1 * Tin + 0,08723 FBP * IBP + 0,02649 FBP * API + 0,10948 FBP * Sin \\ & - 0,12443 FBP * FH - 0,04707 IBP * P - 0,05013 IBP * Tin - 0,03083 API * API \\ & + 0,13885 API * Sin - 0,04324 API * Sout - 0,12562 API * FH \\ & + 0,12054 API * P - 0,14367 API * Tin + 0,13827 Sin * Sin - 0,23968 Sin * Sout \\ & - 0,12053 Sin * FH - 0,11547 Sin * P + 0,18183 Sin * Tin + 0,13213 Sout * Sout \\ & + 0,30684 Sout * FH + 0,03048 Sout * P - 0,24525 Sout * Tin - 0,05955 FH * FH \\ & + 0,06507 FH * P - 0,010436 P * P + 0,08089 P * Tin + 0,01276 Tin * Tin \end{aligned}$$

ANEXO I. GRAFICAS DE SELECCIÓN MODELO DE REGRESIÓN, DE RESPUESTA DE REGRESIÓN PLS Y DE COEFICIENTES. CASO DE ESTUDIO II.

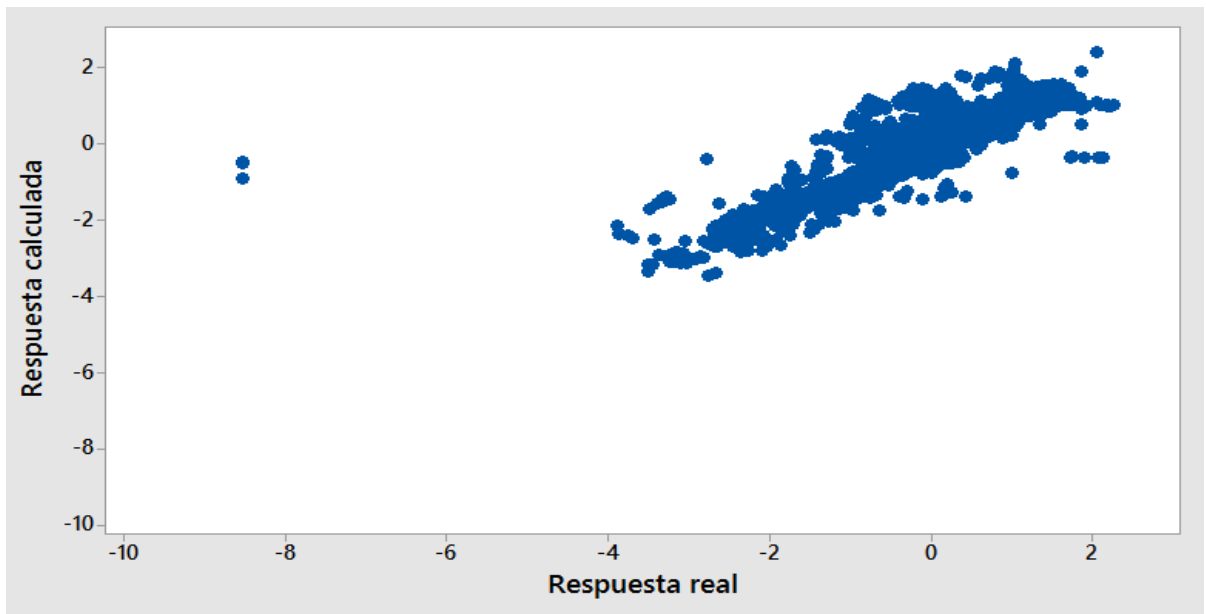
Grafica coeficientes caso de estudio II.



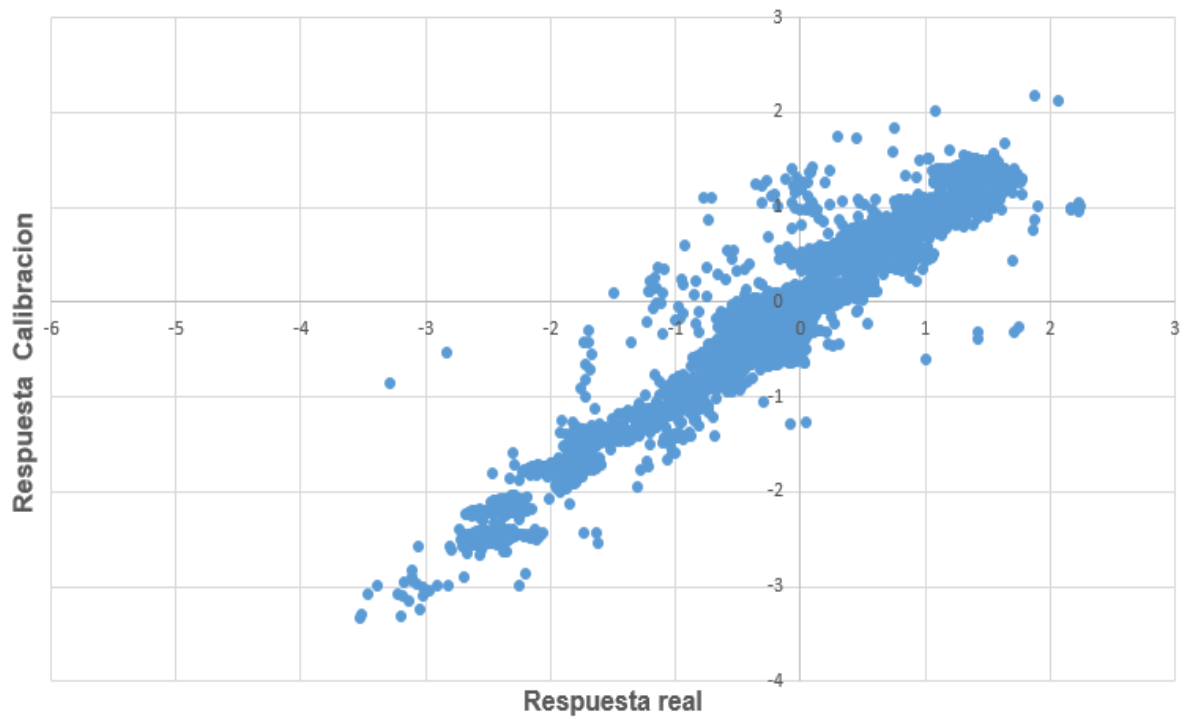
Grafica de selección de modelos regresión PLS caso de estudio II.



Grafica de respuesta planteamiento regresión PLS caso de estudio II.

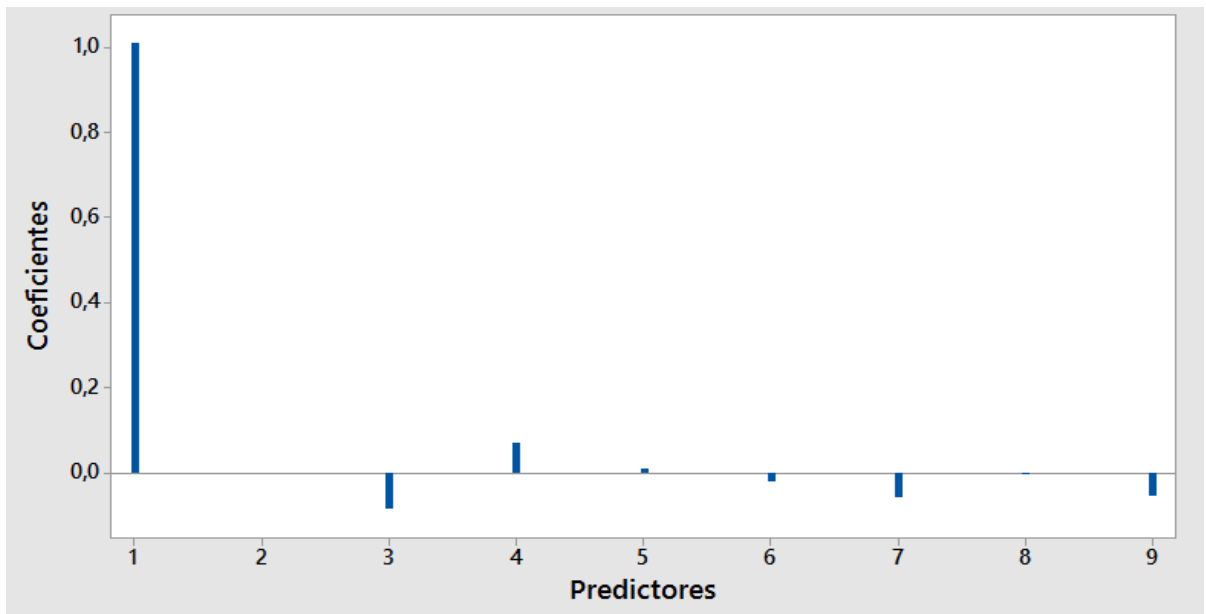


Grafica de respuesta calibración regresión PLS caso de estudio II.

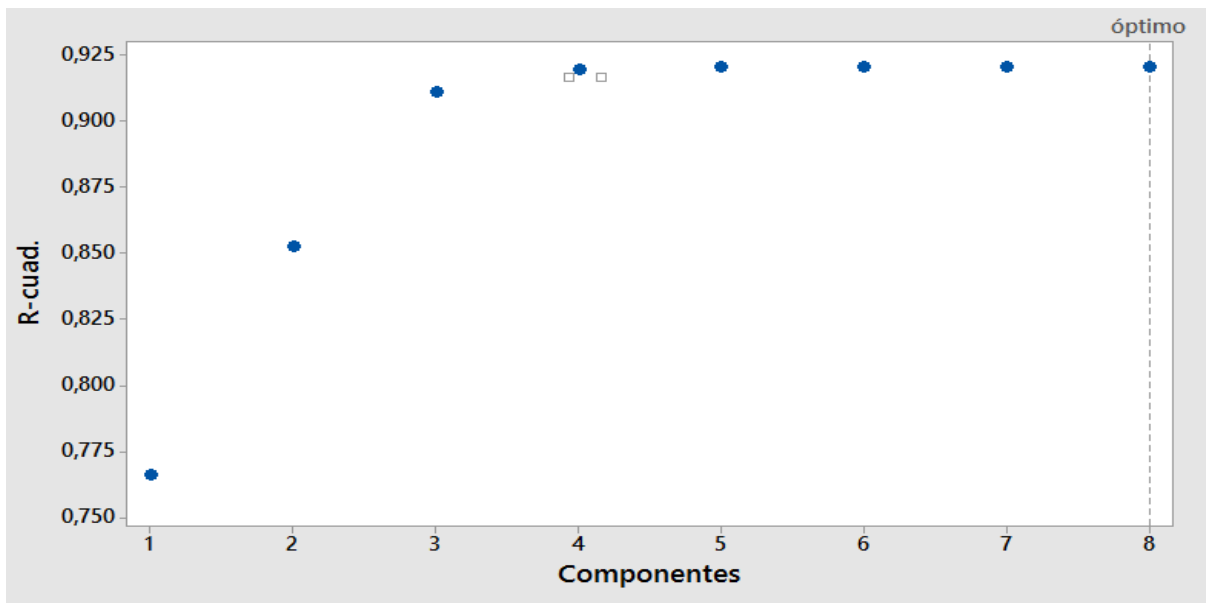


ANEXO J. GRAFICAS DE SELECCIÓN MODELO DE REGRESION, DE RESPUESTA DE REGRESION PLS Y DE COEFICIENTES. CASO DE ESTUDIO III.

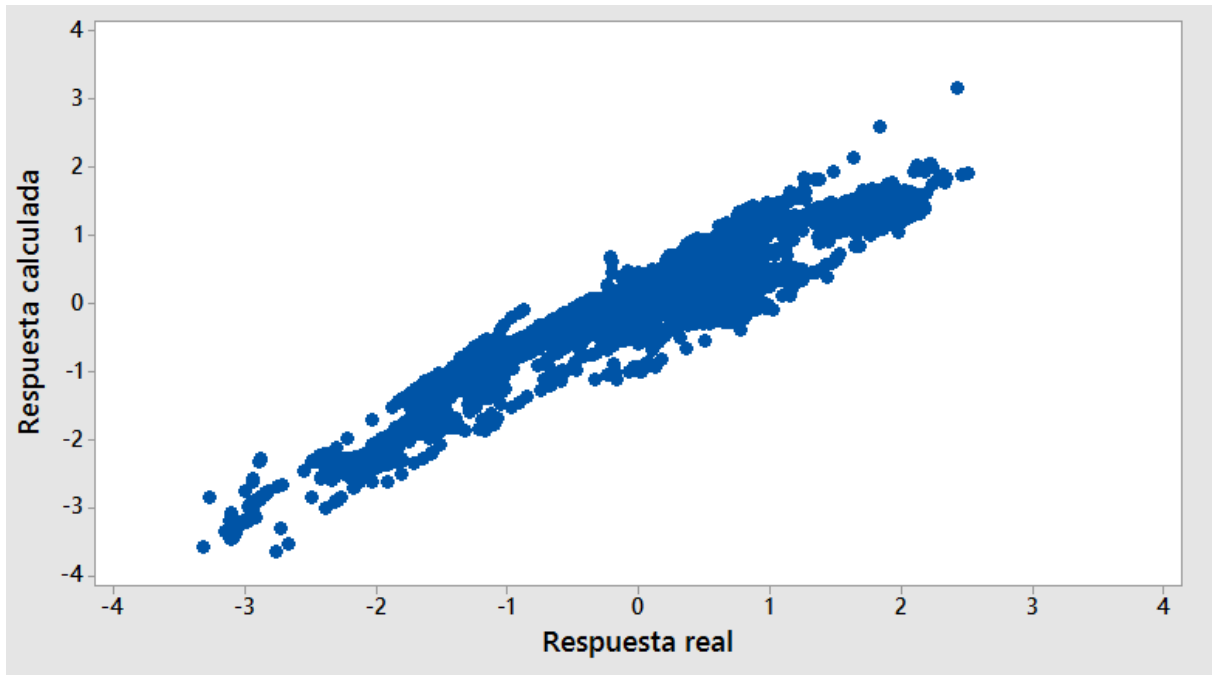
Grafica coeficientes caso de estudio III.



Grafica de selección de modelos regresión PLS caso de estudio III.



Grafica de respuesta planteamiento regresión PLS caso de estudio III.



Grafica de respuesta planteamiento regresión PLS caso de estudio III.

