

**REGRESIÓN PARA LA PRODUCCIÓN DE GASOLINA DE AVIACIÓN (JET  
FUEL) EN UNA TORRE DE DESTILACIÓN INDUSTRIAL DE CRUDOS A  
PARTIR DE DATOS HISTÓRICOS DE PROCESO**

**EDWIN LEONARDO CASTAÑEDA ROBAYO  
ALBA JAZMIN JIMÉNEZ HERRERA**

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

**2017**

**REGRESIÓN PARA LA PRODUCCIÓN DE GASOLINA DE AVIACIÓN (JET  
FUEL) EN UNA TORRE DE DESTILACIÓN INDUSTRIAL DE CRUDOS A  
PARTIR DE DATOS HISTÓRICOS DE PROCESO**

**EDWIN LEONARDO CASTAÑEDA ROBAYO  
ALBA JAZMIN JIMÉNEZ HERRERA**

**Trabajo de grado para optar por el título de  
INGENIERO QUÍMICO**

**Director**

**GIOVANNI MORALES MEDINA, I.Q., Dr.  
Profesor escuela de ingeniería química**

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

**2017**

Dedicatoria.

Dedico esta tesis a Dios primeramente por haberme permitido culminar esta etapa de mi vida como profesional

A mi madre María Luisa Robayo por su apoyo incondicional en cada momento presentado en este proyecto.

A mi padre y a mis hermanos por estar junto a mí en todo este proceso

A mi novia quien me apoyo y alentó para continuar, cuando parecía que me iba a rendir.

A mis padrinos quienes con consejos y compañía hicieron que este proceso fuera de la mejor forma posible

A mis maestros quienes nunca desistieron al enseñarme, aun sin importar que muchas veces no ponía atención en clase, a ellos que continuaron depositando su esperanza en mí.

A todos los que me apoyaron para escribir y concluir esta tesis.

Para ellos es esta dedicatoria de tesis, pues es a ellos a quienes se las debo por su apoyo incondicional.

Ing. EDWIN LEONARDO CASTAÑEDA ROBAYO

La presente tesis está dedicada a Dios ya que gracias a él he logrado concluir este gran logro: mi carrera profesional.

A mis padres y hermanos, por sus palabras de apoyo, compañía, motivación y esfuerzos por verme hecha profesional.

A mis profesores, compañeros de estudio y amigos que con sus consejos y enseñanzas que quedaron guardadas en mi mente y mi corazón, aportaron un granito de arena en cada paso de mi formación profesional.

Al profesor Giovanni por su confianza, orientación y apoyo para culminar con éxito ésta etapa de mi carrera.

A mi novio que siempre ha estado presente con su inmenso cariño, paciencia y compañía.

A aquellos que hicieron parte de este trayecto con su presencia y voz de apoyo en los momentos difíciles.

Ing. ALBA JAZMIN JIMENEZ HERRERA.

## CONTENIDO

|  | <b>Pág.</b> |
|--|-------------|
| INTRODUCCIÓN .....   | 14          |
| 1. OBJETIVOS.....  | 16          |
| 1.1. OBJETIVO GENERAL .....  | 16          |
| 1.2. OBJETIVOS ESPECÍFICOS.....  | 16          |
| 2. MARCO TEÓRICO .....   | 17          |
| 2.1 PRODUCCIÓN DE COMBUSTIBLES .....   | 17          |
| 2.2. MÉTODOS DE ESTADÍSTICA MULTIVARIADA .....   | 19          |
| 3. METODOLOGIA .....   | 20          |
| 3.1. ETAPA 1 CONSTRUCCIÓN DE LA BASE DE DATOS, PROCESAMIENTO<br>Y ANÁLISIS CUALITATIVO Y CUANTITATIVO DE LA INFORMACIÓN..... | 20          |
| 3.2. ETAPA 2 APLICACIÓN DE LA REGRESION LINEAL MULTIPLE .....  | 21          |
| 3.3. ETAPA 3 MEJORA MODELOS .....  | 22          |
| 4. ANÁLISIS DE RESULTADOS.....   | 23          |
| 4.1. DATOS.....  | 23          |
| 4.2. ANÁLISIS ESTADÍSTICO .....  | 23          |
| 4.3. REGRESIONES EN LOS SEIS ESTADOS PSEUDO-ESTACIONARIOS.....   | 24          |
| 4.4. ANÁLISIS POR K-MEANS Y COMPONENTES PRINCIPALES.....   | 27          |
| 4.5. MODOS OPERACIONALES.....  | 31          |
| 4.5.1. Hotteling.....  | 33          |
| 5. CONCLUSIONES .....  | 35          |
| 6. RECOMENDACIONES.....  | 36          |
| CITAS BIBLIOGRÁFICAS.....  | 37          |
| BIBLIOGRAFÍA .....   | 40          |
| ANEXOS .....   | 44          |

## LISTA DE FIGURAS

|   | <b>Pág.</b> |
|---|-------------|
| Figura 1. Diagrama de flujo general del proceso de refinado en la U-200. .... | 18          |
| Figura 2. Descripción DE la metodología.....                                  | 20          |
| Figura 3. Grafica de auto-escalado de todos los datos sin atípicos. ....      | 24          |
| Figura 4. Gráfico de K-Means A 2, 3, 4 Y 5 grupos. ....                       | 27          |
| Figura 5. Distribución de 4 K-Means.....                                      | 28          |
| Figura 6. Pls del grupo 2.....  | 29          |
| Figura 7. Predicción del grupo 2 Por Pls.....                                 | 30          |
| Figura 8. Hotelling de grupos 1 Y 2.....                                      | 33          |
| Figura 9. Hotelling De grupos 3 Y 4.....                                      | 33          |

## LISTA DE ECUACIONES

|   | <b>Pág.</b> |
|---|-------------|
| Ecuación 1. Ecuación de auto escalado. ....     | 21          |
| Ecuación 2. Regresión para todos los datos..... | 25          |
| Ecuación 3. Regresión del estado 2013-1. ....   | 25          |
| Ecuación 4.Regresión del estado 2013-2. ....    | 25          |
| Ecuación 5. Regresión del estado 2014-1. ....   | 25          |
| Ecuación 6.Regresión del estado 2015-1 ....     | 25          |
| Ecuación 7. Regresión del estado 2016-1. ....   | 26          |
| Ecuación 8.Regresión del estado 2016-2. ....    | 26          |
| Ecuación 9. Regresión del estado 1.1. ....      | 32          |
| Ecuación 10. Regresión del estado 2.2. ....     | 32          |
| Ecuación 11.Regresión del estado 3.2. ....      | 32          |
| Ecuación 12.Regresión del estado 4.2. ....      | 32          |

## LISTA DE TABLAS

|  | <b>Pág.</b> |
|--|-------------|
| Tabla 1. Tabla resultante de la cantidad de datos de los estados estacionarios. .. | 24          |
| Tabla 2. Reporte de los R ajustados.....   | 26          |
| Tabla 3. Modificación de R eliminando variables .....                              | 28          |
| Tabla 4. Reporte por medio de PIs a los estados. ....                              | 29          |
| Tabla 5. Tabla de regresiones de sub-grupos .....                                  | 31          |
| Tabla 6. Eliminación de datos atípicos en las subdivisiones.....                   | 31          |

## LISTA DE ANEXOS

|   | <b>Pág.</b> |
|---|-------------|
| Anexo A. Tabla de identificación de todas las variables iniciales.....                        | 44          |
| Anexo B. Sesgo y curtosis variables influyentes.....  | 46          |
| Anexo C. Estadísticos de subgrupos. ....  | 47          |
| Anexo D. Validación, hotelling y distribución de residuales a subgrupos sin<br>atípicos ..... | 49          |
| Anexo E. Prueba de linealidad variables pendiente e independiente .....                       | 51          |
| Anexo F. Regresión con efectos dobles de las variables iniciales. ....                        | 55          |
| Anexo G. Regresión por PLS y PCR del grupo 1.....   | 56          |
| Anexo H. Regresión por PLS y PCR del grupo 3.....   | 58          |
| Anexo I. Regresión por PLS y PCR del grupo 4. ....  | 60          |

## RESUMEN

**TITULO:** REGRESIÓN PARA LA PRODUCCIÓN DE GASOLINA DE AVIACIÓN (JET FUEL) EN UNA TORRE DE DESTILACIÓN INDUSTRIAL DE CRUDOS A PARTIR DE DATOS HISTÓRICOS DE PROCESO\*

**AUTORES:** EDWIN LEONARDO CASTAÑEDA ROBAYO; ALBA JAZMIN JIMÉNEZ HERRERA\*\*.

**PALABRAS CLAVE:** Regresión lineal multivariada, Refinación de crudo, análisis estadístico, agrupación clúster y variables de mayor influencia.

### DESCRIPCIÓN

El ajuste de modelos para la predicción de respuestas en los procesos es una herramienta muy útil en la mayoría de las industrias químicas y petroquímicas, un método fundamental para llegar a dichos modelos es por medio del análisis estadístico de las variables directamente implicadas en el proceso de refinación de crudo para el caso de la gasolina de aviación (JET FUEL). El presente trabajo plantea una metodología para la aplicación de regresión múltiple multivariada sobre DHP comprendidos entre los años 2013 a 2016 obtenidos de la base de datos de ECOPETROL S.A con el objetivo de sugerir un modelo para cada estado pseudo-estacionario y los modos de operación presentes en él, esto con el menor error posible de la predicción. Con un análisis inicial se obtuvo un modelo general para la producción de JET con  $R^2 = 0.78$  en el cual se aplicó la eliminación de datos atípicos mediante análisis clúster definiendo una agrupación de 4 grupos y cada grupo en 2 subgrupos de los cuales salieron 4 modelos a diferentes condiciones de operación donde 2 pertenecen al modo viscorreductor y 2 al modo crudo con un margen aproximado de error del 5%, esto constata que pueden generarse varias predicciones dependiendo del modo operativo y las condiciones de alimentación de la dieta del crudo a la torre.

---

\* Trabajo de grado

\*\* Facultad de ingeniería Físicoquímicas. Escuela de ingeniería química. Director Giovanni Morales

## ABSTRACT

**TITLE:** REGRESSION FOR THE JET FUEL PRODUCTION OF GASOLINES IN AN INDUSTRIAL DISTILLATION COLUMN BASED ON HISTORICAL PROCESS DATA\* .

**AUTHORS:** EDWIN LEONARDO CASTAÑEDA ROBAYO; ALBA JAZMIN JIMENEZ HERRERA \*\*

**KEY WORDS:** Multivariate linear regression, crude oil refining, statistical analysis, cluster aggrupation and greater influence variables.

### DESCRIPTION

Model adjustment for response prediction in processes is a really helpful tool in the chemical and petrochemical industries; an essential method used to achieve said models is the statistical analysis of the directly involved variables in the refining process of crude oil for the production of jet fuel. This project proposes a methodology for the application of multivariate multiple regression of HPD ranging from 2013 to 2016, this data was acquired from ECOPETROL S.A database with the objective of suggesting a model for each pseudo-stationary state and the operation modes present within it, with the least possible error for each prediction. An early analysis led to a general model for JET production with  $R^2 = 0.78$ , in which, atypical data was eliminated using cluster analysis defining a set of 4 groups, each group was subdivided in 2 subgroups from which 4 models were defined at different operating conditions, 2 of those 4 models belong to the visbreaking mode and the other 2 to the crude oil mode with an approximated error of 5%, this notes that there can be generated several predictions depending on the operative mode and the crude oil feed conditions to the column.

---

\* Degree work

\*\* Faculty of Physico-chemical Engineering. School of Chemical Engineering. Director: Giovanni Morales medina.

## INTRODUCCIÓN

Colombia es uno de los fabricantes de gasolina de avión o jet fuel más acreditados en el negocio, con ventas en países como Perú, Ecuador, Brasil y EEUU (en La florida donde llegan a tanquear las aerolíneas del mundo debido a su punto geográfico estratégico); En el año 2001 cuando un barril de petróleo costaba 21 dólares, un barril de combustible JET costaba cerca de 29 dólares, con ventas anuales aproximadas de 3,07 millones de barriles, generando ingresos cercanos a los 230.000 millones de pesos al país [1]; el sector aéreo representa un 40% los costos de operación de las torres [2]. Actualmente, las refinerías en Colombia tienen una capacidad de 415 kbpd (miles de barriles por día) [3] y el precio del combustible jet alcanza US\$101,52 por barril [4].

La mayor parte de los combustibles de origen fósil es generado en las torres de destilación, por lo cual, las propiedades de los crudos procesados y las variables operacionales influyen fuertemente en el rendimiento y en las propiedades de las gasolinas generadas. La predicción de los rendimientos de las gasolinas ayuda en la planeación de la refinería y en la definición de las dietas de crudos a las unidades; un aumento en los flujos de productos livianos en las unidades de destilación impacta positivamente en el margen de refinación en cuanto a un mayor valor comercial [5]. Una fuente importante de datos para el desarrollo de modelos de predicción se encuentra en los datos históricos de proceso (DHP); estos DHP permiten la proposición de regresiones para los rendimientos producidos, en términos de las propiedades del crudo procesado y las variables de operación.

La predicción de las características y flujos de las gasolinas producidas en las torres de destilación ha sido tema en diversos reportes. *Salazar M.* analizó datos de consumo energético de la unidad de refinación de crudos por medio de regresión

lineal múltiple [6]; de igual manera, *Rubio R.* identifica mejoras en el proceso de destilación de crudos de la U-200 utilizando estadística multivariada [7]; asimismo estudios anteriores realizados por *López C. et al.*, reportan modelos con nivel de confianza de 80, 90 y 95% en rendimiento del jet mediante diseño de experimentos de filtrado en un análisis jerárquico de variables [8].

La presente propuesta establece un protocolo para el tratamiento de los DHP de una torre de destilación industrial proponiendo modelos de regresión para la producción de los rendimientos de los flujos de gasolina de aviación.

## **1. OBJETIVOS**

### **1.1. OBJETIVO GENERAL**

Proponer y validar un modelo de regresión basado en datos históricos de proceso para la predicción del desempeño de una unidad de destilación industrial en la generación de gasolina de aviación.

### **1.2. OBJETIVOS ESPECÍFICOS**

- Obtener los estadísticos representativos de los datos históricos de una unidad de destilación de crudos.
- Clasificar los modos operativos de la torre de destilación de crudo según datos históricos.
- Validar regresiones para la predicción del rendimiento de gasolina de aviación según los modos operativos.

## 2. MARCO TEÓRICO

### 2.1 PRODUCCIÓN DE COMBUSTIBLES

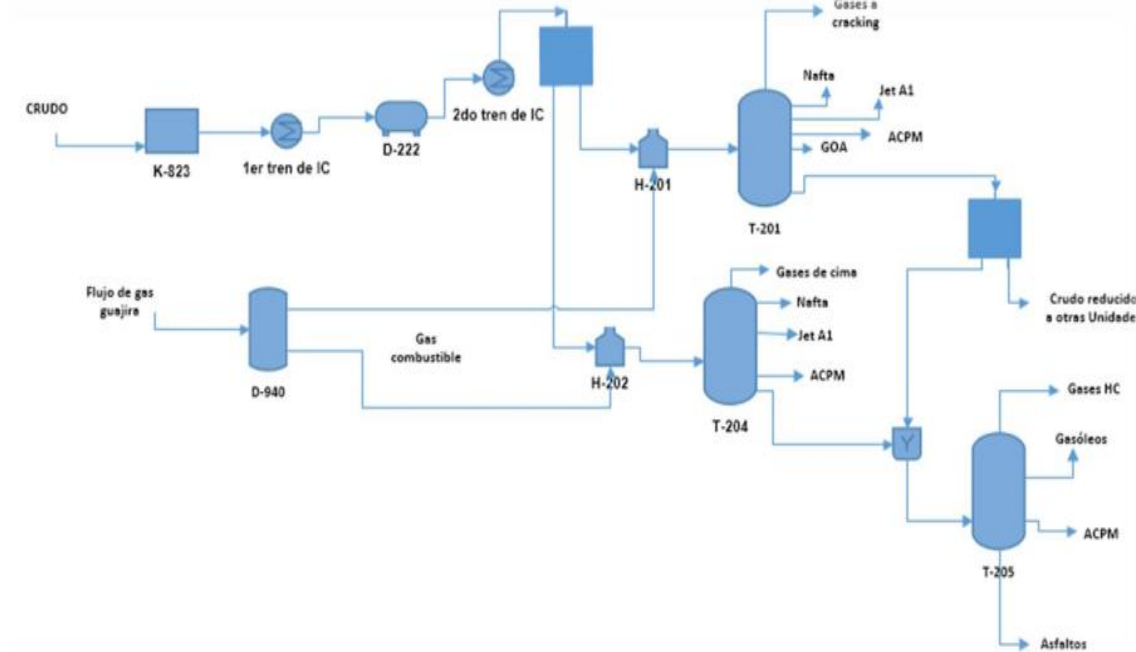
El petróleo crudo es una de las principales fuentes económicas en varios países. La transformación a productos de alto valor agregado está ligada a las propiedades de los crudos; elementos como el azufre, el nitrógeno y ciertos metales, así como la relación Carbono/Hidrógeno repercuten en la severidad de los tratamientos y por lo tanto en los costos de refinación [9].

El tratamiento de los crudos envuelve su separación en fracciones por medio de la operación de destilación. Las fracciones resultantes de la destilación son gases, destilados ligeros (naftas), destilados medios (queroseno, combustible pesado, diésel y combustible de calefacción) y residuos [10]; debido a su valor económico y su elevado consumo, una parte de la nafta es utilizada para producir gasolina de alto octanaje que puede dirigirse a automóviles o aviones.

El Jet fuel es el combustible utilizado en aviación para turbina de reacción o aviación (ATF); asimismo, es un tipo de combustible diseñado para uso en aviones propulsados por motores de turbina de gas. Este combustible está constituido por hidrocarburos no olefínicos entre C9 a C12, exhibiendo un intervalo de ebullición entre 150 y 250 °C [11]. En la refinería de Barrancabermeja (GRB), una de las unidades de mayor capacidad de tratamiento de crudos corresponde a la denominada U-200 (Figura 1). Aunque inicialmente la U-200 fue construida como unidad viscorreductora para procesar fondos de vacío, también se ha dispuesto para operar como unidad de refinación; la U-200 puede operar con dos tipos de escenario, como viscorreductora y modo unidad de destilación de crudos. Los dos

modos envuelven diferentes requerimientos (temperaturas y presiones de operación).

**Figura 1. Diagrama de flujo general del proceso de refinado en la U-200.**



Fuente: Salazar M. Diagrama de flujo de proceso U-200

Debido a la gran cantidad de variables que intervienen en la operación de la U-200, éste proceso es catalogado como de alta complejidad; debido a esta complejidad la predicción de los productos como el Jet requiere tiempo para la solución de un modelo de tipo fenomenológico. Alternativamente, el análisis de datos históricos de proceso (DHP) puede conllevar a la proposición de un modelo de regresión para el Jet, a partir de las variables importantes de entrada al proceso. [12]

Los DHP proporcionados por sensores instalados en planta proporcionan valores para los flujos, las concentraciones, las temperaturas y las presiones, para determinado intervalo de tiempo (ventana de operación).

## 2.2. MÉTODOS DE ESTADÍSTICA MULTIVARIADA

Regresión Lineal Multivariada (RLM): la regresión lineal multivariada permite el estudio de varias variables de respuesta, describiendo un modelo que relaciona un conjunto de  $m$  respuestas múltiples a un conjunto de  $n$  variables predictoras [13]

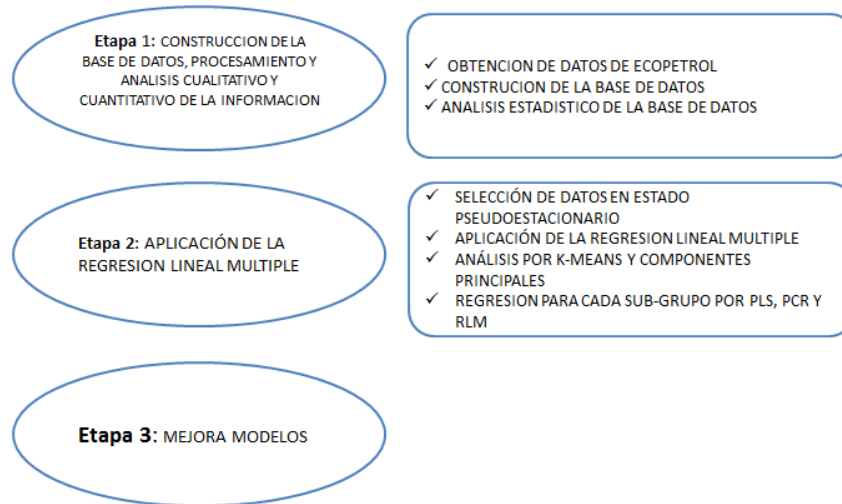
Análisis de Componentes Principales (PCA): El análisis de componentes principales es una forma de identificar patrones en los datos, expresando éstos de tal manera que resalten sus similitudes y diferencias. Dado que los patrones en los datos pueden ser difíciles de encontrar en datos de alta dimensión, donde no se cuenta con representación gráfica, PCA es una poderosa herramienta para analizar los datos, ya que una vez encontrados los patrones, comprime los datos reduciendo las dimensiones sin grandes pérdidas de información [14]

Parcial Least Square (PLS): La regresión de PLS es una técnica que generaliza y combina características de análisis de componentes principales y regresión múltiple. Su objetivo es predecir o analizar un conjunto de variables dependientes de un conjunto de variables independientes o predictoras y describir su estructura común. Esta predicción se logra extrayendo de las variables predictoras un conjunto de factores ortogonales llamados variables latentes que tienen el mejor poder predictivo, debido a la multi-colinealidad que se presenta en ocasiones es necesario utilizar este método en conjunto con métodos como PCA. [15]

### 3. METODOLOGIA

Las 3 etapas mostradas en la figura 2, enmarcan la secuencia del proceso aplicado a los datos para la obtención del modelo de predicción para el jet. Los programas utilizados para la aplicación de cada procedimiento del presente trabajo fueron el paquete libre computacional *R* versión *i386* y el paquete comercial *THE UNSCRAMBLER X* licenciado para ECOPETROL y la UIS.

**Figura 2. Descripción de la metodología.**



#### 3.1. ETAPA 1 CONSTRUCCIÓN DE LA BASE DE DATOS, PROCESAMIENTO Y ANÁLISIS CUALITATIVO Y CUANTITATIVO DE LA INFORMACIÓN

- **OBTENCIÓN DE DATOS DE ECOPETROL:** En esta etapa se solicitaron las mediciones de todas las variables de proceso de la de la U-200 disponibles en la base de datos de ECOPETROL Barrancabermeja denominada PI

- **CONSTRUCCIÓN DE LA BASE DE DATOS:** se construyó una tabla de datos en Excel como un soporte de información para la realización de este trabajo en cuanto al análisis y aplicación de técnicas estadísticas.
- **ANÁLISIS ESTADÍSTICO DE LA BASE DE DATOS:** Se les aplicó un análisis estadístico univariado (valores promedio, desviaciones estándar y varianzas) para determinar datos incongruentes, datos erróneos y datos atípicos (gráfico *boxplot*).

### 3.2. ETAPA 2 APLICACIÓN DE LA REGRESION LINEAL MULTIPLE

- **SELECCIÓN DE DATOS EN ESTADO PSEUDOESTACIONARIO.** Los datos representativos para esquemas de operación en estado pseudoestacionario, fueron localizados visualmente utilizando la gráfica de dispersión carga versus tiempo. Los datos en estado pseudoestacionario fueron escalados por medio de la media y la desviación estándar, según.

#### Ecuación 1. Ecuación de auto escalado.

$$x_{aut} = \frac{x - \bar{x}}{x_{desv}}$$

Siendo:

$x_{aut}$  : La variable auto-escalada.

$x$ : Variable a auto-escalar.

$\bar{x}$ : Media de la variable a auto-escalar.

$x_{desv}$ : Desviación estándar de la variable a auto-escalar.

- **APLICACIÓN DE LA REGRESION LINEAL MULTIPLE:** luego del análisis y acoplamiento de los datos se procede a aplicar RLM para la predicción de flujos

de Jet. La reducción de variables fue realizada por medio del método de eliminación hacia atrás, utilizando el criterio de información de Akaike (AIC).

- **ANÁLISIS POR K-MEANS Y COMPONENTES PRINCIPALES:** El siguiente análisis aplicado a los datos consistió en tomar los estados pseudo-estacionarios y determinar agrupaciones o sub-grupos por medio del método *k-means* y el análisis de componentes principales. Los diferentes subgrupos fueron identificados con los modos de operación de la unidad.
- **REGRESION PARA CADA AGRUPACIÓN POR RLM Y PLS:** Luego de elegir la cantidad de grupos de los *k-means* más adecuadas, a cada subgrupo se le aplican regresiones por RLM y PLS (*partial least square regression*).

### 3.3. ETAPA 3 MEJORA MODELOS

- En esta etapa final las agrupaciones identificadas por *k-means* y PCA fueron subdivididas para mejorar los desempeños de los modelos de predicción RLM.

## **4. ANÁLISIS DE RESULTADOS**

### **4.1. DATOS**

Los datos históricos de proceso de la unidad de destilación U-200 fueron suministrados por la GRB. La ventana de operación de los datos históricos consideró desde enero del año 2013 hasta septiembre del 2016 reportados cada hora para 25 variables, lo cual conduce a un total de 30057 muestras totales. La base de datos fue depurada para eliminar muestras con datos faltantes debido a lecturas erróneas en los sensores de lectura. Con esto la base de datos fue reducida en un 3%. (Ver anexo A).

### **4.2. ANÁLISIS ESTADÍSTICO**

Se realizó un análisis de los datos recopilados para cada variable, hallando así estadísticos importantes como: los valores de los cuartiles, valor máximo y mínimo, desviación estándar, media, mediana, moda y varianza, diagramas de caja y bigotes (anexo B); en éste último se indican los valores atípicos contenidos en el conjunto de datos lo cual es una herramienta clave para identificarlos. Los valores atípicos representan desviaciones muy grandes para el conjunto de datos a tratar debido a que pueden ser datos tomados en funcionamiento anormal de los equipos o posibles errores de medición. En total fueron eliminados 3410 datos atípicos reduciendo la matriz a 25647 muestras.

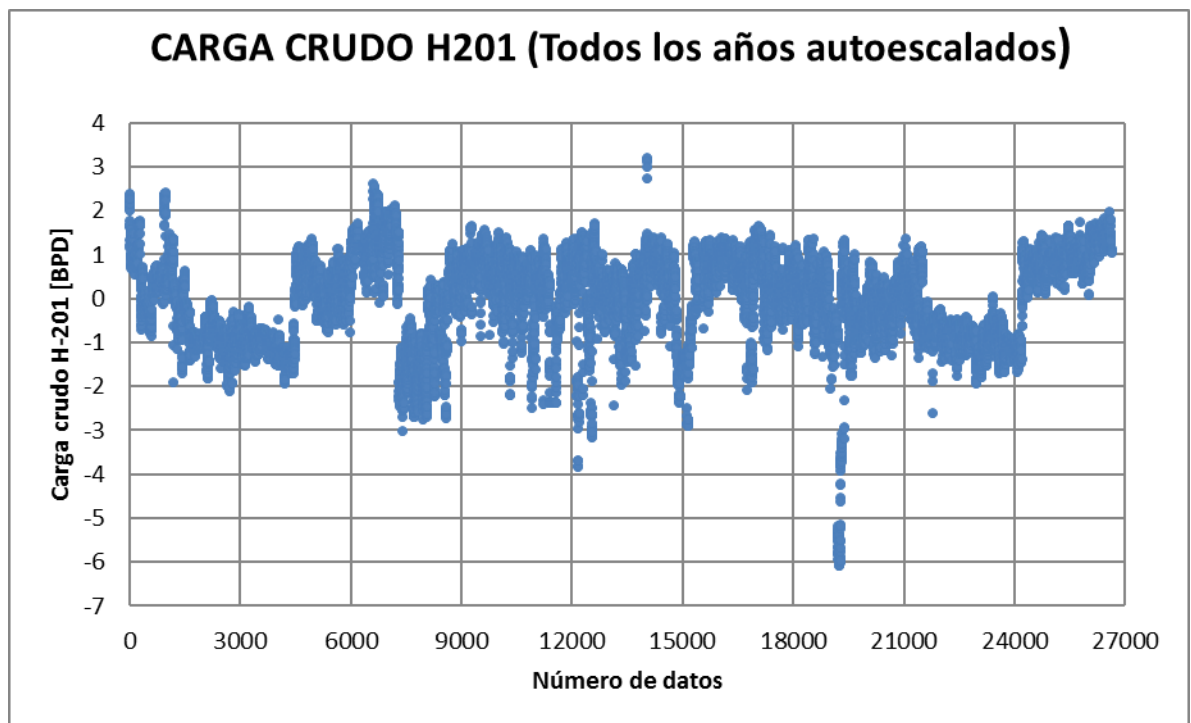
Con la matriz sin datos atípicos, se identificaron 6 estados pseudo-estacionarios; la Figura 4 ilustra las tendencias en la dispersión de la carga al horno H-201. La Tabla

1 presenta el número de estados pseudo-estacionarios por año en la ventana de operación considerada.

**Tabla 1. Tabla resultante de la cantidad de datos de los estados estacionarios.**

| Estados  | 2013-1 | 2013-2 | 2014-1 | 2015-1 | 2016-1 | 2016-2 |
|----------|--------|--------|--------|--------|--------|--------|
| Muestras | 2901   | 2501   | 5901   | 5601   | 2751   | 2373   |

**Figura 3. Grafica de auto-escalado de todos los datos sin atípicos.**



#### **4.3. REGRESIONES EN LOS SEIS ESTADOS PSEUDO-ESTACIONARIOS**

A continuación se presentan las regresiones obtenidas utilizando el método de eliminación hacia atrás (*backward elimination*) en los seis estados pseudo-estacionarios identificados por medio del análisis de la Figura 4; en las regresiones x18 corresponde al flujo de Jet producido.

**Ecuación 2. Regresión para todos los datos.**

$$X_{18} = (9.568x_1 - 2.430x_2 + 9.834x_3 + 1.400x_4 - 1.652x_5 + 4.935x_6 + 7.194x_7 \\ + 7.988x_8 - 10.36x_9 + 3.966x_{10} + 30.98x_{11} + 65.74x_{12} + 1.630x_{13} \\ + 18.88x_{14} - 1.839x_{15} - 60.74x_{16} + 3.502x_{17} - 12.43x_{22} + 1.107x_{23} \\ - 23.96x_{24} + 26.76x_{47} + 1.405x_{49} - 24.75x_{50} + 67.19x_{53}) * 10^{-2}$$

**Ecuación 3. Regresión del estado 2013-1.**

$$x_{18} = (3.421x_1 - 0.2296x_2 + 2.784x_3 + 14.64x_4 - 63.37x_5 + 84.57x_6 - 34.16x_7 \\ - 2.390x_8 + 5.339x_9 + 0.3046x_{10} + 0.9274x_{11} + 9.565x_{12} - 5.507x_{13} \\ - 13.62x_{14} + 2.856x_{15} + 0.9457x_{17} - 3.649x_{22} - 1.450x_{23} - 2.255x_{24} \\ + 15.21x_{47} + 9.744x_{49} - 4.081x_{50} - 0.3256x_{53}) * 10^{-1}$$

**Ecuación 4. Regresión del estado 2013-2.**

$$x_{18} = (1.568 x_1 + 2.905 x_2 + 2.951 x_3 + 6.619 x_4 - 12.17 x_5 + 19.32 x_6 - 13.52 x_7 \\ + 9.040 x_8 + 1.694 x_9 + 1.697 x_{10} + 3.932 x_{12} - 4.900 x_{13} \\ + 2.998 x_{14} + 2.320 x_{15} + 0.8554 x_{16} + 0.3658 x_{17} - 2.192 x_{22} \\ + 0.3144 x_{23} + 0.6153 x_{24} + 7.452 x_{49} - 6.996 x_{50}) * 10^{-1}$$

**Ecuación 5. Regresión del estado 2014-1.**

$$x_{18} = (-0.8292 x_1 + 0.4596x_2 - 1.898x_3 - 3.370x_5 + 2.524x_7 + 0.9518x_8 \\ + 3.571x_{10} + 0.7877x_{11} - 0.6640x_{13} - 0.6703x_{14} + 2.666x_{15} \\ + 2.870x_{16} + 1.520x_{17} - 5.110x_{22} - 3.034x_{23} - 0.2510x_{24} \\ + 2.124x_{49} - 3.333x_{50}) * 10^{-1}$$

**Ecuación 6. Regresión del estado 2015-1**

$$x_{18} = (0.6398 x_1 + 1.297 x_2 + 1.762 x_4 - 2.257 x_5 + 1.402 x_6 - 1.021 x_8 \\ - 0.2558 x_9 + 3.006 x_{10} - 0.2870 x_{11} - 0.5005 x_{13} - 22.99 x_{16} \\ - 0.5963 x_{17} - 2.474 x_{22} - 1.287 x_{24} - 1.469 x_{47} - 0.6880 x_{49} \\ + 22.24 x_{53}) * 10^{-1}$$

**Ecuación 7. Regresión del estado 2016-1.**

$$x_{18} = (5.460x_1 + 0.4846x_2 + 8.097x_3 + 22.26x_4 - 73.02x_5 + 55.33x_6 - 5.385x_9 + 4.587x_{10} - 0.6708x_{11} + 4.251x_{12} + 0.5763x_{13} - 25.76x_{14} + 4.406x_{15} + 18.93x_{16} - 1.166x_{17} + 1.782x_{22} - 1.429x_{23} - 2.128x_{24} + 30.95x_{47} - 0.4780x_{49} - 6.847x_{50} - 19.31x_{53}) * 10^{-1}$$

**Ecuación 8. Regresión del estado 2016-2.**

$$x_{18} = (0.6775x_1 - 9.269x_2 - 5.694x_3 - 4.899x_9 + 3.541x_{10} + 0.8178x_{11} - 12.59x_{12} + 2.528x_{13} - 10.27x_{14} + 0.7062x_{15} + 2.777x_{16} + 1.049x_{17} - 1.406x_{22} - 1.283x_{23} - 1.108x_{24} - 27.64x_{49} + 24.74x_{50}) * 10^{-1}$$

Según la Tabla 2, los desempeños de las regresiones en cada estado pseudo-estacionario son menores que el desempeño mostrado con el modelo, considerando todos los estados pseudo-estacionarios en la regresión.

**Tabla 2. Reporte de los R ajustados.**

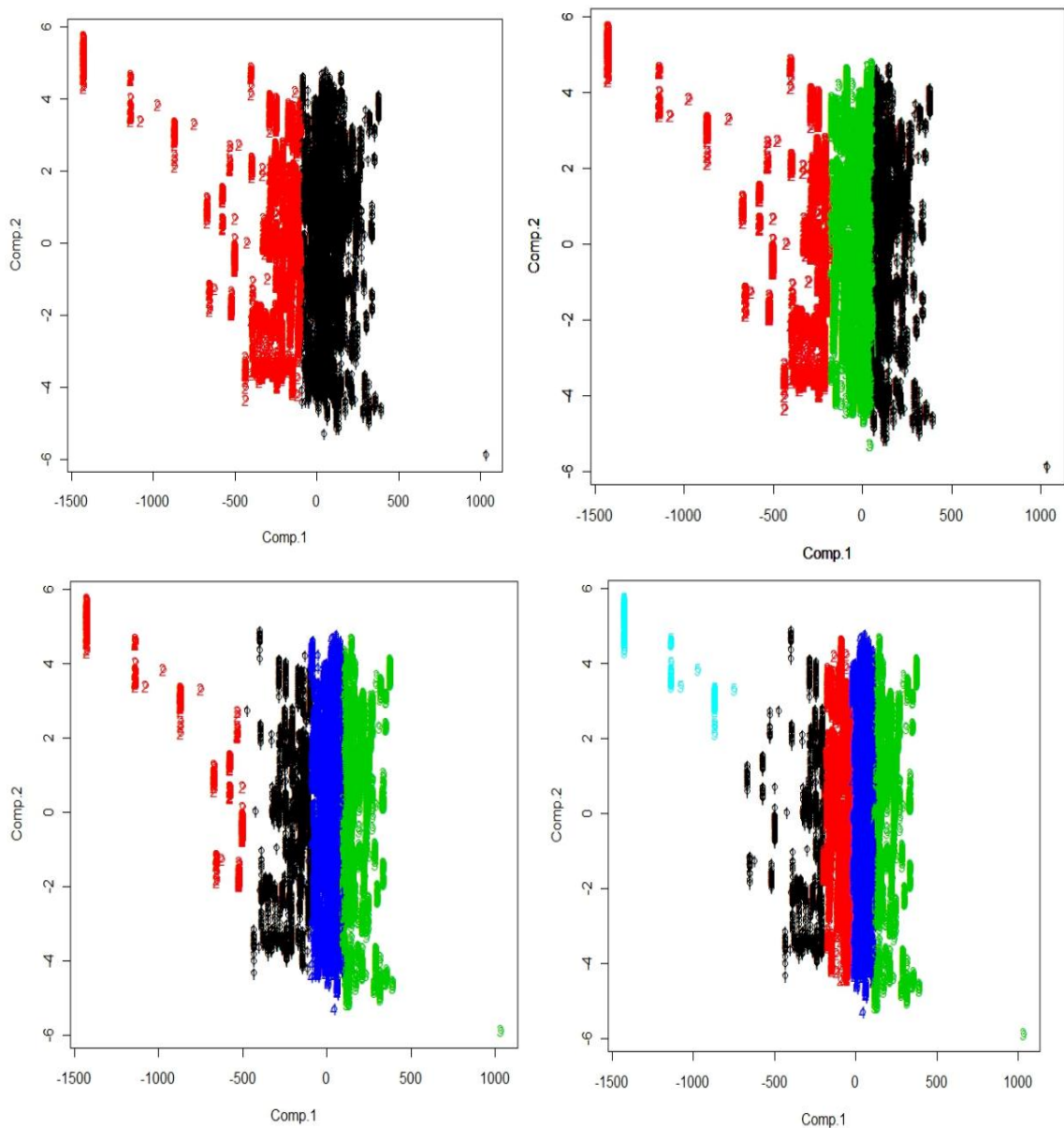
| ESTADO         | TODOS  | 2013-1 | 2013-2 | 2014-1 | 2015-1 | 2016-1 | 2016-2 |
|----------------|--------|--------|--------|--------|--------|--------|--------|
| R <sup>2</sup> | 0.7823 | 0.677  | 0.7686 | 0.5135 | 0.4854 | 0.6782 | 0.5087 |

Para mejorar los desempeños, se incluyeron factores de interacción cuadráticos, obteniendo un R<sup>2</sup> de 0.8864 para todos los datos (Ver anexo F). La mejora obtenida con la inclusión de términos cuadráticos conlleva a soportar el modelo para aplicaciones semi-cuantitativas; sin embargo, su aplicación está restringida por el número de entradas del modelo el cual contiene 61 variables estando en ellas solo 22 de las iniciales.

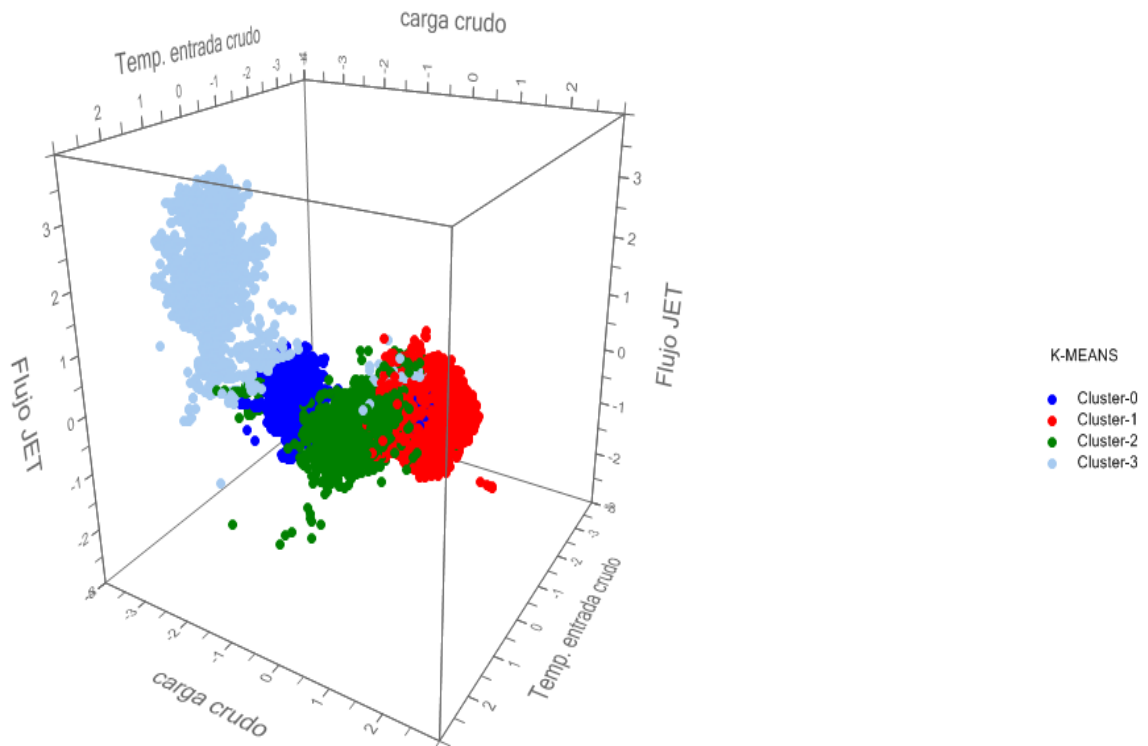
#### 4.4. ANÁLISIS POR K-MEANS Y COMPONENTES PRINCIPALES

El proceso inició con la aplicación del método de agrupamiento *k-means*, dividiendo el conjunto de 22028 muestras en 2, 3, 4 y 5 subgrupos (Ver figuras 4 y 5). A partir de las figuras 5 y 6, cuatro grupos fueron considerados para realizar nuevas regresiones RLM para la predicción del flujo de salida de Jet (Tabla 3).

Figura 4. Gráfico de k-means a 2, 3, 4 y 5 grupos.



**Figura 5. Distribución de 4 k-means.**



**Tabla 3. Modificación de R eliminando variables**

| GRUPO        | VARIABLES | R      | RSE    |
|--------------|-----------|--------|--------|
| <b>TODOS</b> | 24        | 0.7843 | 0.4647 |
| <b>1</b>     | 24        | 0.8277 | 0.4155 |
| <b>2</b>     | 16        | 0.9706 | 0.1755 |
| <b>3</b>     | 24        | 0.7199 | 0.5305 |
| <b>4</b>     | 21        | 0.8582 | 0.3776 |

Según la Tabla 3, los desempeños de las regresiones RLM de predicción para las agrupaciones aumentan respecto a los desempeños de las regresiones ajustadas para los seis estados pseudo-estacionarios, basados en la carga a la unidad. Esta mejora en el desempeño es consecuencia de la clasificación en 4 grupos por *k-mean* y PCA (*i.e.* clasificación de tipo multivariable). Asimismo, las regresiones para los 4 grupos exhiben mejor desempeño que la regresión considerando todos los datos ( $R^2=0.7843$ ). A pesar en la mejora obtenida con las agrupaciones y RLM, las

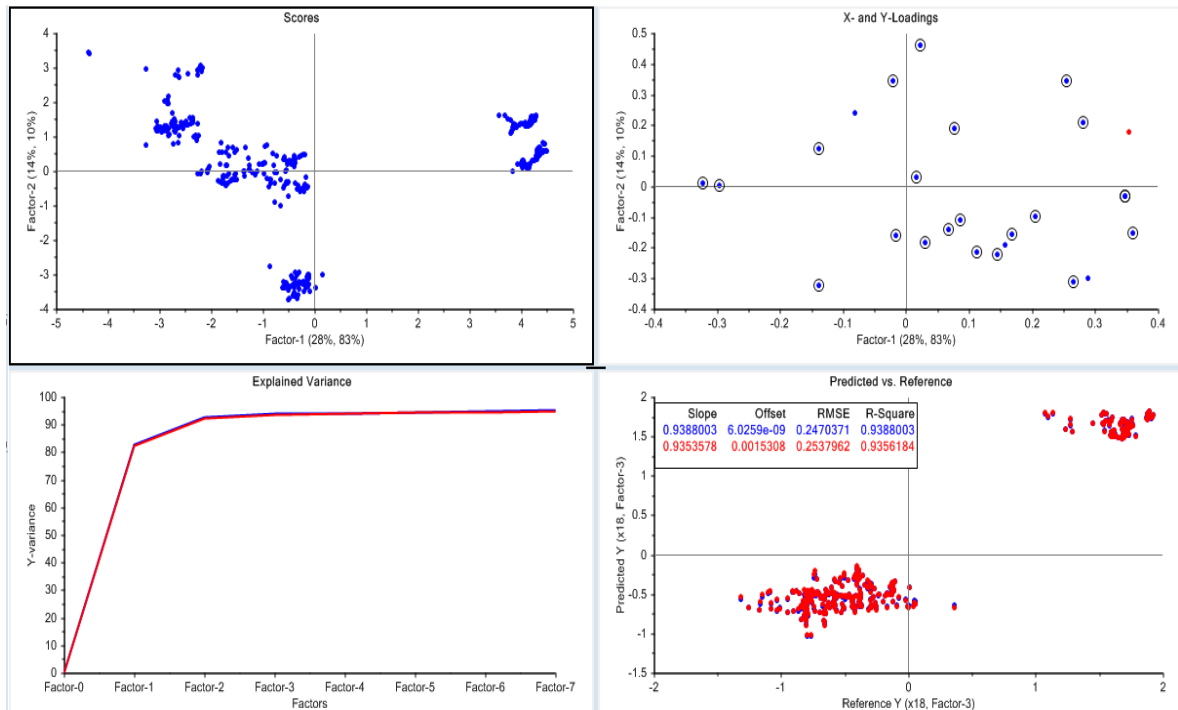
regresiones obtenidas no soportan su aplicación para un análisis cuantitativo del proceso.

Con lo anterior, la regresión por PLS fue efectuada en cada uno de las agrupaciones. La Tabla 4 presenta los resultados de  $R^2$  y RMSE para las regresiones PLS en cada agrupación. Según esta tabla, el mejor desempeño se obtuvo en el 2 modelo reportado. Asimismo, cada ajuste por PLS exhibe una cantidad de datos atípicos que sugieren la división de las agrupaciones (Figuras 6 y 7).

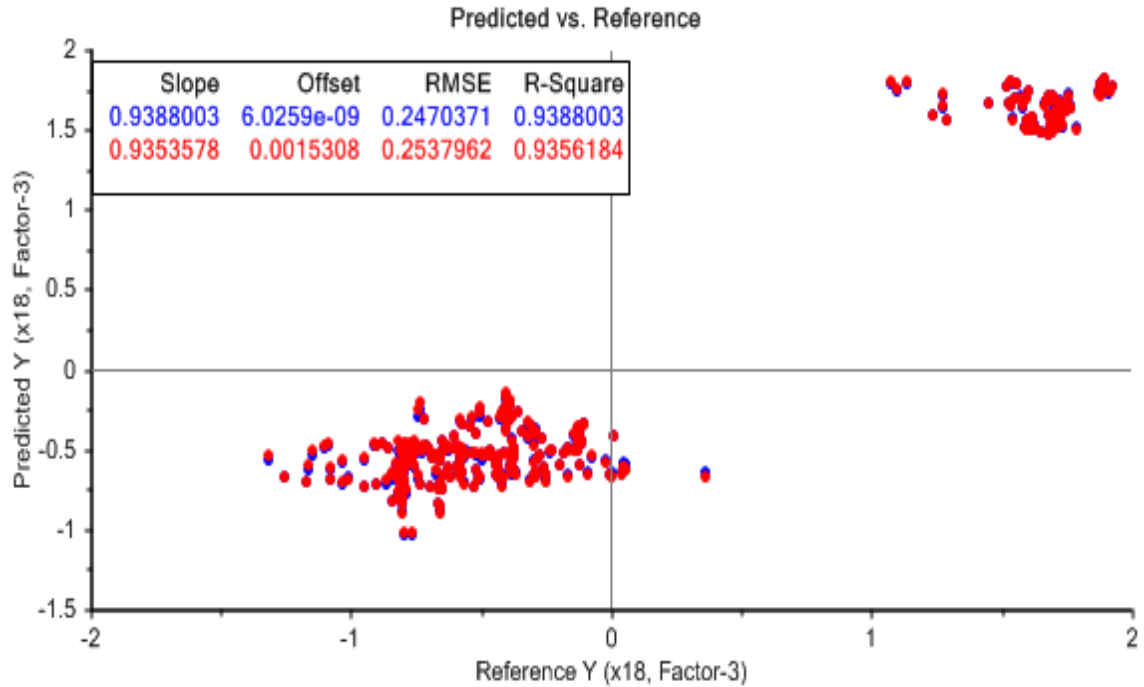
**Tabla 4. Reporte por medio de PLS a los estados.**

| PLS | R      | RMSE   |
|-----|--------|--------|
| 1   | 0.8013 | 0.4458 |
| 2   | 0.9388 | 0.247  |
| 3   | 0.6825 | 0.5635 |
| 4   | 0.8013 | 0.4458 |

**Figura 6. PLS del grupo 2.**



**Figura 7. Predicción del grupo 2 por PLS.**



Bajo esta nueva perspectiva, cada agrupación fue subdividida en 2. En cada subdivisión se aplicó de nuevo el proceso de regresión por RLM y por PLS. La Tabla 5 presenta los desempeños obtenidos con la regresión RLM aplicada a cada subdivisión en las agrupaciones. Según esta tabla, las subdivisiones 1 para el grupo 1, 2 para el grupo 2, 2 para el grupo 3 y 2 para el grupo 4, presentan los mejores desempeños para la predicción de Jet. El análisis de la dispersión de los errores condujo al mejoramiento de los desempeños en las regresiones anteriores (Tabla 6)

**Tabla 5. Tabla de regresiones de sub-grupos**

| grupo | Sub | # datos | #variables | R      | RSE    |
|-------|-----|---------|------------|--------|--------|
| 1     | 1   | 4208    | 24         | 0.929  | 0.2671 |
| 1     | 2   | 8366    | 24         | 0.4042 | 0.773  |
| 2     | 1   | 61      | 3          | 0.6505 | 0.6066 |
| 2     | 2   | 294     | 12         | 0.9744 | 0.1633 |
| 3     | 1   | 3093    | 17         | 0.5012 | 0.7082 |
| 3     | 2   | 1868    | 17         | 0.7838 | 0.4671 |
| 4     | 1   | 2492    | 20         | 0.4261 | 0.7606 |
| 4     | 2   | 1426    | 19         | 0.9167 | 0.2891 |

**Tabla 6. Eliminación de datos atípicos en las subdivisiones.**

| sub-grupo | # datos | # var | R             | RSE    | eliminación de datos |         |       |               |        |
|-----------|---------|-------|---------------|--------|----------------------|---------|-------|---------------|--------|
|           |         |       |               |        | sub-grupo            | # datos | # var | R             | RSE    |
| 1.1       | 4208    | 24    | <b>0.929</b>  | 0.2671 | 1.1                  | 4031    | 20    | <b>0.9486</b> | 0.2296 |
| 2.2       | 294     | 14    | <b>0.9744</b> | 0.1633 | 2.2                  | 249     | 14    | <b>0.9905</b> | 0.1026 |
| 3.2       | 1868    | 17    | <b>0.7838</b> | 0.4671 | 3.2                  | 1614    | 11    | <b>0.8587</b> | 0.2984 |
| 4.2       | 1426    | 19    | <b>0.9167</b> | 0.2891 | 4.2                  | 1284    | 17    | <b>0.9506</b> | 0.2186 |

La mejora reportada en la Tabla 6 para los desempeños de las regresiones conlleva a la afirmación de que los datos históricos contienen diferentes estados pseudo-estacionarios; la eliminación de datos con errores y datos en estado dinámico realizada por medio de análisis multivariado condujo a la definición de 4 estados pseudo-estacionarios contenidos en un total de 8406 muestras.

#### **4.5. MODOS OPERACIONALES**

Según los valores de temperatura para los cuatro estados pseudo-estacionarios, se puede sugerir que las subdivisiones 1.1 y 3.2 corresponden a la operación de la U-200 en modo viscorreductora (Ecuaciones 9 y 10). De otro lado, los estados 2.2 y 4.2 presentaron en general temperaturas menores a las del modo viscorreductora, con lo cual es posible sugerir a las subdivisiones 2.2 y 4.2 (ecuaciones 11 y 12)

como representativas del modo operacional crudo en la U-200. Los modos operacionales también pueden ser diferenciados por los valores de los flujos de salida de la torre de destilación; muestras con mayor flujo de fondos resulta característico para el modo crudo, mientras que muestras con mayores flujos de productos livianos (respecto a la carga) resulta característico para el modo viscorreductora.

**Ecuación 9. Regresión del estado 1.1.**

$$\begin{aligned}
 x_{18} = & 0.0122 * x_2 + 0.1812 * x_3 + 0.3029 * x_4 - 3.6545 * x_5 + 6.4348 * x_6 - 2.6623 \\
 & * x_7 - 0.1972 * x_8 + 0.1635 * x_9 - 0.8893 * x_{10} + 0.3299 * x_{11} \\
 & - 1.7401 * x_{14} - 9.0502 * x_{15} + 9.8901 * x_{16} - 0.0420 * x_{22} - 0.1229 \\
 & * x_{24} + 1.2070 * x_{47} - 0.6999 * x_{49} + 0.7606 * x_{50} - 1.64 * x_{53} \\
 & - 13.6586 * x_{55}
 \end{aligned}$$

**Ecuación 10. Regresión del estado 2.2.**

$$\begin{aligned}
 x_{18} = & 2.2006 * x_1 - 0.5722 * x_2 + 2.3463 * x_3 + 0.3898 * x_4 - 0.3820 * x_9 + 0.9082 \\
 & * x_{10} - 0.2997 * x_{11} - 0.6871 * x_{12} - 0.9454 * x_{14} + 0.1552 * x_{22} \\
 & - 0.4181 * x_{23} + 1.2707 * x_{47} - 0.8330 * x_{49} - 0.3781 * x_{50}
 \end{aligned}$$

**Ecuación 11. Regresión del estado 3.2.**

$$\begin{aligned}
 x_{18} = & -0.0481 * x_1 + 0.0583 * x_2 - 0.1930 * x_6 + 0.2979 * x_7 + 1.1964 * x_9 - 2.1272 \\
 & * x_{10} + 0.1320 * x_{11} + 2.5362 * x_{12} - 1.3660 * x_{13} - 0.1084 * x_{24} \\
 & + 3.4959 * x_{49}
 \end{aligned}$$

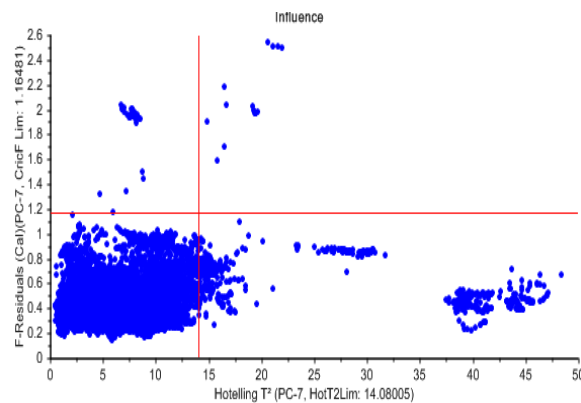
**Ecuación 12. Regresión del estado 4.2.**

$$\begin{aligned}
 x_{18} = & -0.1559 * x_1 + 0.1409 * x_4 - 0.5967 * x_5 + 0.5787 * x_6 + 0.3215 * x_9 \\
 & - 0.9374 * x_{10} + 0.0560 * x_{11} + 0.5239 * x_{12} - 0.1943 * x_{13} \\
 & - 0.5566 * x_{14} - 19.9953 * x_{15} + 41.3205 * x_{16} + 0.0821 * x_{22} \\
 & - 0.0953 * x_{24} + 0.5841 * x_{50} - 6.3917 * x_{53} - 40.7357 * x_{55}
 \end{aligned}$$

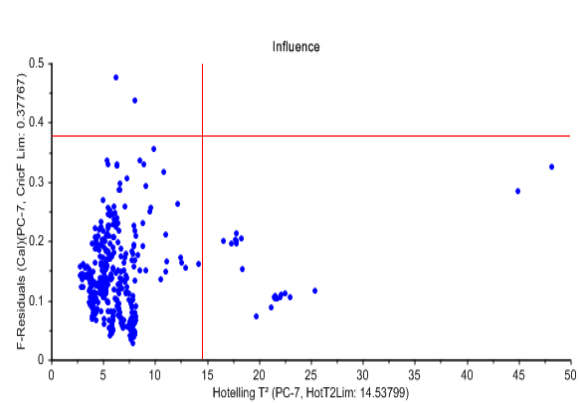
**4.5.1. Hottelling.** Al realizar un ajuste a cada sub-grupo se observó que aún era posible mejorar el  $R^2$  ya que no se había llegado al óptimo por lo cual se procedió a analizar los datos aislados de cada conjunto lo que se reporta en las siguientes gráficas (ver figuras 8 y 9), éstos estadísticos de resumen permiten mejorar el modelo mediante la eliminación de datos aislados.

**Figura 8. Hotelling de grupos 1 y 2.**

Grupo 1

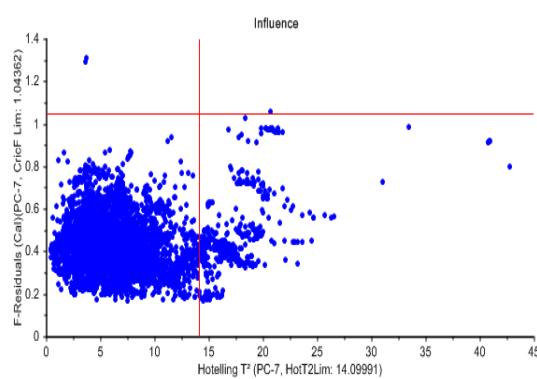


Grupo 2

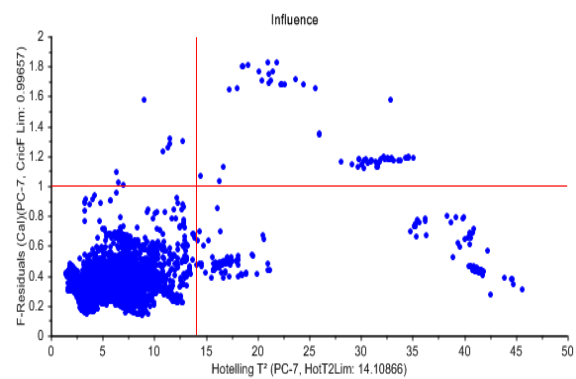


**Figura 9. Hotelling de grupos 3 y 4.**

Grupo 3



Grupo 4



Resumiendo desde el riguroso tratamiento aplicado a los DHP y el análisis a los estados pseudo-estacionarios de los grupos obtenidos mediante el análisis clúster

de operación de la torre de la U-200, los resultados obtenidos se tomaron los 4 mejores modelos validados que representan estados operacionales de la torre en su modo viscorreductor y crudo de acuerdo al rango de temperaturas y flujos de la torre, representando estos el mejor desempeño ( $R^2$ ).

## 5. CONCLUSIONES

Según los resultados obtenidos con la metodología propuesta en el presente trabajo de grado, los datos históricos de proceso de la U-200 pueden ser depurados respecto a muestras en estado dinámico, utilizando una discriminación inicial por medio de la dispersión de la carga de entrada a la unidad. En una segunda etapa, la depuración continua con estadística pmultivariada con *k-means* y PCA; la tercera etapa involucra de nuevo el uso de *k-means* con 2 agrupaciones y PCA. Lo anterior condujo a la proposición de regresiones con suficientes desempeños para la predicción de Jet a la salida de la torre en las subdivisiones 1.1, 2.2 y 4.2.

Los modos de operación obtenidos por medio de la metodología propuesta en el presente trabajo fueron identificados con la operación de la unidad U-200 en modo viscorreductora y en modo crudo, por medio del análisis de las temperaturas y los flujos de destilados.

## **6. RECOMENDACIONES**

Debido a la cantidad de datos en estado dinámico, se recomienda implementar otros métodos de agrupamiento para verificar los estados pseudo-estacionarios identificados.

Según los desempeños de las regresiones, se recomienda continuar con el proceso de regresión para otros productos de torres de destilación de crudos.

Tener en cuenta los rangos operativos válidos para cada modelo.

## CITAS BIBLIOGRÁFICAS

- [1] J. G. Londoño, «El tiempo,» 24 febrero 2002. [En línea]. Available: <http://www.eltiempo.com/archivo/documento/MAM-1379739>. [Último acceso: 25 mayo 2017].
- [2] R. Gutiérrez Zapata, «EL ESPECTADOR- ECONOMÍA,» 12 Mayo 2008. [En línea]. Available: <http://www.elespectador.com/impreso/negocios/articuloimpreso-cruda-realidad-los-costos-de-aerolineas>. [Último acceso: 10 marzo 2017].
- [3] C. A. Gutiérrez B. y A. F. Londoño B., «Dirección de segmentos Empresas y Gobierno,» Grupo Bancolombia, 9 septiembre 2016. [En línea]. Available: <https://www.grupobancolombia.com/wps/portal/empresas/capital-inteligente/actualidad-economica-sectorial/cual-es-la-importancia-del-sector-combustibles-en-colombia/>. [Último acceso: 20 mayo 2017].
- [4] R. Económica, «Vanguardia Nacional,» VANGUARDIA, 15 junio 2011. [En línea]. Available: <http://www.vanguardia.com/historico/108766-combustible-para-aviones-es-111-mas-carro-en-colombia> . [Último acceso: 22 junio 2017].
- [5] M. Marquez D, «OLADE/CEPAL/GTZ “Energía y desarrollo sustentable en América Latina y el Caribe”». Santiago de Chile Patente p-14, Diciembre 2000.
- [6] M. Salazar Cachopo, «ANÁLISIS DE DATO DE CONSUMO ENERGETICO DA LA UNIDAD DE REFINACION DE CRUDOS U-200 DEL COMPLEJO INDUSTRIAL BARRANCABERMEJA :REGRESION LINEAL MULTIPLE». Bucaramanaga 2015.

- [7] R. I. Rubio Lizcano, «LINEA BASE ENERGETICA PARA LA UNIDAD U-200 DE LA GRB: ANÁLISIS Y REGRESIÓN UTILIZANDO ESTADÍSTICA UNIVARIADA». Barranacabermeja Enero 2017.
- [8] C. López, C. Mahecha, L. Acevedo y J. Hoyos, «ANÁLISIS JERÁRQUICO DE VARIABLES DE OPERACIÓN DE UNA TORRE DE DESTILACIÓN DE CRUDO USANDO DATOS DE PLANTA, MODELOS DE SIMULACIÓN Y DISEÑO DE EXPERIMENTOS». Bucaramanaga 22 Abril 2008.
- [9] The international council on clean transportation (ICCT), «Introducción a la refinación del petróleo y producción de gasolina y diesel con contenidos ultra bajo de azufre.» 2011.
- [10] J. C. Campos Avella, E. Lora Figueroa y L. Meriño Stand, «Ahorro de energía en la industria del refino y petroquímica,» Unidad de planeación minero energética de Colombia, Atlántico.
- [11] C. P. Company, «Aviation Fuels technical review,» 2007.
- [12] G. RUIZ MONCADA, Análisis y optimización del sistema de entrega de propileno en el departamento de materias primas de la gerencia refinería de Barrancabermeja., Universidad industrial de Santander: Tesis de maestría en ingeniería química, 2015.
- [13] G. Reinsel y V. Raja P., «Multivariate Reduced-Rank Regression: Theory and Applications,» Springer-Verlag New York, 2013-08-01, pp. p-1.
- [14] L. L. Smith, «Principal Components Analysis,» 26 febrero 2002 . p-12. [En línea]. Available: [http://www.cs.otago.ac.nz/cosc453/student\\_tutorials/principal\\_components.pdf](http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf) f. [Último acceso: 20 febrero 2017].
- [15] A. Hervé, «Partial Least Square Regression-PLS-Regression,» *Enciclopedia of Measurement and statistics*, pp. 1-2, 2007.

- [16] j. Gorgas Garcia , N. Cardiel Lopez y J. Zamorano Calvo, 17 febrero 2011.  
] [En línea]. Available:  
[http://pendientedemigracion.ucm.es/info/Astrof/ESTADISTICA/libro\\_GCZ2009.pdf](http://pendientedemigracion.ucm.es/info/Astrof/ESTADISTICA/libro_GCZ2009.pdf). [Último acceso: 10 marzo 2017].
- [17] P. Gonzales, «Desarrollo de un programa para estudiar el comportamiento de una columna de fraccionamiento etano/etileno de una planta de oleofinas,» 2009.
- [18] J. Fernández, «formulacion de un modelo matematico para estimar el volumen recobrado de los principales productos en torres de destilación atmosféricas,» 2012.
- [19] J. Devore, «Probabilidad y estadística para ingeniería y ciencias,» 2007.  
]
- [20] A. A. Cerutti, La refinación del petróleo. Tomo II, Cuyo: Facultad de ciencias aplicadas a la industria, Universidad nacional de cuyo, 2001.
- [21] R. R. Ricardo Paredes, «Petroleum refining,» Atlantic International University, Honolulu,Hawai, 2011.
- [22] C. Huerga Castro, P. Blanco Alonso y J. Abad Gonzales, Aplicación de los gráficos de control en análisis de calidad., Pecunia: Revista de la facultad de ciencias económicas y empresariales Universidad de León, vol 1, 2005.
- [23] J. Mina y C. Verde, «Deteccion de fallas usando análisis de componentes principales,» instituto de ingeniería UNAM.
- [24] A. C. RENCHER, Methods of Multivariate Analysis, Canadá: Willey Interscience, 2002.
- [25] C. PEREZ LOPEZ, técnicas de Análisis Multivariante de Datos, Aplicaciones en SPSS, Madrid: Pearson, 2004.

## BIBLIOGRAFÍA

J. G. Londoño, «El tiempo,» 24 febrero 2002. [En línea]. Available: <http://www.eltiempo.com/archivo/documento/MAM-1379739>. [Último acceso: 25 mayo 2017].

R. Gutiérrez Zapata, «EL ESPECTADOR- ECONOMÍA,» 12 Mayo 2008. [En línea]. Available: <http://www.elespectador.com/impreso/negocios/articuloimpreso-cruda-realidad-los-costos-de-aerolineas>. [Último acceso: 10 marzo 2017].

C. A. Gutiérrez B. y A. F. Londoño B., «Dirección de segmentos Empresas y Gobierno,» Grupo Bancolombia, 9 septiembre 2016. [En línea]. Available: <https://www.grupobancolombia.com/wps/portal/empresas/capital-inteligente/actualidad-economica-sectorial/cual-es-la-importancia-del-sector-combustibles-en-colombia/>. [Último acceso: 20 mayo 2017].

R. Económica, «Vanguardia Nacional,» VANGUARDIA, 15 junio 2011. [En línea]. Available: <http://www.vanguardia.com/historico/108766-combustible-para-aviones-es-111-mas-carro-en-colombia> . [Último acceso: 22 junio 2017].

M. Marquez D, «OLADE/CEPAL/GTZ “Energía y desarrollo sustentable en América Latina y el Caribe”». Santiago de Chile Patente p-14, Diciembre 2000.

M. Salazar Cachopo, «ANÁLISIS DE DATO DE CONSUMO ENERGETICO DA LA UNIDAD DE REFINACION DE CRUDOS U-200 DEL COMPLEJO INDUSTRIAL BARRANCABERMEJA :REGRESION LINEAL MULTIPLE». Bucaramanaga 2015.

R. I. Rubio Lizcano, «LINEA BASE ENERGETICA PARA LA UNIDAD U-200 DE LA GRB: ANÁLISIS Y REGRESIÓN UTILIZANDO ESTADÍSTICA UNIVARIADA». Barranacabermeja Enero 2017.

C. López, C. Mahecha, L. Acevedo y J. Hoyos, «ANÁLISIS JERÁRQUICO DE VARIABLES DE OPERACIÓN DE UNA TORRE DE DESTILACIÓN DE CRUDO USANDO DATOS DE PLANTA, MODELOS DE SIMULACIÓN Y DISEÑO DE EXPERIMENTOS». Bucaramanaga 22 Abril 2008.

The international council on clean transportation (ICCT), «Introducción a la refinación del petróleo y producción de gasolina y diesel con contenidos ultra bajo de azufre.,» 2011.

J. C. Campos Avella, E. Lora Figueroa y L. Meriño Stand, «Ahorro de energía en la industria del refino y petroquímica,» Unidad de planeación minero energética de Colombia, Atlántico.

C. P. Company, «Aviation Fuels technical review,» 2007.

G. RUIZ MONCADA, Análisis y optimización del sistema de entrega de propileno en el departamento de materias primas de la gerencia refinera de Barranacabermeja., Universidad industrial de Santander: Tesis de maestría en ingeniería química, 2015.

G. Reinsel y V. Raja P., «Multivariate Reduced-Rank Regression: Theory and Applications,» Springer-Verlag New York, 2013-08-01, pp. p-1.

L. L. Smith, «Principal Components Analysis,» 26 febrero 2002 . p-12. [En línea]. Available:  
[http://www.cs.otago.ac.nz/cosc453/student\\_tutorials/principal\\_components.pdf](http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf).  
[Último acceso: 20 febrero 2017].

A. Hervé, «Partial Least Square Regression-PLS-Regression,» Enciclopedia of Measurement and statistics, pp. 1-2, 2007.

j. Gorgas Garcia , N. Cardiel Lopez y J. Zamorano Calvo, 17 febrero 2011. [En línea]. Available:  
[http://pendientedemigracion.ucm.es/info/Astrof/ESTADISTICA/libro\\_GCZ2009.pdf](http://pendientedemigracion.ucm.es/info/Astrof/ESTADISTICA/libro_GCZ2009.pdf). [Último acceso: 10 marzo 2017].

P. Gonzales, «Desarrollo de un programa para estudiar el comportamiento de una columna de fraccionamiento etano/etileno de una planta de oleofinas,» 2009.

J. Fernández, «formulacion de un modelo matematico para estimar el volumen recobrado de los principales productos en torres de destilación atmosféricas,» 2012.

J. Devore, «Probabilidad y estadística para ingeniería y ciencias,» 2007.

A. A. Cerutti, La refinación del petróleo. Tomo II, Cuyo: Facultad de ciencias aplicadas a la industria, Universidad nacional de cuyo, 2001.

R. R. Ricardo Paredes, «Petroleum refining.,» Atlantic International University, Honolulu,Hawai, 2011.

C. Huerga Castro, P. Blanco Alonso y J. Abad Gonzales, Aplicación de los gráficos de control en análisis de calidad., Pecunia: Revista de la facultad de ciencias económicas y empresariales Universidad de León, vol 1, 2005.

J. Mina y C. Verde, «Deteccion de fallas usando análisis de componentes principales,» instituto de ingeniería UNAM.

A. C. RENCHER, *Methods of Multivariate Analysis*, Canadá: Willey Interscience, 2002.

C. PEREZ LOPEZ, *técnicas de Análisis Multivariante de Datos, Aplicaciones en SPSS*, Madrid: Pearson, 2004.

## ANEXOS

### Anexo A. Tabla de identificación de todas las variables iniciales.

**Tabla 7. Reporte de cada variable adquirida en los datos históricos de proceso.**

| VARIABLE | DENOMINACIÓN                                    | UNIDADE  | DEFINICIÓN                            |
|----------|---|----------|---------------------------------------|
| $X_0$    | <i>energía</i><br>$(carga\ H201 + H202 + T205)$ |          |                                       |
| $X_1$    | GRADAPI   | °API     | Gravedad API del crudo                |
| $X_2$    | BSW   | mL/1000m | Sedimentos y agua presentes en el     |
| $X_3$    | AZUFRE  | %        | Azufre en el crudo                    |
| $X_4$    | TBP10   | °F       | Temperatura verdadera de ebullición a |
| $X_5$    | TBP30   | °F       | Temperatura verdadera de ebullición a |
| $X_6$    | TBP50   | °F       | Temperatura verdadera de ebullición a |
| $X_7$    | TBP70   | °F       | Temperatura verdadera de ebullición a |
| $X_8$    | TBP90   | °F       | Temperatura verdadera de ebullición a |
| $X_9$    | LHV   |          | Gas combustible a topping             |
| $X_{10}$ | CARGA H-201                                     | BPD      | Carga crudo a Horno 201               |
| $X_{11}$ | GAS COMBUSTIBLE H201                            | KPC      | Gas combustible horno 201             |
| $X_{12}$ | T ENTRADA H201                                  | °F       | Temperatura de entrada carga crudo    |
| $X_{13}$ | T SALIDA H 201                                  | °F       | Temperatura de salida crudo Horno     |
| $X_{14}$ | P ENTRADA H-201                                 | PSI      | Presión entrada a H201                |
| $X_{15}$ | TCH H201  | °F       | Temperatura de chimenea Horno         |
| $X_{16}$ | ANALIZADOR O2 H201                              | %        | Temperatura zona flash torre 201      |
| $X_{17}$ | T FLASH T201                                    | °F       | Temperatura zona flash torre 201      |
| $X_{18}$ | FLUJO JET T201                                  | BPD      | Flujo JET torre 201                   |
| $X_{19}$ | FLUJO ACPM T201                                 | BPD      | Flujo ACPM torre 201                  |
| $X_{20}$ | FLUJO GASOLEO T201                              | BPD      | Flujo gasóleo torre 201               |
| $X_{21}$ | FONDOS T201                                     | BPD      | Fondos torre 201                      |
| $X_{22}$ | REFLUJO CIMA T201                               | BPD      | Reflujo de cima torre 201             |
| $X_{23}$ | REFLUJO MEDIO T201                              | BPD      | Reflujo medio torre 201               |
| $X_{24}$ | REFLUJO INFERIOR T201                           | BPD      | Reflujo inferior torre 201            |
| $X_{25}$ | CARGA H202                                      | BPD      | Carga crudo a horno 202               |
| $X_{26}$ | T ENTRADA H202                                  | °F       | Temperatura de entrada horno 202      |
| $X_{27}$ | T SALIDA H202                                   | °F       | Temperatura de salida Horno 202       |
| $X_{28}$ | PENTRADA H 202                                  | psi      | Presión entrada Horno 202             |

|                 |                        |                      |  |
|-----------------|------------------------|----------------------|--|
| X <sub>29</sub> | T CH H202              | °F                   | Temperatura de chimenea horno 202      |
| X <sub>30</sub> | ANALIZADOR O2 H-202    | %                    |  |
| X <sub>31</sub> | T zona flash T-204     | °F                   | Temperatura zona flash torre 204       |
| X <sub>32</sub> | Flujo Jet T-204        | BPD                  | Flujo JET torre 204                    |
| X <sub>33</sub> | Flujo ACPM T-204       | BPD                  | Flujo de ACPKMM torre 204              |
| X <sub>34</sub> | Flujo Gasoleo T-204    | BPD                  | Flujo gasóleo torre 2014               |
| X <sub>35</sub> | Fondos T-204           | BPD                  | Fondos torre 204                       |
| X <sub>36</sub> | REFLUJO CIMA T204      | BPD                  |  |
| X <sub>37</sub> | REFLUJO MEDIO T204     | BPD                  |  |
| X <sub>38</sub> | REFLUJO INFERIOR T204  | BPD                  |  |
| X <sub>39</sub> | CARGA T-205            | BPD                  |  |
| X <sub>40</sub> | Flujo Gasoleo 1 T-205  | BPD                  |  |
| X <sub>41</sub> | Flujo Gasoleo 2 T-205  | BPD                  |  |
| X <sub>42</sub> | REFLUJO INFERIOR T-205 | BPD                  | Reciclo del reflujo inferior torre 205 |
| X <sub>43</sub> | Fondos T-205           | BPD                  |  |
| X <sub>44</sub> | REFLUJO CIMA T205      | BPD                  |  |
| X <sub>45</sub> | REFLUJO MEDIO T205     | BPD                  |  |
| X <sub>46</sub> | REFLUJO INFERIOR T205  | BPD                  |  |
| X <sub>47</sub> | Delta P H201           | Pentrada -<br>Pflash |  |
| X <sub>48</sub> | Delta P H202           | Pentrada -<br>Pflash |  |
| X <sub>49</sub> | Carga crudo*DT/PC      |                      |  |
| X <sub>50</sub> | Carga crudo*DP/PC      |                      |  |
| X <sub>51</sub> | Carga crudo*DT/PC      |                      |  |
| X <sub>52</sub> | Carga crudo*DP/PC      |                      |  |
| X <sub>53</sub> | Temp. Chimenea * O2    |                      |  |
| X <sub>54</sub> | Temp. Chimenea * O2    |                      |  |
| X <sub>55</sub> | eficiencia             |                      | Eficiencia operativa T-201             |

**Anexo B. Sesgo y curtosis variables influyentes**

**Tabla 8. Reporte de sesgo y curtosis de las variables de T-201**

| VARIABLE INFLUYENTE | SESGO   | CURTOSIS |
|---------------------|---------|----------|
| $X_1$               | 0.354   | 0.344    |
| $X_2$               | 2.439   | 8.110    |
| $X_3$               | 0.467   | (0.410)  |
| $X_4$               | (0.225) | 0.898    |
| $X_5$               | (2.037) | 10.834   |
| $X_6$               | 0.818   | 1.622    |
| $X_7$               | 1.796   | 2.760    |
| $X_8$               | (0.079) | 0.316    |
| $X_9$               | 2.704   | 16.631   |
| $X_{10}$            | 1.460   | 1.153    |
| $X_{11}$            | (1.460) | 1.687    |
| $X_{12}$            | 0.798   | (0.225)  |
| $X_{13}$            | (3.262) | 35.378   |
| $X_{14}$            | 0.673   | (0.002)  |
| $X_{15}$            | 0.071   | (0.450)  |
| $X_{16}$            | 3.692   | 20.684   |
| $X_{17}$            | 2.302   | 5.964    |
| $X_{18}$            | 1.965   | 3.406    |
| $X_{22}$            | 0.542   | (0.549)  |
| $X_{23}$            | (3.873) | 19.699   |
| $X_{24}$            | (0.450) | 2.495    |
| $X_{47}$            | 0.692   | 0.036    |
| $X_{49}$            | (1.800) | 11.815   |
| $X_{50}$            | 0.047   | 0.493    |
| $X_{53}$            | 3.595   | 19.848   |
| $X_{55}$            | (0.078) | (0.454)  |

## Anexo C. Estadísticos de subgrupos.

**Tabla9. Estadísticos de importancia de los mejores subgrupos.**

| sub-grupo 1.1    |      |      |        |        |        |        |         |        |          |        |        |        |      |          |          |        |           |          |       |      |
|------------------|------|------|--------|--------|--------|--------|---------|--------|----------|--------|--------|--------|------|----------|----------|--------|-----------|----------|-------|------|
|                  | x2   | x3   | x4     | x5     | x6     | x7     | x8      | x9     | x10      | x11    | x14    | x15    | x16  | x22      | x24      | x47    | x49       | x50      | x53   | x55  |
| <b>minimo</b>    | 0.05 | 1    | 168.5  | 472.4  | 646.3  | 817.7  | 1045.1  | 902.75 | 24401.08 | 55.47  | 194.94 | 680.26 | 1.79 | 6117.01  | 1536.58  | 180.91 | -22479.88 | 4856.71  | 13.05 | 0.82 |
| <b>maximo</b>    | 0.6  | 1.2  | 349.3  | 587.3  | 773.1  | 1028.8 | 1281.1  | 1073.6 | 51727.47 | 147.58 | 308    | 787.95 | 7.15 | 37113.26 | 17971.97 | 291.59 | 19369.23  | 12561.08 | 51.95 | 0.87 |
| <b>media</b>     | 0.13 | 1.08 | 266.03 | 538.82 | 722.05 | 913.83 | 1164.69 | 975.29 | 39646.12 | 117.89 | 253.84 | 741.97 | 3.52 | 22295.48 | 13208.31 | 237.43 | 7738.86   | 9587.4   | 26.05 | 0.84 |
| <b>mediana</b>   | 0.1  | 1.1  | 284.3  | 549.7  | 735.5  | 929    | 1194.8  | 968.1  | 37514.34 | 124.5  | 249.17 | 746.96 | 3.39 | 22915.21 | 13166.61 | 231.99 | 7511.82   | 9503.27  | 25.17 | 0.84 |
| <b>cuartil 1</b> | 0.05 | 1    | 271.8  | 539.3  | 717.5  | 916.1  | 1161.5  | 942.3  | 35223.7  | 101.94 | 235.24 | 724.59 | 2.96 | 17645.23 | 11741.96 | 218.97 | 5800.93   | 8765.52  | 21.98 | 0.84 |
| <b>cuartil 3</b> | 0.2  | 1.1  | 284.3  | 549.7  | 735.5  | 929    | 1194.8  | 1001.4 | 45133.9  | 134.95 | 272.73 | 758.97 | 4    | 26830.37 | 14666.42 | 256.32 | 10148.64  | 10218.67 | 29.56 | 0.85 |

| sub-grupo 2.2    |       |      |      |       |         |          |        |        |        |          |          |        |         |         |
|------------------|-------|------|------|-------|---------|----------|--------|--------|--------|----------|----------|--------|---------|---------|
|                  | x1    | x2   | x3   | x4    | x9      | x10      | x11    | x12    | x14    | x22      | x23      | x47    | x49     | x50     |
| <b>minimo</b>    | 22.1  | 0.1  | 1    | 205.6 | 1453.8  | 31470.56 | 66.45  | 403.81 | 161.95 | 7736.65  | 19877    | 149.48 | 3297.23 | 2916.74 |
| <b>maximo</b>    | 24.4  | 0.2  | 1.2  | 284.6 | 2049.9  | 51479.94 | 151.67 | 561.96 | 275.74 | 35470.72 | 22058.21 | 258.55 | 9716.68 | 7163.82 |
| <b>media</b>     | 23.01 | 0.14 | 1.08 | 277.7 | 1590.27 | 41992.26 | 118.57 | 492.49 | 225.48 | 18380.37 | 21128.78 | 209.77 | 5851.35 | 5575.14 |
| <b>mediana</b>   | 22.7  | 0.1  | 1.1  | 284.3 | 1479.8  | 43359.9  | 128.19 | 502.73 | 223.93 | 14504.47 | 21486.8  | 208.71 | 6242.27 | 5942.88 |
| <b>cuartil 1</b> | 22.5  | 0.1  | 1    | 284.3 | 1470.8  | 36320.24 | 116.66 | 439.9  | 206.96 | 11713.99 | 20690.83 | 191.39 | 3870.11 | 5583.87 |
| <b>cuartil 3</b> | 23.2  | 0.2  | 1.1  | 284.3 | 1612.8  | 44690.71 | 133.45 | 553.33 | 259.44 | 26585.79 | 21569.14 | 241.84 | 6646.6  | 6217.44 |

| sub-grupo 3.2    |       |      |        |        |       |          |        |        |        |          |          |
|------------------|-------|------|--------|--------|-------|----------|--------|--------|--------|----------|----------|
|                  | x1    | x2   | x6     | x7     | x9    | x10      | x11    | x12    | x13    | x24      | x49      |
| <b>minimo</b>    | 20.3  | 0.05 | 594.8  | 811.1  | 621.3 | 29148.98 | 28.62  | 412.08 | 619.92 | 8266.54  | 5067.21  |
| <b>maximo</b>    | 24.5  | 0.5  | 744.1  | 955.1  | 959.9 | 50378.77 | 148.86 | 562.5  | 726.82 | 16686.29 | 17368.92 |
| <b>media</b>     | 22.76 | 0.11 | 721.58 | 911.23 | 843   | 39754.04 | 120.06 | 504.54 | 705.16 | 13527.98 | 9588.13  |
| <b>mediana</b>   | 22.7  | 0.1  | 735.5  | 929    | 864.3 | 38254.18 | 128.67 | 497.68 | 714.11 | 13540.26 | 8999.36  |
| <b>cuartil 1</b> | 22.2  | 0.1  | 717.5  | 896.6  | 834   | 35662.42 | 100.56 | 485.69 | 691.38 | 12418.81 | 8005.17  |
| <b>cuartil 3</b> | 23.6  | 0.1  | 735.5  | 929    | 882   | 45048.23 | 137.91 | 514.86 | 718.53 | 14944.97 | 11328.4  |

sub-grupo 4.2

|           | x1    | x4     | x5     | x6     | x9      | x10      | x11    | x12    | x13    | x14    | x15    | x16   | x22      | x24      | x50      | x53   | x55  |
|-----------|-------|--------|--------|--------|---------|----------|--------|--------|--------|--------|--------|-------|----------|----------|----------|-------|------|
| minimo    | 21.8  | 233    | 499.3  | 672.9  | 1075.7  | 28344.77 | 70.26  | 402.63 | 530.95 | 157.84 | 685.06 | 1.79  | 7474.67  | 5068.88  | 5461.7   | 12.38 | 0.83 |
| maximo    | 24.5  | 303.3  | 578.5  | 777.3  | 1382.78 | 52143.18 | 158.24 | 572.93 | 726.69 | 302.04 | 775.96 | 10.19 | 36498.39 | 16510.48 | 11057.25 | 74.29 | 0.89 |
| media     | 22.79 | 276.43 | 544.27 | 725.01 | 1175.26 | 42115.9  | 127.61 | 497.16 | 707.29 | 248.23 | 735.2  | 3.84  | 20596.48 | 11904.83 | 8281.84  | 28.23 | 0.85 |
| mediana   | 22.2  | 273.4  | 549.7  | 735.5  | 1169.2  | 44716.29 | 130.48 | 468.52 | 711.14 | 243.53 | 730.32 | 3.47  | 21261.31 | 11592.83 | 8230.76  | 25.66 | 0.84 |
| cuartil 1 | 22.1  | 271.8  | 529.5  | 697.1  | 1118.1  | 36328.83 | 117.7  | 453.52 | 700.28 | 231.44 | 717.05 | 2.96  | 13981.55 | 10976.27 | 7580.39  | 21.59 | 0.84 |
| cuartil 3 | 23.8  | 284.3  | 549.7  | 735.5  | 1218.4  | 47060.27 | 139.26 | 556.35 | 715.82 | 270.88 | 759.33 | 4.17  | 26328.08 | 12523.84 | 9024.35  | 30.83 | 0.85 |

## Anexo D. Validación, hotelling y distribución de residuales a subgrupos sin atípicos

Figura 10. Validación por hotelling grupo 1.1.

### Subgrupo 1.1

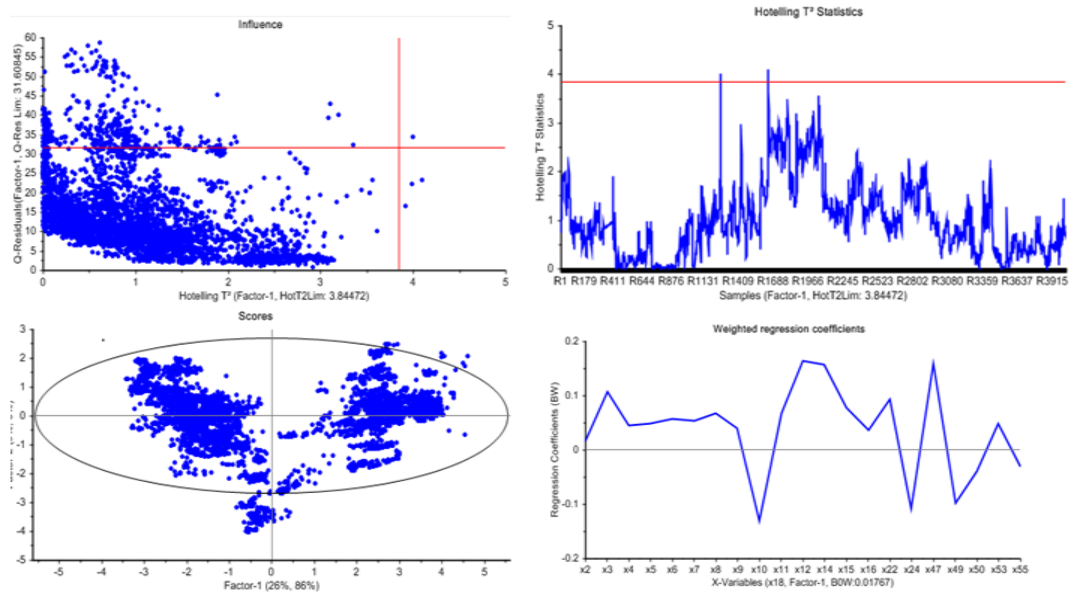


Figura 11. Validación por hotelling grupo 2.2

### Subgrupo 2.2

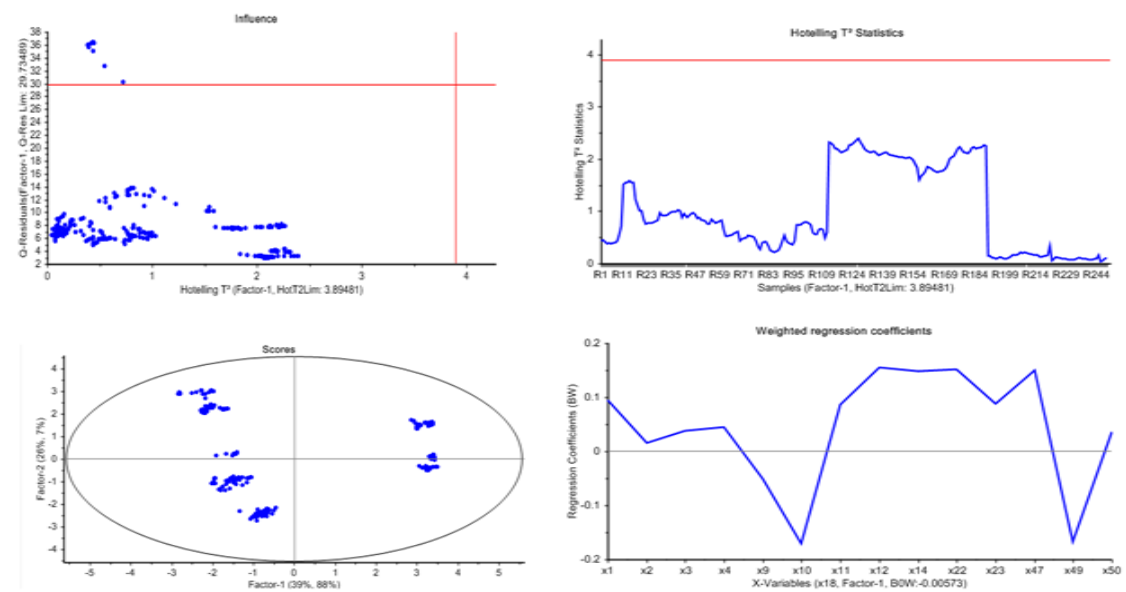


Figura 12. Validación por hotelling grupo 3.2.

Subgrupo 3.2

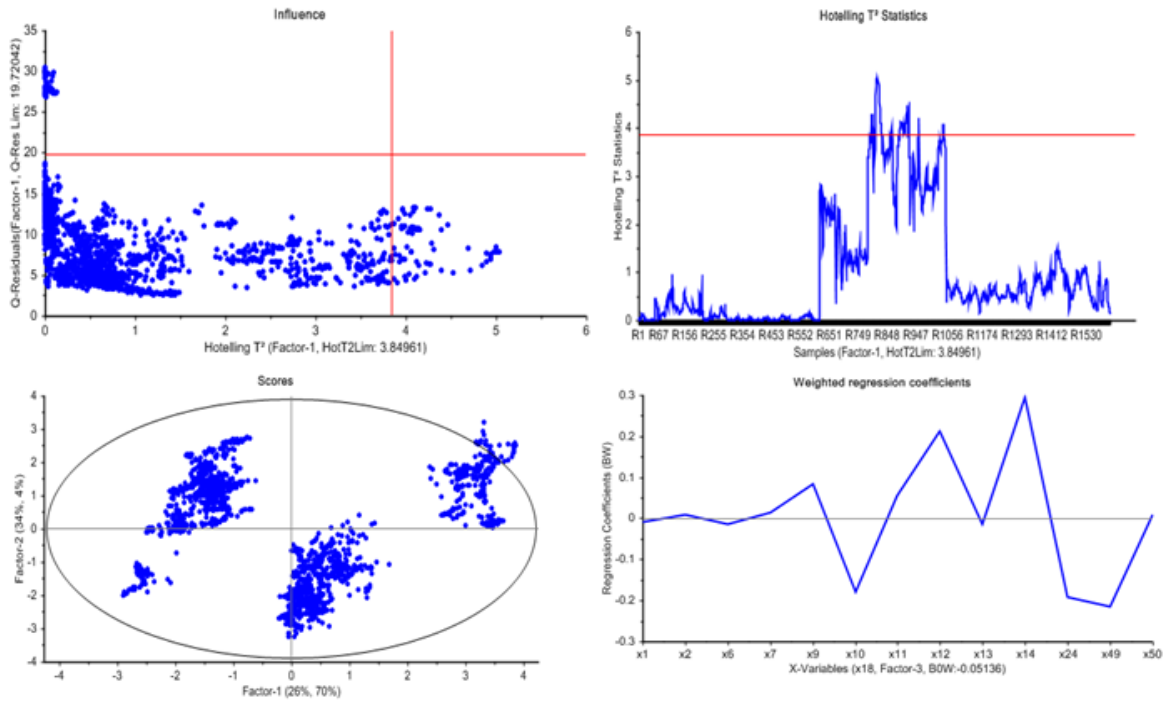
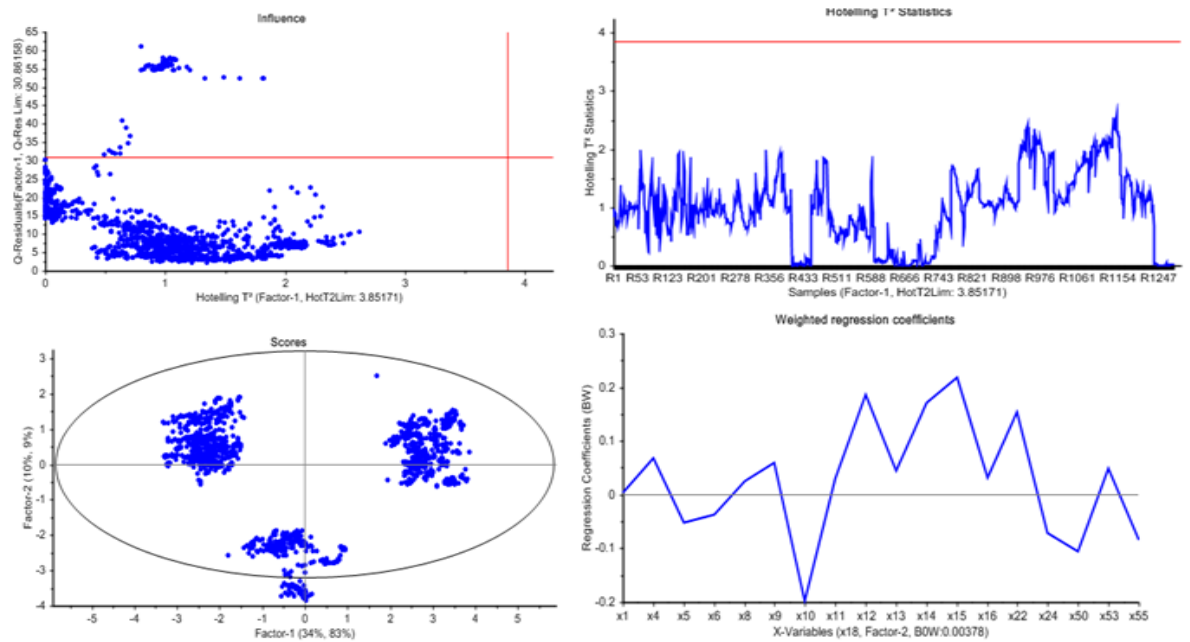


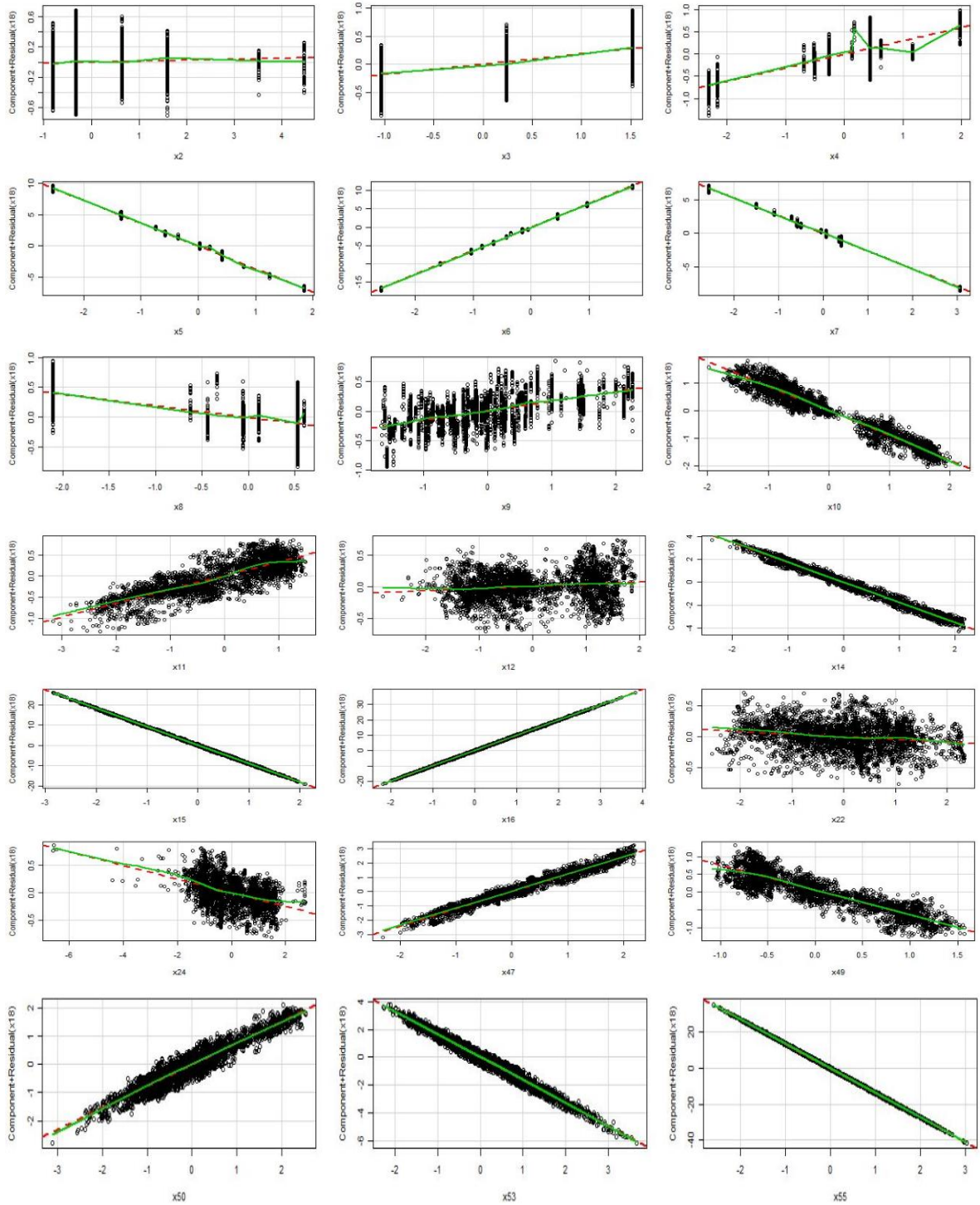
Figura 13. Validación por hotelling grupo 4.2.

Subgrupo 4.2

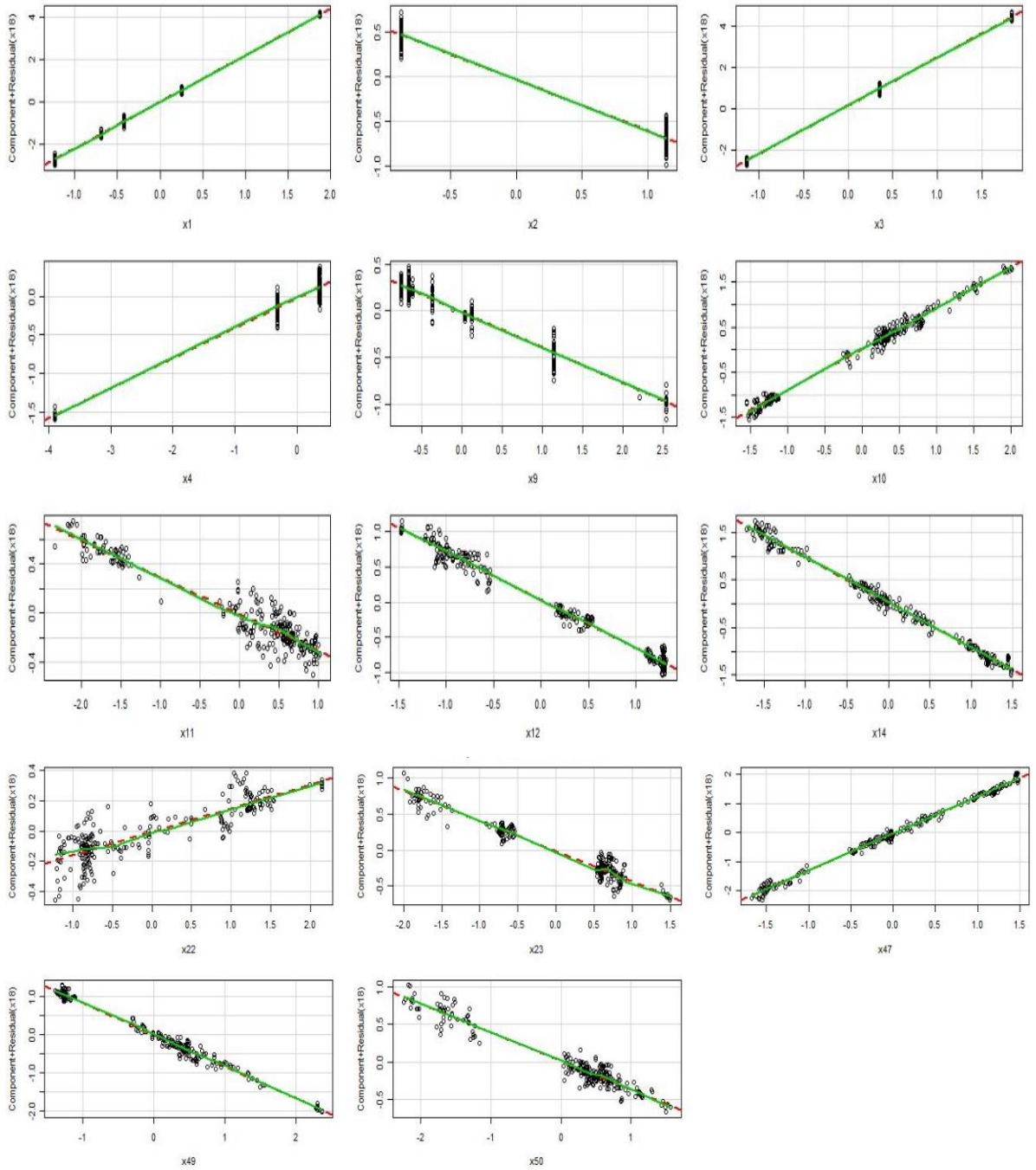


# Anexo E. Prueba de linealidad variables pendiente e independiente

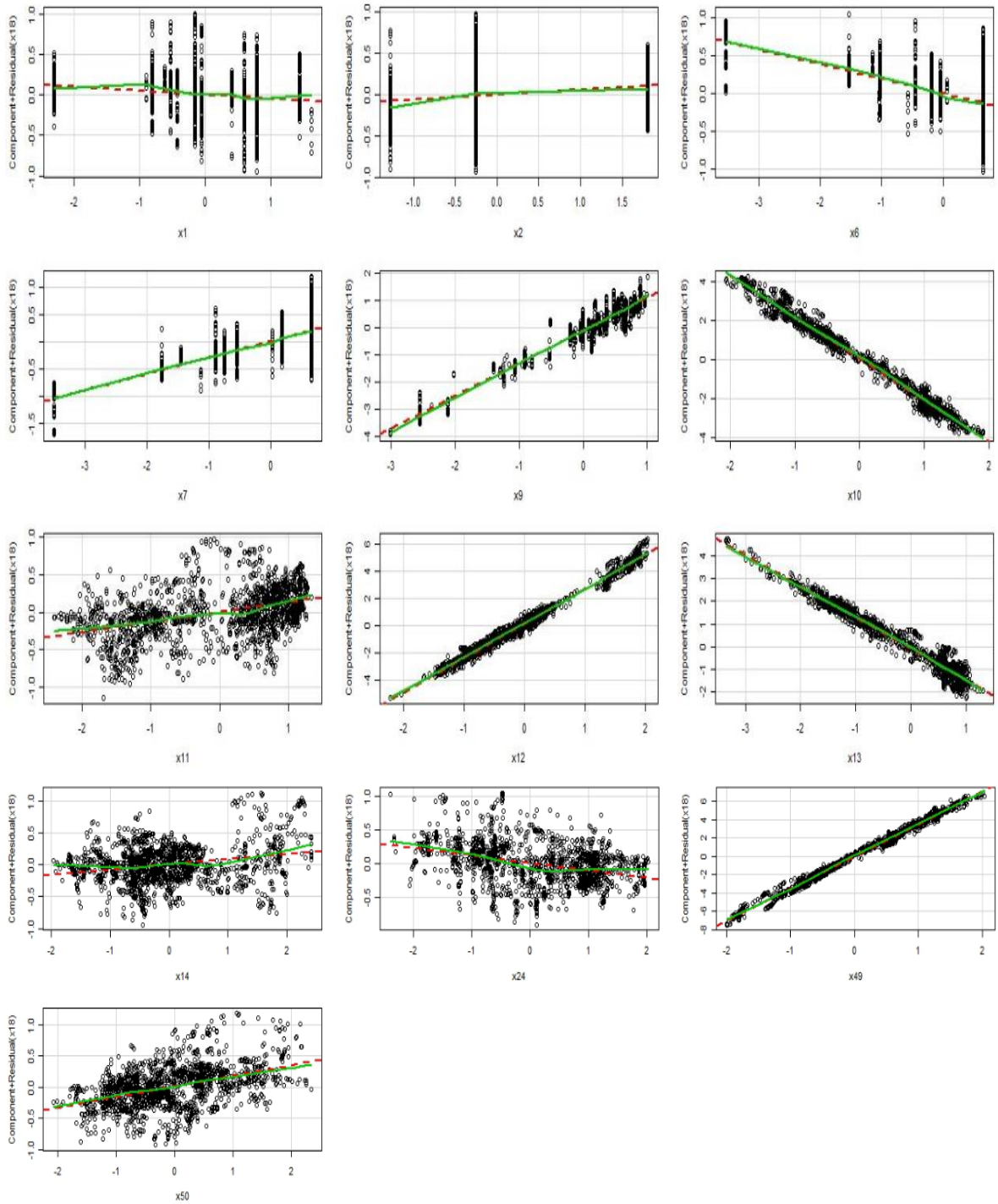
## Subgrupo 1.1



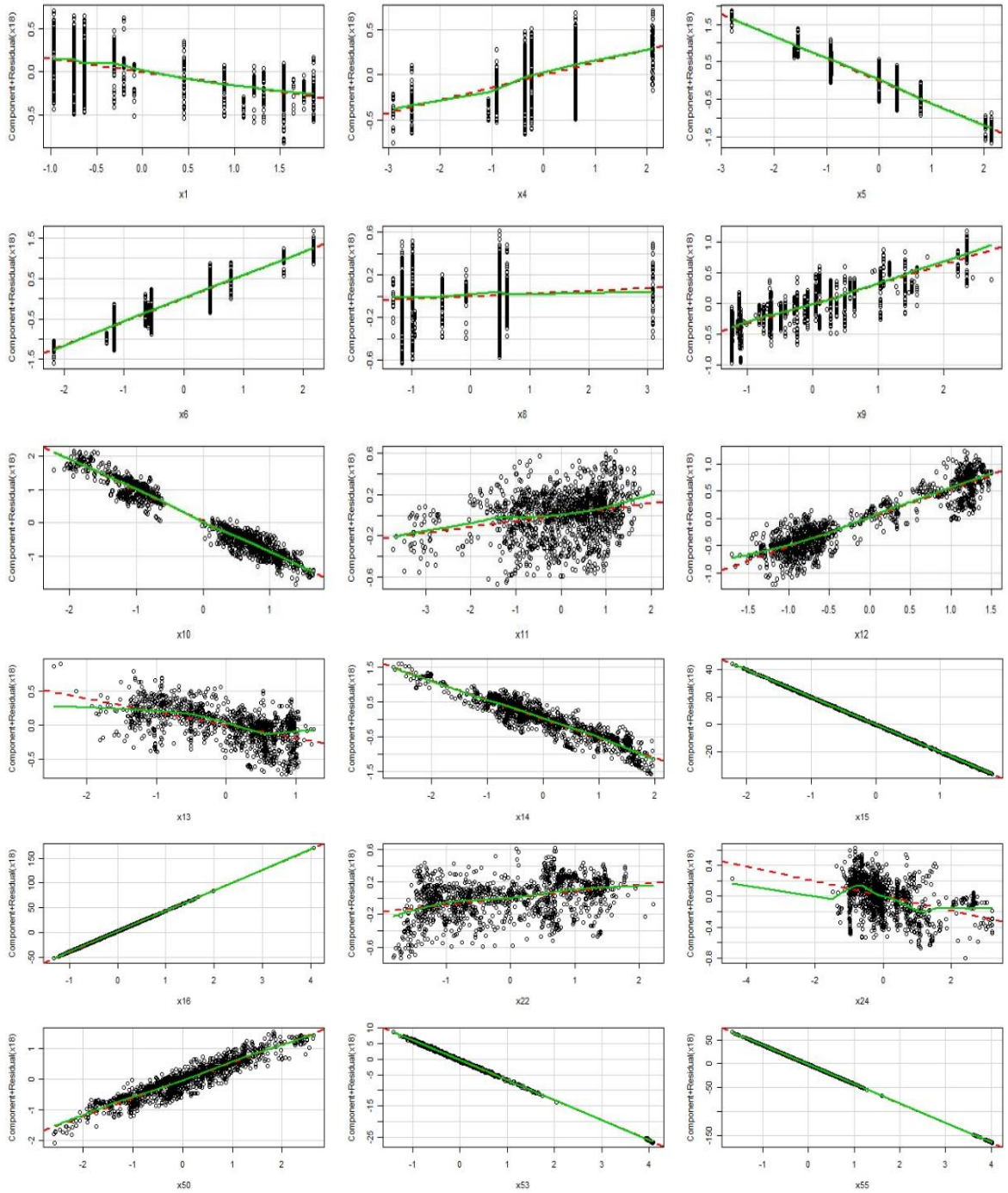
## Subgrupo 2.2



## Subgrupo 3.2



## Subgrupo 4.2

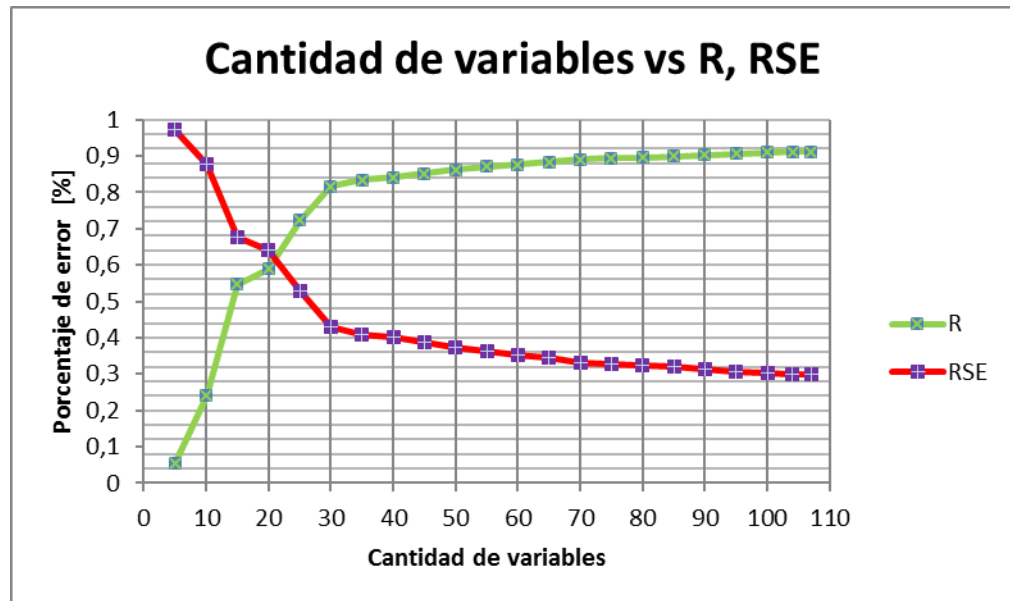


**Anexo F. Regresión con efectos dobles de las variables iniciales.**

**Ecuación 13. Regresión del modelo de efectos dobles**

$$\begin{aligned}
 x_{18} = & (-1.56x_1x_1 + 3.29x_2 - 1.29x_2x_2 - 4.71x_2x_3 - 1.90x_3x_3 + 2.91x_3x_4 + 13.6x_4 \\
 & - 4.19x_4x_4 - 99.5x_5 - 30.2x_5x_5 + 81.4x_5x_6 + 0.0155x_6 - 70.7x_6x_6 \\
 & - 86.2x_7 + 23.4x_7x_7 + 4.41x_7x_8 - 2.76x_8 - 2.10x_8x_8 + 2.92x_8x_9 \\
 & + 17.0x_9 - 1.32x_9x_9 - 5.10x_9x_{10} + 1.19x_{10} - 1.75x_{10}x_{10} + 6.77x_{11} \\
 & - 3.58x_{11}x_{11} + 5.25x_{11}x_{12} + 46.9x_{12} + 29.4x_{12}x_{12} - 8.77x_{12}x_{13} \\
 & - 12.6x_{13} - 769x_{13}x_{13} + 5.08x_{13}x_{14} - 44.2x_{14} - 38.8x_{14}x_{14} + 61.3x_{15} \\
 & + 80.9x_{15}x_{15} + 0.0523x_{16} + 0.0199x_{16}x_{16} - 1.79x_{17}x_{17} + 5.89x_{17}x_{22} \\
 & - 13.0x_{22} + 3.32x_{22}x_{22} - 3.96x_{22}x_{23} - 4.02x_{23} - 975x_{23}x_{23} \\
 & - 6.21x_{23}x_{24} - 13.5x_{24} - 646x_{24}x_{24} - 7.90x_{24}x_{47} + 35.1x_{47} \\
 & + 47.0x_{47}x_{47} - 4.47x_{47}x_{49} + 2.95x_{49}x_{50} + 17.0x_{50} - 5.12x_{50}x_{50} \\
 & - 2.61x_{50}x_{53} - 73.7x_{53} - 86.6x_{53}x_{55} + 97.5x_{55} - 0.0244x_{55}x_{55}) * 10^{-2}
 \end{aligned}$$

**Figura 14. Efecto de la cantidad de variables en la regresión.**



## Anexo G. Regresión por PLS y PCR del grupo 1

Figura 15. PCR de grupo 1.

### PCR GRUPO 1

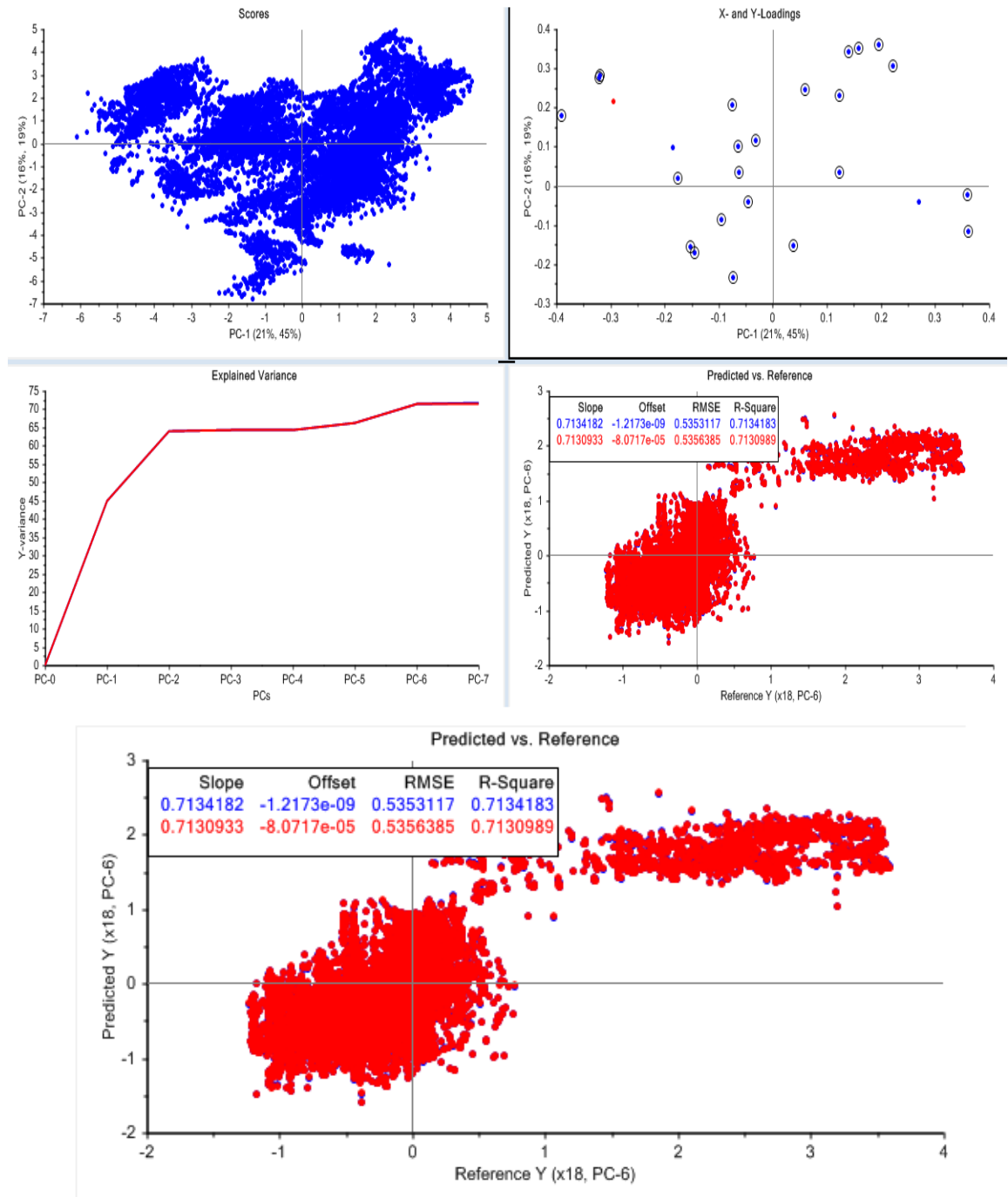
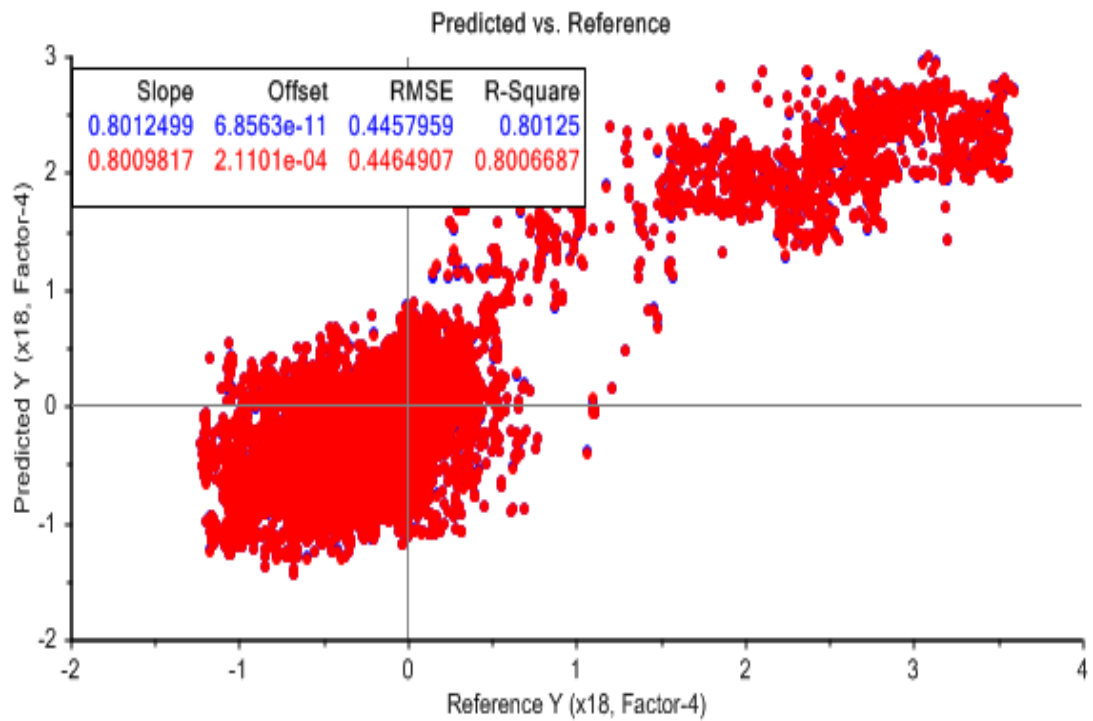
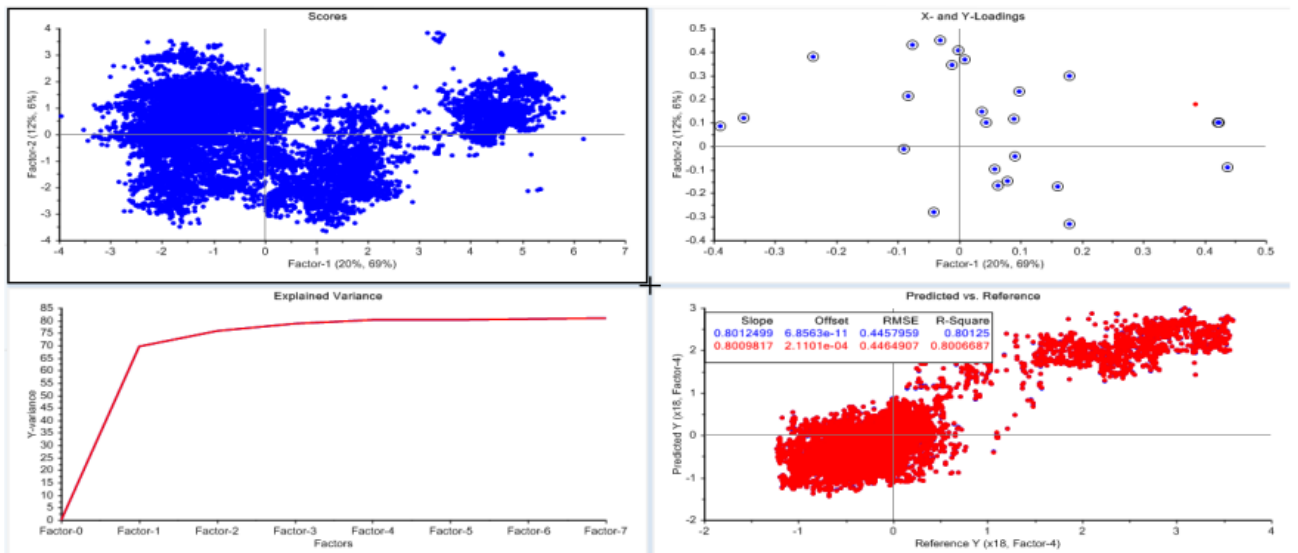


Figura 16. PLS de grupo 1.

PLS



## Anexo H. Regresión por PLS y PCR del grupo 3.

Figura 17. PCR de grupo 3.

### PCR GRUPO 3

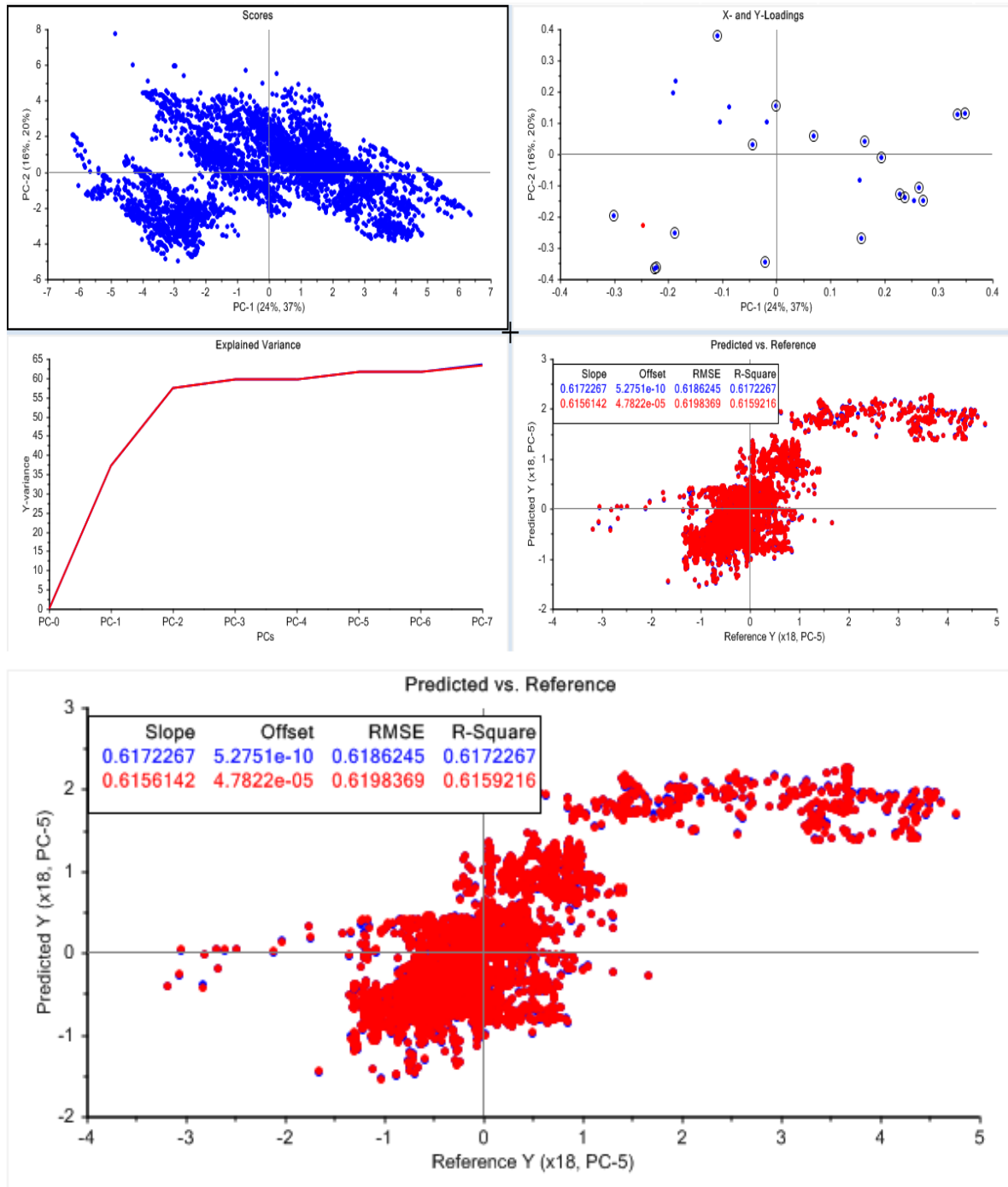
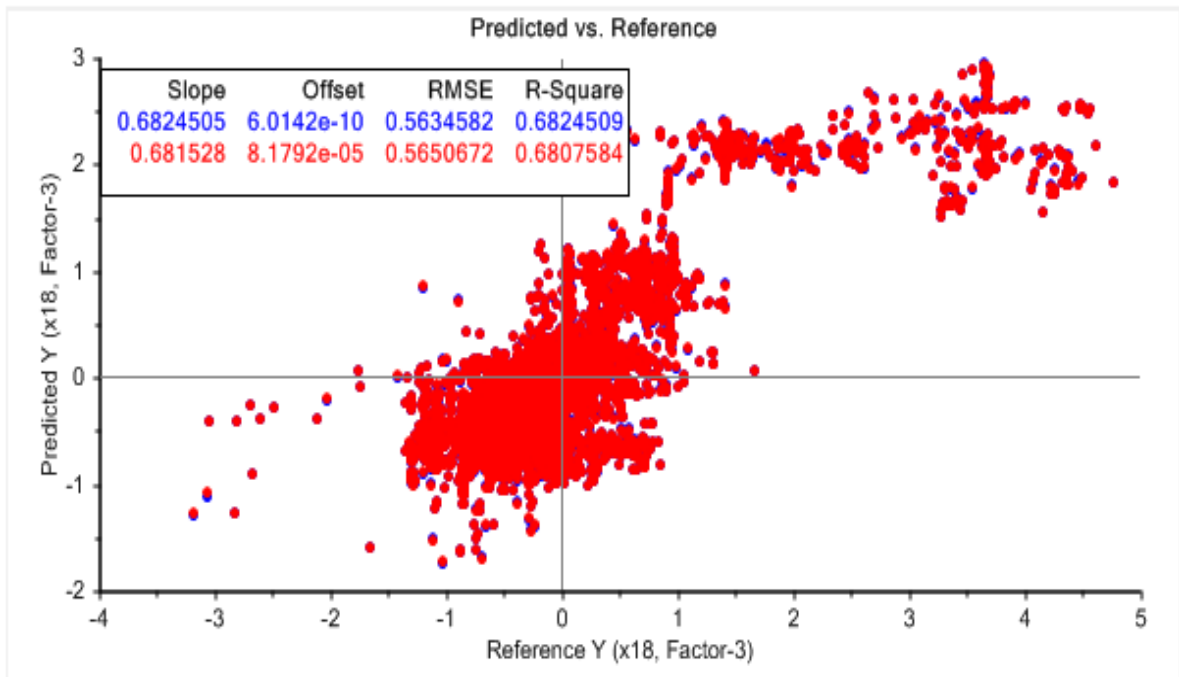
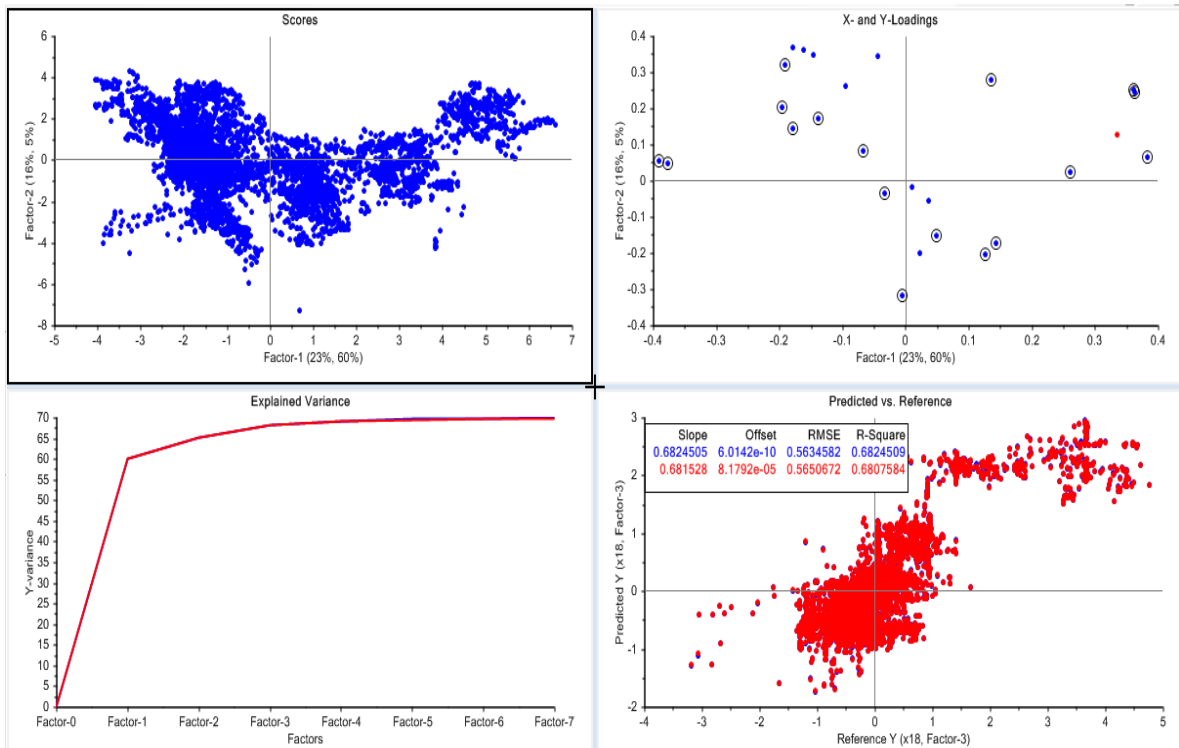


Figura 18. PLS de grupo 3.

PLS



# Anexo I. Regresión por PLS y PCR del grupo 4.

## Figura 19. PCR de grupo 4.

### PCR GRUPO 4

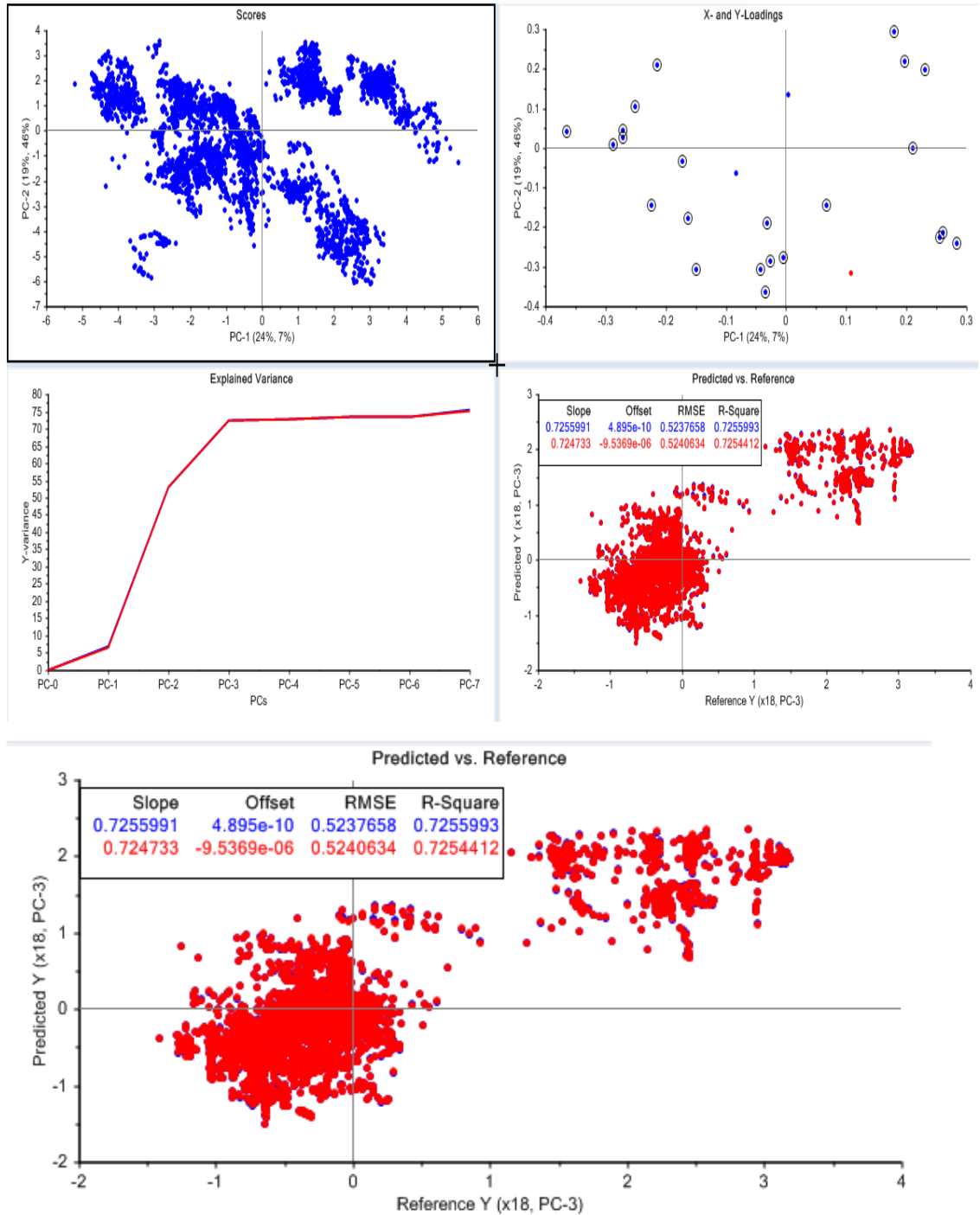


Figura 20. PLS de grupo 4.

PSL

