

**ANÁLISIS DE DATOS DE CONSUMO ENERGÉTICO DE LA UNIDAD DE
REFINACIÓN DE CRUDOS U-200 DEL COMPLEJO INDUSTRIAL
BARRANCABERMEJA: REGRESIÓN LINEAL MÚLTIPLE.**

MANUEL SALVADOR SALAZAR CACHOPO

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

2015

**ANÁLISIS DE DATOS DE CONSUMO ENERGÉTICO DE LA UNIDAD DE
REFINACIÓN DE CRUDOS U-200 DEL COMPLEJO INDUSTRIAL
BARRANCABERMEJA: REGRESIÓN LINEAL MÚLTIPLE.**

MANUEL SALVADOR SALAZAR CACHOPO

**TRABAJO DE GRADO PRESENTADO PARA OPTAR POR EL TÍTULO DE
INGENIERO QUÍMICO**

DIRECTOR:

**GIOVANNI MORALES MEDINA
INGENIERO QUÍMICO, Dr.**

CODIRECTOR:

**RICARDO IVÁN RUBIO LIZCANO
INGENIERO QUÍMICO.**

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

2015

DEDICATORIA

Quiero dedicar esta importante etapa de mi vida, llena de crecimiento personal y profesional a mis padres, Lázaro y Luz Mery, a mi hermanos Jorge Luis y Juan Pablo; por ser ese apoyo, ese motor incondicional y esa fuente de motivación de cada mañana, cada día y cada noche, y en cada peldaño que superé para llegar a ser quién soy, un excelente Ingeniero Químico.

A mi familia en general, aquellos amigos y profesores que siempre confiaron en mis capacidades desde el colegio, de la Universidad y de ECOPETROL durante mis prácticas industriales

AGRADECIMIENTOS

A Dios por este gran logro, por permitirme vivir la vida y vivir cada experiencia en la Universidad Industrial de Santander (UIS).

A mis padres que con mucha paciencia y mucho amor han hecho de mis hermanos y de mí hombres de valor en Cristo Jesús.

A la familia Giorgi Ortíz por su apoyo y colaboración incondicional en esta última etapa académica de mi vida.

A los compañeros que dejé y los que continuaron conmigo en cada semestre de mi carrera que con su presencia fueron aportando consejos y palabras que quedarán siempre guardados en mi mente y corazón.

A Pato, Al Saya, A Moni y A Clau, que a pesar de los momentos amargos siempre predominó la madurez, el respeto y la fraternidad a la hora de trabajar y contar con el apoyo mutuo.

Al profesor Giovanni por su apoyo, paciencia y compañía en el desarrollo de esta tesis. Déjeme decirle: Es Usted una calidad de persona.

A la Ingeniera Lilia que desde el primer día como practicante en ECOPETROL me enseñó que la calidad humana y la humildad no dependen de la experiencia profesional, ni los títulos, ni los diplomas... todo está en el corazón de cada persona. Gracias por cada enseñanza Lili... Todos mis respetos siempre estarán hacia tí.

A Mauro, el popular millos, es Ud. Una excelente persona, gracias por cada consejo mi compadre.

Al Ingeniero Ricardo, por enseñarme que conocer bien a una persona es importante para desenvolverse bien en un equipo de trabajo.

Manuel Salvador Salazar Cachopo

CONTENIDO

	Pág.
INTRODUCCIÓN	15
1. MARCO TEÓRICO	18
1.1 DESCRIPCIÓN GENERAL DEL PROCESO DE LA U-200	18
1.2 ANÁLISIS EXPLORATORIO DE DATOS	20
1.2.1 Análisis estadístico unidimensional.	21
1.2.2 Análisis estadístico bidimensional.....	22
1.3 ANÁLISIS ESTADÍSTICO MULTIVARIABLE.....	23
2. METODOLOGÍA	26
2.1 ETAPA I: RECOLECCIÓN, PROCESAMIENTO Y ANÁLISIS DE LA INFORMACIÓN	27
2.2 ETAPA II: APLICACIÓN DE TÉCNICA MULTIVARIABLE.....	30

3.	ANÁLISIS Y RESULTADOS	32
3.1	ANÁLISIS DE DATOS DE CONSUMO ENERGÉTICO	32
3.2	APLICACIÓN DE LA REGRESIÓN LINEAL MÚLTIPLE (RLM)	36
4.	CONCLUSIONES	48
5.	RECOMENDACIONES	49
6.	BIBLIOGRAFÍA	52
7.	ANEXOS	55

LISTA DE TABLAS

	Pág.
Tabla 1. Variables potenciales de mayor influencia PIVs. Fuente: ECOPETROL .	33
Tabla 2. Resumen estadístico del modelo 1 con variables autoescaladas.	38
Tabla 3. Resumen estadístico modelo 1 con variables normalizadas.....	39
Tabla 4. Resumen estadístico modelo 2 autoescalado.....	40
Tabla 5. Resumen estadístico modelo 2 con variables normalizadas.....	40
Tabla 6. Resumen estadístico para modelo 3 con variables autoescaladas.....	41
Tabla 7. Resumen estadístico para modelo 3 con variables normalizadas.	42
Tabla 8. Resumen de parámetros estadísticos descriptivos y de valoración de modelos 1y 3.....	42
Tabla 9. Modelo 1 reducido con variables autoescaladas.	43
Tabla 10. Modelo 1 reducido con variables normalizadas.	44
Tabla 11. Datos atípicos que influyen en el ajuste del seleccionado.	45
Tabla 12. Estadísticos H-201	59
Tabla 13. Estadísticos H-202.....	60
Tabla 14. Estadísticos U-200.....	60
Tabla 15. Estadísticos T-204	61
Tabla 16. Estadísticos T-201	61
Tabla 17. Resumen estadístico de modelo reducido seleccionado.	75
Tabla 18. Comparación de modelos para determinar desempeño de predicción. .	75

LISTA DE FIGURAS

	Pág.
Figura 1. Diagrama de flujo general del proceso en la U-200	19
Figura 2. Identificación de un posible datos atípicos univariado por medio del gráfico boxplot.....	21
Figura 3. Ejemplo matriz de dispersión-correlación. Tomado Hair J. et al séptima edición.	23
Figura 4. Clasificación de las técnicas multivariantes de acuerdo al objetivo de estudio. Tomado y modificado de De la Garza J. et al, primera edición.	24
Figura 5. Metodología seguida en el presente trabajo.....	26
Figura 6. Tendencia de consumo de energía eléctrica de la U-200.....	27
Figura 7. Tendencia de consumo de energía en forma de gas combustible de la U-200.....	28
Figura 8. Tendencia de consumo de energía en forma de vapor de agua.....	28
Figura 9. Estado cuasi-estacionario utilizando variables de cargas a las unidades mostradas.	29
Figura 10. Matriz de dispersión y coeficiente de correlación para variables de proceso seleccionadas para el estudio estadístico bivariado.	36
Figura 11. Diagrama Boxplot para variables de proceso utilizadas para el análisis univariado de la U-200.....	59
Figura 12. Distribución normal de los residuos del modelo 1 normalizado.	62
Figura 13. Dispersión de los residuos con respecto a los valores predichos por el modelo.	63
Figura 14. Prueba de linealidad del modelo 1 normalizado.	64
Figura 15. Prueba de normalidad para modelo 1 autoescalado.	65
Figura 16. Prueba de varianza constante de residuos del modelo 1 autoescalado.	66
Figura 17. Prueba de linealidad de variables dependiente e independiente.....	66

Figura 18. Prueba de normalidad modelo 3 normalizado.	67
Figura 19. Prueba de varianza constante de los residuos modelo 3 normalizado.	68
Figura 20. Prueba de linealidad de variables dependiente e independiente.	69
Figura 21. Prueba de normalidad modelo 3 autoescalado.	70
Figura 22. Prueba de varianza constante modelo 3 autoescalado.	71
Figura 23. Prueba de linealidad de variables dependiente e independiente.	72
Figura 24. Diagrama de distancias Cook de puntos atípicos del modelo seleccionado (modelo 1 autoescalado).	73
Figura 25. Diagrama de puntos influyentes del modelo seleccionado (modelo 1 autoescalado).	74
Figura 26. Prueba de normalidad modelo final seleccionado.	76
Figura 27. Prueba de varianza constante de residuos del modelo final seleccionado.	77
Figura 28. Prueba de linealidad de variables dependiente e independiente.	78

LISTA DE ANEXOS

	Pág
Anexo A: CÓDIGO BASE UTILIZADO EN RI386 PARA EL MODELADO ESTADÍSTICO.....	55
Anexo B: METODOLOGÍA PARA EL CÁLCULO DEL CONSUMO DE ENERGÍA EN TÉRMINOS DE VAPOR, ELECTRICIDAD Y GAS COMBUSTIBLE.....	57
Anexo C: ESTADÍSTICA UNIVARIADA SOBRE VARIABLES DE PROCESO DE LA U-200.....	59
Anexo D: VERIFICACIÓN DE SUPUESTOS ESTADÍSTICOS PARA MODELOS CON MEJOR DESEMPEÑO.....	62
Anexo E: GRÁFICOS ESTADÍSTICOS DE PUNTOS ATÍPICOS.....	73
Anexo F: ESTADÍSTICO DEL MODELO SELECCIONADO REDUCIDO MEDIANTE LA FILTRACIÓN DE DATOS ALTAMENTE INFLUYENTES.	75
Anexo G: VERIFICACIÓN DE SUPUESTOS ESTADÍSTICOS MODELO REDUCIDO FINAL.....	76

RESUMEN

TÍTULO: ANÁLISIS DE DATOS DE CONSUMO ENERGÉTICO EN LA UNIDAD DE REFINACIÓN DE CRUDOS U-200 DEL COMPLEJO INDUSTRIAL BARRANCABERMEJA: REGRESIÓN LINEAL MÚLTIPLE*.

AUTOR: MANUEL SALVADOR SALAZAR CACHOPO**

PALABRAS CLAVES: Unidad de refinación de crudos, variables potenciales de mayor influencia (PIVs), Análisis multivariable, Regresión lineal múltiple.

CONTENIDO: El análisis estadístico multivariado de datos se ha constituido en una poderosa herramienta para predecir el consumo de energía de las industrias petroquímicas. Identificar variables operativas que influyen potencialmente en el consumo de energía es el propósito fundamental del cual se centra el análisis estadístico de datos para predecir dicho consumo. El presente trabajo provee una metodología para la aplicación de una regresión lineal múltiple sobre un conjunto de datos históricos de consumo energético obtenidos de la base de datos de ECOPETROL SA con el objetivo de sugerir bajo tres escenarios diferentes un modelo que pueda predecir con el menor error posible el consumo de energía. El resultado de este tratamiento reportó, a un nivel de confianza del 95%, un modelo generalizado y reducido con muy buena capacidad predictiva ($R^2 = 0.85$ para el modelo generalizado y $R^2 = 0.82$ para el modelo reducido). Asimismo, se concluyó que las regresiones obtenidas sugieren que la operación de los hornos en la unidad de refinación, en general, se establece sin considerar los valores del poder calorífico del gas combustible. Por lo anterior, un aumento en el poder calorífico del gas combustible reporta un mayor consumo energético de la unidad. Esto se contrasta a nivel de proceso puesto que en la planta un mayor poder calorífico garantizaría menor consumo de gas combustible, por ende, menor consumo de energía de la U- 200.

* Trabajo de grado

** Facultad de Ingenierías Físicoquímica. Escuela de Ingeniería Química. Director: Giovanni Morales

ABSTRACT

TITLE: DATA ANALYSIS OF ENERGY CONSUMPTION FOR THE CRUDE DISTILLATION UNIT U-200 OF REFINERY OF BARRANCABERMEJA: APPLICATION OF MULTIPLE LINEAR REGRESSION*.

AUTHOR: Manuel Salvador Salazar Cachopo**

CONTENT: Multivariate data analysis is a useful tool to assess energy consumption in the petroleum industry. Identification of potential influence variables (PIVs) is the main concern in data analysis to predict the energy consumption. In this document, a methodology is proposed to analyse the energy requirements for the operation of the crude distillation unit U-200 of ECOPETROL's refinery located in Barrancabermeja. The methodology considered three models to adjust the energy consumption required during steady state operation of U-200. The model of best performance considered 14 independent variables and explained 85% data variance with a statistical significance of 5%. This model was reduced to consider 8 independent variables with the loss of only 3% explained variance. According to the regression models obtained, the energy consumption in U-200 increases with the increasing in the calorific value of the fuel gas. This suggests that operation in U-200 can be carried out with less energy consumption by updating the calorific value before manipulating the fuel gas. In terms of process, when the calorific value go up, the fuel gas consumption should decrease for maintaining the levels of energy consumption during the operation. The models calculated herein can be utilized as a baseline useful for energy consumption analysis.

* Thesis

** Faculty of Phisico-Chemical Engineering. School of Chemical Engineering. Director: Giovanni Morales Medina

INTRODUCCIÓN

Las proyecciones mundiales establecen un crecimiento en el consumo de la energía a una tasa promedio anual del 1.5% en los próximos 25 años [1]; se reporta que el sector industrial fisicoquímico contribuyó con el 28 % del consumo mundial de energía en el año 2012 [2]. Esta contribución se debe principalmente a operaciones relacionadas con el procesamiento de materias primas, la purificación de productos finales, los espacios acondicionados y la iluminación. El consumo energético en la industria fisicoquímica también incluye los procesos de tratamiento del gas natural y del crudo del petróleo [3].

Un enorme potencial de ahorro energético presentan las refinerías en su operación diaria; este potencial puede ser aprovechado para contribuir con el incremento en el margen de ganancia y en la reducción en el impacto ambiental. Por ejemplo, en Colombia, el consumo energético del sector petrolero representó el 10,9% del consumo final de energía del país en el año 2009 [4]. La implementación de un sistema de gestión energética permite minimizar el costo asociado al uso de la energía y el impacto ambiental relacionado con las actividades industriales sin afectar la calidad y la tasa de producción [5]; lo anterior a través de un ciclo de mejora continua del desempeño, de la eficiencia y la conservación energética [2]. La determinación de indicadores de desempeño es indispensable en la evaluación de la eficiencia a través de un sistema de gestión con el objetivo de diagnosticar, plantear una política, definir metas y monitorear el cumplimiento de los objetivos energéticos establecidos [6].

En la refinería de Barrancabermeja (GRB), la U-200 es la unidad de destilación de crudos que cuenta con la mayor capacidad (hasta 83 KBPD) y por lo tanto la que presenta los mayores consumos energéticos (gas combustible, vapor y energía

eléctrica) principalmente en los hornos y en las torres de destilación. Según los indicadores de desempeño energético obtenidos por la coordinación de energía y pérdidas de la GRB, pueden existir escenarios operativos para la U-200 que conlleven a reducciones en consumos de energía de hasta un 3% (de un total de ca. 3000 MMBTU/año-3 millones de MBTU/año-), para un potencial de disminución de gasto de 283 \$MUSD anuales-283 000 \$USD/año. Asimismo, la reducción en el consumo energético en la U-200 también se vería reflejada en la reducción de 1.43 ton/año de CO₂ [7].

Por otra parte, los datos de procesos recolectados por los instrumentos de medición en la industria pueden ser utilizados para analizar el comportamiento de los procesos ante situaciones operacionales diferentes. Con esta información y por medio del análisis estadístico es posible visualizar tendencias y patrones operacionales relevantes para el análisis del desempeño, la interrelación de variables, el diagnóstico de fallas y las oportunidades factibles de mejora y optimización en los procesos respectivos. Asimismo, el análisis estadístico de la información de procesos permite estimar el consumo energético en las refinerías ante algún cambio en las condiciones operacionales [8, 9]. Un ejemplo de regresión lo presentan Velázquez *et al.*, quienes desarrollaron un modelo basado en datos para evaluar el desempeño energético de una planta de reformado de nafta a partir de la identificación de variables claves de mayor influencia [10]. Asimismo, se tiene un precedente de la obtención de la línea base energética para la U-250 de la refinería de ECOPETROL [11]. Puerto & Rodríguez, analizaron un número reducido de datos operativos de la U-250 (filtrados) y propusieron dos regresiones simples para las respectivas secciones de destilación con un coeficiente de determinación menor al 90%; las regresiones fueron orientadas en la dependencia del consumo energético con la carga a la unidad [11].

El presente trabajo utiliza métodos estadísticos univariados, bivariados y de regresión lineal múltiple para el análisis y el desarrollo de un modelo de consumo energético basado en datos históricos de operación de la unidad de refinación de crudos U-200 de la GRB. Con este modelo se puede facilitar la planeación y la programación de la producción de la U-200 en cuanto a la selección de variables operacionales que conduzcan a una menor demanda de energía [12,13].

1. MARCO TEÓRICO

1.1 DESCRIPCIÓN GENERAL DEL PROCESO DE LA U-200

En las unidades de refinación de crudos, el petróleo transportado desde los campos es separado en fracciones, cortes o productos que poseen una mayor calidad y rentabilidad comercial que la materia prima inicial [14]. En la refinería de Barrancabermeja (GRB), el departamento de Refinación está conformado por 5 unidades de destilación de crudos que procesan alrededor de 240000 barriles por día (BPD). De estas unidades, la U-200 es la unidad que cuenta con la mayor capacidad de procesamiento de crudo. La U-200 se encuentra dividida en dos secciones: una sección atmosférica y una zona de viscorreducción en modo operación de crudos. En la Figura 1 se presenta el diagrama de flujo general del proceso de refinación de la U-200.

El procesamiento del crudo en la U-200 comienza cuando el crudo es impulsado por unas bombas hacia un primer tren de intercambiadores de calor, los cuales suministran energía para aumentar la temperatura de la carga antes del desalador (D-222). En este equipo, el crudo es lavado con agua y productos químicos con el propósito de remover las impurezas que pueden causar daños, fallas y afectaciones, en equipos posteriores. Después, el crudo pasa por un segundo tren de intercambiadores de calor y es conducido al horno atmosférico (H-201). En este horno, el crudo desalado es calentado hasta su temperatura de dos fases para efectuar su separación en la torre atmosférica (T-201).

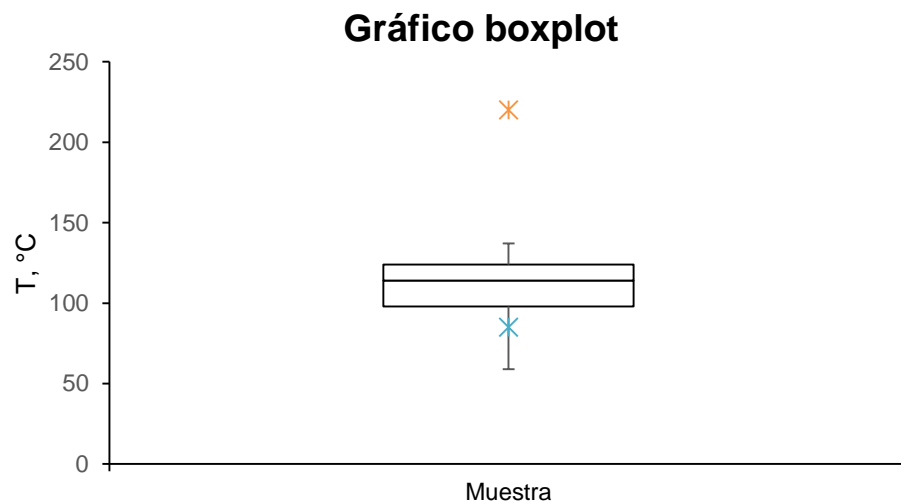
La U-200 fue construida inicialmente como una unidad de viscorreducción para procesar fondos de vacío. En los últimos años ha estado operando como una unidad de refinación, es por ello que presenta una disposición de equipos totalmente diferente a una planta de refinación común. El procesamiento del crudo en la zona viscorreductora inicia con la llegada del crudo reducido de la torre T-201 a un tambor de carga. Luego, este crudo reducido es transportado hacia los serpentines del horno H-202, donde se mezcla con crudo proveniente del segundo tren de intercambiadores de calor, antes de entrar a la torre fraccionadora T-204. En modo de operación denominado crudo, la torre T-204 trabaja como una torre destiladora normal, donde se separan: gases, nafta, jet, ACPM y crudo reducido que pasa por vasos comunicantes. Este último sale por el fondo de la torre y es enviado a la zona flash de la torre de vacío (T-205). La parte vaporizada asciende por la torre y a medida que se va enfriando produce corrientes laterales líquidas que se recuperan en los colectores de salida de la torre, la parte líquida va a los platos del fondo [15].

1.2 ANÁLISIS EXPLORATORIO DE DATOS

El análisis exploratorio de datos (AED) corresponde al uso de procedimientos de estadística descriptiva que permiten examinar un conjunto de datos antes de aplicarle cualquier técnica de estadística inferencial. De esta manera, el analista consigue un entendimiento básico de la información disponible y de las relaciones existentes entre las variables analizadas. Entre los procedimientos del AED se incluyen: definición del conjunto de datos, análisis estadístico unidimensional y análisis estadístico bidimensional [16].

1.2.1 Análisis estadístico unidimensional. Definido el conjunto de datos, un análisis unidimensional es efectuado con el fin de obtener una descripción inicial de la información contenida; con esto es posible definir los valores mínimos, máximos y promedios, los cuartiles, la desviación estándar y los gráficos de caja y bigotes (*boxplot*) para cada variable del conjunto [16]. En esta etapa se define el número de datos faltantes y los posibles datos atípicos desde una mirada univariada. Los valores atípicos, que pueden corresponder a posibles errores de medición o diferentes tendencias de consumo (diferentes escenarios), se identifican mediante el gráfico *boxplot* definiendo un límite superior como $Q_3+1.5*(Q_3-Q_1)$ y un límite inferior como $Q_1-1.5*(Q_3-Q_1)$, donde Q_i corresponde al cuartil i [12,17]; los puntos atípicos se encontrarán fuera de los límites definidos. La Figura 2 ilustra un gráfico *boxplot* que presenta unos posibles datos atípicos.

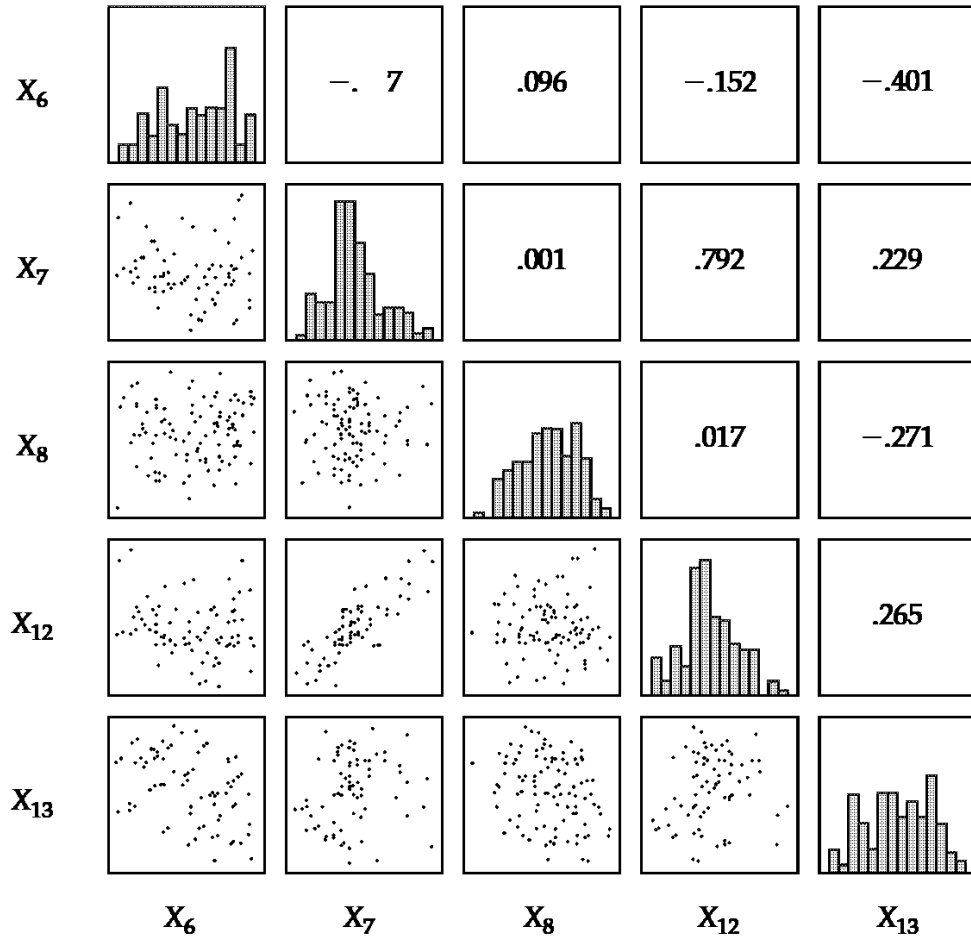
Figura 2. Identificación de posibles datos atípicos univariado por medio del gráfico *boxplot*.



1.2.2 Análisis estadístico bidimensional. La distribución conjunta entre pares de variables puede representarse gráficamente mediante un gráfico de dispersión. Con este gráfico es posible visualizar una posible relación lineal entre las variables asociadas. La relación entre las variables también puede expresarse de forma cuantitativa mediante el coeficiente de correlación lineal de Spearman (r). El valor de este coeficiente r oscila entre -1 y 1. Un valor cercano a -1 o 1 implica una correlación fuertemente lineal, mientras que su signo indica si esta relación es inversa o directa, respectivamente. De otro lado, si r es cercano a cero indica una correlación nula entre las respectivas variables [16].

Las distribuciones y los coeficientes de correlación pueden ser visualizados de manera conjunta para los diferentes pares de variables en un problema en estudio, mediante la denominada matriz de dispersión-correlación. En esta matriz los gráficos de dispersión bivariados se ubican debajo de la diagonal principal, mientras que los coeficientes r son presentados sobre la diagonal principal [16], de acuerdo a la Figura 3. Esta matriz puede permitir una mejor visualización de las relaciones lineales y no lineales entre las variables, la distribución de la varianza entre las variables y al mismo tiempo una estimación cuantitativa de estas relaciones.

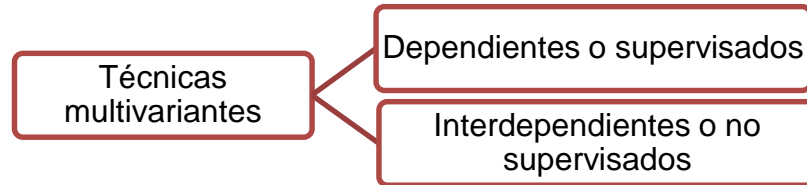
Figura 3. Ejemplo matriz de dispersión-correlación. Tomado Hair J. et al séptima edición.



1.3 ANÁLISIS ESTADÍSTICO MULTIVARIABLE

El análisis multivariado es el conjunto de técnicas estadísticas que analizan simultáneamente múltiples medidas entre numerosas variables sometidas a investigación. Los métodos empleados se clasifican de acuerdo con la finalidad del estudio en dos grandes grupos como lo muestra la Figura 4. En este libro se aplica la técnica dependiente en el análisis energético de la U-200.

Figura 4. Clasificación de las técnicas multivariantes de acuerdo al objetivo de estudio. Tomado y modificado de De la Garza J. et al, primera edición.



Las técnicas dependientes (también denominadas supervisadas o funcionales), implican un modelo, ecuación o función formada por las variables involucradas. En esta técnica intervienen variables dependientes e independientes. El objetivo de las técnicas funcionales es predecir una o más variables dependientes a partir de cambios en las independientes. Entre estas técnicas se encuentra la regresión lineal múltiple establece una relación entre las variables (Ec.1) por medio de unos coeficientes (β) obtenidos utilizando la minimización de la varianza residual resuelta por el algoritmo de los mínimos cuadrados [18].

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n$$

Ec. 1.

La ecuación 1 se puede representar en notación matricial de mínimos cuadrados de la siguiente manera [16]:

Para $n > k$, donde n es el número de observaciones y k es número de variables.

$$Y = X\beta + \varepsilon$$

Con

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} X_{11} & \dots & X_{1k} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{nk} \end{bmatrix}, \beta = \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_k \end{bmatrix}$$

Donde,

Y: es el vector de observaciones.

X: Matriz de nxp de niveles de la variable regresora.

β : es un vector px1 de coeficientes de regresión

ε : es el vector aleatorio error de orden px1

La suma de cuadrados del error es dada por la Ec. 2:

$$SC_E = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon' \varepsilon = (Y - X\beta)'(Y - X\beta)$$

Ec. 2.

La representación matricial de un modelo RLM y la solución de los coeficientes β es:

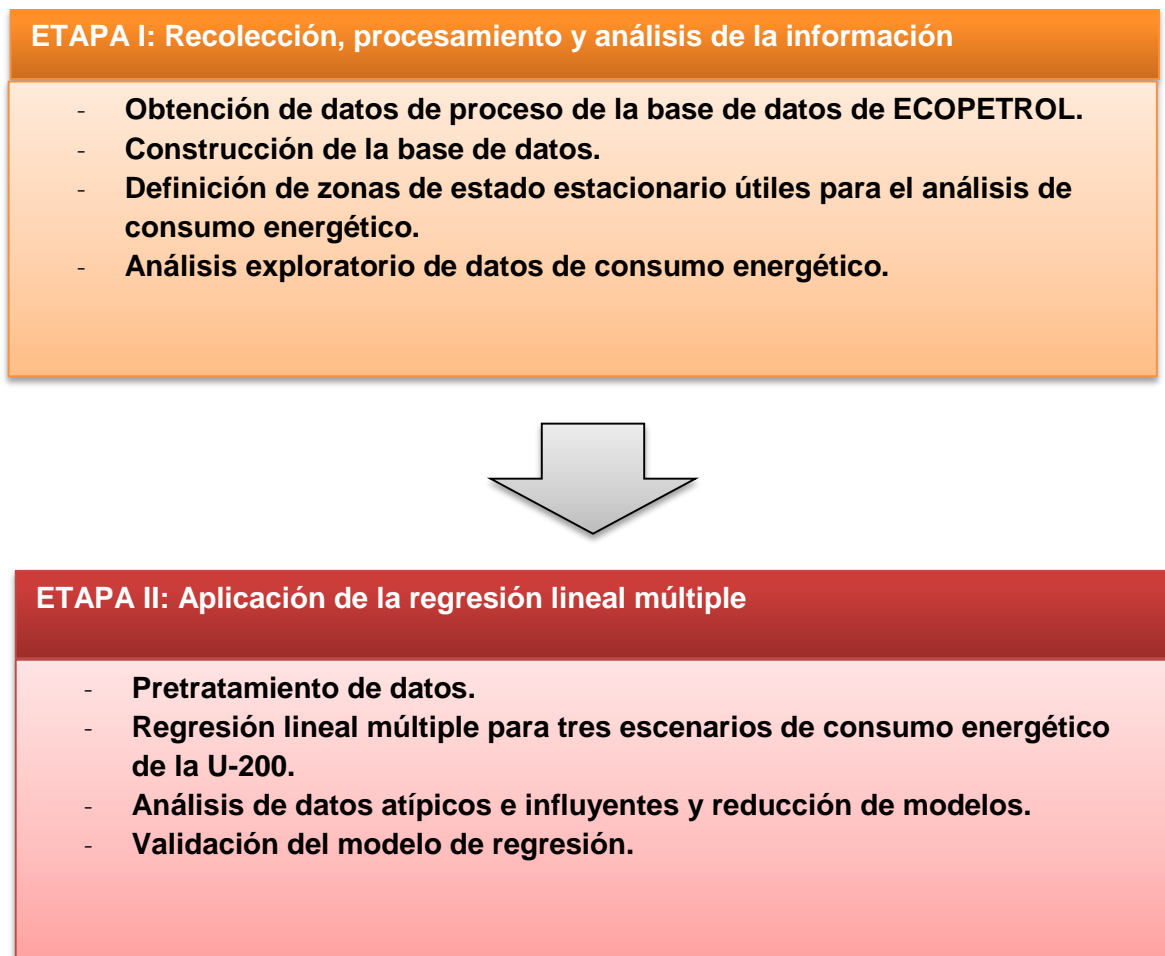
$$X'X\beta = X'Y$$

$$\begin{bmatrix} \sum_{i=1}^n X_{i1} & \cdots & \sum_{i=1}^n X_{ik} X_{i1} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n X_{ik} X_{i1} & \cdots & \sum_{i=1}^n X_{ik}^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n Y_i \\ \vdots \\ \sum_{i=1}^n Y_i X_{ik} \end{bmatrix}$$

2. METODOLOGÍA

El análisis de la información histórica y el desarrollo de un modelo de consumo energético para la unidad de refinación de crudos U-200 involucraron las etapas mostradas en la Figura 5. El paquete computacional libre denominado *R* versión *i386* y el paquete comercial *The Unscrambler* versión *X* fueron utilizados para aplicar los procedimientos estadísticos mencionados en el presente trabajo.

Figura 5. Metodología seguida en el presente trabajo.



2.1 ETAPA I: RECOLECCIÓN, PROCESAMIENTO Y ANÁLISIS DE LA INFORMACIÓN

- **Obtención de datos de proceso de la base de datos de ECOPETROL:** En este paso se realizó la recolección de las mediciones de todas las variables de proceso de la U-200 disponibles en la base de datos de ECOPETROL SA, denominada PI.
- **Construcción de la base de datos:** Se construyó una tabla de datos en Microsoft Excel como soporte de información necesaria para el análisis y aplicación de técnicas estadísticas. La base de datos se obtuvo de un proceso de selección de variables potenciales de mayor influencia (PIVs) en el consumo energético de la U-200. De otro lado, se definieron las tendencias de consumo energético por medio del cálculo de los requerimientos de vapor, electricidad y gas combustible (Ver Figuras 6,7,8).

Figura 6. Tendencia de consumo de energía eléctrica de la U-200.

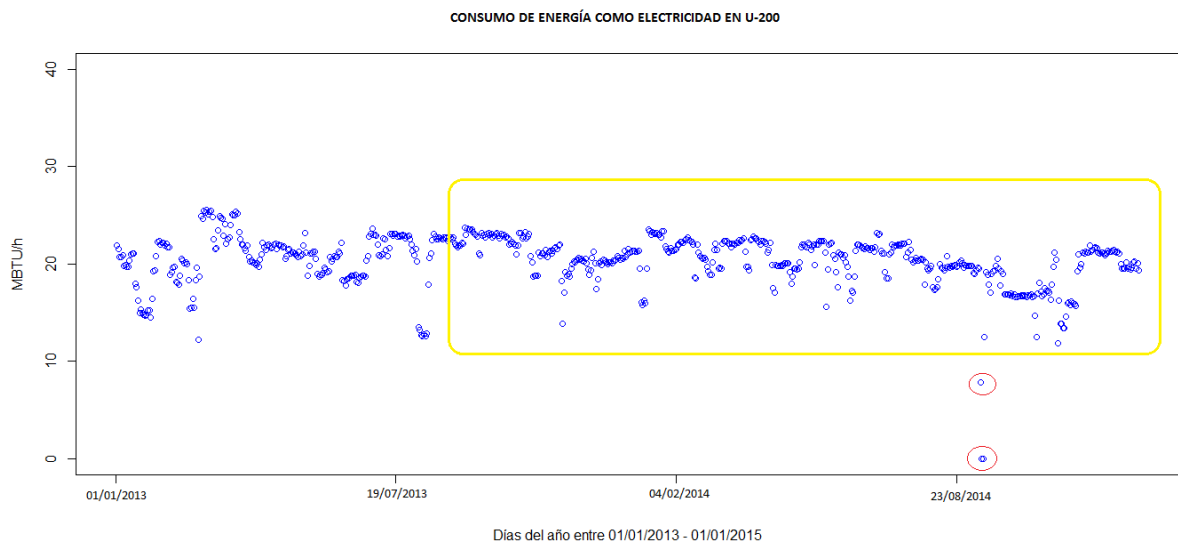


Figura 7. Tendencia de consumo de energía en forma de gas combustible de la U-200.

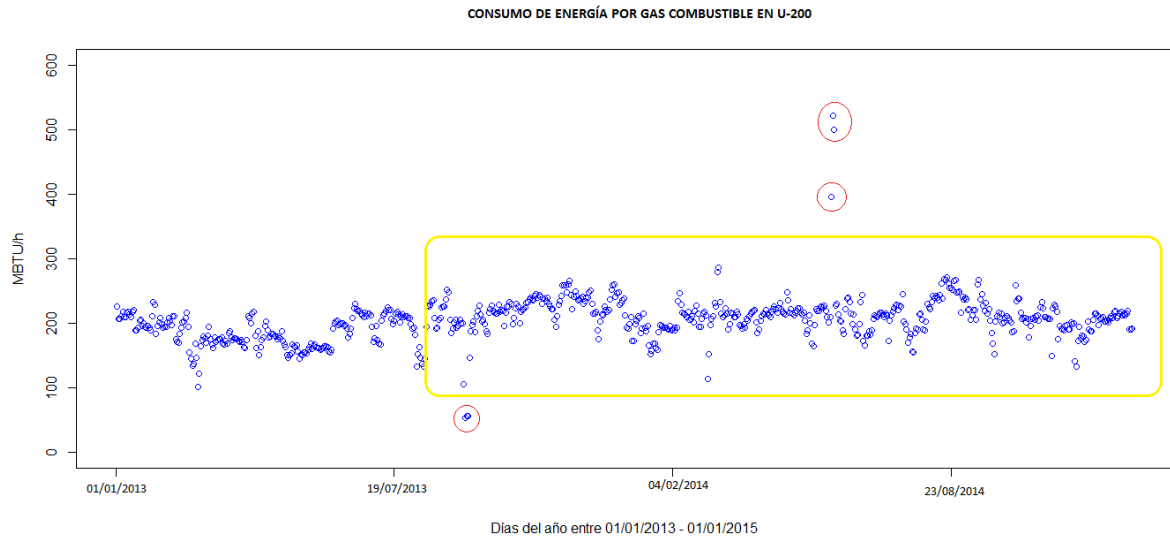
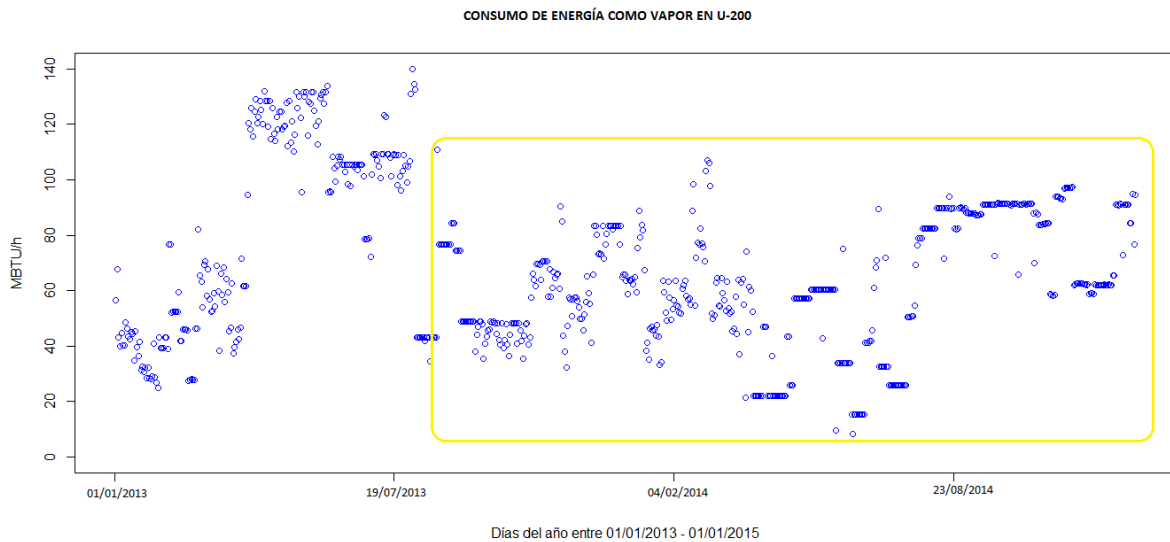
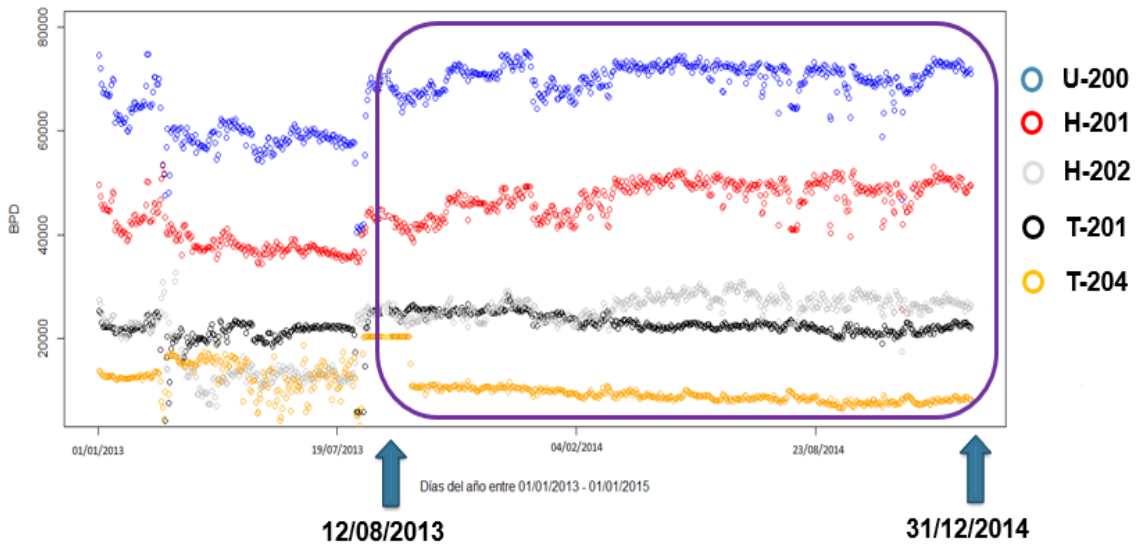


Figura 8. Tendencia de consumo de energía en forma de vapor de agua.



- **Definición de zonas de estado estacionario útiles para el análisis de consumo energético:** La tendencia de los datos de cada variable fue representada gráficamente con el propósito de identificar el intervalo de tiempo en donde el sistema se comportó en forma cuasi-estacionaria.

Figura 9. Estado cuasi-estacionario utilizando variables de cargas a las unidades mostradas.



- **Análisis exploratorio de datos de consumo energético:** Se aplicó un análisis estadístico univariado y bivariado sobre los datos PIVs en la zona cuasi-estacionaria para determinar correlaciones y analizar condiciones de normalidad en las observaciones de cada variable. Asimismo, se emplearon estas herramientas para detectar y analizar datos atípicos que tienden a sesgar la tendencia normal del conjunto de datos.

2.2 ETAPA II: APLICACIÓN DE TÉCNICA MULTIVARIABLE

El método de la regresión lineal múltiple fue aplicado al conjunto de datos tabulado en Excel. Esta etapa se subdivide en dos partes: pretratamiento del conjunto de datos, ajuste y validación de los modelos de regresión.

- **Pretratamiento de datos:** Se procedió a transformar los datos iniciales mediante el procedimiento de normalización utilizando el rango de cada variable y el procedimiento de autoescalado mediante la diferencia con la media y la división con la desviación estándar [19]. Este pretratamiento se realizó con el objetivo de centrar los datos y disminuir posibles reportes erróneos en los resultados estadísticos.

-**Regresión lineal múltiple:** Se aplicó la técnica de análisis de regresión lineal múltiple al conjunto de datos de consumo de energía de la U-200. El ajuste de los modelos fue efectuado sobre un subconjunto de calibración equivalente al 70% de los datos iniciales, seleccionados de forma aleatoria. El 30% de los restantes datos fue utilizado para la validación externa de cada modelo. Asimismo, se calcularon los estadísticos de desempeño para cada etapa: Para calibración $R^2 = 1 - \text{RSS}/\text{SST}$, donde RSS corresponde a la suma de los cuadrados de los residuos de calibración y SST representa la varianza total del conjunto. Para validación $R^2 = 1 - \text{PRESS}/\text{SST}$, donde PRESS es la suma de los cuadrados de los residuos de predicción según el conjunto externo (30%) y SST representa la varianza total del conjunto. Para calibración $RMSE = \sqrt{\text{RSS}/N}$, donde N es el número total de muestras de calibración. Para validación $RMSE = \sqrt{\text{PRESS}/N}$ donde N es el número de muestras de la validación externa.

- **Identificación de datos atípicos e influyentes y reducción de modelos:** Una vez obtenido el modelo generalizado que abarcará las variables independientes de mayor significancia estadística se procedió a analizar y eliminar los valores atípicos no influyentes para determinar su participación en el desempeño de cada regresión. Asimismo, cada modelo generalizado fue reducido al menor número de variables de tal manera de que se explicará hasta un mínimo del 75 % de la varianza total. Lo anterior, como alternativa de uso de la regresión con un número menor de variables.

- **Validación de los modelos de regresión:** Se llevaron a cabo dos procesos para verificar los supuestos en cada ajuste reportado por los modelos de regresión.

- Análisis de supuestos estadísticos: Una vez obtenidos el modelo de regresión, se aplicó el análisis de residuos (error de predicción), respecto a las variables individuales (dependientes e independientes).

- Análisis de validación cruzada (Cross Validation): Se llevó a cabo un diagnóstico final para la aceptación del modelo obtenido que garantizara el grado de generalidad del modelo a través de métodos de validación cruzada.

En el Anexo 1 se presenta el código seguido en el paquete *R* para obtener el modelo ajustado entre las variables de proceso y la variable de consumo de energía.

3. ANÁLISIS Y RESULTADOS

3.1 ANÁLISIS DE DATOS DE CONSUMO ENERGÉTICO

En el presente trabajo se asignó como variable independiente todas las variables de proceso disponibles para la unidad de refinación de crudos U-200. De otro lado, el consumo energético fue asignado como la variable dependiente y fue calculada como el aporte del consumo de gas combustible, el aporte de vapor y el aporte de electricidad (Anexo 2). Según los datos de consumo energético, es posible establecer que el consumo de gas combustible, el de vapor y el de electricidad corresponden a 72%, 21% y 7%, respectivamente, del consumo total de la energía para la operación de la U-200.

El consumo de energía como gas combustible se encuentra asociado a la cantidad de gas utilizado en los hornos (H-201 y H-202). El consumo de gas depende de su poder calorífico, del flujo de entrada de crudo y de la temperatura de salida de crudo de los hornos. En la U-200, el gas utilizado proviene de una mezcla de gas guajira con gases de refinería. Estos últimos, son desvíos de GLP (gas licuado del petróleo) de las unidades de cracking: Modelo IV y Orthoflow. Altos desvíos conllevan a un alto poder calorífico o energía por unidad de volumen de gas (BTU/ pie³) para calentar la carga que ingresa, por lo tanto, se necesitará de menos flujo de gas combustible en los quemadores para calentar la misma cantidad de carga. De otro lado, el caso de consumo de energía por electricidad depende de la confiabilidad de los transformadores que suministran energía a la planta, así como de equipos de bombeo a motor. Por su parte, el consumo de energía como vapor en la U-200 se define en función del consumo en las turbinas de relevo empleadas cuando el equipo de bombeo a motor se encuentra fuera de servicio. Cabe destacar que el consumo de una turbina, en promedio, oscila entre 15 a 24 veces el consumo de energía de una bomba a motor eléctrico.

Posteriormente, se identificaron y recolectaron las mediciones diarias registradas para la U-200 entre las fechas 01/01/2013 hasta el 31/12/2014. Estos datos conformaron una matriz de 4,015 variables y 729 casos o mediciones. Esta matriz fue reducida a 39 variables x 474 mediciones considerando sólo las variables con impacto directo en el consumo energético (Ver Tabla 1); estas variables están asociadas a cargas de crudo a las unidades, propiedades del crudo de carga, poder calorífico del gas combustible, características de los hornos y características de operación de las torres. También, las mediciones consideradas se ubicaron en las fechas comprendidas entre 12/08/2013 hasta 31/12/2014, ya que las variables presentaban pocas fluctuaciones relativas evidenciando un comportamiento cuasi-estacionario del proceso. Debido a enorme cantidad de datos, se tomó únicamente el comportamiento en variables de carga a las unidades como se muestra en el Anexo 3.

Tabla 1. Variables potenciales de mayor influencia PIVs. Fuente: ECOPETROL

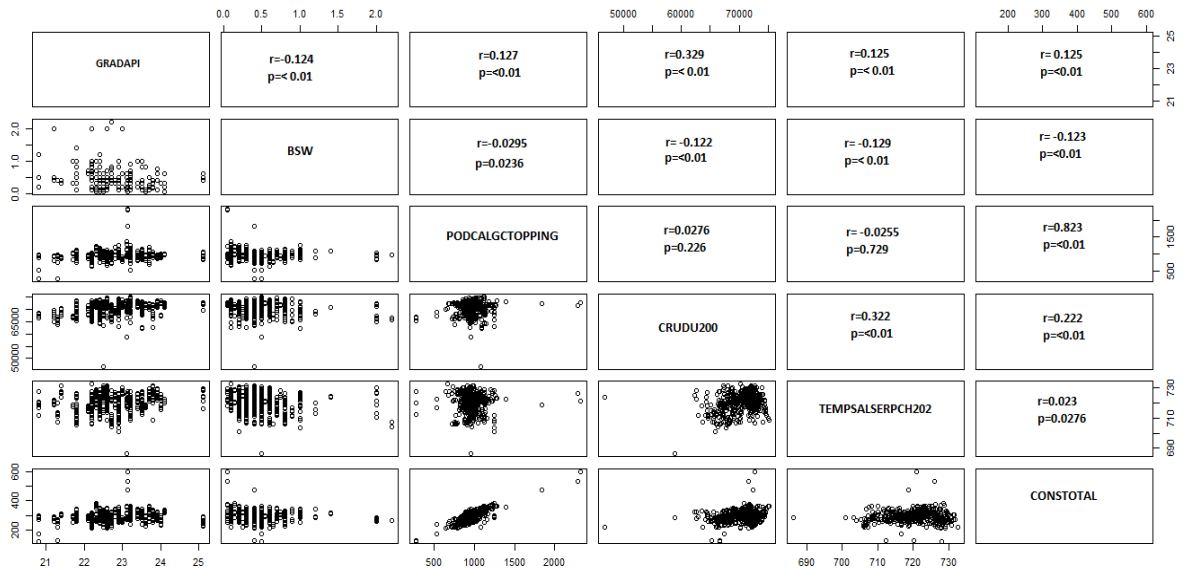
Variable	Etiqueta	Unidad	Descripción
x1	GRADAPI	°API	Gravedad API del crudo
x2	BSW	ml/100ml	Sedimentos y agua en el crudo
x3	AZUFRE	g/100g	Azufre presente en el crudo
x4	PRESTATM	PSI	Presión en la torre atmosférica
x5	PRESZONFLASHT201	PSI	Presión en la zona flash Torre 201
x6	PRESZOFLASHT204	PSI	Presión en la zona flash Torre 204
x7	TBP10	°F	Temperatura verdadera de ebullición al 10%
x8	TBP30	°F	Temperatura verdadera de ebullición al 30%
x9	TBP50	°F	Temperatura verdadera de ebullición al 50%
x10	TBP70	°F	Temperatura verdadera de ebullición al 70%

x11	TBP90	°F	Temperatura verdadera de ebullición al 90%
x12	PODCALGCTOPPING	BTU/PC	Poder calorífico de gas combustible a Topping
x13	CRUDU200	BPD	Carga de crudo a la U-200
x14	CRUDAH201	BPD	Crudo a la Horno 201
x15	CRURET201AT205	BPD	Crudo reducido de T-201 a T-205
x16	CRUDRET201	BPD	Crudo reducido de T-201
x17	CRUDMEZH202	BPD	Crudo mezclado al Horno 202
X18	SALFONT204	BPD	Salida fondo de la Torre 204
X19	GASPEST204	BPD	Gasóleo pesado de la Torre 204
x20	REFCIMT201	BPD	Reflujo de cima de la Torre 201
x21	KEROP238/239	BPD	Queroseno producto de la Torre 201
x22	GAST202C	BPD	Gasóleo de Torre despojadora T-202C
x23	ACPMPROD	BPD	ACPM producto de la U-200
x24	REFINFT201	BPD	Reflujo inferior de la Torre 201
x25	GASOTRAT	BPD	Gasolina a tratamiento
x26	NAFTOTU200	BPD	Nafta total producida en la U-200
x27	ENTSERPAH202	°F	Temperatura de entrada serpentín A al H-202
X28	TEMPSALSERPAH202	°F	Temperatura de salida serpentín A al H-202
X29	ENTSERPBH202	°F	Temperatura de entrada serpentín B al H-202
x30	TEMPSALSERPBH202	°F	Temperatura de salida serpentín B al H-202
x31	ENTCARGSERPCH202	°F	Temperatura de entrada serpentín C al H-202
x32	TEMPSALSERPCH202	°F	Temperatura de salida serpentín C al H-202
x33	ENTCARGSERPDH202	°F	Temperatura de entrada serpentín D al H-202
x34	TEMPSALSERPDH202	°F	Temperatura de salida serpentín D al H-202
x35	TEMPSALCRUDH201	°F	Temperatura de salida serpentín A al H-201
x36	TEMPENTCRUDH201	°F	Temperatura de entrada serpentín A al H-201
x37	EFOPERH201	%	Eficiencia operativa del Horno 201
X38	EFOPERH202	%	Eficiencia operativa del Horno 202

Según los diagramas *boxplot* las variables de proceso que conforman la base de datos presentan posibles valores atípicos. En el Anexo 3 se presentan los diagramas de *boxplot* para las variables $x_{13}, x_{14}, x_{16}, x_{17}$ & x_{18} de proceso. La presencia de estos puntos atípicos se presentan con círculos de color rojo y están asociados a paradas de planta, arrancadas de la unidad y problemas de baja carga y confiabilidad de equipos; también, pueden estar asociadas con errores en las lecturas en los instrumentos de control, errores en los reportes de los operadores y errores de código del sistema de registro de información de ECOPETROL.

De otro lado, se eligieron dos variables que representaran las propiedades del crudo (x_1, x_2), una asociada a la carga a la unidad (x_{13}), el poder calorífico (x_{12}) y la temperatura de salida del crudo de uno de los hornos (x_{32}) para aplicar una matriz de correlación con respecto al consumo total de energía, pues representan de manera global el tipo de información de proceso que fue posible obtener de la base de datos. Se puede observar que el poder calorífico experimenta una fuerte correlación con un r igual a 0.82, en contraste al resto de grupo, las cuáles presentan una correlación cuasi-nula (Ver Figura 10).

Figura 10. Matriz de dispersión y coeficiente de correlación para variables de proceso seleccionadas para el estudio estadístico bivariado.



3.2 APLICACIÓN DE LA REGRESIÓN LINEAL MÚLTIPLE (RLM)

El autoescalado y la normalización fueron aplicadas para resolver los inconvenientes de conflicto entre las diferentes unidades de las variables de la base de datos. El autoescalado fue aplicado según,

$$x_i^* = \frac{x_i - \bar{x}}{\delta}$$

Ec. 3

La Ec. 3 representa matemáticamente la estandarización de una variable i , un donde x_i se le resta el valor promedio de sus de datos y se divide entre se desviación estándar. Los modelos de regresión fueron ajustados considerando los datos con pretratamiento.

Tres modelos fueron propuestos para representar los consumos energéticos de la U-200. Los modelos corresponden a:

- Modelo 1: regresión del consumo total de energía para la U-200.
- Modelo 2: regresión del consumo total por unidad de poder calorífico.
- Modelo 3: regresión del consumo de gas combustible.

Los resultados del ajuste de cada modelo por RLM son descritos a continuación.

Modelo 1: En las Tablas 2 y 3 se presenta el resumen estadístico mostrado en R para las autoescaladas y las normalizadas. La regresión utilizando las variables autoescaladas y normalizadas reportaron un desempeño similar ($R^2=0.85$), sin embargo, teniendo en cuenta que la regresión con variables normalizadas incluye el término independiente (valor asociado al intercepto) el modelo se descarta desde el punto de vista real del proceso de la U-200, pues esto significaría que la planta consume energía cuando está apagada (variables independientes igual a 0). De lo contrario, si el modelo se ajusta sin incluir el término independiente la varianza explicada disminuye a 84%, reportando un desempeño ligeramente menor respecto al modelo con variables autoescaladas.

Tabla 2. Resumen estadístico del modelo 1 con variables autoescaladas.

Variable	Coefficiente	% Aporte	Intervalo de confianza	Error estándar	P-valor
x5	-0.215	6%	[-0.283, -0.147]	0.0346	1.67E-09
x6	0.085	2%	[0.028, 0.142]	0.0290	0.00369
x12	0.770	20%	[0.725, 0.817]	0.0235	< 2e-16
x14	0.692	18%	[0.522, 0.863]	0.0870	3.31E-14
x15	-0.0658	2%	[-0.121, -0.011]	0.0279	0.019146
x16	-0.207	6%	[-0.324, -0.094]	0.0585	0.000421
x17	0.541	14%	[0.436, 0.647]	0.0539	< 2e-16
x18	-0.114	3%	[-0.176, -0.052]	0.0315	0.000341
x19	-0.113	3%	[-0.184, -0.043]	0.0360	0.001845
x22	0.0597	2%	[0.010, 0.110]	0.0254	0.019249
x23	-0.188	5%	[-0.284, -0.093]	0.0486	0.000133
x25	-0.125	3%	[-0.204, -0.046]	0.0401	0.001985
x28	0.173	5%	[0.118, 0.229]	0.0285	3.34E-09
x32	-0.161	4%	[-0.227, -0.094]	0.0339	3.42E-06
x33	0.257	7%	[0.162, 0.353]	0.0485	2.07E-07

La Ecuación 3 representa el modelo 1 generalizado con variables auto escaladas. El intervalo de confianza para cada coeficiente se muestra en la Tabla 2.

$$\begin{aligned}
 Y_{total} = & -0.215 * x_5 + 0.085 * x_6 + 0.770 * x_{12} + 0.692 * x_{14} - 0.0658 * x_{15} - 0.207 * x_{16} \\
 & + 0.541 * x_{17} - 0.114 * x_{18} - 0.113 * x_{19} + 0.0597 * x_{22} - 0.188 * x_{23} \\
 & - 0.173 * x_{25} - 0.161 * x_{32} + 0.257 * x_{33}
 \end{aligned}$$

Ec. 2

Tabla 3. Resumen estadístico modelo 1 con variables normalizadas.

Variable	Coeficientes	Intervalo de confianza	Error estándar	P-valor
x5	-0.096	[-0.129,-0.063]	0.0169	3.29E-08
x6	0.0564	[0.027,0.086]	0.0151	0.000221
x12	0.852	[0.800,0.906]	0.027	< 2e-16
x14	0.481	[0.349,0.614]	0.0675	6.96E-12
x15	-0.0317	[-0.053,-0.010]	0.0108	0.003668
x16	-0.126	[-0.192,-0.059]	0.0340	0.000267
x17	0.215	[0.171,0.258]	0.0223	< 2e-16
x18	-0.047	[-0.074,-0.020]	0.0138	0.000755
x19	-0.089	[-0.120,-0.059]	0.0155	1.93E-08
x21	-0.035	[-0.067,-0.003]	0.0164	0.0352
x23	-0.168	[-0.244,-0.094]	0.0383	1.42E-05
x24	-0.061	[-0.087,-0.034]	0.0136	1.27E-05
x25	-0.094	[-0.124,-0.063]	0.0155	4.40E-09
x28	0.096	[0.061,0.131]	0.0177	1.17E-07
x29	-0.768	[-1.380,-0.156]	0.3123	0.014451
x32	-0.088	[-0.117,-0.058]	0.0152	2.21E-08
x33	0.845	[0.240,1.451]	0.3088	0.006545

Modelo 2: los modelos lineales obtenidos no explican de manera significativa el consumo de energía total de la U-200. En la Tabla 4 se evidencia este hecho teniendo en cuenta que la varianza explicada es igual a 1% y que las variables de proceso no cumplen en criterio de significancia de P-valor menor a 0.05. Asimismo, el modelo ajustado con variables normalizadas significativas (P-valor < 0.05) explican con muy bajo desempeño el consumo energético de la U-200 ($R^2 = 0.34$ sin intercepto) Ver Tabla 5.

Tabla 4. Resumen estadístico modelo 2 autoescalado.

Variables	Coefficientes	Error estándar	P-valor
x2	-0.10076	0.05938	0.0907
x28	0.04687	0.05833	0.4223

Tabla 5. Resumen estadístico modelo 2 con variables normalizadas.

Variables	Coefficientes	Error estándar	P-valor
x4	0.1053	0.0427	0.0143
x5	-0.1067	0.0304	0.0005
x10	0.0117	0.0051	0.0209
x12	-0.1509	0.0187	0.0000
x14	0.2686	0.0437	0.0000
x16	-0.0666	0.0241	0.0061
x17	0.0933	0.0147	0.0000
x19	0.0436	0.0113	0.0001
x22	0.0218	0.0087	0.0125
x23	-0.0659	0.0241	0.0066
x25	0.0274	0.0113	0.0160
x28	0.092	0.0119	0.0000
x33	0.0438	0.0138	0.0017
x36	0.0536	0.0132	0.0001

Modelo 3: En las Tablas 6 y 7 se presenta el resumen estadístico para las variables autoescaladas y normalizadas respectivamente. La regresión utilizando las variables autoescaladas y normalizadas reportaron un desempeño de predicción igual al 81%.

Tabla 6. Resumen estadístico para modelo 3 con variables autoescaladas.

Variable	Coficiente	Intervalo de confianza	Error estándar	P-valor
x3	0.08716	[0.034,0.139]	0.02709	0.001429
x4	0.23039	[0.0305, 0.428]	0.10219	0.02485
x5	-0.20932	[-0.393, -0.023]	0.09509	0.028429
x8	-0.10745	[-0.173, -0.041]	0.03414	0.001803
x12	-0.1414	[-0.190, -0.090]	0.02569	7.65E-08
x13	0.60516	[0.536, 0.669]	0.03426	< 2e-16
x17	0.32743	[0.228, 0.424]	0.05036	3.10E-10
X18	-0.33847	[-0.409, -0.265]	0.03711	< 2e-16
x22	0.07195	[0.010, 0.133]	0.03166	0.023704
x24	0.12746	[0.050, 0.203]	0.03948	0.001376
x25	-0.15792	[-0.235, -0.079]	0.0399	9.33E-05
X28	0.11581	[0.056,0.174]	0.03032	0.000161
X29	-0.26763	[-0.361, -0.171]	0.04871	8.06E-08
x30	0.1805	[0.118, 0.241]	0.03177	3.02E-08
x36	-0.24298	[-0.304, -0.180]	0.03185	2.81E-13
x37	0.08456	[0.024, 0.143]	0.03047	0.005856

La Tabla 8 presenta los resultados del desempeño estadístico de la regresión seleccionada para cada modelo ajustado. En esta tabla se muestra también los resultados de los estadísticos referentes a la validación externa (con el 30% de los datos) y la validación cruzada para cada modelo. Según los anteriores resultados, el caso 1 (autoescalado y normalizado) es el que mejor se ajusta al consumo energético de la U-200 pues posee mejor desempeño en la predicción y cumple todos los supuestos estadísticos. La validez de los supuestos de las RLM en todos los modelos ajustados fue verificada¹, según lo reportado en el Anexo 4.

¹ Normalidad de residuos, linealidad de variables dependiente e independiente, independencia de los errores y homocedasticidad de los residuos.

Tabla 7. Resumen estadístico para modelo 3 con variables normalizadas.

Variables	Coeficientes	Intervalo de confianza	Error Estándar	P-valor
x3	0.0586	[0.026,0.091]	0.0166	0.0005
x4	0.187	[0.016,0.358]	0.0873	0.033
x5	-0.1221	[-0.243,-0.001]	0.0618	0.0491
x8	-0.0644	[-0.122,-0.007]	0.0295	0.0299
x12	-0.1878	[-0.263,-0.112]	0.0385	0.00E+00
x13	0.6873	[0.613,0.762]	0.038	0.00E+00
x17	0.1786	[0.127,0.230]	0.0261	0.00E+00
x18	-0.1679	[-0.204,-0.132]	0.0185	0.00E+00
x22	0.0443	[0.007,0.081]	0.0188	0.0191
x24	0.0737	[0.024,0.124]	0.0254	0.004
x25	-0.0929	[-0.139,-0.047]	0.0234	1.00E-04
x28	0.1183	[0.069,0.168]	0.0254	0.0000
x29	-0.1352	[-0.185,-0.086]	0.0253	0.0000
x30	0.1016	[0.068,0.135]	0.0169	0.0000
x35	0.1173	[0.044,0.191]	0.0375	1.90E-03
x36	-0.2248	[-0.284,-0.166]	0.03	0.0000
x37	0.1465	[0.060,0.233]	0.04	0.001

Tabla 8. Resumen de parámetros estadísticos descriptivos y de valoración de modelos 1y 3.

	Modelo 1		Modelo 3	
	Autoescalado	Normalizado	Autoescalado	Normalizado
R² 70%	0.85	0.85	0.81	0.80
R² ajus 70%	0.84	0.83	0.80	0.79
R² 30%	0.83	0.83	0.81	0.70
R² ajus 30%	0.82	0.82	0.79	0.68
R² VC	0.83	0.82	0.72	0.76
RMSE 70%	0.39	0.33	0.44	0.51
RMSE 30%	0.41	0.34	0.68	0.368
RMSE VC	0.40	0.35	0.53	0.541
DW ESTADISTICO	2.037	1.99	2.036	2.068

Para reafirmar el desempeño del modelo 1 con variables autoescaladas se procedió a calcular el coeficiente de determinación múltiple con un número de variables reducido. Una vez reducido los dos modelos del caso 1, se apreció que el modelo con variables autoescaladas pudo predecir el consumo energético con un R^2 equivalente al 76% con solo 5 variables, mientras que el modelo reducido de variables normalizadas explicó igual valor de varianza para el consumo energético utilizando un número mayor de entradas (9 variables independientes). Según lo anterior, el modelo autoescalado es sugerido para predecir el consumo de energía en la U-200. En las Tablas 9 y 10 se presentan los estadísticos de los modelo 1 reducido para los dos casos.

Tabla 9. Modelo 1 reducido con variables autoescaladas.

	Coefficientes	Error estándar	P-valor	Estadísticos	Valor
x5	-0.30916	0.03229	< 2e-16	R²	0.76
x12	0.82729	0.02712	< 2e-16	R² ajust	0.76
x14	0.51594	0.04997	< 2e-16	P-valor	< 2.2e-16
x16	-0.14378	0.04438	0.00132	RMSE 70%	8.77
x17	0.17716	0.03198	6.25E-08	RMSE 30%	7.56

Tabla 10. Modelo 1 reducido con variables normalizadas.

	Coefficientes	Error estándar	P-valor	Estadísticos	Valor
x5	-0.094	0.0189	0.0000	R²	0.76
x12	0.844	0.0299	0.0000	R² ajust	0.74
x14	0.201	0.0358	0.0000	P-valor	0.0000
x16	-0.081	0.0262	0.0021	RMSE 70%	0.72
x17	0.0302	0.0142	0.0344	RMSE 30%	0.54
x25	-0.0442	0.0163	0.0070		
x28	0.0623	0.0179	0.0006		
x29	-0.898	0.3655	0.0145		
x33	0.876	0.3619	0.0160		

Seguidamente, el modelo seleccionado (modelo 1 autoescalado) fue analizado teniendo en cuenta la presencia de datos atípicos influyentes y no influyentes utilizando los estadísticos de *Leverage* y la distancia de Cook. Los resultados de los estadísticos son mostrados en el Anexo 5. Los datos atípicos e influyentes encontrados son mostrados en la Tabla 11. El modelo de regresión fue recalibrado sin considerar los correspondientes puntos atípicos no influyentes; según esta recalibración, los casos con ID 134 y 207 ejercen una mayor influencia en la varianza (R² disminuye de 85% a 82% para el modelo 1). Estos casos corresponden a estados inusuales presentados en la Unidad 200 debido a altos desvíos de GLP a la red de gas combustible hacia los Hornos y a valores inusuales del poder calorífico de gas combustible. La supresión de los datos atípicos influyentes fue rechazada debido a que a nivel operativo, es posible contener estos casos sin que involucren fallas operativas ni errores de medición.

Con el modelo de regresión 1 (Ec. 3) es posible representar el consumo energético total presentado por la U-200 y predecir los mismos con cambios en las variables independientes. Un análisis de la influencia de cada variable en el modelo de regresión 1 muestra que las variables con mayor peso en la predicción

corresponden a x_{12} , le sigue x_{14} , y x_{17} (Tabla 2). Las anteriores variables en conjunto tienen una influencia del 53% en la predicción del modelo. Sin embargo, no sólo con estas 3 variables (poder calorífico del gas combustible y flujos de crudo a los hornos) es posible la representación y la predicción de los consumos energéticos de la U-200. Con el objetivo de proponer un modelo reducido con menor número de variables independientes, el modelo de regresión 1 fue reajustado sin considerar las variables de poca contribución (Tabla 13 del Anexo 8).

Tabla 11. Datos atípicos que influyen en el ajuste del seleccionado.

ID Datos atípicos	Eventos operacionales
311	La baja carga al H201 generó un bajo consumo de gas combustible.
134	Alto consumo de energía debido a un gas combustible de alto poder calorífico.
240	Aumento del poder calorífico del gas combustible al H202.
100	Aumento del poder calorífico del gas combustible al H202.
207	Falso registro del poder calorífico. Un poder calorífico cercano a 200 BTU/Pies cúbicos no se maneja en refinería.
32	La baja confiabilidad en equipos eléctricos ocasiona la operación de turbinas, aumentando consumo de vapor.
83	Alto inventario de GLP genera desvíos que ocasionan un aumento en el poder calorífico para un alto caudal de gas al H202.
97	Aumento de poder calorífico del gas combustible al H202.
159	Aumento de poder calorífico del gas combustible al H202.
190	Aumento de poder calorífico del gas combustible al H202.
54	Aumenta el consumo de vapor en la U200 por la puesta en marcha de turbinas para relevar bombas a motor eléctrico fuera de servicio.

Según la reducción del modelo de regresión 1, un modelo reducido representativo del consumo energético para la U-200 posee 8 variables para una explicación de la varianza del 82 % (Ec.4); la validez estadística de los supuestos del modelo reducido se presenta en el Anexo 9. Desde luego, el modelo reducido (Ec.4) presenta un desempeño menor que el modelo de la ecuación Ec.3. Sin embargo, el modelo reducido corresponde a una alternativa útil cuando se requiere una estimación aproximada del consumo energético y se cuenta con sólo las variables

requeridas para esta estimación. La Tabla 14 del Anexo 8 resume los desempeños estadísticos para el modelo de regresión 1 y su forma reducida.

$$Y_{\text{total reducido}} = -0.195 * x_5 + 0.795 * x_{12} + 0.452 * x_{14} - 0.102 * x_{16} + 0.290 * x_{17} + 0.234 * x_{28} - 0.098 * x_{32} + 0.139 * x_{33}$$

Ec. 3.

La regresión dada por la ecuación 4 permite inferir que el consumo energético total de la U-200 disminuye con la presión en la zona de flash de la T-201 (x_5), aumenta con el poder calorífico del gas utilizado (x_{12}), aumenta con la carga a los hornos (x_{14}, x_{17}), disminuye con respecto al crudo reducido producido en la T-201 (x_{16}), aumenta con respecto a la temperatura de salida por el serpentín A del H-202 (x_{28}), disminuye con respecto a la temperatura de salida del H-202 por el serpentín C y aumenta con la temperatura de entrada al H-202 por el serpentín D (x_{33}). Asimismo, las variables que más influyen corresponden a x_{12}, x_{14}, x_{17} con un aporte de 34%, 20% y 13%, respectivamente. Lo anterior tiene explicación desde el punto de vista del proceso ya que al parecer la operación de la U-200 no considera valores elevados de poder calorífico en el gas combustible (x_{12}). Lo anterior se refleja también en la baja correlación presentada entre la cantidad de gas combustible y las variables de proceso obtenida con el modelo ajustado 2; el coeficiente de correlación para las variables de poder calorífico y el flujo de gas combustible total corresponde a -0.24. Con el desconocimiento en el poder calorífico actual de operación, el consumo de gas de combustible en la U-200 puede ser incrementado para mantener la operación al procesar mayores cantidades de crudo o al disminuir las variables de salida de los hornos de sus respectivos valores de *set point*.

Asimismo, el análisis del modelo de la ecuación 4 sugiere que a una mayor cantidad de crudo se incrementa la demanda de energía, particularmente en los hornos; adicionalmente, una mayor carga representa un mayor consumo de energía en las bombas y turbinas para impulsarlo de un punto a otro de la Unidad. De igual manera, el menor valor de carga al H-201 (x_{14}) fue de 25,607.5 BPD, en este punto se reportó un consumo de energía de la planta bajo de 219.22 MBTU/h promedio en ese día. Este valor es significativo para una tasa promedio hora en el día de 312.5 MBTU/h esperada por los auditores en energía de la empresa para la U-200.

De otro lado, el desempeño de la regresión descrita por la ecuación 3 (85% de explicación de la varianza del consumo energético), sugiere que existen diferentes errores en los datos recolectados en planta que evitan una mejor representación del proceso de la U-200; más aún, el verdadero aporte de alguna variable puede estar camuflado en esos errores. Uno de estos errores puede encontrarse en la estimación del consumo de vapor de las turbinas, ya que su valor se estima a partir del consumo teórico de las mismas por carencia de la instrumentación respectiva. De igual forma, los instrumentos de medición podrían no estar calibrados o presentar ruidos al momento de registrar los valores respectivos. Una de las maneras de mejorar el desempeño de las regresiones desarrolladas en el presente documento consiste en aplicar un método de reconciliación a los datos de planta antes de efectuar los RLM.

4. CONCLUSIONES

- El presente documento provee una metodología para analizar y predecir el consumo de energía en una unidad de refinación de crudos basada en la regresión estadística de datos históricos de consumo energético. Esto mediante el ajuste de un modelo de regresión lineal múltiple para obtener las variables de mayor influencia en el consumo energético.
- Según los resultados del presente proyecto de grado, las variables que más influyen en el consumo de energía de la U-200 teniendo en cuenta la magnitud de su aporte de mayor a menor son: poder calorífico del gas combustible, el crudo al H201, el crudo al H202, la temperatura de salida del serpentín A del H202, la presión en la zona flash de la T201, la temperatura de entrada de crudo por el serpentín D del H202, crudo reducido de T-201 y Temperatura de salida de crudo del serpentín C H202.
- Las tendencias y las regresiones obtenidas en el presente trabajo sugieren que la operación de los hornos en la U-200, en general, se establece sin considerar los valores del poder calorífico del gas combustible. Por lo anterior, un aumento en el poder calorífico del gas combustible reporta un mayor consumo energético de la unidad, lo cual a nivel de proceso no es posible pues el consumo de energía realmente disminuye en términos de gas combustible al aumentar el poder calorífico.

5. RECOMENDACIONES

- Se recomienda aplicar técnicas de análisis multivariado como regresión por mínimos cuadrados parciales para obtener un ajuste de mejor desempeño.
- En términos operativos, se recomienda diseñar una metodología para obtener el consumo de energía en forma de vapor por parte de las turbinas.
- Aplicar un algoritmo de reconciliación de datos para disminuir los errores de medición antes de aplicar una regresión para el consumo de energía.

BIBLIOGRAFÍA

- [1] ISO 50001:2011. Energy management systems-requirements with guidance for use; 2011.
- [2] International Energy Agency. Online energy statistics. Energy balances; 2012. <<http://www.iea.org/statistics>>
- [3] US Energy Information Administration. International energy Outlook 2011; September, 2011. <<http://www.eia.gov/forecast/aeo>>
- [4] Romero, L. Carlos. León, I. Andrade, J. Proyección de demanda de energía en Colombia. Revisión 2010. UPME.
http://www.upme.gov.co/Docs/Energia/PROYECC_DEMANDA_ENERGIA_OCTUBRE_2010.pdf
- [5] Abdelaziz EA, Saidur R, Mekhilef S. A review on energy saving strategies in industrial sector. Renew Sust Energy Rev 2011:150-68.
- [6] Van Gorp John C, CEM Services Marketing. Using key performance indicators to manage energy costs. Strategic planning for energy and the environment; 2005. Pp 9-25.
- [7] ECOPETROL S.A. Informe de GEI (Gases de efecto invernadero) de la GRB. Coordinación de Energía y Pérdidas, GRB, 2015.
- [8] Morales Giovanni. (2015). Curso introductorio de análisis y correlación estadística de datos. Universidad Industrial de Santander.
- [9] Monedero, I. Biscarri, F. León, C. Guerrero, J. González, R. Pérez-Lombard, L. (2012). Decision system based on neural networks to optimize the energy efficiency of a petrochemical plant.

- [10] Velázquez, D. González-Falcón, R. Pérez-Lombard, L. Gallego, L. Monedero, I. Biscarri, F. (2013). Development of an energy management system for a naphtha reforming plant: a data mining approach. 217-225.
- [11] Puerto, S. Rodríguez, E. (2015). Línea de base energética de la sección atmosférica y la sección de vacío de la unidad U-250 de refinación de crudo en la refinería de ECOPETROL S.A (BARRANCABERMEJA).
- [12] Hair J., Black W., Babin B., Anderson R. Multivariate Data Analysis. Seventh Edition.
- [13] Rodríguez Rosa. Aportaciones al conocimiento del estado medioambiental de hidrosistemas de interés internacional situados en Castilla-De la mancha. Universidad Castilla-La Mancha. (2000).
- [14] Biscarri Félix et al. (2012). A decision support system for consumption optimization in a naphtha reforming plant.
- [15] ECOPETROL S.A. Manual de operación U-200. Barrancabermeja: ECOPETROL S.A. 2010.
- [16] Salvador F. y Gargello, P. (2003): "Análisis exploratorio de datos", [en línea] 5campus.com, Estadística <<http://www.5campus.com/leccion/aed>>
- [17] Montgomery D. (1992). Introduction To Linear Regression Analysis. Second Edition. John Wiley & Sons, Inc.
- [18] De la Garza J., Morales B., González B. Análisis estadístico Multivariante: Un enfoque teórico y práctico. Primera edición.

6. BIBLIOGRAFÍA

ABDELAZIZ EA, SAIDUR R, MEKHILEF S. A review on energy saving strategies in industrial sector. 2011.

BISCARRI Félix et al. A decision support system for consumption optimization in a naphtha reforming plant. 2012.

DE LA GARZA, Jorge. MORALES, Blanca. GONZÁLEZ, Beatriz. Análisis estadístico Multivariante: Un enfoque teórico y práctico. Primera edición.

ECOPETROL S.A. Informe de GEI (Gases de efecto invernadero) de la GRB. Coordinación de Energía y Pérdidas, GRB, 2015.

ECOPETROL S.A. Manual de operación U-200. Barrancabermeja: ECOPETROL S.A. 2010.

HAIR, Joseph. BLACK, William. BABIN, Barry. ANDERSON, Rolph. Multivariate Data Analysis. Seventh Edition.

INTERNATIONAL ENERGY AGENCY. Online energy statistics. Energy balances. 2012. <http://www.iea.org/statistics>

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 50001:2011. Energy management systems-requirements with guidance for use. 2011.

MONEDERO, Iñigo et al. Decision system based on neural networks to optimize the energy efficiency of a petrochemical plant. (2012).

MORALES, Giovanni. Curso introductorio de análisis y correlación estadística de datos. Universidad Industrial de Santander. (2015).

MONTGOMERY Douglas. PECK, Elizabeth. WINING, Geoffrey. Introduction to Linear Regression Analysis. John Wiley & Sons, Inc. (1992).

RODRÍGUEZ, Rosa. Aportaciones al conocimiento del estado medioambiental de hidrosistemas de interés internacional situados en Castilla-La Mancha. Universidad Castilla-La Mancha. (2000).

PUERTO, Sindy. RODRÍGUEZ, Erika. Línea de base energética de la sección atmosférica y la sección de vacío de la unidad U-250 de refinación de crudo en la refinería de ECOPETROL S.A (BARRANCABERMEJA). Universidad Industrial de Santander. 2015.

ROMERO, Luis. LEÓN, Ismael. ANDRADE, Jaime. Proyección de demanda de energía en Colombia. UPME. 2010.

http://www.upme.gov.co/Docs/Energia/PROYECC_DEMANDA_ENERGIA_OCTUBRE_2010.pdf

SALVADOR, F. GARGELLO, P. Análisis exploratorio de datos. (2003).
<http://www.5campus.com/leccion/aed>

US ENERGY INFORMATION ADMINISTRATION. International energy Outlook 2011. September, 2011. <http://www.eia.gov/forecast/aeo>

VAN GORP, John. CEM Services Marketing. Using key performance indicators to manage energy costs. Strategic planning for energy and the environment. 2005. Pp 9-25.

VELÁZQUEZ, D. et al. Development of an energy management system for a naphtha reforming plant: a data mining approach. (2013). Pp. 217-225.

7. ANEXOS

ANEXO A: CÓDIGO BASE UTILIZADO EN R1386 PARA EL MODELADO ESTADÍSTICO.

```
# Importar datos de Excel
read.excel<-function (header=TRUE,...)
+ {read.table("clipboard",sep="\t",header=header,dec=",",...)}
> Datos=read.excel()
> fix (Datos)
#Generación de muestra aleatoria
> set.seed (474)
#La muestra equivale al 70% de la información
> val<-sample(1:474,size=332)
#Datos para el ajuste del modelo
> valsimu<- Datos [val,]
#Datos para la validación del modelo
> valvali<- Datos [-val,]
#Método stepwise
> modelo1<-lm(x41~., Datos)
> summary (modelo1)
>step.modelo1<-step (modelo1) {}
#Modelo ajustado
>modelo<-lm(x41~ x3 + x5 + x6 + x12+ 1, data = valsimu)
#Estadística del modelo Ajustado
> summary (modelo)
#Datos ajustados del modelo
>Ysimu<-fitted (modelo)
#Validación del modelo con datos restantes de la matriz
>Yvali<-predict(modelo,valvali)
#AIC modelo
AIC (modelo)
#RMSE de datos ajustados
Sqrt (mean(Ysimu-Yobs)^2)
#RMSE de datos de validación
Sqrt (mean(Ysimu-Yob)^2)
#R^2 de validación cruzada utilizando 10 folds
```

```
library(bootstrap)
#Definición de funciones
theta.fit <- function(x,x41){lsfit(x,x41)}
theta.predict <- function(modelo,x){cbind(1,x)%*%modelo$coef}
#Matriz de variables Predictoras
X <- as.matrix(Datos[c("x3","x5","x6","x12")])
#Matriz de variable de respuesta
Y <- as.matrix(Datos[c("x41")])
results <- crossval(X,y,theta.fit,theta.predict,ngroup=10)
# R^2 de validación Cruzada
cor(y,results$cv.fit)**2
```

ANEXO B: METODOLOGÍA SEGUIDA PARA EL CÁLCULO DEL CONSUMO DE ENERGÍA EN TÉRMINOS DE VAPOR, ELECTRICIDAD Y GAS COMBUSTIBLE.

-VARIABLE CONSUMO ENERGÍA ELÉCTRICA

De la base de datos PI se obtuvo la cantidad de energía en MW promedio hora al día consumida en la U-200, y teniendo en cuenta que el índice registrado *Heat Rate* promedio, se obtuvo la cantidad de energía en MBTU promedio hora consumida a diario mediante la siguiente expresión matemática:

$$C.E.E = MW_{promedio\ hora} * Heat\ Rate \left(\frac{MBTU}{MW} \right)$$

Donde,

C.E.E : variable de consumo asociado a la electricidad en MBTU/H al día.

Heat Rate promedio hora al día= 1.894 MBTU/MW

- VARIABLE CONSUMO GAS COMBUSTIBLE

Se obtuvieron los datos de consumo de gas combustible promedio hora en el día y del poder calorífico del gas promedio a través de PI-ECOPETROL. Se siguió la siguiente expresión matemática para el cálculo de esta variable de consumo:

$$C.E.G.C = Flujo\ de\ gas\ combustible \left(\frac{kiloft^3}{hora} \right) * Poder\ calorífico\ del\ gas \left(\frac{BTU}{ft^3} \right) * \frac{1}{1000}$$

Donde,

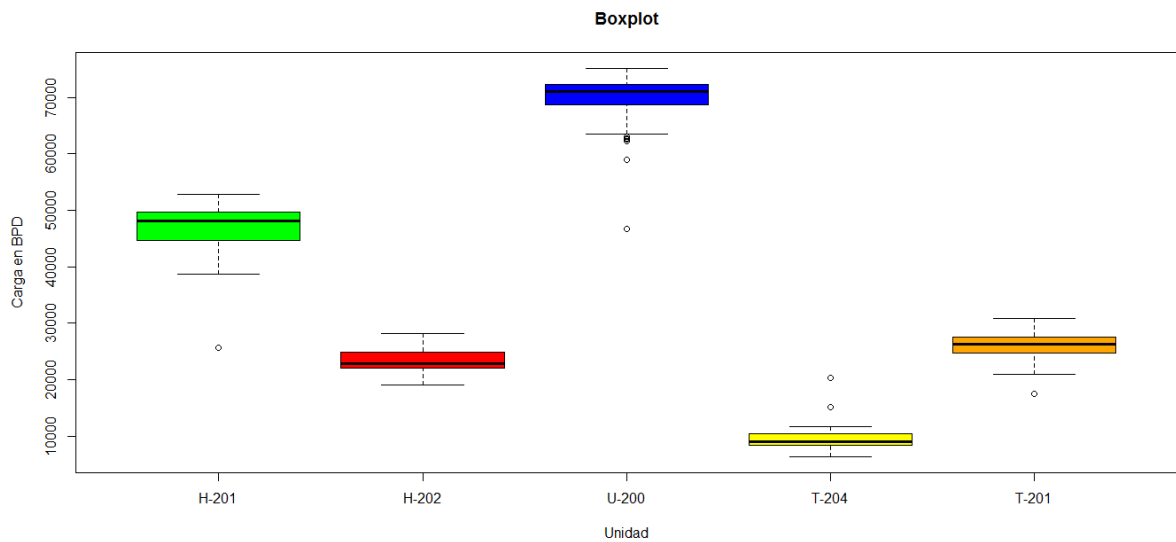
C.E.G.C: consumo de energía como gas combustible.

- VARIABLE DE CONSUMO VAPOR.

El mayor consumo de vapor en la U-200 es debido a las turbinas en servicio y los eyectores de vacío. En este estudio se encontró que las turbinas que se pusieron en servicio entre 12/08/2013 y 31/12/2014 fueron: NP-212D/ 223B/ 225B/ 207D/ 204D/ 237D/ 231B. El consumo se tomó como el gasto de vapor teórico multiplicando el flujo de diseño de cada equipo por el índice de generación de vapor (IGV) promedio en estas fechas. Finalmente cada consumo de vapor por turbina se suma al consumo correspondiente diario de eyectores.

ANEXO C: ESTADÍSTICA UNIVARIADA SOBRE VARIABLES DE PROCESO DE LA U-200.

Figura 11. Diagrama Boxplot para variables de proceso utilizadas para el análisis univariado de la U-200.



- Estadísticos nivel H-201

Tabla 12. Estadísticos H-201

Variable carga al H-201	CRUDAH201
Parámetro estadístico	Valor BPD
Valor mínimo	25,608
1er cuartil	44,720
Mediana	48,056
Media	47,122
3er cuartil	49,750
Máximo	52,889
Desv. estándar	1,673.661

- Estadísticos nivel H-202

Tabla 13. Estadísticos H-202

Variable carga al H-202	CRUDMEZH202
Parámetro estadístico	Valor BPD
Valor mínimo	19,035
1er cuartil	22,067
Mediana	22,873
Media	23,244
3er cuartil	24,791
Máximo	28,240
Desv. estándar	3,600.368

- Estadísticos nivel U-200

Tabla 14. Estadísticos U-200

Variable carga a la U-200	CRUDU200
Parámetro estadístico	Valor BPD
Valor mínimo	46,781
1er cuartil	68,632
Mediana	71,080
Media	70,325
3er cuartil	72,355
Máximo	75,178
Desv. estándar	4,516.128

- Estadísticos nivel T-204

Tabla 15. Estadísticos T-204

Variable Fondos T-204	SALFONT204
Parámetro estadístico	Valor BPD
Valor mínimo	6,323
1er cuartil	8,434
Mediana	8,980
Media	9,955
3er cuartil	10,337
Máximo	20,298
Desv. estándar	3,424.839

- Estadísticos nivel T-201

Tabla 16. Estadísticos T-201

Carga reducida de T-201	CRUDRET201
Parámetro estadístico	Valor BPD
Valor mínimo	17,461
1er cuartil	24,638
Mediana	26,297
Media	26,153
3er cuartil	27,566
Máximo	30,858
Desv. estándar	2,242.5

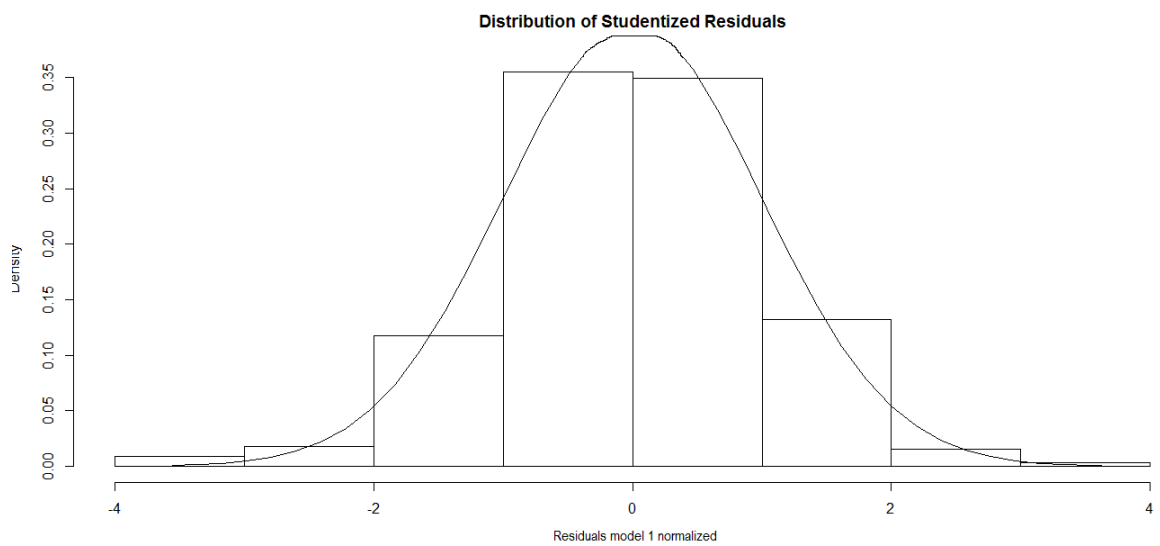
ANEXO D: VERIFICACIÓN DE SUPUESTOS ESTADÍSTICOS PARA MODELOS CON MEJOR DESEMPEÑO.

- Modelo 1:

Normalizado:

Supuesto de normalidad de residuos: Sí Cumple [11].

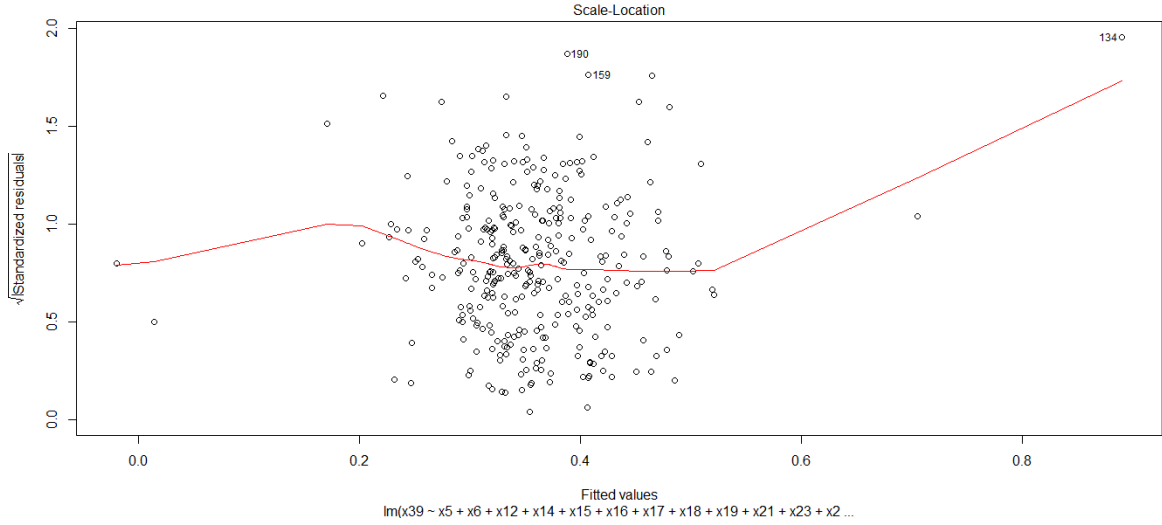
Figura 12. Distribución normal de los residuos del modelo 1 normalizado.



Supuesto de independencia de errores: De acuerdo al estadístico de Durbin-Watson para el modelo, este es igual a 2, por lo tanto el modelo sí cumple [11].

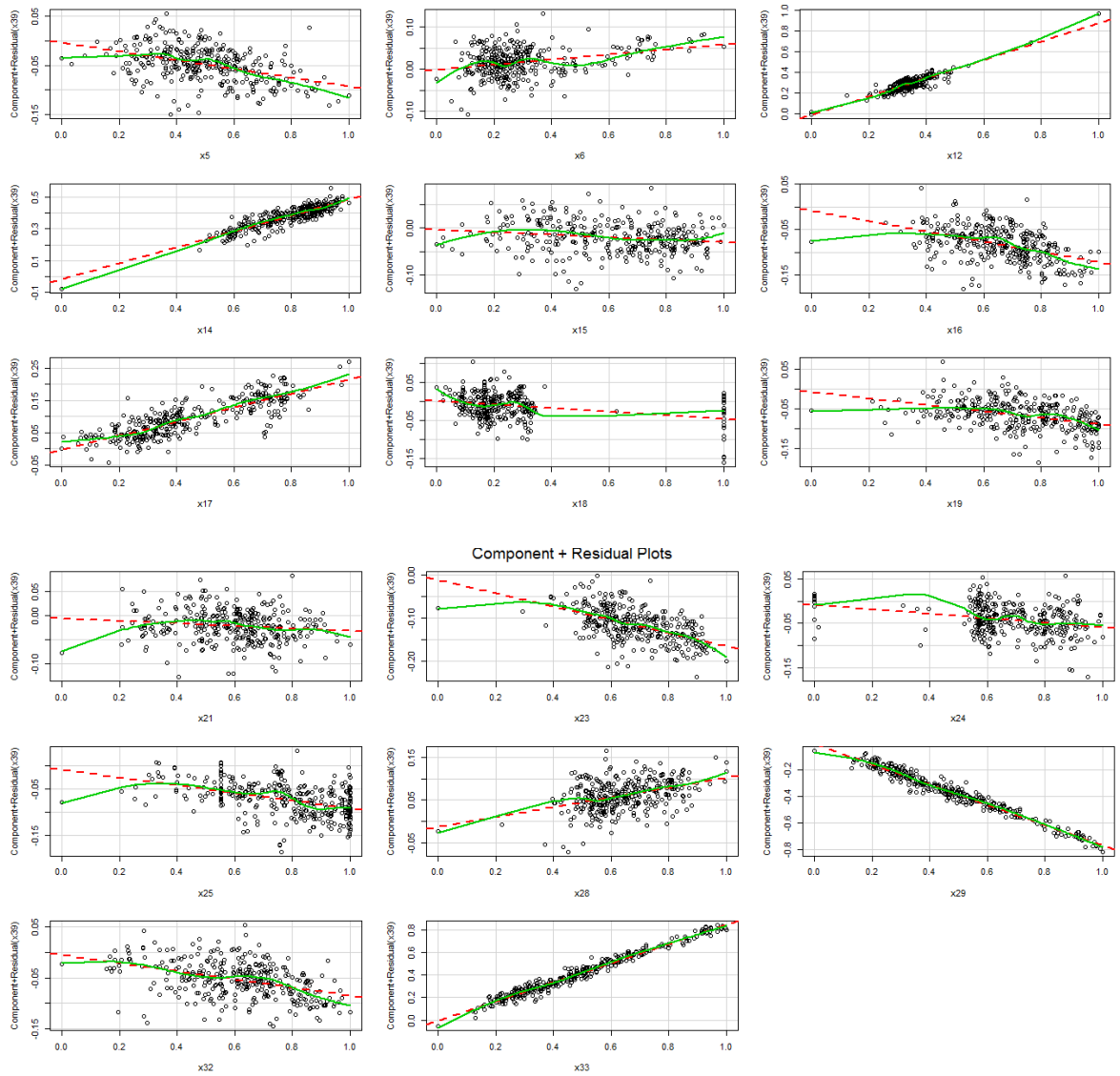
Supuesto de homocedasticidad o de varianza constante de los residuos: Sí cumple [11].

Figura 13. Dispersión de los residuos con respecto a los valores predichos por el modelo.



Supuesto de linealidad entre variable dependiente e independiente: Sí cumple [11].

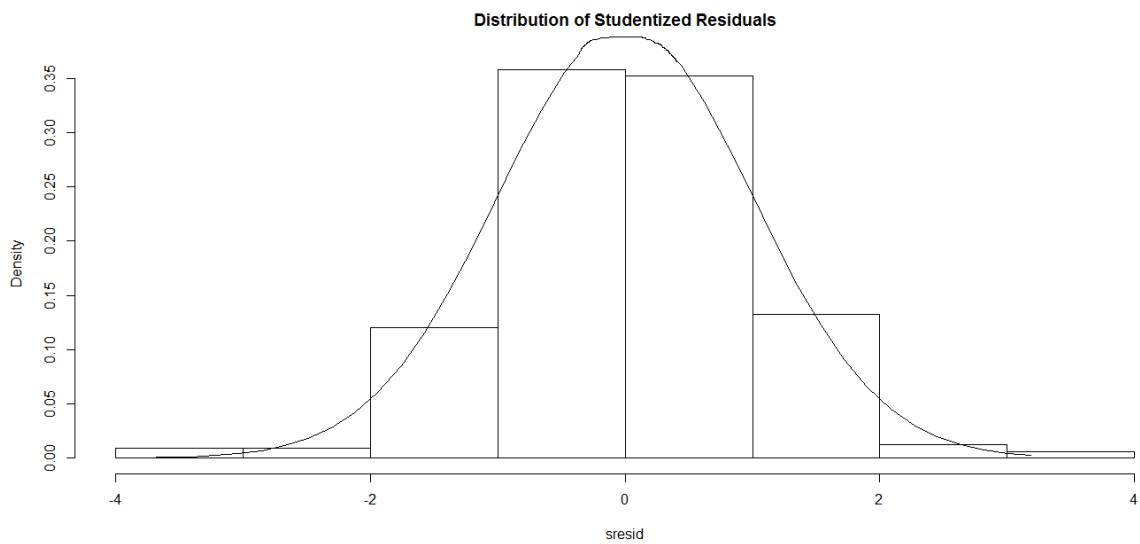
Figura 14. Prueba de linealidad del modelo 1 normalizado.



Autoescalado:

Supuesto de normalidad de residuos: Sí Cumple [11].

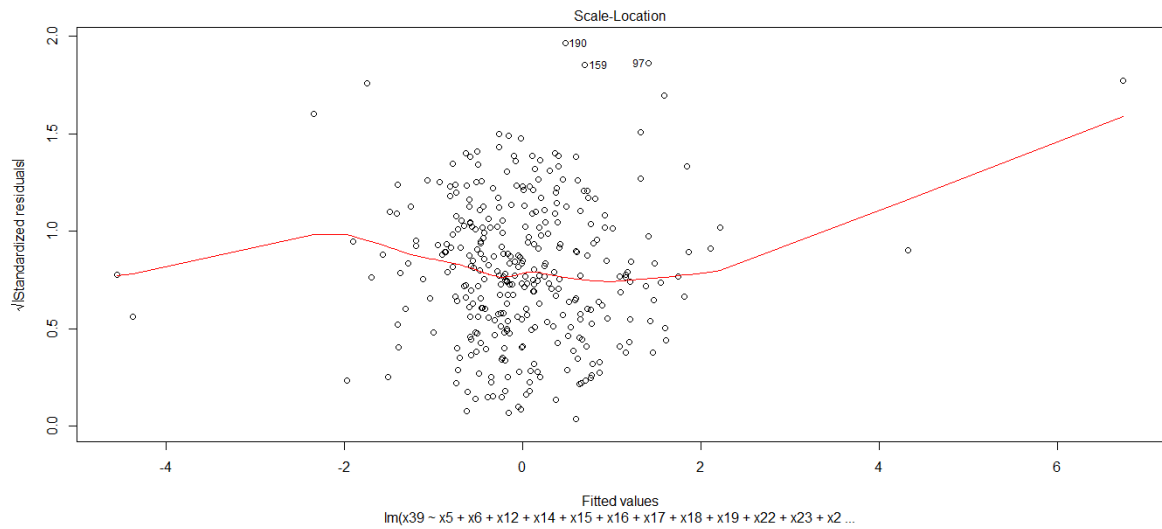
Figura 15. Prueba de normalidad para modelo 1 autoescalado.



Supuesto de independencia de errores: De acuerdo al estadístico de Durbin-Watson para el modelo, este es igual a 2, por lo tanto el modelo sí cumple [11].

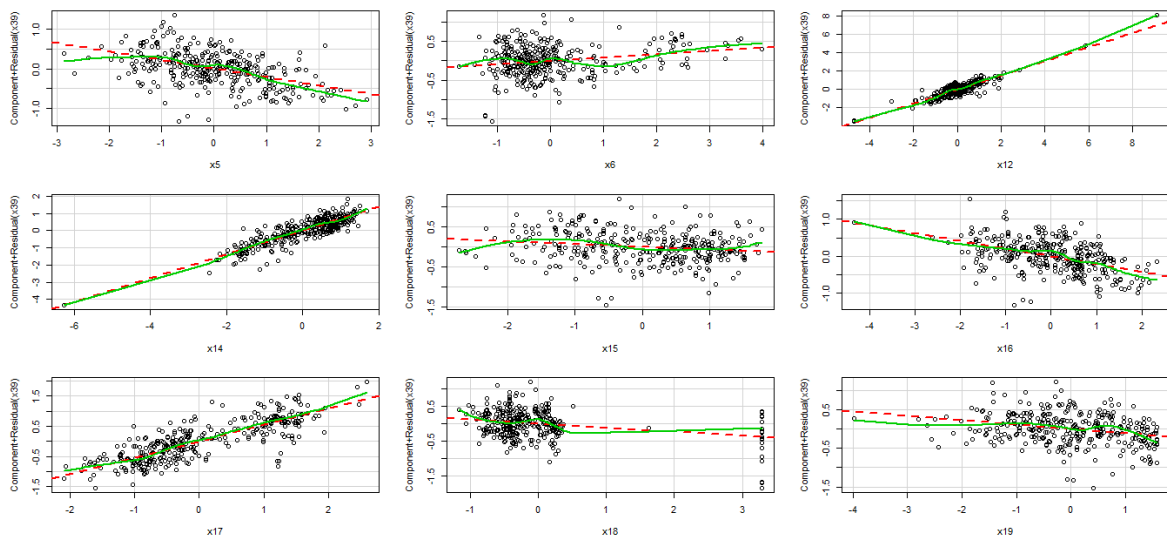
Supuesto de homocedasticidad o de varianza constante de los residuos: Sí cumple [11].

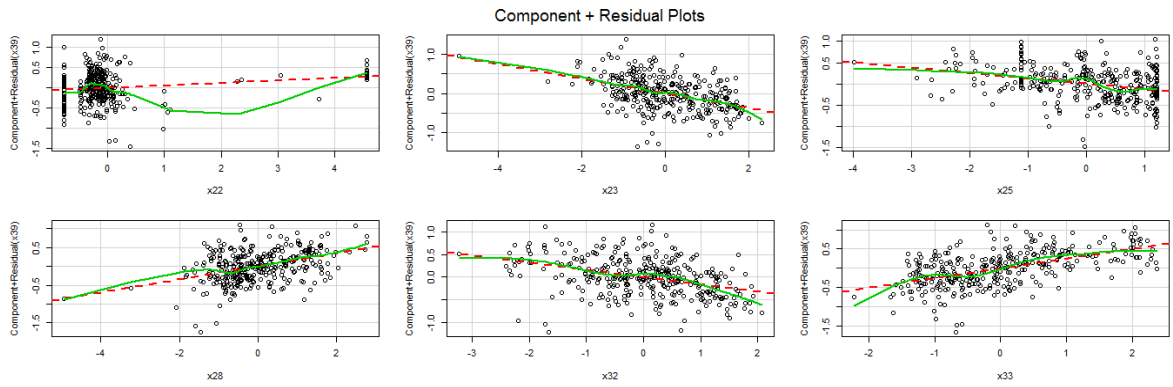
Figura 16. Prueba de varianza constante de residuos del modelo 1 autoescalado.



Supuesto de linealidad entre variable dependiente e independiente: Sí cumple [11].

Figura 17. Prueba de linealidad de variables dependiente e independiente.



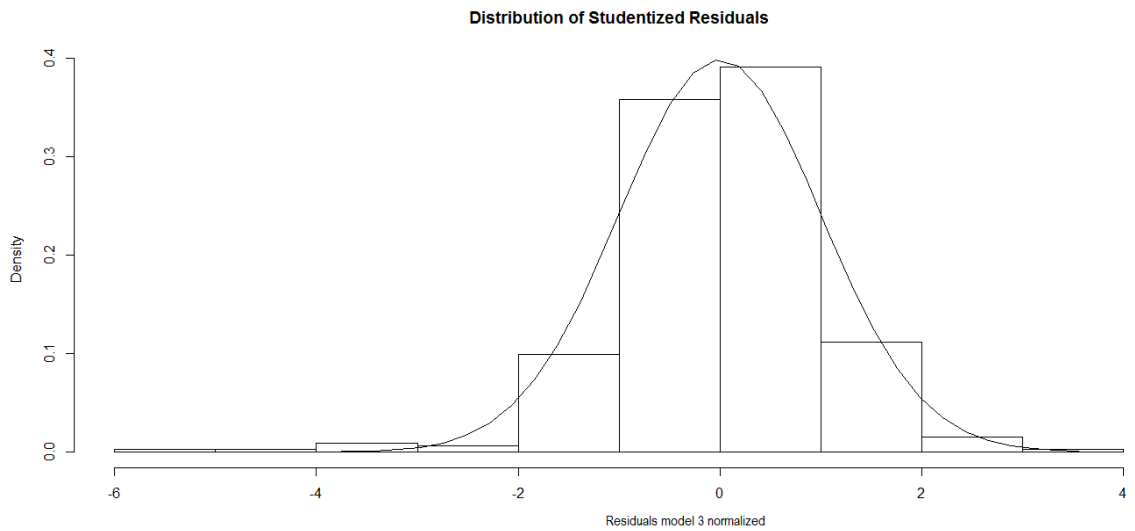


- **Modelo 3**

Normalizado:

Supuesto de normalidad de residuos: Sí Cumple [11].

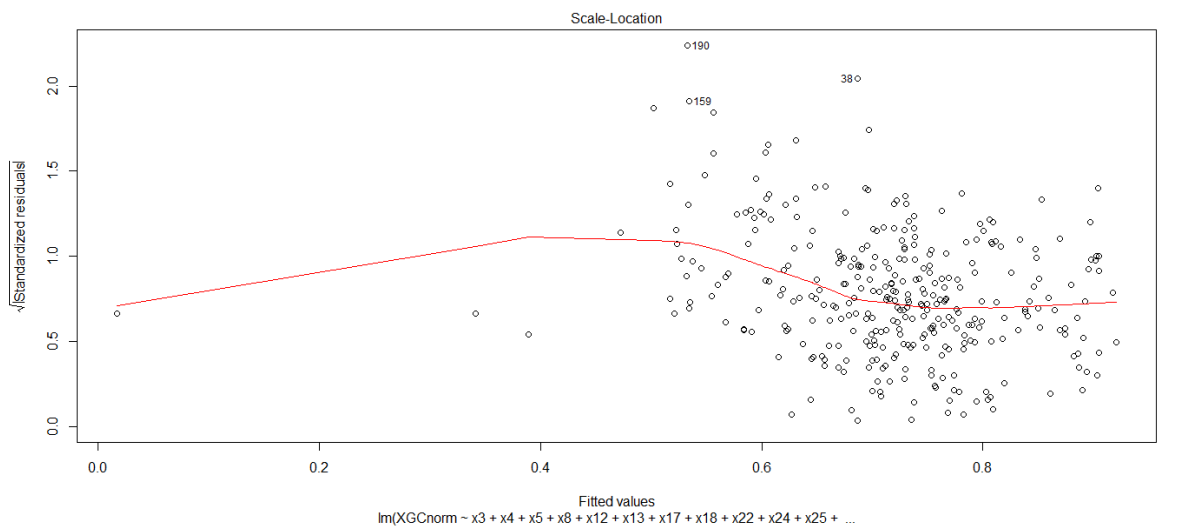
Figura 18. Prueba de normalidad modelo 3 normalizado.



Supuesto de independencia de errores: De acuerdo al estadístico de Durbin-Watson para el modelo, este es igual a 2, por lo tanto el modelo sí cumple [11].

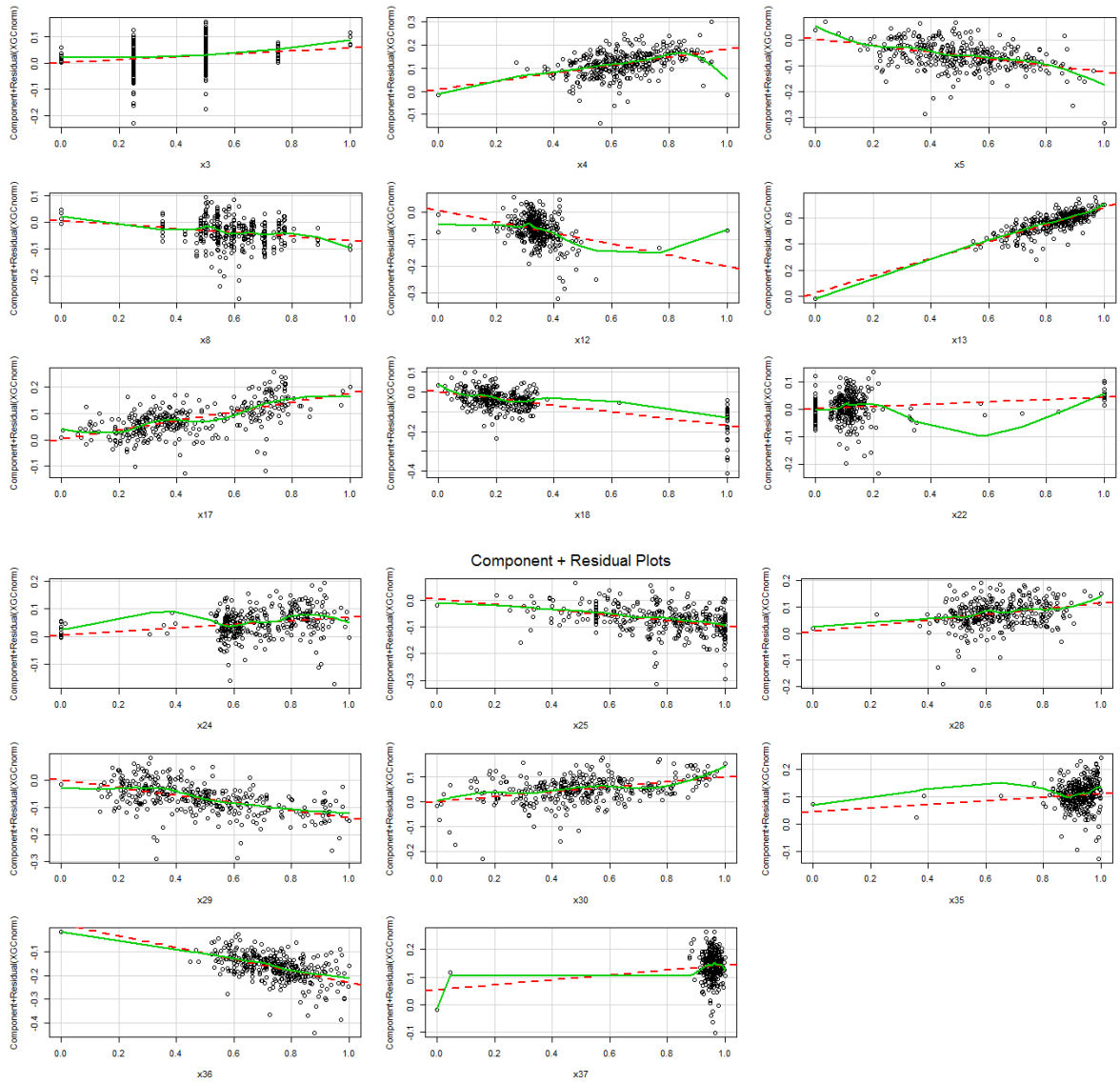
Supuesto de homocedasticidad o de varianza constante de los residuos: No cumple, los residuos experimentan comportamiento Heterocedástico [11].

Figura 19. Prueba de varianza constante de los residuos modelo 3 normalizado.



Supuesto de linealidad entre variable dependiente e independiente: Sí cumple [11].

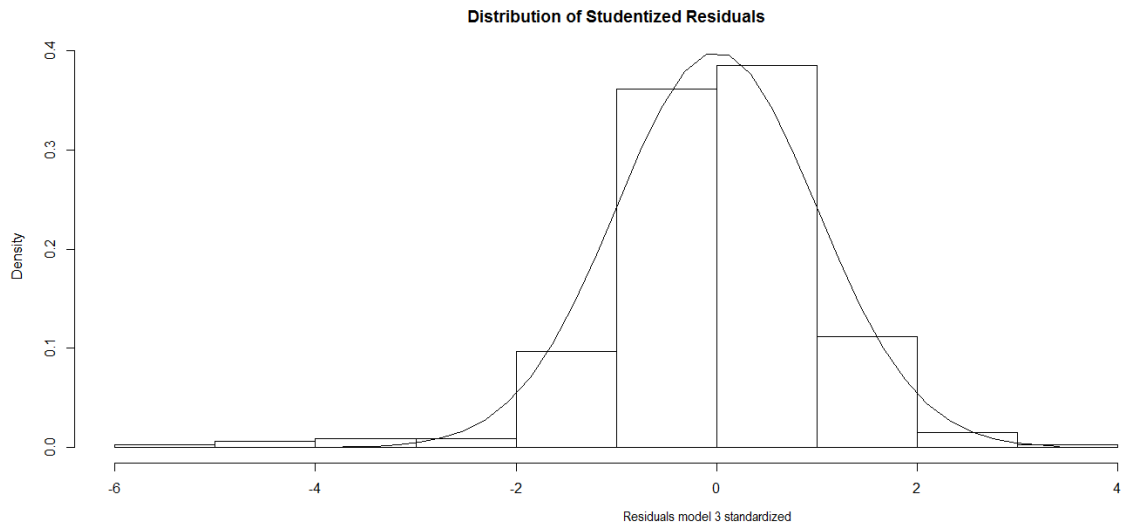
Figura 20. Prueba de linealidad de variables dependiente e independiente.



Autoescalado:

Supuesto de normalidad de residuos: Sí Cumple [11].

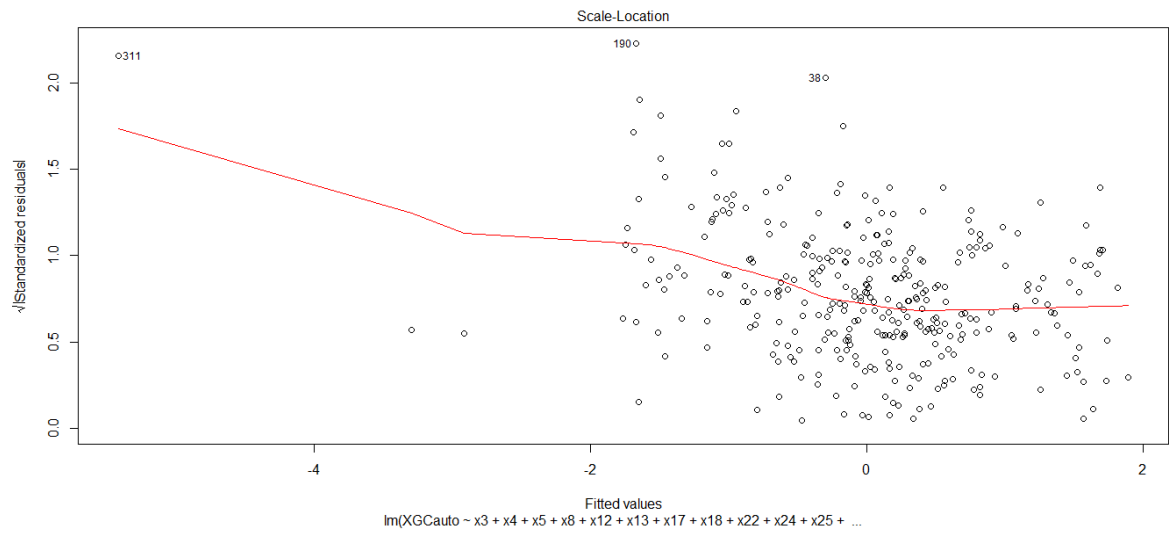
Figura 21. Prueba de normalidad modelo 3 autoescalado.



Supuesto de independencia de errores: De acuerdo al estadístico de Durbin-Watson para el modelo, este es igual a 2, por lo tanto el modelo sí cumple [11].

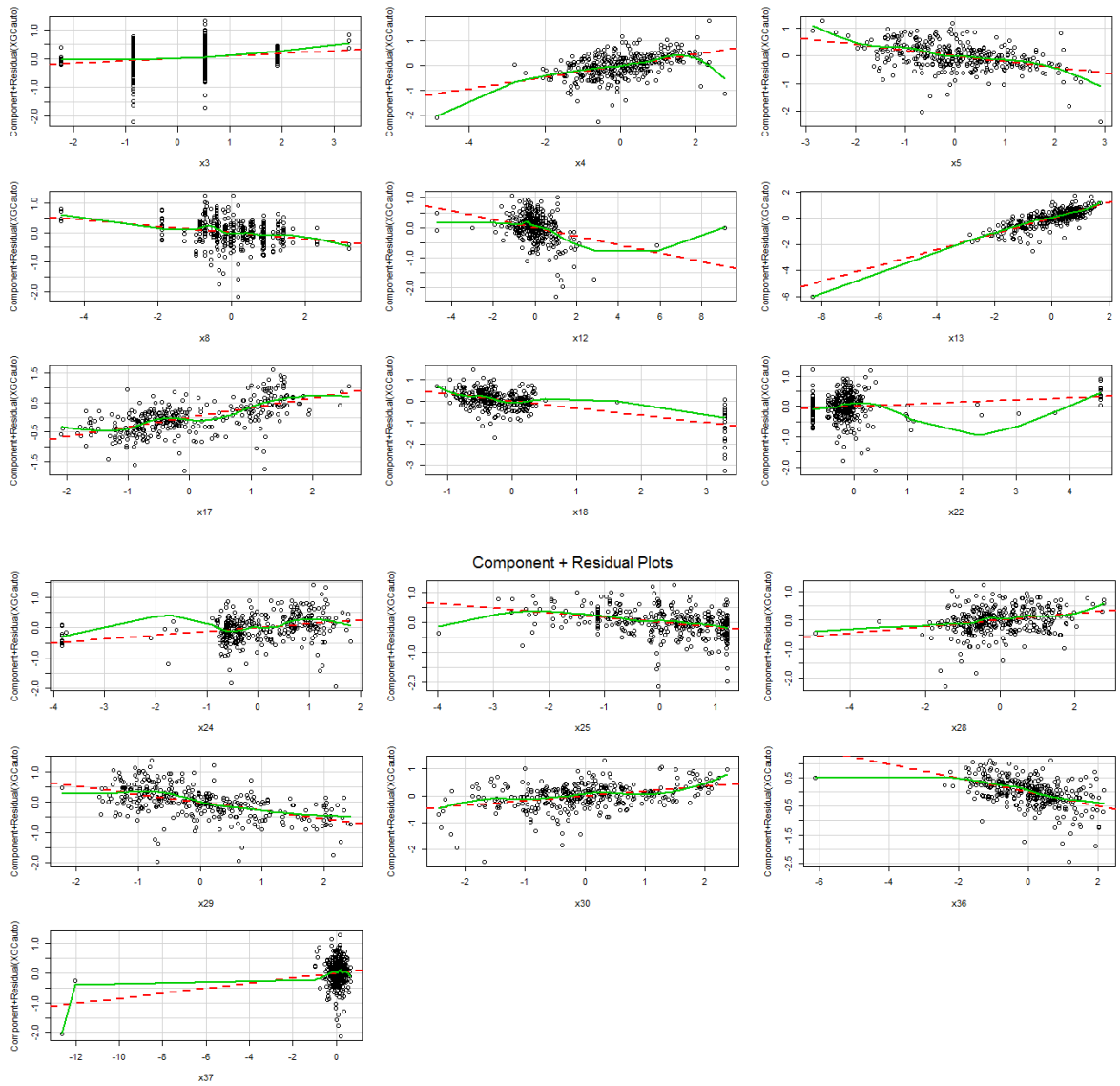
Supuesto de homocedasticidad o de varianza constante de los residuos: No cumple, los residuos experimentan comportamiento Heterocedástico [11].

Figura 22. Prueba de varianza constante modelo 3 autoescalado.



Supuesto de linealidad entre variable dependiente e independiente: Sí cumple [11].

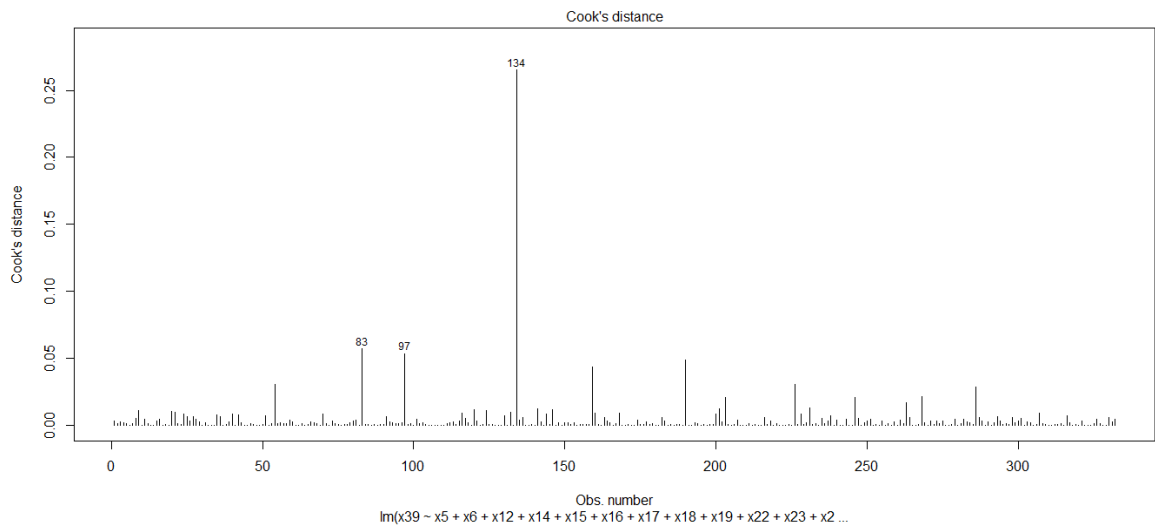
Figura 23. Prueba de linealidad de variables dependiente e independiente.



ANEXO E: GRÁFICOS ESTADÍSTICOS DE PUNTOS ATÍPICOS.

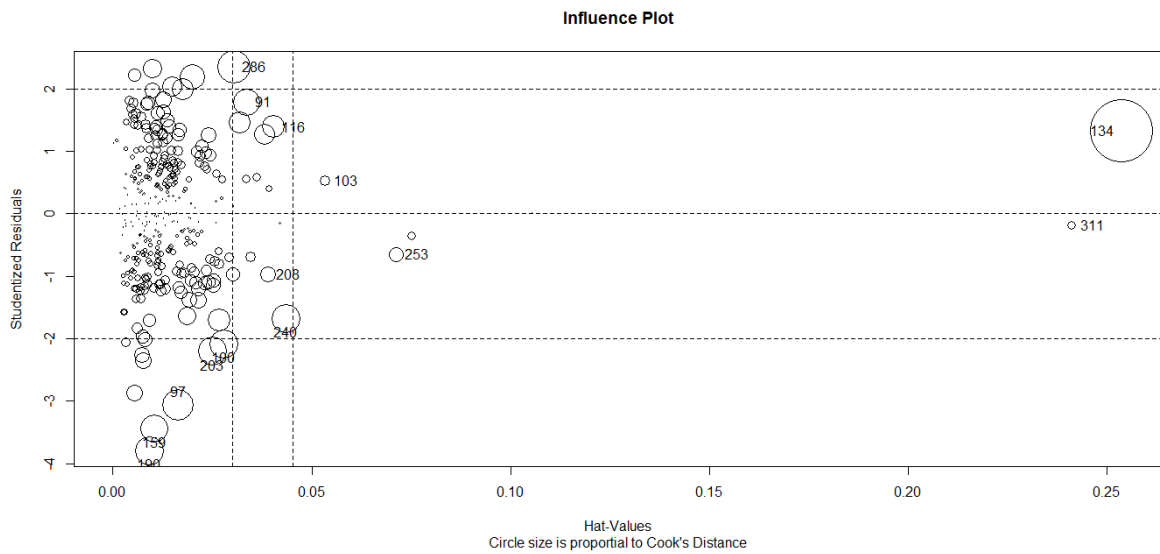
- Distancia de Cook de las variables de proceso

Figura 24. Diagrama de distancias Cook de puntos atípicos del modelo seleccionado (modelo 1 autoescalado).



- Puntos influyentes de acuerdo a la dimensión del círculo (relación directa tamaño- distancia cook)

Figura 25. Diagrama de puntos influyentes del modelo seleccionado (modelo 1 autoescalado).



ANEXO F: ESTADÍSTICO DEL MODELO SELECCIONADO REDUCIDO MEDIANTE LA FILTRACIÓN DE DATOS ALTAMENTE INFLUYENTES.

Tabla 17. Resumen estadístico de modelo reducido seleccionado.

	Coeficientes	% aporte	Error est.	P -valor	Estadístico	Valor
x12	0.795±0.048	34%	0.024	< 2e-16	R ²	0.82
x14	0.452±0.094	20%	0.048	< 2e-16	R ² <i>ajus</i>	0.81
x17	0.290±0.075	13%	0.038	3.60E-13	D-W	2.14
x28	0.234±0.057	10%	0.029	8.78E-15	R ² VC	0.81
x5	-0.194±0.062	8%	0.032	2.15E-09	R ² 30%	0.77
x33	0.139±0.083	6%	0.042	0.00104	R ² <i>ajus</i> 30%	0.75
x16	-0.102±0.08	4%	0.041	0.0119	RMSE 70%	0.16
x32	-0.098±0.06	4%	0.031	0.00144	RMSE 30 %	0.30
					RMSE VC	0.21

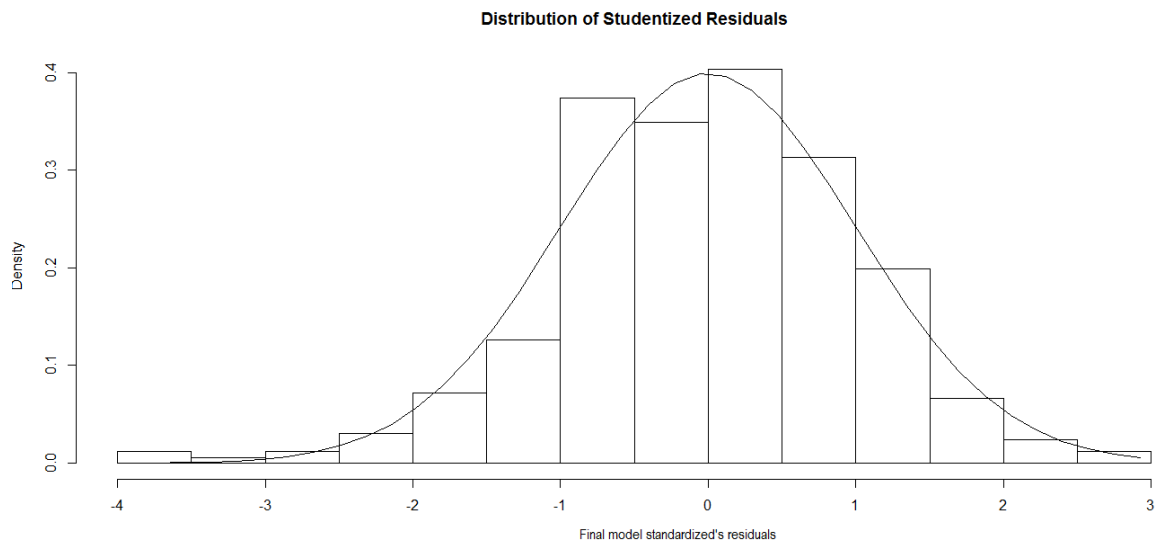
Tabla 18. Comparación de modelos para determinar desempeño de predicción.

Modelo 1	General	Reducido
R² 70%	0.85	0.82
R² <i>ajus</i> 70%	0.84	0.81
R² 30%	0.85	0.77
R² <i>ajus</i> 30%	0.83	0.75
R² VC	0.83	0.81
RMSE 70%	0.39	0.16
RMSE 30%	0.41	0.30
RMSE VC	0.40	0.21
DW estadístico	2.037	2.140

ANEXO G: VERIFICACIÓN DE SUPUESTOS ESTADÍSTICOS MODELO REDUCIDO FINAL.

Supuesto de normalidad de residuos: Sí Cumple [11].

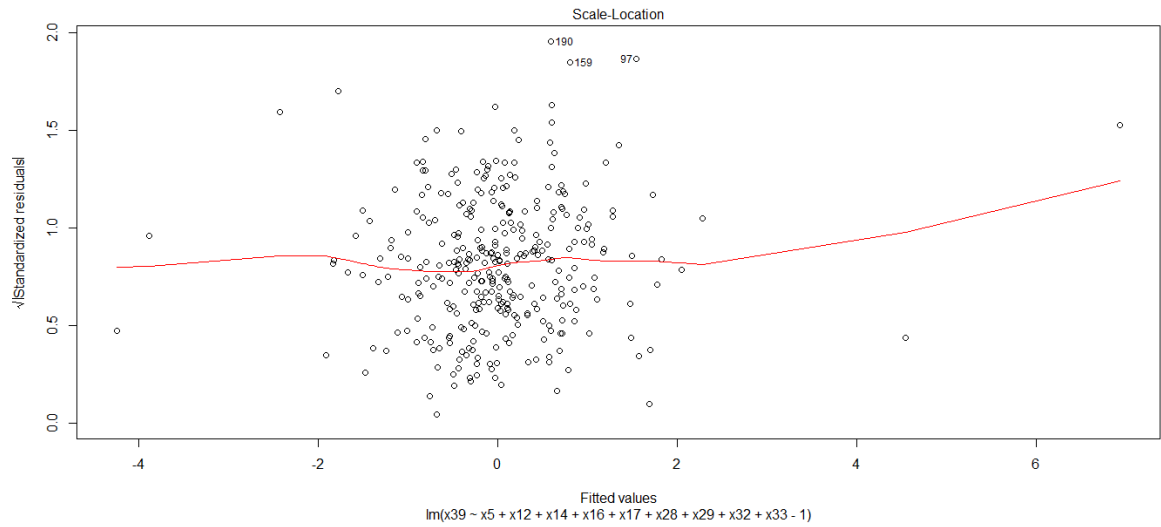
Figura 26. Prueba de normalidad modelo final seleccionado.



Supuesto de independencia de errores: De acuerdo al estadístico de Durbin-Watson para el modelo, este es igual a 2, por lo tanto el modelo sí cumple [11].

Supuesto de homocedasticidad o de varianza constante de los residuos: Sí cumple [11].

Figura 27. Prueba de varianza constante de residuos del modelo final seleccionado.



Supuesto de linealidad entre variable dependiente e independiente: Sí cumple [11].

Figura 28. Prueba de linealidad de variables dependiente e independiente.

