

**ANÁLISIS DE DATOS DE PROCESO DE LA UNIDAD DE LLENADO DE PGR
DE LA GRB UTILIZANDO PCA**

LEIDY YURANY GUALDRÓN SANDOVAL

JULIÁN FRANCISCO JIMÉNEZ BELLO

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICO QUÍMICAS
ESCUELA DE INGENIERIA QUÍMICA
BUCARAMANGA**

2016

**ANÁLISIS DE DATOS DE PROCESO DE LA UNIDAD DE LLENADO DE PGR
DE LA GRB UTILIZANDO PCA**

LEIDY YURANY GUALDRÓN SANDOVAL

JULIÁN FRANCISCO JIMÉNEZ BELLO

**Trabajo de grado para optar al título de
Ingeniero Químico**

Director:

GIOVANNI MORALES MEDINA

Ingeniero Químico, Dr.

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICO QUÍMICAS
ESCUELA DE INGENIERIA QUÍMICA
BUCARAMANGA**

2016

DEDICATORIA

Dedico esta tesis primeramente a Dios por forjar mi camino y permitirme culminar uno de mis más grandes logros. A mis padres Alejo y Nelly por ser el fundamento y la razón de mi vida, a mis tres hermanos por su inigualable compañía y motivación, a la familia Gualdrón por su incondicional apoyo y amor. A mis amigas por todo el cariño brindado a lo largo de todos estos años.

Leidy Yurany Gualdrón Sandoval

Dedicado a mis padres cuyos esfuerzos, consejos y valores fueron de vital importancia para realizar este proyecto, a mi hermano por su compañía y su apoyo durante toda mi carrera, a mis amigos que me brindaron una voz de apoyo en tiempos difíciles.

Julián Francisco Jiménez Bello

TABLA DE CONTENIDO

	Pág.
INTRODUCCIÓN.....	16
1. DESCRIPCIÓN DEL PROBLEMA Y PRECEDENTES.....	19
1.1 UNIDAD DE ENTREGA DE PROPILENO.....	19
1.1.1 Fallas operativas.....	21
1.2 MONITOREO DE PROCESOS CON MÉTODOS DE ANÁLISIS ESTADÍSTICOMULTIVARIADO.....	22
1.3 AGRUPACIÓN POR EL ALGORITMO K-MEANS.....	23
2. METODOLOGÍA.....	25
2.1 CONSECUCIÓN DE DATOS DE PROCESO.....	25
2.2 ANÁLISIS ESTADÍSTICO DE LOS DATOS.....	26
2.3 CÁLCULOS DE MASA EN LOS CARGUES.....	26
2.4 DEFINICIÓN DE CARGUES ESTÁNDAR Y CARGUES SUBESTÁNDAR.....	27
2.5 ANÁLISIS POR COMPONENTES PRINCIPALES Y AGRUPACIONES DE CARGUES.....	27
3. RESULTADOS Y ANÁLISIS DE RESULTADOS.....	29
3.1 DESCRIPCIÓN Y ANÁLISIS DE LOS DATOS HISTÓRICOS DEL PROCESO.....	29
3.2 CARGUES ESTÁNDAR Y MASA DE LLENADO DE LAS CISTERNAS.....	32
3.3 AGRUPAMIENTOS Y ANÁLISIS PCA.....	35
4. CONCLUSIONES.....	41
5. RECOMENDACIONES.....	42

BIBLIOGRAFÍA.....43

ANEXOS.....46

LISTA DE FIGURAS

	Pág.
Figura 1. Diagrama de proceso de la unidad de llenado de PGR (U-3160).....	18
Figura 2. Diagrama resumido sistema de almacenamiento y entrega de PGR (U-3160).....	19
Figura 3. Diagrama de flujo de la metodología.	25
Figura 4. Tendencia de presión de descarga de las bombas en cargues con sensores en línea (08/01/2014).	30
Figura 5. Tendencia de presión de descarga de las bombas y flujo manejado en cargues estándar (02/01/2104).....	30
Figura 6. Tendencias de presión en los trenes y presión a la entrada de los trenes con los sensores en línea (07/02/2014).	32
Figura 7. Tendencias de cargues comparando la presión de retorno con flujo másico (04/02/2014).	34
Figura 8. Comparativa de la masa reportada en ventas y la masa calculada por bombeo.	34
Figura 9. Distribución de la masa calculada cargada según los grupos identificados por el PCA y método k-mean (grupo 0 y grupo 1).....	35
Figura 10. Distribución de la masa reportada en ventas según los grupos identificados por el PCA y método k-mean (grupo 0 y grupo 1).....	36
Figura 11. Varianza explicada por los 8 primeros componentes principales en las etapas de calibración y validación cruzada.....	37
Figura 12. Gráfico de los score de los dos primeros componentes principales; pretratamientos de normalización y derivación según Savitzky Golay. La notación de 0 y 1 representa a las clases obtenidas por medio del método <i>k-mean</i> y la distancia euclidiana.....	37
Figura 13. Gráfico de dispersión biplot de los dos primeros componentes principales; en azul los cargues, en rojo las variables.	38

Figura 14. Presión inicial descarga de la bomba vs presión de entrada al tren; los cargues utilizan la notación de 0 y 1 según la clasificación obtenida por el método *k-mean*.....39

Figura 15. Presión final del tren de llenado vs presión de entrada al tren; los cargues utilizan la notación de 0 y 1 según la clasificación obtenida por el método *k-mean*.....39

LISTA DE TABLAS

	Pág.
Tabla 1. Relación presión de llegada cisterna y cargues estándar.....	33

LISTA DE ANEXOS

	Pág.
ANEXO A. CARGUE COMPORTAMIENTO ATÍPICO DE FLUJO MÁSSICO.	46
ANEXO B. CARGUE COMPORTAMIENTO ATÍPICO DE FLUJO MÁSSICO	46
ANEXO C. GRÁFICA DE TENDENCIA DE PRESIÓN DE DESCARGA DE LAS BOMBAS EN CARGUES CON SENSORES FUERA LÍNEA.....	47
ANEXO D. GRÁFICA DE TENDENCIAS DE PRESIÓN EN LOS TRENES Y PRESIÓN A LA ENTRADA DE LOS TRENES CON LOS SENSORES FUERA DE LÍNEA.	47
ANEXO E. GRÁFICA DE TENDENCIAS DE TEMPERATURA EN EL TREN UNO	48
ANEXO F. GRÁFICA DE LA TENDENCIA EN UN CARGUE ESTÁNDAR DE LA PRESIÓN DE RETORNO.....	48
ANEXO G. PCA	49
ANEXO H. ESTADISTICA DESCRIPTIVA MATRIZ DEL SISTEMA Y SOPORTE DE LOS DATOS INGRESADOS PARA APLICACIÓN DE PCA	51
ANEXO I.	54

LISTA DE ABREVIATURAS

PGR	Propileno grado refinería
GRB	Gerencia refinería de Barrancabermeja
PCA	Análisis de componentes principales
ca	Aproximadamente
e.g.	por ejemplo

RESUMEN

TÍTULO: ANALISIS DE DATOS DE PROCESO DE LA UNIDAD DE LLENADO DE PGR DE LA GRB UTILIZANDO PCA*.

AUTORES: Leidy Yurany Gualdrón Sandoval
Julián Francisco Jiménez Bello**

PALABRAS CLAVE: Estación de cargue, PCA, K-mean, pseudoestacionario.

DESCRIPCIÓN:

En este trabajo de investigación se estudia la problemática de la GRB (Gerencia Refinería Barrancabermeja) con la estación de carga de cisternas con propileno U-3160, en la que se presentan múltiples fallas en su operación debido principalmente a sobrepresión de líneas de cargue y daños de los dispositivos de accionamiento de la estación de carga, lo que ha ocasionado incumplimientos en los despachos de ventas de propileno. Para ayudar a estandarizar la operación de cargues en la estación de carga, se realiza un estudio estadístico cualitativo de las tendencias en las variables medidas en las líneas de llenado por parte de los sensores de la unidad de llenado, y se observa el comportamiento de las variables en el tiempo para así encontrar y definir la tendencia de un cargue estándar. Se encontró un comportamiento del sistema pseudoestacionario (PSSH). Una vez definidas las tendencias y características estándar se realiza una matriz representativa de la operación de la estación de carga, la matriz es analizada con la herramienta estadística análisis por componentes principales (PCA) aplicando el método K-mean, donde se agrupan las características, los rangos operativos de un cargue estándar y los umbrales de las variables en un cargue sub-estándar.

* Proyecto de grado.

** Facultad de Físicoquímicas Escuela de Ingeniería Química Director: Ph.D. Giovanni Morales Medina.

ABSTRACT

TITLE: ANALYSIS OF DATA OF FILLING UNITY PROCESSING OF PGR OF GRB USING PCA.*

AUTHORS: Leidy Yurany Gualdrón Sandoval
Julián Francisco Jiménez Bello**

KEYWORDS: charging station, PCA, K-mean, pseudo-steady.

In this research the problem of GRB (Barrancabermeja refinery management) was studied with the charging station tanks with propylene U-3160, in which occurs multiple faults in its operation mainly because overpressure load lines and damage of the actuators of the charging station, which have caused breaches in the sales offices of propylene. To help standardize the upload operation onto the loading station, a qualitative statistical study of trends was done in the variables measured in the filling lines by the sensors of the filling unit, and the behavior of the variables was observed in time to find and define the trend of a standard load. Behavior of the system found pseudo-steady (PSSH). Once defined trends and standard features a representative matrix operation of the charging station was performed, the matrix was analyzed with statistical tool principal component analysis (PCA) using the K-mean method, where the characteristics, operating ranges of a standard load and thresholds of the variables in a sub-standard were grouped.

* Project of grade

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

INTRODUCCIÓN

El propileno o propeno es un alqueno que a condiciones normales es un gas incoloro y casi inoloro; el propileno se licúa a aproximadamente 147.5 psi a temperatura ambiente¹. El propileno es utilizado en la producción de numerosas sustancias químicas orgánicas, entre las cuales se encuentran plásticos, resinas, caucho sintético, gasolina, textiles y componentes automotrices². En lo concerniente a la industria del petróleo, el propileno se utiliza como aditivo para gasolina (20% del volumen), en la generación de polipropileno, síntesis de isopropanol, en mezclas de gases de calibración para la industria petroquímica, entre otras aplicaciones.

Debido a sus aplicaciones, el requerimiento de propileno se encuentra en aumento y se estima que en pocos años la demanda global podría superar la producción de las refinerías; las principales unidades generadoras de propileno en la industria del petróleo corresponden a las unidades de ruptura catalítica³. El propileno derivado de las unidades de craqueo catalítico tipo FCC puede ser comercializado dentro del gas licuado del petróleo (GLP, pureza menor al 50%) o como grado refinería (PGR, pureza de hasta el 80%) con una diferencia marcada en su valor comercial⁴; e.g. en el segundo semestre del 2014 el precio del PGR estaba en \$83 U.S. por barril, mientras que el precio del GLP se encontraba \$50 U.S. por barril por debajo del PGR.

¹ AIR LIQUIDE, Enciclopedia de los gases, [Consultado julio 2016], disponible en internet: <<http://encyclopedia.airliquide.com/encyclopedia.asp>>.

² NEW JERSEY DEPARTMENT OF HEALTH AND SENIOR SERVICES. Hoja informativa sobre sustancias peligrosas. New Jersey. 2004. 6 p.

³ COUCH, K., GLAVIN, J.P., WEGERER, D.A., QAFISHEH J.A. Factors influencing the selection of propylene yield design points and a comparative study of an absorption-based gas concentration unit relative to cold-box recovery design show how refiners can leverage existing assets to close the market gap for propylene production. UOP LLC Q3 (2007); p. 33-43.

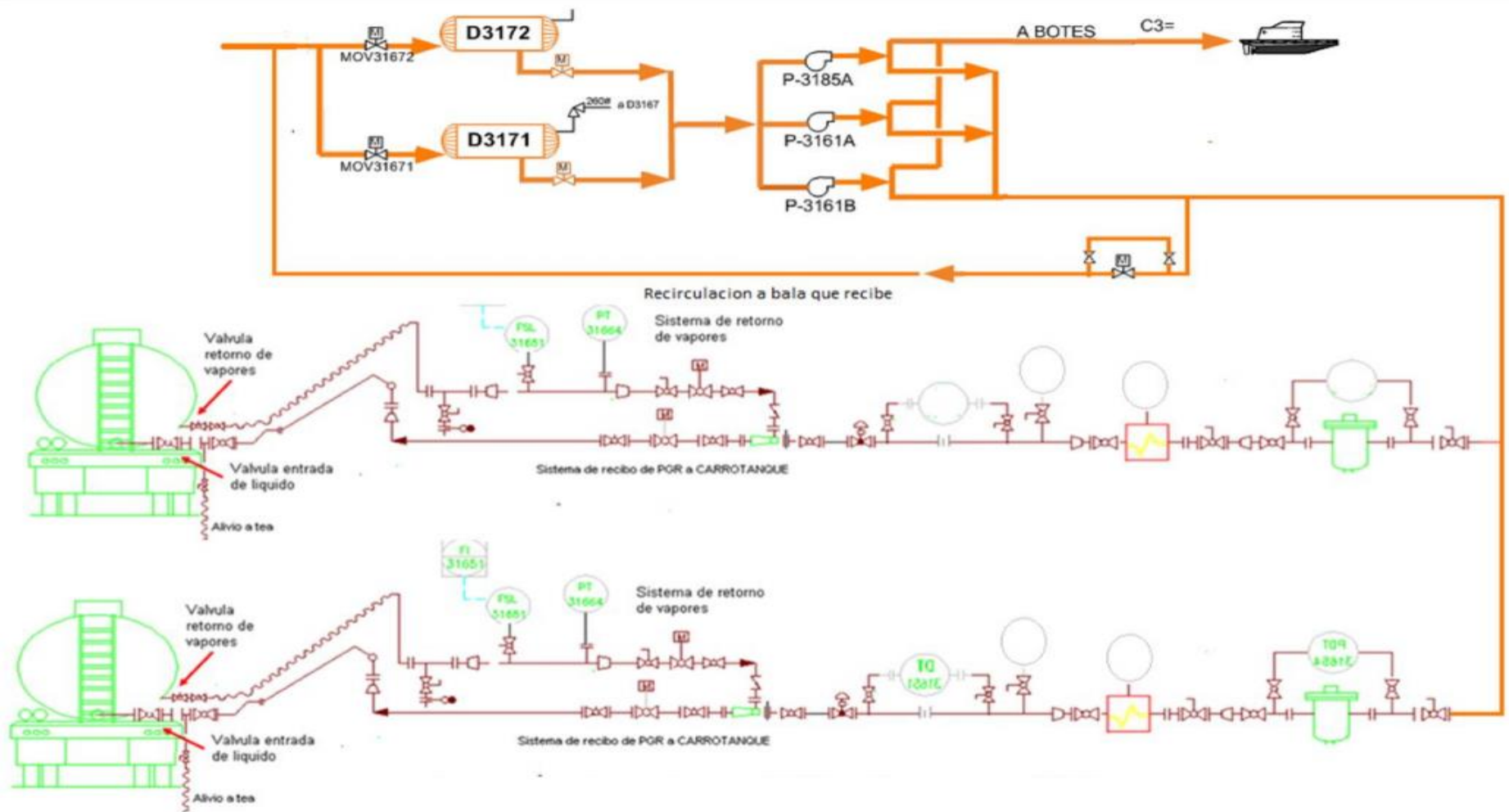
⁴ KNIGHT, J. & MEHLBERG, R. Maximize propylene from your FCC unit. Innovative use of catalyst and operating conditions increases on-purpose olefin production. *Hydrocarbon Processing* (2011); p. 91-95.

En la gerencia refinería Barrancabermeja (GRB) el PGR es producido en las unidades FCC y posteriormente almacenado en la unidad U-3160 (comúnmente denominada como llenadero, Figura 1), desde donde es despachado por cisternas y embarcaciones a la empresa Propilco⁵. El llenadero presenta una serie de problemas relacionados con la sección de cargue de cisternas ocasionando disparos en los dispositivos de seguridad a la tea y fallas recurrentes en diferentes componentes del sistema. Los anteriores inconvenientes disminuyen las entregas de PGR comprometidas debido a la no disponibilidad del sistema, causan el desvío a corrientes de menor valor (GLP) y ocasionan retrasos en el despacho de cisternas.

Debido a la importancia del PGR el análisis de las causas de falla en el sistema de llenado (U-3160) conducirá a la definición de estrategias que conlleven a una mejora económica en la actividad de la GRB. El presente trabajo describe una metodología para el análisis estadístico de los datos históricos correspondientes a la operación del llenadero en el año 2014. Los resultados de este análisis buscan la identificación de comportamientos recurrentes en los cargues estándar y en los cargues con características de operación sub-estándar.

⁵ FERNÁNDEZ, M.R. Ingeniería Conceptual para despacho de PGR por botes y carrotanques simultáneamente. Ecopetrol S.A. 2008. 21 p.

Figura 1. Diagrama de proceso de la unidad de llenado de PGR (U-3160).



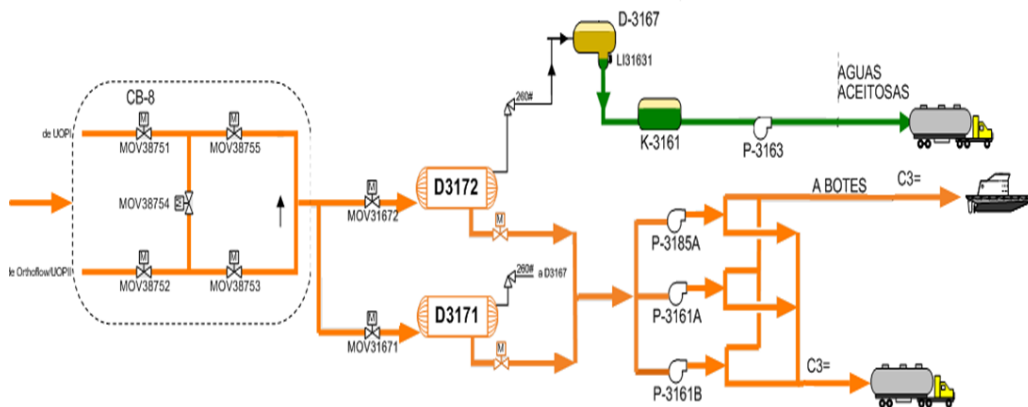
Fuente: Ecopetrol S.A.

1. DESCRIPCIÓN DEL PROBLEMA Y PRECEDENTES

1.1 UNIDAD DE ENTREGA DE PROPILENO

El PGR que se produce en la GRB, es enviado desde las plantas de craqueo catalítico UOP I, UOP II y Orthoflow, hasta la estación de GLP NEGLP (U-3160) en donde se hace la medición del flujo, se regula la presión y se envía hacia una de las balsas de almacenamiento (D-3171/3172). El PGR se almacena en condiciones de 190 psi y 100 °F. Subsecuentemente, el PGR es bombeado y medido antes de su entrega a cisternas y/o a botes en el muelle Galán (Figura 1)⁶.

Figura 2. Diagrama resumido sistema de almacenamiento y entrega de PGR (U-3160).



Fuente: Ecopetrol S.A.

El control operacional de la planta de llenado es realizado a través de válvulas motorizadas, que monitorean el nivel de inventario de los tanques, los flujos, presiones y temperaturas de las corrientes de llegada a la estación desde las

⁶ FILOSOFIA DE CONTROL-NEGLP. Construcción, montaje, gestión de compras, compras, ingeniería de detalle y puesta en marcha de las obras requeridas para la ampliación de la capacidad y almacenamiento de propileno grado refinería de la Gerencia Refinería Barrancabermeja. Ecopetrol S.A. Barrancabermeja: 2011. 53 p. (GRB-5209011-210001-ID-IN-MD-005).

unidades de proceso. El monitoreo de PGR se efectúa con el sistema de control distribuido, “DCS I/A Foxboro” existente en la planta. Desde las balas D-3171/D-3172 pasa el PGR a través de los analizadores AT31653A/B y de ahí a las líneas de succión con las válvulas MOV-31673 (D-3171) y MOV-31674(D-3172). Posteriormente, se encuentran las bombas P-3161 A/B y P-3185A que direccionan el PGR hacia los brazos número uno y dos en el llenadero de cisternas o carro-tanques⁷.

Las bombas P-3161A/B despachan 680 kg/min. Si el caudal de descarga de las bombas es bajo (menor a 230 kg/min), las líneas de recirculación (FIC-31652 ó FIC31653) son habilitadas para elevar el flujo de llenado. Así mismo, los brazos de llenado cuentan con dos válvulas motorizadas (MOV-31677 y MOV31678) para bloquear el flujo en caso de una situación anormal o de emergencia. Estas válvulas MOV solo habilitan el paso cuando el sistema de aterrizaje de la cisterna está conectado para iniciar su operación de llenado.

Adicionalmente, los brazos de llenado cuentan con un transmisor de presión PT31667 en la línea de retorno, el cual genera una alarma y detiene el llenado para evitar el disparo de la válvula de seguridad de la cisterna en caso de sobrepresión. La información generada por la instrumentación de los brazos de llenado es recibida en el DCS (Distributed Control System) a través de un módulo de comunicaciones⁸.

A pesar de la instrumentación, las válvulas y los accesorios instalados y activos en el sistema de llenado, la planta U-3160 presenta diferentes eventos operativos que disminuyen su desempeño y ocasionan retrasos en las entregas pactadas con Propilco.

⁷ Ibíd.

⁸ Ibíd.

1.1.1 Fallas operativas

Entre los eventos que han ocasionado el incumplimiento de las entregas de PGR a Propilco se tienen:

- Sobrepresión en el tren de llenado, ocasionando disparo del dispositivo de seguridad a la tea y fallas recurrentes en diferentes componentes del sistema (válvulas, eductores y brazos de llenado). Igualmente, la sobrepresión en las cisternas ha ocasionado el descargue a la atmosfera y la terminación de los cargues sin completar el volumen programado. Desde la puesta en servicio en julio del 2012, se han presentado 8 eventos de seguridad de proceso.
- Fallas en los computadores de control de flujo, lo cual ha permitido el cargue de cantidades mayores de lo programado, problemas de control de flujo y retrasos en las liquidaciones (Generación de los registros de los volúmenes entregados).
- Daños en las válvulas de solenoide (20 válvulas con un costo de 5 millones de pesos por cada válvula) y accesorios, también se han presentado múltiples fallas de las válvulas “set stop” y la válvula de control de presión del sistema de recirculación, todo esto impacta el cumplimiento de las entregas de PGR.
- Control inadecuado por tiempos muertos entre el DCS y la respuesta de la válvula de recirculación la cual controla la presión del sistema.

Las paradas de planta y el mantenimiento del sistema debido a los anteriores eventos han causado el desvío de la producción de propileno a GLP o gas combustible, con grandes pérdidas económicas. Además, estos eventos retrasan las entregas y ocasionan el incumplimiento de los volúmenes comprometidos de PGR, con las consecuentes penalizaciones de orden pecuniario por parte del transportador⁹.

⁹ RUIZ MONCADA, G. Análisis y optimización del sistema de entrega de Propileno en el departamento de materias primas de la gerencia refinería de Barrancabermeja. Tesis de maestría en Ingeniería química. Universidad Industrial de Santander, 2015; p. 5-10.

1.2 MONITOREO DE PROCESOS CON MÉTODOS DE ANÁLISIS ESTADÍSTICO MULTIVARIADO

En los procesos industriales se dispone de un número considerable de variables asociadas ya sea al control, monitoreo o simplemente como indicadores del proceso, este tipo de procesos se conocen como multivariados, y es común que la mayoría de las variables se encuentren correlacionadas. Para el llenadero U-3160 las variables medidas por diferentes sensores en los dos brazos de llenado corresponden a las mencionadas a continuación: presión de descarga de la bomba, presión de entrada del PGR al tren de llenado, presión del PGR en el tren de llenado, temperatura del PGR en el tren de llenado, flujo másico del PGR en el tren de llenado y presión de los gases de retorno de la cisterna. Sin embargo, dado el número de variables disponibles, el volumen de datos puede ser demasiado extenso para su análisis, por lo tanto, conviene una reducción del número de datos de estudio, buscando conservar la mayor variabilidad de los datos originales.

Los procesos de monitoreo por estadística multivariada emplean modelos para análisis de datos históricos, como lo son el análisis de componentes principales (PCA) y los mínimos cuadrados parciales (PLS). Herramientas como el PCA / PLS han permitido la supervisión de los procesos con comportamientos complejos (por ejemplo, no gaussiano, no lineal y no estacionario) y el diagnóstico y detección de fallas; también se han usado métodos como PCA kernel, análisis de componentes independiente, PCA dinámico, entre otros¹⁰.

El uso de la estadística multivariada para la detección de situaciones anormales se ha estudiado intensamente en el área de control de calidad multivariado (MQC). La distribución estadística T^2 del Hotelling y la distribución estadística Q, que también se conoce como el error de predicción al cuadrado (SPE), se

¹⁰ HE, B., CHEN, T., YANG, X., Root cause analysis in multivariate statistical process monitoring: Integrating reconstruction-bases multivariate contribution analysis with fuzzy-signed directed graphs. 2014. [Citado Julio 2016], 1 p.

utilizan para la detección de una situación fuera de control. Estas dos distribuciones estadísticas, calculadas en base a un modelo director como el análisis por componente (PCA), dan una visión de conjunto de situaciones anormales de manera complementaria¹¹. Este monitoreo se aplica a procesos con registro de datos que operan en estado estacionario; sin embargo, las variables de influencia en el proceso de llenado de las cisterna presentan un comportamiento cuasi estacionario durante la mayor parte de cargue, en la zona de estabilización, es decir en donde en cada cargue se da la formación de mesetas, con lo cual se justifica la aplicación del PCA para el caso de análisis que se trabaja en el presente documento¹².

El análisis de componentes principales (PCA), es una herramienta estadística aplicable a sistemas multivariados, que permite la transformación de los datos a un espacio de menor dimensión conteniendo la información más relevante sobre del proceso. El PCA transforma el conjunto de variables originales en un conjunto reducido de k-componentes principales y combinaciones lineales, que contienen la mayoría de la varianza del conjunto original a partir de una matriz de observaciones de $n \times m$ ¹³.

1.3 AGRUPACIÓN POR EL ALGORITMO K-MEANS

El procedimiento sigue una manera simple y fácil de clasificar un determinado conjunto de datos a través de un cierto número de grupos (asumir k conglomerados) fijado a priori. La idea principal es definir centroides k, uno para cada grupo y así tomar cada punto que pertenece a un determinado conjunto de datos para asociarlo con el centroide más cercano. Este algoritmo tiene por objeto reducir al mínimo una función objetivo identificando el centroide para cada

¹¹ QIN, J. Survey on data-driven industrial process monitoring and diagnosis. 2012. [Citado Julio 2016], 1 p.

¹² YIN, S., DING, S., XIE, X., LUO, H. A review on Basic Data-Driven Approaches for Industrial Process Monitoring. Vol. 61, No 11, 2014. [Citado Julio 2016], 1 p.

¹³ MINA, J. VERDE, C. Detección de Fallas usando Análisis de Componentes Principales. Instituto de ingeniería UNAM Apdo. p. 1-2.

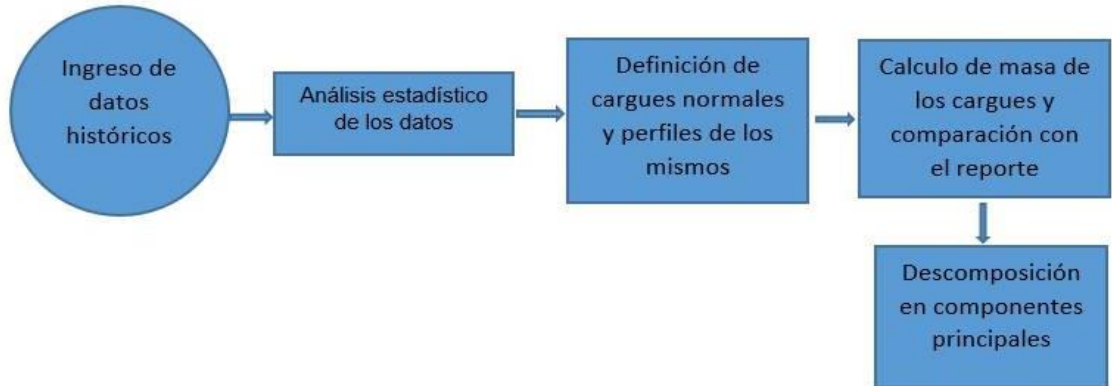
muestra según una distancia establecida (e.g. distancia euclidiana)¹⁴. Las clases establecidas según los centroides identifican muestras con características similares según la distancia establecida.

¹⁴ MACQUEEN, J. Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability. Vol. 1, Berkeley, University of California Press, 1967. p. 281-297.

2. METODOLOGÍA

La metodología seguida durante el desarrollo del presente trabajo se muestra en la siguiente figura.

Figura 3. Diagrama de flujo de la metodología.



Fuente: Los autores.

2.1 CONSECUCIÓN DE DATOS DE PROCESO

Los datos históricos de proceso fueron suministrados por la unidad U-3160 de la GRB. Los datos históricos presentan la operación del llenadero para el año 2014 reportando por minuto los valores de las variables medidas en los dos brazos de llenado (tren de llenado) disponibles para cargar PGR a cisternas. Son seis las variables estudiadas en general en el proceso, pero se cuentan con datos reportados para cada uno de los brazos de llenado por aparte. Así, es necesario aclarar que las variables influyentes en el proceso se denotan de la siguiente manera: presión de descarga de las bombas P-3161A y P-3161B, presión de cisterna del tren uno (PI-31664) y tren dos (PI-31667), flujo tren uno (FI-31656) y flujo tren dos (FI-31657), presión tren uno y presión tren dos, presión entrada en los dos trenes de llenado y temperaturas en los dos trenes de llenado.

Para efectos de simplicidad del estudio se tomará que los datos reportados se analizaran carga a carga y como un solo tren de llenado.

2.2 ANÁLISIS ESTADÍSTICO DE LOS DATOS

Los datos históricos fueron analizados para obtener tendencias representativas, dispersión, valores máximos y mínimos, promedios y desviaciones estándar para las variables de proceso, con el fin de detectar o clasificar comportamientos característicos en situaciones de carga normal y carga con fallas de manera cualitativa. Estos análisis se realizaron usando el programa Excel, con sus herramientas gráficas y análisis de datos estadística descriptiva. (Ver anexo H).

2.3 CÁLCULOS DE MASA EN LOS CARGUES

El flujo másico fue analizado para identificar los lapsos de tiempo en donde se presentan cargues estándar y cargues subestándar.

La cantidad de masa bombeada a las cisternas durante los cargues fue determinada empleando los datos de flujo másico según el método de integración Simpson 1/3.¹⁵

$$\int_a^b f(x) dx \approx \frac{h}{3} \left[f(x_0) + 2 \sum_{j=1}^{n/2-1} f(x_{2j}) + 4 \sum_{j=1}^{n/2} f(x_{2j-1}) + f(x_n) \right] \quad (1)$$

$$h = (b - a)/n \quad (2)$$

Los valores calculados para la masa en las cisternas (masa bombeada) fueron comparados con los reportados por la sección de ventas de PGR para el año 2014. Se buscó establecer umbrales entre la diferencia de la masa reportada en ventas con respecto a una masa calculada a partir del flujo másico mediante la integral. Así el dato obtenido en la diferencia de las masas indicaría las pérdidas de masa.

¹⁵ CHAPRA, S., CANALE, R. Métodos numéricos para ingenieros. Ciudad de México. McGraw-Hill, 2007. p. 640-643.

A partir de lo anterior, se procede a seleccionar cargues para la matriz del sistema tomando como heurística que la diferencia entre la masa bombeada (calculada) y la masa reportada no debe exceder los 1.500 Kg equivalentes a una pérdida de masa del 5,7% del cargue promedio de 26.500 Kg. Después de determinar los cargues que cumplían estas condiciones se procede a aplicarse el PCA y el método k-mean.

2.4 DEFINICIÓN DE CARGUES ESTÁNDAR Y CARGUES SUBESTÁNDAR

Los cargues fueron clasificados como cargues estándar y cargues subestándar. Los primeros corresponden a cargues en donde los datos reportados para cada una de las variables cumplen exitosamente con las condiciones óptimas de llenado en la cisterna, las tendencias de los datos son acordes al comportamiento normal del proceso, por el contrario, los cargues subestándar son aquellos cargues con fallas en el proceso, en donde el comportamiento de los datos en el tiempo para las variables tiene continua variación, se presentan oscilaciones, ruido e inestabilidad en la tendencia de los datos en el tiempo.

2.5 ANÁLISIS POR COMPONENTES PRINCIPALES Y AGRUPACIONES DE CARGUES

Las curvas representativas de 89 cargues fueron seleccionadas para conformar una matriz de trabajo con datos multivariados. Las filas y las columnas de la matriz multivariada representaban los cargues y las variables operacionales medidas en el tiempo (hasta 59 min), respectivamente. Los pretratamientos de normalización y la primera derivada (Savitzky-Golay) fueron aplicados a la matriz con el fin de homogenizar los rangos de las variables y permitir la aplicación

eficiente del análisis por componentes principales (PCA) y el método de agrupación *k-mean*¹⁶.

Las clases obtenidas por el método *k-mean* fueron visualizadas sobre el espacio reportado por los *scores* de los componentes principales para determinar la cercanía relativa entre las determinadas muestras; asimismo, la relación entre cargues y variables de proceso fue inferida por medio de la figura *bi-plot*¹⁷. Los métodos PCA y el método de clasificación *k-mean* fueron aplicados según los algoritmos codificados en el programa *The Unscrambler® X* de CAMO software AS.

¹⁶ SMITH, L. Principal Components Analysis. 2002, p. 12-20. [Citado Julio 2016]. Disponible en http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

¹⁷ PÉREZ, D. Estudio de la robustez del estadístico T^2 de Hotelling para el caso de una y dos poblaciones cuando los datos provienen de una distribución normal. Tesis de maestría en estadística. Universidad Nacional de Colombia. 2009. p. 10-12.

3. RESULTADOS Y ANÁLISIS DE RESULTADOS

3.1 DESCRIPCIÓN Y ANÁLISIS DE LOS DATOS HISTÓRICOS DEL PROCESO

Los datos históricos de proceso para la operación del llenadero durante el año 2014 fueron descargados del sistema de proceso de la GRB denominado *PI*. Estos datos que incluyen 6 variables y 2251 cargues conforman una base de alrededor de 878000 datos que describen la operación de *ca.* 7 cargues diarios por tren de llenado. Las variaciones del flujo másico y de la presión de descarga de las bombas MP-3161A/B son mostradas en las Figura 4 y 5, respectivamente, para diversos cargues reportados en la ventana operativa. Según la Figura 4, tres tendencias de comportamiento de flujo másico pueden ser diferenciados, según:

1. Se presenta inicialmente un pico de flujo que puede llegar a valores de 680 kg/min, aunque suele estar entre 480 y 500 kg/min, luego desciende hasta estabilizarse en valores cercanos a los 450 kg/min para luego caer abruptamente.
2. Ocurre un comportamiento casi idéntico al anterior, pero al finalizar el cargue el flujo presenta oscilaciones, es decir disminuye y aumenta hasta valores de 530 a 550 kg/min y finaliza con una caída abrupta.
3. Cargues con mucho ruido, inestables, con interrupciones y de duración prolongada que no es posible definir un rango o valores precisos que describan el comportamiento. (Ver anexo A y B). Estos cargues no se analizan puesto que no es posible reconocer si corresponden a datos registrados por un sensor descalibrado o fuera de línea.

El cambio de la pendiente de las curvas de flujo en un cargue, reporta la existencia de tres zonas correspondientes a un inicio, meseta y final. La zona inicial puede durar entre 7 y 20 minutos con un promedio de 19 minutos; esta zona presenta pendientes pronunciadas. La zona de la meseta puede durar entre 20 y 50 minutos con un promedio de 37 minutos; las pendientes de esta zona son cercanas a cero. La zona final del cargue puede durar entre 3 y 20 minutos

con un promedio en 9; en esta zona se puede presentar la formación de un pico, con una caída abrupta del flujo. Al revisar en el tiempo la elevación de flujo en los trenes y comparándolos con los datos de presión de descarga de las bombas se concluye que una sola bomba se encarga de brindarle la presión al PGR para su circulación en los dos trenes y que las bombas funcionan de manera alterna en los cargues diarios.

Figura 4. Tendencia de presión de descarga de las bombas en cargues con sensores en línea (08/01/2014).

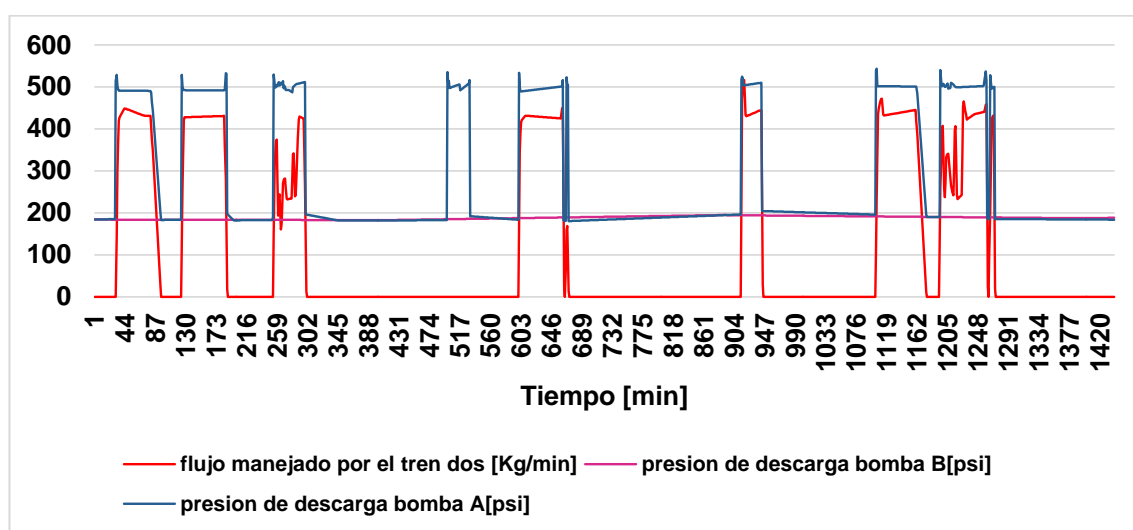
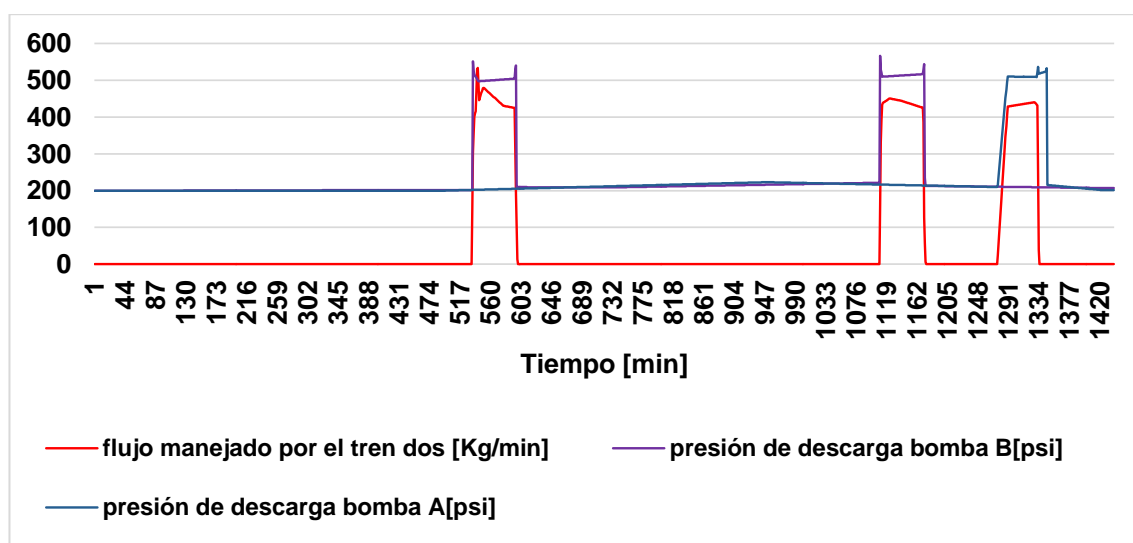


Figura 5. Tendencia de presión de descarga de las bombas y flujo manejado en cargues estándar (02/01/2104).

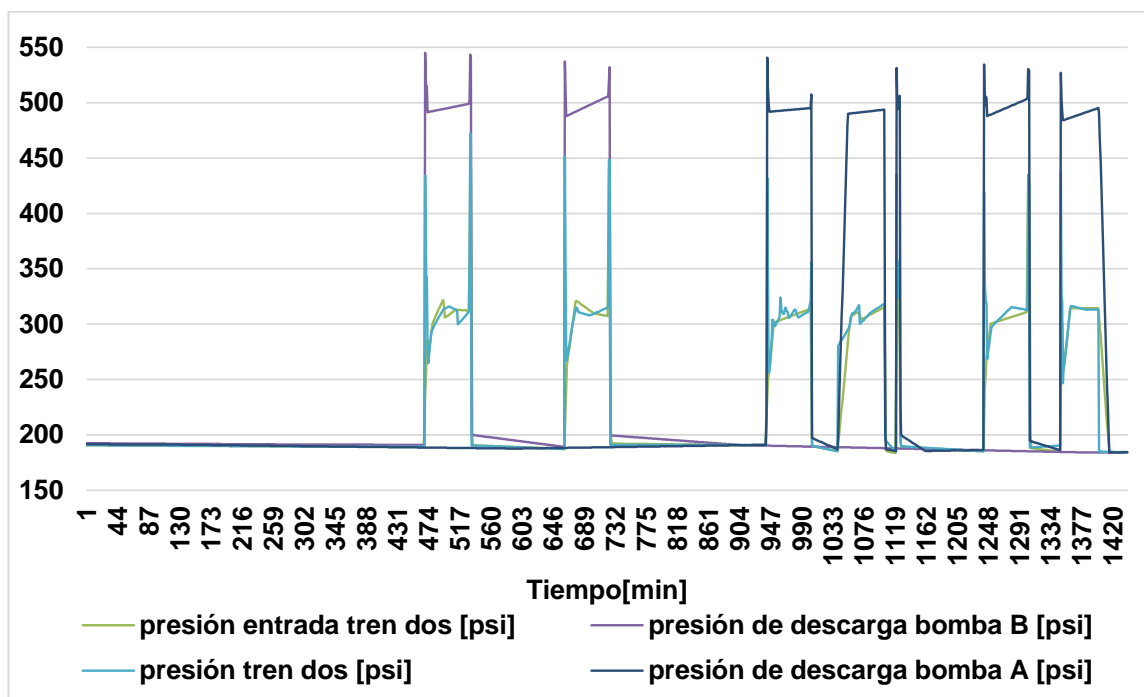


De otro lado, según la Figura 5, los cargues pueden agruparse de forma cualitativa según un comportamiento estándar o normal, en donde los puntos máximos de presión se presentan al inicio y al final de cada cargue, con un valor promedio en 550 psi. En general, la tendencia del cargue es un pico máximo de presión al inicio, después desciende hasta estabilizarse en un valor de ca. 510 psi y finaliza con otro pico máximo de presión. Lo anterior se debe principalmente a la acción de las válvulas de control en la apertura y en el cierre del flujo en cada cargue. También, los sensores pueden presentar periodos de tiempo en los que se encuentran descalibrados o fuera de línea como ocurre para el mes de marzo. (Ver anexo C).

Respecto a la presión de entrada a los trenes y la presión en los trenes se observa en la Figura 6 que la diferencia entre estas dos variables es muy pequeña y tienen comportamiento, tendencia y lapso de tiempo iguales en los cargues. También, la presión de entrada al tren y la presión del tren exhiben tendencias similares a la presión de descarga de la bomba, con valores máximos de 465 psi, valor promedio de 300 psi y valores iniciales de presión al comienzo del cargue de 182 psi. En periodos de tiempo los sensores de presión de entrada a los trenes y presión en los trenes se encuentran descalibrados o fuera de línea como ocurre en enero y abril. (Ver anexo D).

Por otro lado, en los datos históricos reportados no se encuentran cambios abruptos de temperatura, pero sí ligeras variaciones que ocurren en periodos de tiempo largos con una temperatura promedio de 96°F, en un intervalo entre 80 y 103°F. (Ver anexo E)

Figura 6. Tendencias de presión en los trenes y presión a la entrada de los trenes con los sensores en línea (07/02/2014).



La presión de la cisterna a lo largo de los cargues puede ser inferida directamente de la presión de retorno en los brazos de los trenes. Según la Figura 7, durante la operación del llenadero existen cargues que reportan datos de presión después del corte del flujo másico a la cisterna. Estos datos se estudiaron buscando usarlos para obtener a partir de ellos y los históricos de temperatura el balance de masa en el reciclo, con el fin de determinar si estas lecturas eran por pérdidas de PGR, los resultados obtenidos y junto con la verificación de los planos del llenadero se determinó que es la presión del fluido que queda confinado entre las válvulas MOV del reciclo correspondiente a una de 3 a 4 kilogramos.

3.2 CARGUES ESTÁNDAR Y MASA DE LLENADO DE LAS CISTERNAS

De las Figuras de tendencia se definió un comportamiento de las variables en condiciones estándar para el despacho exitoso del PGR por cargues a cisternas.

La tendencia en un cargue estándar para la presión de la cisterna, inferida por la presión en la línea de retorno de vapores, es una curva creciente continua, que inicia en la presión de llegada del carro-tanque; la presión promedio de ingreso de las cisternas es de 78 psi. La presión de la cisterna durante un cargue exitoso aumenta hasta una presión de ca. 250 psi y una vez alcanzado el punto de presión máxima en el cargue esta cae abruptamente cuando por procedimiento se desconecta la cisterna tan pronto termina el cargue. (Ver anexo F). Del análisis de la pendiente de la presión de retorno se encontró que existen entre 3 y 4 zonas, siendo la primer zona la de pendiente más pronunciada que puede durar los primeros 20 minutos, en secuencia cada zona va teniendo una pendiente menos pronunciada para en la última zona tener una pendiente negativa pronunciada o una caída abrupta.

De otro lado, la presión de llegada de las cisternas puede influir en el porcentaje de éxito en un cargue. Según el comportamiento de las variables existe una tendencia de fallo cuando la cisterna llega con presiones superiores a los 80 psi (Tabla 1). Estos fallos impactan en la masa cargada en cada cisterna despachada y en la frecuencia de cargue diaria. La masa de llenado calculada y la masa reportada por la oficina de despacho de la GRB para los cargues pueden ser comparadas utilizando la Figura 8. Según esta figura, se puede observar que en general la masa cargada o que se halló por flujo másico es mayor a la masa reportada en ventas que suele permanecer constante de ca. 26500 Kg.

Tabla 1. Relación presión de llegada cisterna y cargues estándar. Fuente: Los autores.

# Cargues estándar	# Cargues con presión llegada cisterna mayor a 80 psi
100	20
# Cargues fuera del estándar	# Cargues con presión llegada cisterna mayor a 80 psi
100	40

Figura 7. Tendencias de cargues comparando la presión de retorno con flujo másico (04/02/2014).

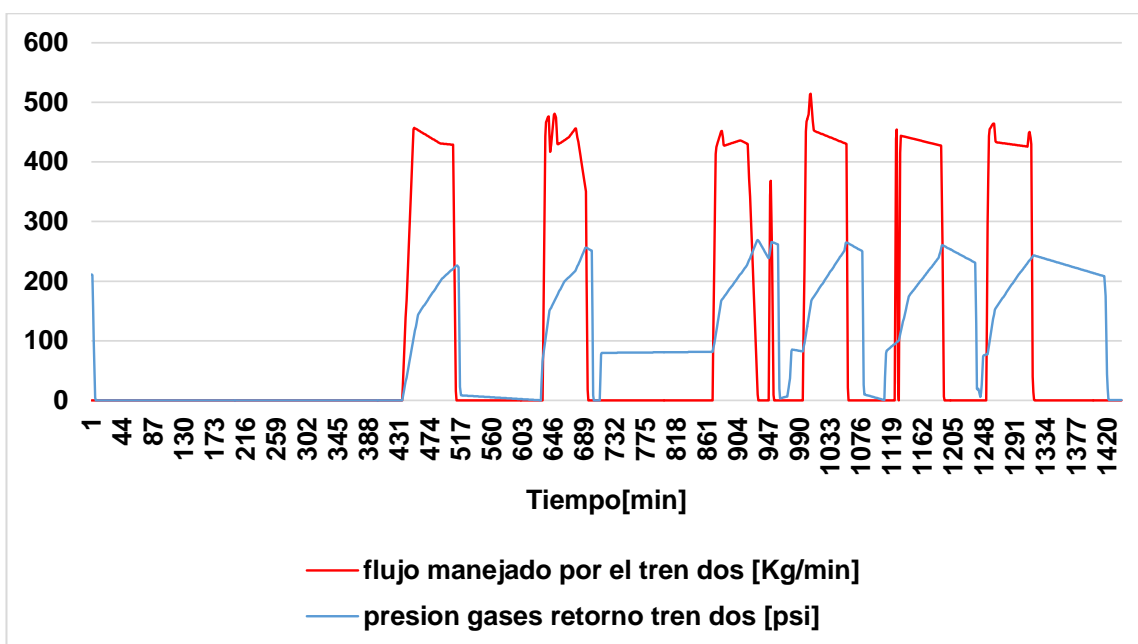
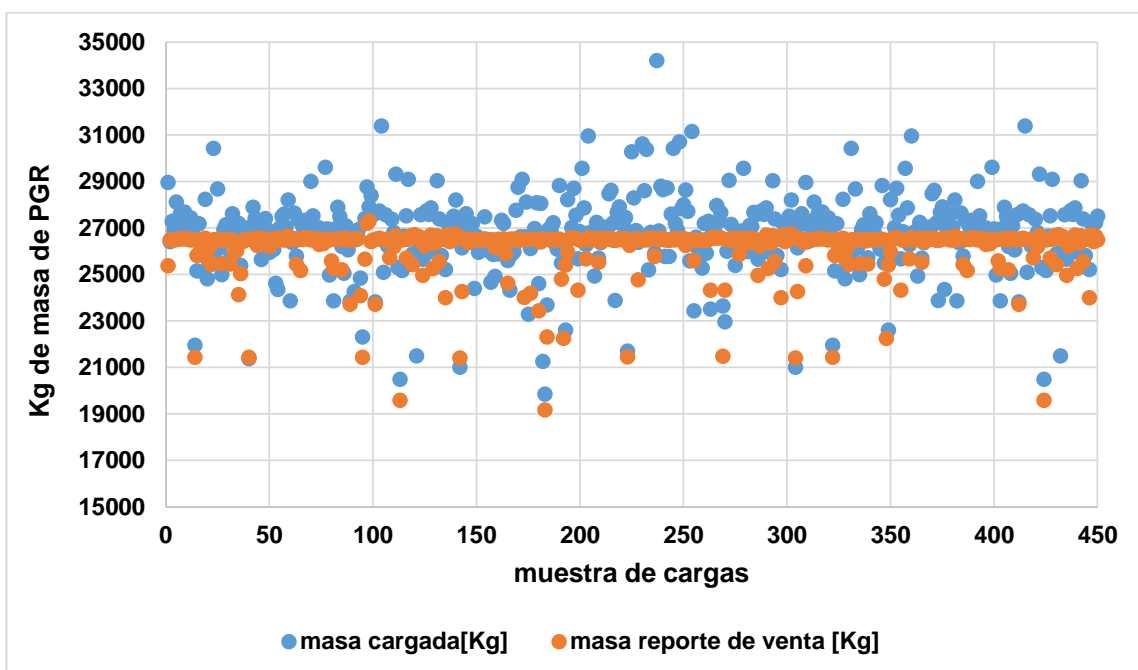


Figura 8. Comparativa de la masa reportada en ventas y la masa calculada por bombeo.



Los datos y tendencias que se encontraron de sensores fuera de línea y los cargues en los cuales las variables no se pudiesen verificar se omitieron en el

análisis del comportamiento para la matriz de aplicación del análisis de componentes principales PCA.

3.3 AGRUPAMIENTOS Y ANÁLISIS PCA

La matriz de cargues a analizar con estadística multivariada fue establecida con los datos reportados para 89 cargues considerando la información hasta el minuto 59; estos cargues fueron seleccionados según la continuidad y la disponibilidad de su información. El método de agrupamiento *k-mean* fue aplicado utilizando la distancia euclidiana para clasificar la información en dos grupos, denominados clase 0 y clase 1. Las masas de llenado calculadas y reportadas de acuerdo a las clases 0 y 1 son mostradas en la Figura 9 y en la Figura 10, respectivamente. El análisis de las tendencias reportadas en estas figuras conlleva a establecer que no existen diferencias marcadas respecto a las masas de llenado para los cargues de la clase 0 y los cargues de la clase 1; con esto la diferencia entre las clases debe radicar en la variación de los datos operativos durante los cargues y a inconvenientes durante por fallas de componentes del sistema.

Figura 9. Distribución de la masa calculada cargada según los grupos identificados por el PCA y método k-mean (grupo 0 y grupo 1).

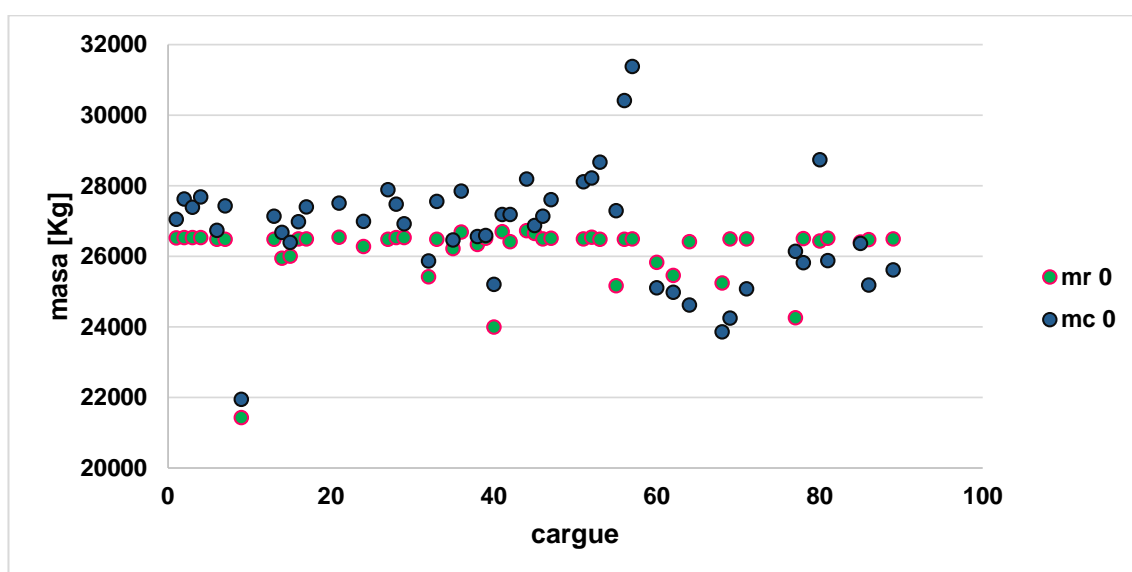
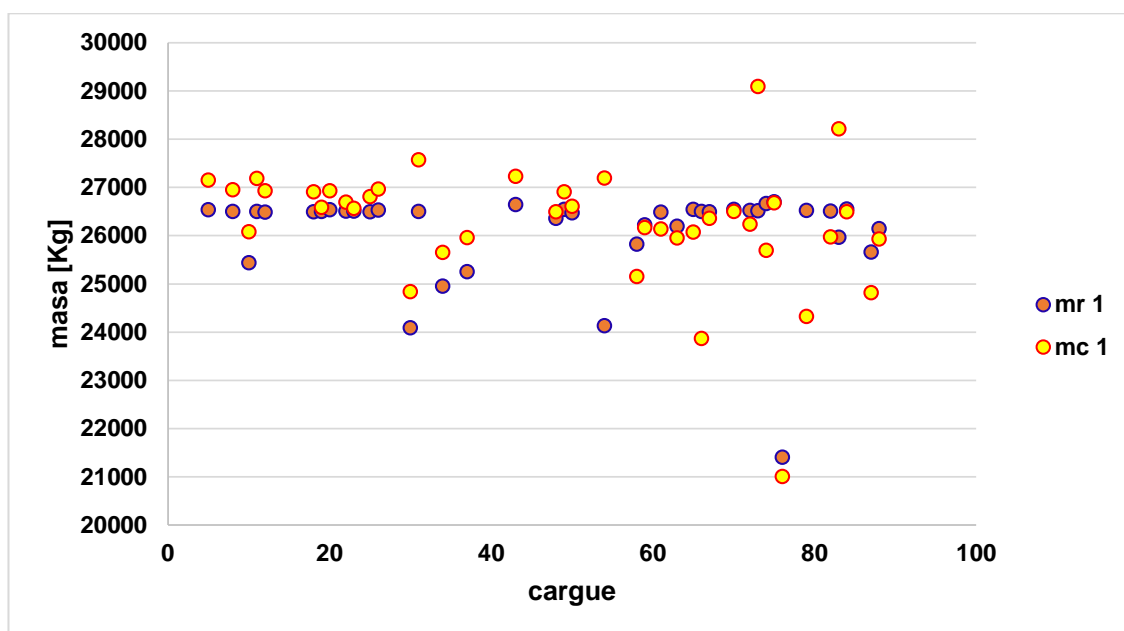


Figura 10. Distribución de la masa reportada en ventas según los grupos identificados por el PCA y método k-mean (grupo 0 y grupo 1).



El análisis por componentes principales fue aplicado a la matriz con los cargues etiquetados según los resultados del método *k-mean* con la distancia euclidiana; la matriz fue sujeta a los pretratamientos de normalización y derivación. El resultado del PCA indica que el 90% de la varianza de los datos iniciales pretratados es reproducida utilizando 8 componentes principales (Figura 11); los resultados de la validación cruzada indican que el 80% de la varianza es reproducida con igual número de componentes principales. Los resultados de los score para los dos primeros componentes principales (58% de varianza) son presentados en forma gráfica en la Figura 12. En esta figura se presentan los diferentes cargues de las clases 0 y 1; según las distancias entre los conglomerados en la gráfica de scores es posible proponer que el conjunto total de cargues se divide en cargues clase 0 y cargues clase 1; con esto, los resultados del PCA dan soporte adicional a los resultados obtenidos con el método *k-mean*.

Figura 11. Varianza explicada por los 8 primeros componentes principales en las etapas de calibración y validación cruzada.

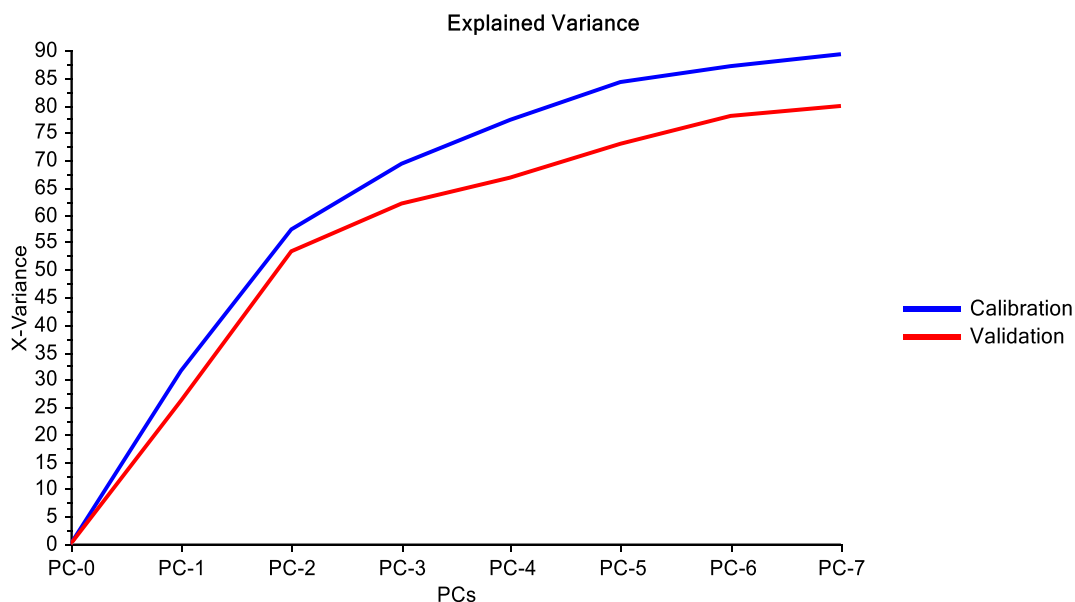
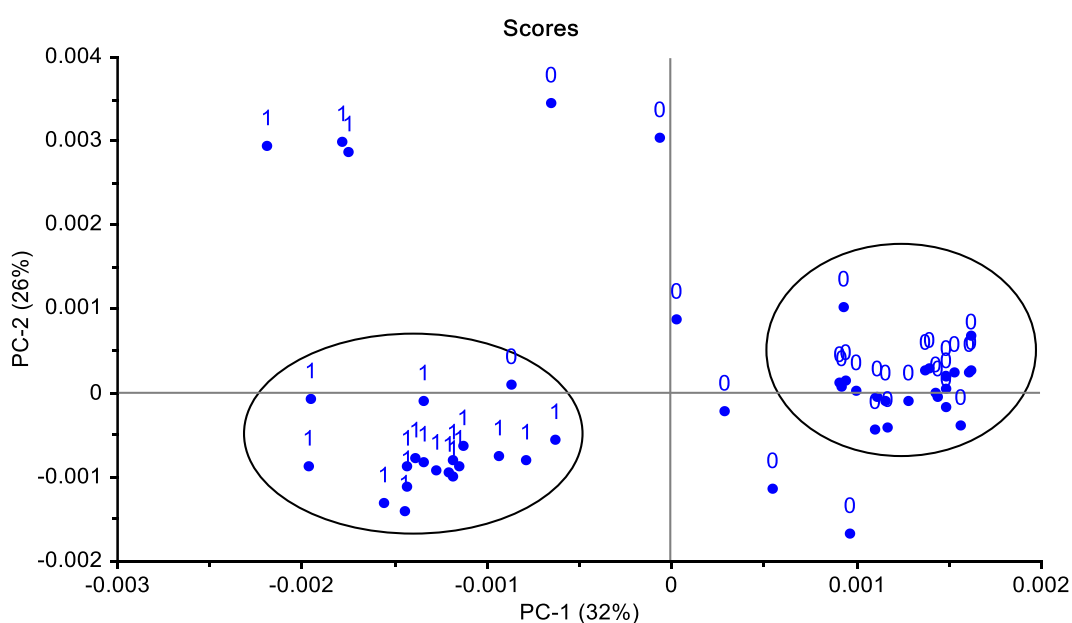


Figura 12. Gráfico de los score de los dos primeros componentes principales; pretratamientos de normalización y derivación según Savitzky Golay. La notación de 0 y 1 representa a las clases obtenidas por medio del método *k-mean* y la distancia euclidiana.



Según la definición dada para los cargues estándar que cumplen con la masa de llenado de las cisternas, la clase 0 se encuentra conformada por un 65% de cargues estándar, mientras que la clase 1 se encuentra conformada en su mayoría por cargues con características sub-estándar. Con el objetivo de identificar las variables que más influyen en el agrupamiento de los cargues, los resultados del análisis PCA fueron graficados en la denominada gráfica de dispersión *biplot*, en la cual se relacionan tanto los *scores* como los *loadings* de los dos primeros componentes principales. La Figura 13 presenta la gráfica de dispersión *biplot*, según la cual, los cargues son agrupados en las clases 0 o 1 debido a los signos de los *loading* referentes a las variables 120 y 179 y, 118 y 177, respectivamente. Estas variables están relacionadas con los valores iniciales de la presión de descarga de la bomba (Figura 14) y los valores finales de la presión del tren de llenado (Figura 15).

Figura 13. Gráfico de dispersión biplot de los dos primeros componentes principales; en azul los cargues, en rojo las variables.

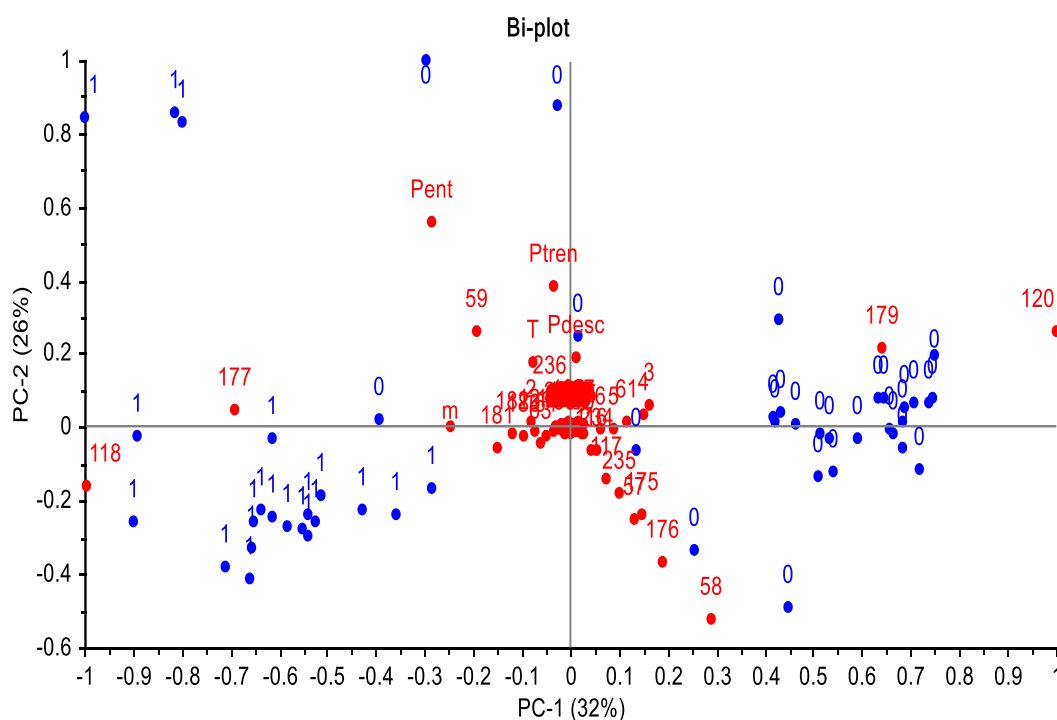


Figura 14. Presión inicial descarga de la bomba vs presión de entrada al tren; los cargues utilizan la notación de 0 y 1 según la clasificación obtenida por el método *k-mean*.

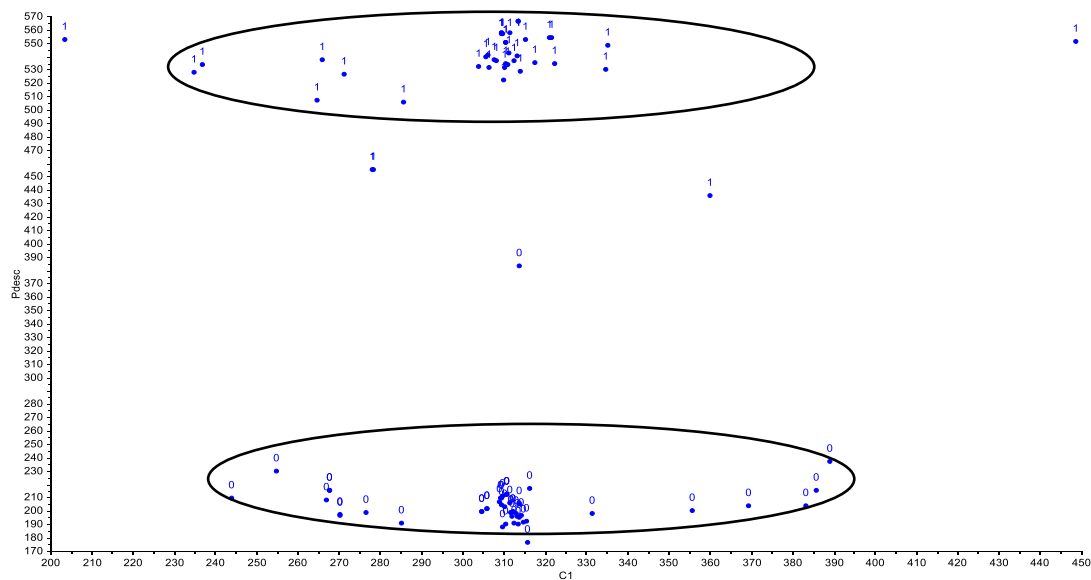
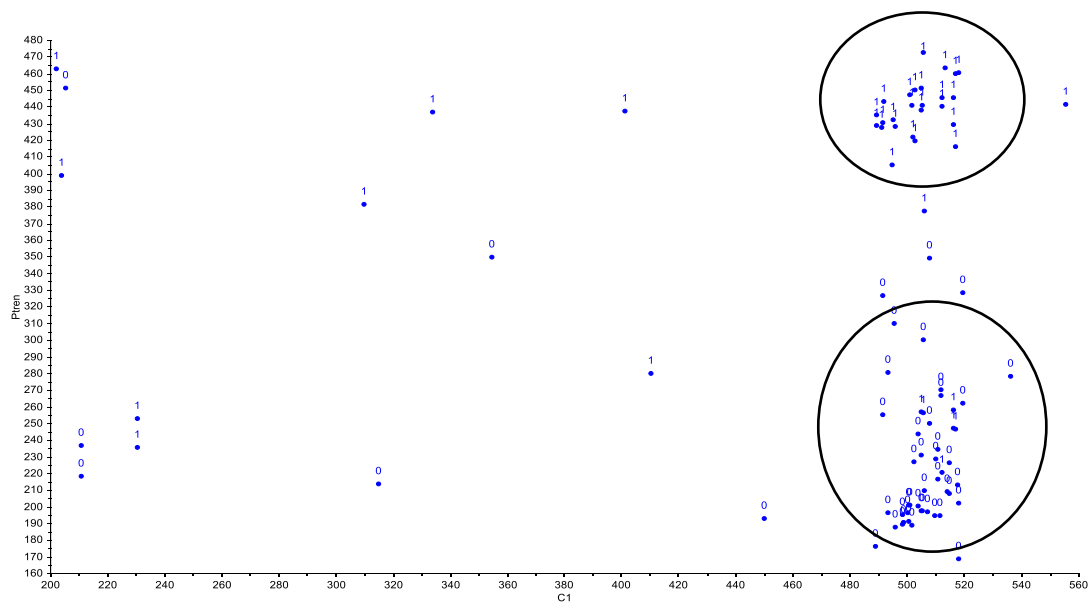


Figura 15. Presión final del tren de llenado vs presión de entrada al tren; los cargues utilizan la notación de 0 y 1 según la clasificación obtenida por el método *k-mean*.



Según las Figura 14 y 15, los cargues de la clase 0 presentan los menores valores para la presión inicial de descarga de la bomba y la presión final del tren de llenado; los cargues de la clase 1 presentan el comportamiento inverso a los cargues de la clase 0 para las anteriores variables.

Según lo obtenido anteriormente, aquellos cargues que inicien con una presión de descarga de la bomba inferior a 230 psi y finalicen con una presión del tren de llenado inferior a 280 psi, se comportarán según los estándares de llenado de las cisternas con un porcentaje de éxito superior al 65 % (cargues de la clase 0). De otro lado, los cargues cuya presión inicial de descarga de la bomba sea superior a 490 psi y el valor final del tren de llenado presente un valor superior a 400 psi, habrán exhibido un comportamiento sub-estándar (clase 1) con una duración muy superior a los cargues estándar y con posibles fallas operativas.

4. CONCLUSIONES

La aplicación conjunta de los métodos k-mean y análisis por componentes principales permite identificar los cargues con características estándar de aquellos con características sub-estándar. Las presiones de descarga de la bomba y las presiones de los trenes de llenado fueron identificadas como las partes sensibles del proceso de llenado de las cisternas. Estos métodos permiten identificar y estandarizar la operación del proceso de llenado aun cuando su aplicación se realiza a un proceso en estado cuasi-estacionario.

Según el análisis realizado en el presente trabajo, para garantizar la operación de un cargue estándar, se debe operar bajo la condición de una presión inicial de descarga de la bomba inferior a los 230 psi y una presión del tren de llenado final inferior a los 280 psi. Asimismo, cuando un cargue supera la presión inicial de descarga de 490 psi y excede la presión final en el tren de llenado del valor de 400 psi, la probabilidad de falla operativa o cargues sub-estándar aumenta.

Múltiples fallas en los sensores de presión de descarga de las bombas, presión de los trenes y presión de entrada a los trenes impiden un seguimiento y control adecuado de variables sensibles de operación del despacho de PGR en la GRB. No hay una alerta en los datos históricos que permita reconocer cuando los sensores fallan registrando datos no representativos del comportamiento.

5. RECOMENDACIONES

El análisis de las tendencias y la información suministrada permiten recomendar una mejora en los procesos de mantenimiento y calibración de los sensores de presión en el llenadero U-3160.

Asimismo, es recomendable la aplicación de un método para analizar datos dinámicos, como el PCA dynamic, para establecer relaciones adicionales entre las variables de proceso y las fallas del sistema. Los resultados de estos métodos dinámicos junto con los resultados del presente trabajo permitirían la implementación de un sistema de control MQS para la detección de fallas en línea en el llenadero U-3160 por perturbaciones del proceso o fallas en los sensores.

BIBLIOGRAFÍA

AIR LIQUIDE, Enciclopedia de los gases, [Consultado julio 2016], disponible en internet: <<http://encyclopedia.airliquide.com/encyclopedia.asp>>.

CHAPRA, S., CANALE, R. Métodos numéricos para ingenieros. Ciudad de México. McGraw-Hill, 2007. p. 640-643.

COUCH, K., GLAVIN, J.P., WEGERER, D.A., QAFISHEH J.A. Factors influencing the selection of propylene yield design points and a comparative study of an absorption-based gas concentration unit relative to cold-box recovery design show how refiners can leverage existing assets to close the market gap for propylene production. UOP LLC Q3 (2007); p. 33-43.

E, LAURA. Análisis multivariado: método de componentes principales. Secretaria general de la OEA, Washington, DC.

FERNÁNDEZ, M.R. Ingeniería Conceptual para despacho de PGR por botes y carrotanques simultáneamente. Ecopetrol S.A. 2008. 21 p.

FILOSOFIA DE CONTROL-NEGLP. Construcción, montaje, gestión de compras, compras, ingeniería de detalle y puesta en marcha de las obras requeridas para la ampliación de la capacidad y almacenamiento de propileno grado refinería de la Gerencia Refinería Barrancabermeja. Ecopetrol S.A. Barrancabermeja: 2011. 53 p. (GRB-5209011-210001-ID-IN-MD-005).

HE, B., CHEN, T., YANG, X., Root cause analysis in multivariate statistical process monitoring: Integrating reconstruction-bases multivariate contribution analysis with fuzzy-signed directed graphs. 2014. [Citado Julio 2016], 1 p.

KNIGHT, J. & MEHLBERG, R. Maximize propylene from your FCC unit. Innovative use of catalyst and operating conditions increases on-purpose olefin production. *Hydrocarbon Processing* (2011); p. 91-95.

MACQUEEN, J. Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability. Vol 1, Berkeley, University of California Press, 1967. p. 281-297.

MINA, J. VERDE, C. Detección de Fallas usando Análisis de Componentes Principales. Instituto de ingeniería UNAM Apdo. p. 1-2.

NEW JERSEY DEPARTMENT OF HEALTH AND SENIOR SERVICES. Hoja informativa sobre sustancias peligrosas. New Jersey. 2004. 6 p.

PÉREZ, D. Estudio de la robustez del estadístico T^2 de Hotelling para el caso de una y dos poblaciones cuando los datos provienen de una distribución normal. Tesis de maestría en estadística. Universidad Nacional de Colombia. 2009. p. 10-12.

QIN, J. Survey on data-driven industrial process monitoring and diagnosis. 2012. [Citado Julio 2016], 1 p.

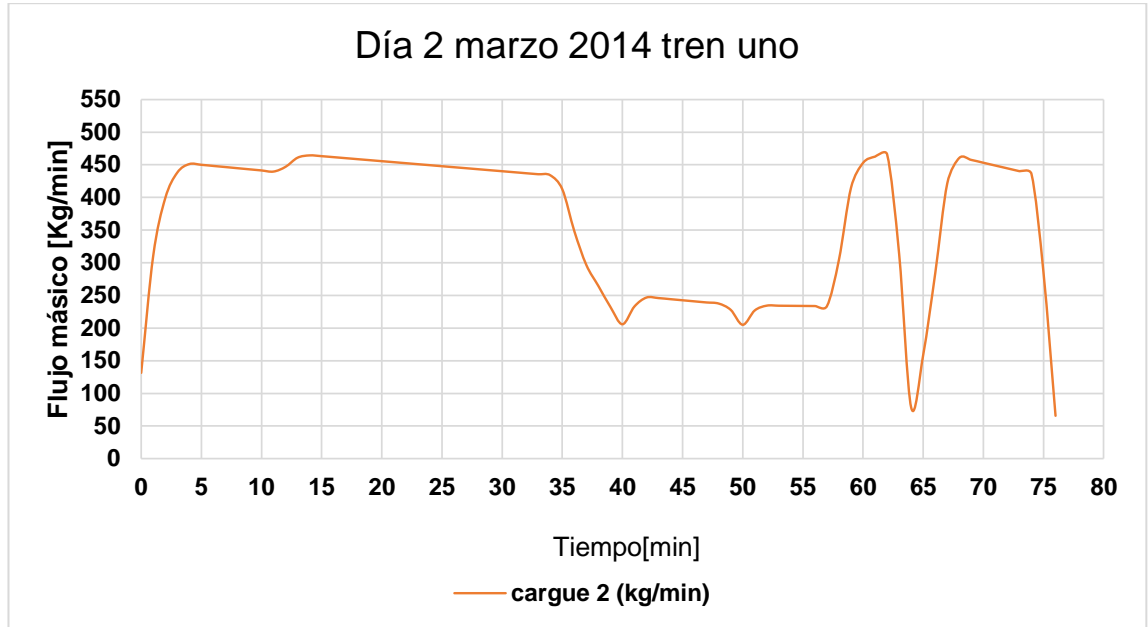
RUIZ MONCADA, G. Análisis y optimización del sistema de entrega de Propileno en el departamento de materias primas de la gerencia refinería de Barrancabermeja. Tesis de maestría en Ingeniería química. Universidad Industrial de Santander, 2015; p. 5-10.

SMITH, L. Principal Components Analysis. 2002, p. 12-20. [Citado Julio 2016]. Disponible en http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

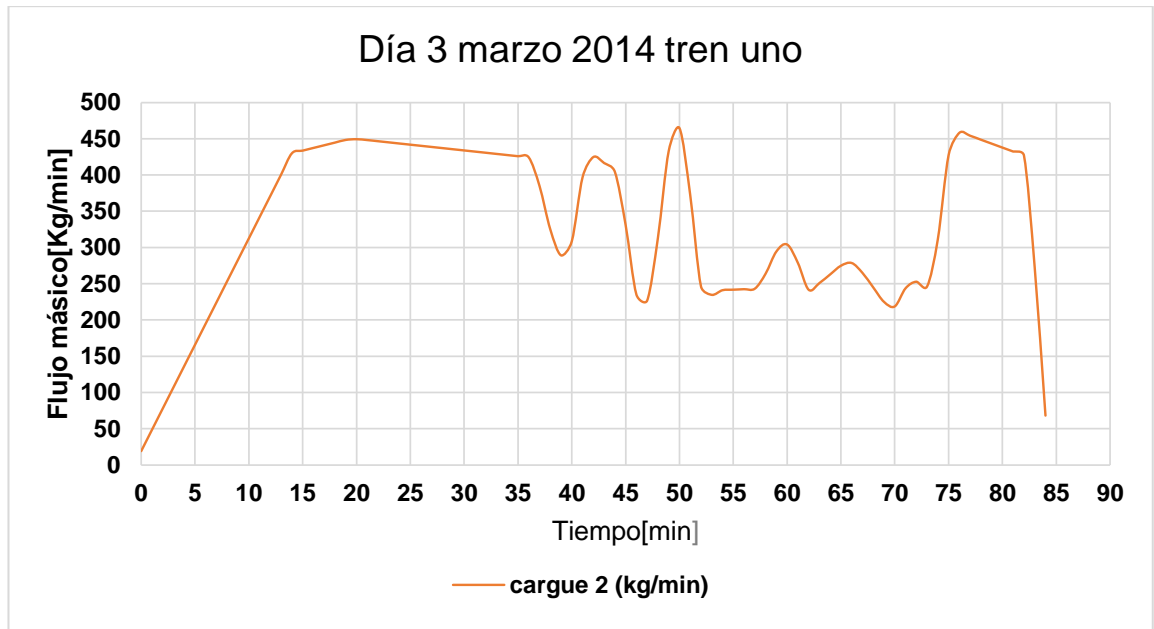
YIN, S., DING, S., XIE, X., LUO, H. A review on Basic Data-Driven Approaches for Industrial Process Monitoring. Vol 61, No 11, 2014. [Citado Julio 2016], 1 p.

ANEXOS

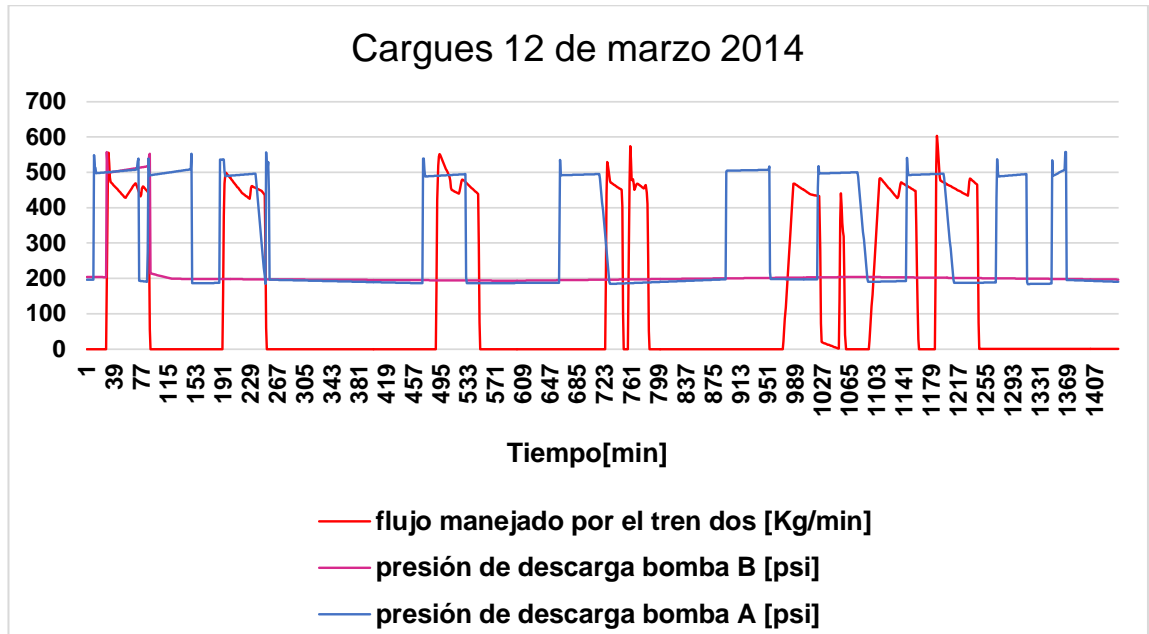
ANEXO A. CARGUE COMPORTAMIENTO ATÍPICO DE FLUJO MÁSSICO.



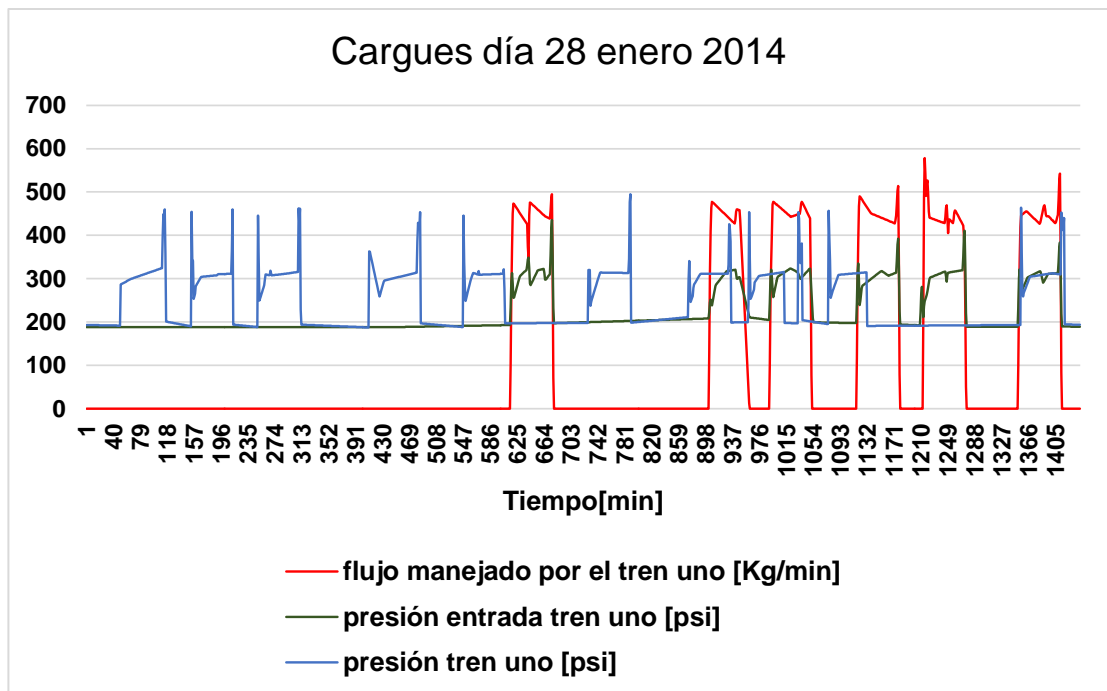
ANEXO B. CARGUE COMPORTAMIENTO ATÍPICO DE FLUJO MÁSSICO



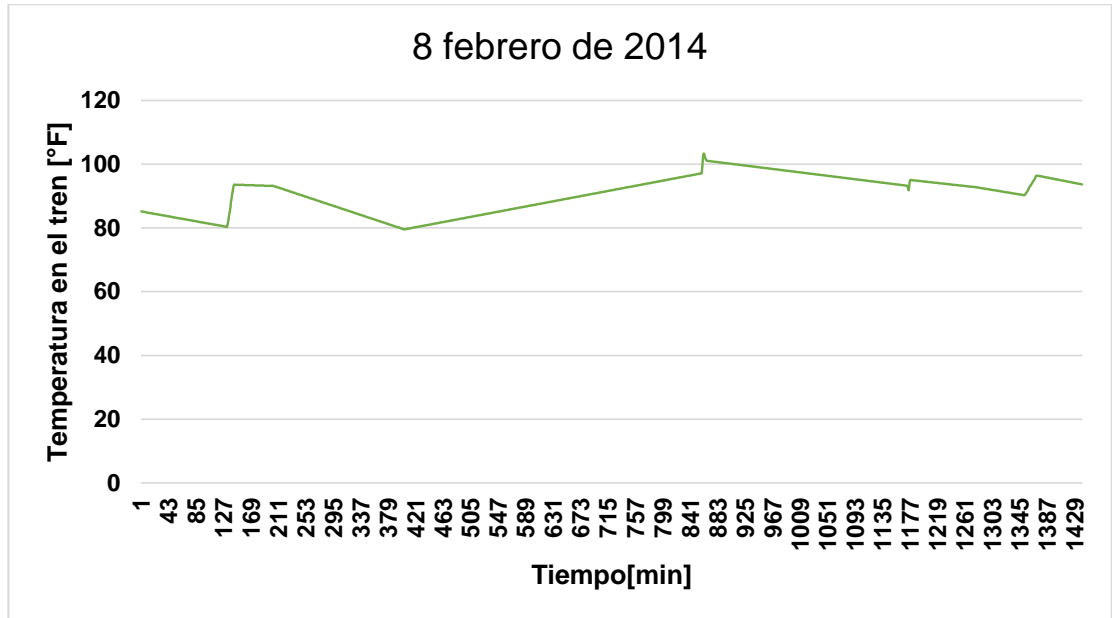
ANEXO C. GRÁFICA DE TENDENCIA DE PRESIÓN DE DESCARGA DE LAS BOMBAS EN CARGUES CON SENSORES FUERA LÍNEA.



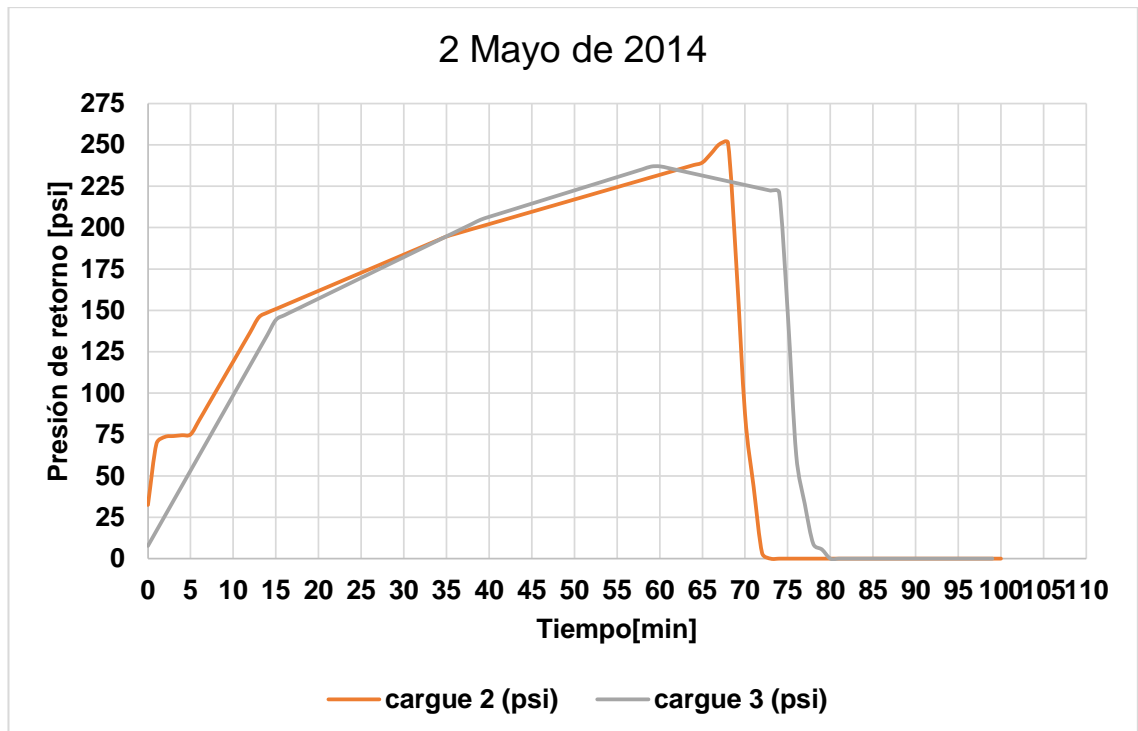
ANEXO D. GRÁFICA DE TENDENCIAS DE PRESIÓN EN LOS TRENES Y PRESIÓN A LA ENTRADA DE LOS TRENES CON LOS SENSORES FUERA DE LÍNEA.



ANEXO E. GRÁFICA DE TENDENCIAS DE TEMPERATURA EN EL TREN UNO



ANEXO F. GRÁFICA DE LA TENDENCIA EN UN CARGUE ESTÁNDAR DE LA PRESIÓN DE RETORNO.



ANEXO G. PCA

El análisis de Componentes Principales (PCA) es un procedimiento estadístico que utiliza una transformación ortogonal para convertir un conjunto de observaciones de variables correlacionadas, en un conjunto de valores posibles con forma lineal y no correlacionadas entre sí, a estas variables se les llama componentes principales. El número de componentes principales es menor o igual al número de variables originales. Esta transformación se define de tal manera que el primer componente principal tiene la mayor varianza posible (es decir, representa como gran parte de la variabilidad en los datos como sea posible), y cada componente subsiguiente a su vez tiene la mayor varianza posible bajo la restricción que es ortogonal a los componentes anteriores. Los vectores resultantes son un no correlacionados conjunto de base ortogonal. El PCA es sensible a la escala relativa de las variables originales.

Los únicos requerimientos previos para la aplicación del ACP son: a) Continuidad en las variables. b) El número n de individuos o elementos observados debe ser mayor que el número p de variables originales. Por otra parte, el ACP tiene la ventaja de no exigir supuestos tales como la normalidad u homoscedasticidad.

Existen dos formas básicas de aplicar el PCA:

Método basado en la matriz de correlación, cuando los datos no son dimensionalmente homogéneos o el orden de magnitud de las variables aleatorias medidas no es el mismo.

Método basado en la matriz de covarianzas, que se usa cuando los datos son dimensionalmente homogéneos y presentan valores medios similares.

Cálculo de los Componentes Principales

Se considera una serie de variables (x_1, x_2, \dots, x_p) sobre un grupo de objetos o individuos y se trata de calcular, a partir de ellas, un nuevo conjunto de variables y_1, y_2, \dots, y_p , incorreladas entre sí, cuyas varianzas vayan decreciendo progresivamente.

Cada y_j (donde $j = 1, \dots, p$) es una combinación lineal de las x_1, x_2, \dots, x_p originales, es decir:

$$\begin{aligned}y_j &= a_{j1}x_1 + a_{j2}x_2 + \dots + a_{jp}x_p = \\ &= \mathbf{a}'_j \mathbf{x}\end{aligned}$$

siendo $\mathbf{a}'_j = (a_{1j}, a_{2j}, \dots, a_{pj})$ un vector de constantes, y

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$$

Obviamente, si lo que queremos es maximizar la varianza, como veremos luego, una forma simple podría ser aumentar los coeficientes a_{ij} . Por ello, para mantener la ortogonalidad de la transformación se impone que el módulo del vector $\mathbf{a}'_j = (a_{1j}, a_{2j}, \dots, a_{pj})$ sea

6

1. Es decir,

$$\mathbf{a}'_j \mathbf{a}_j = \sum_{k=1}^p a_{kj}^2 = 1$$

El primer componente se calcula eligiendo \mathbf{a}_1 de modo que y_1 tenga la mayor varianza posible, sujeta a la restricción de que $\mathbf{a}'_1 \mathbf{a}_1 = 1$. El segundo componente principal se calcula obteniendo \mathbf{a}_2 de modo que la variable obtenida, y_2 esté incorrelada con y_1 .

Del mismo modo se eligen y_1, y_2, \dots, y_p , incorrelados entre sí, de manera que las variables aleatorias obtenidas vayan teniendo cada vez menor varianza.

18

¹⁸ E, LAURA. Análisis multivariado: método de componentes principales. Secretaria general de la OEA, Washington, DC.

ANEXO H. ESTADISTICA DESCRIPTIVA MATRIZ DEL SISTEMA Y SOPORTE DE LOS DATOS INGRESADOS PARA APLICACIÓN DE PCA

flujo masico		presión a la entrada del tren		presión de descarga de la bomba		presión del tren		temperatura del tren		presión de retorno	
<i>cargue1</i>		<i>cargue1</i>		<i>cargue1</i>		<i>cargue1</i>		<i>cargue1</i>		<i>cargue1</i>	
Media	434.557958	Media	434.557958	Media	434.557958	Media	434.557958	Media	434.557958	Media	434.557958
Error típico	10.3934847	Error típico	10.3934847	Error típico	10.3934847	Error típico	10.3934847	Error típico	10.3934847	Error típico	10.3934847
Mediana	453.119919	Mediana	453.119919	Mediana	453.119919	Mediana	453.119919	Mediana	453.119919	Mediana	453.119919
Moda	#N/A	Moda	#N/A	Moda	#N/A	Moda	#N/A	Moda	#N/A	Moda	#N/A
Desviación estándar	79.1544216	Desviación estándar	79.1544216	Desviación estándar	79.1544216	Desviación estándar	79.1544216	Desviación estándar	79.1544216	Desviación estándar	79.1544216
Varianza de la muestra	6265.42246	Varianza de la muestra	6265.42246	Varianza de la muestra	6265.42246	Varianza de la muestra	6265.42246	Varianza de la muestra	6265.42246	Varianza de la muestra	6265.42246
Curtosis	19.8888814	Curtosis	19.8888814	Curtosis	19.8888814	Curtosis	19.8888814	Curtosis	19.8888814	Curtosis	19.8888814
Coefficiente de asimetría	-4.44963096	Coefficiente de asimetría	-4.44963096	Coefficiente de asimetría	-4.44963096	Coefficiente de asimetría	-4.44963096	Coefficiente de asimetría	-4.44963096	Coefficiente de asimetría	-4.44963096
Rango	450.033443	Rango	450.033443	Rango	450.033443	Rango	450.033443	Rango	450.033443	Rango	450.033443
Mínimo	30.1103558	Mínimo	30.1103558	Mínimo	30.1103558	Mínimo	30.1103558	Mínimo	30.1103558	Mínimo	30.1103558
Máximo	480.143799	Máximo	480.143799	Máximo	480.143799	Máximo	480.143799	Máximo	480.143799	Máximo	480.143799
Suma	25204.3616	Suma	25204.3616	Suma	25204.3616	Suma	25204.3616	Suma	25204.3616	Suma	25204.3616
Cuenta	58	Cuenta	58	Cuenta	58	Cuenta	58	Cuenta	58	Cuenta	58
<i>cargue2</i>		<i>cargue2</i>		<i>cargue2</i>		<i>cargue2</i>		<i>cargue2</i>		<i>cargue2</i>	
Media	444.401074	Media	444.401074	Media	444.401074	Media	444.401074	Media	444.401074	Media	444.401074
Error típico	4.6746395	Error típico	4.6746395	Error típico	4.6746395	Error típico	4.6746395	Error típico	4.6746395	Error típico	4.6746395
Mediana	446.408468	Mediana	446.408468	Mediana	446.408468	Mediana	446.408468	Mediana	446.408468	Mediana	446.408468
Moda	#N/A	Moda	#N/A	Moda	#N/A	Moda	#N/A	Moda	#N/A	Moda	#N/A
Desviación estándar	35.6009938	Desviación estándar	35.6009938	Desviación estándar	35.6009938	Desviación estándar	35.6009938	Desviación estándar	35.6009938	Desviación estándar	35.6009938
Varianza de la muestra	1267.43076	Varianza de la muestra	1267.43076	Varianza de la muestra	1267.43076	Varianza de la muestra	1267.43076	Varianza de la muestra	1267.43076	Varianza de la muestra	1267.43076
Curtosis	43.0532134	Curtosis	43.0532134	Curtosis	43.0532134	Curtosis	43.0532134	Curtosis	43.0532134	Curtosis	43.0532134
Coefficiente de asimetría	-6.12612194	Coefficiente de asimetría	-6.12612194	Coefficiente de asimetría	-6.12612194	Coefficiente de asimetría	-6.12612194	Coefficiente de asimetría	-6.12612194	Coefficiente de asimetría	-6.12612194
Rango	275.258016	Rango	275.258016	Rango	275.258016	Rango	275.258016	Rango	275.258016	Rango	275.258016
Mínimo	196.069438	Mínimo	196.069438	Mínimo	196.069438	Mínimo	196.069438	Mínimo	196.069438	Mínimo	196.069438
Máximo	471.327454	Máximo	471.327454	Máximo	471.327454	Máximo	471.327454	Máximo	471.327454	Máximo	471.327454
Suma	25775.2623	Suma	25775.2623	Suma	25775.2623	Suma	25775.2623	Suma	25775.2623	Suma	25775.2623
Cuenta	58	Cuenta	58	Cuenta	58	Cuenta	58	Cuenta	58	Cuenta	58
<i>cargue3</i>		<i>cargue3</i>		<i>cargue3</i>		<i>cargue3</i>		<i>cargue3</i>		<i>cargue3</i>	
Media	436.676229	Media	436.676229	Media	436.676229	Media	436.676229	Media	436.676229	Media	436.676229
Error típico	5.88074962	Error típico	5.88074962	Error típico	5.88074962	Error típico	5.88074962	Error típico	5.88074962	Error típico	5.88074962
Mediana	441.854094	Mediana	441.854094	Mediana	441.854094	Mediana	441.854094	Mediana	441.854094	Mediana	441.854094
Moda	#N/A	Moda	#N/A	Moda	#N/A	Moda	#N/A	Moda	#N/A	Moda	#N/A
Desviación estándar	44.7864548	Desviación estándar	44.7864548	Desviación estándar	44.7864548	Desviación estándar	44.7864548	Desviación estándar	44.7864548	Desviación estándar	44.7864548
Varianza de la muestra	2005.82653	Varianza de la muestra	2005.82653	Varianza de la muestra	2005.82653	Varianza de la muestra	2005.82653	Varianza de la muestra	2005.82653	Varianza de la muestra	2005.82653
Curtosis	30.272779	Curtosis	30.272779	Curtosis	30.272779	Curtosis	30.272779	Curtosis	30.272779	Curtosis	30.272779
Coefficiente de asimetría	-4.75462792	Coefficiente de asimetría	-4.75462792	Coefficiente de asimetría	-4.75462792	Coefficiente de asimetría	-4.75462792	Coefficiente de asimetría	-4.75462792	Coefficiente de asimetría	-4.75462792
Rango	377.12736	Rango	377.12736	Rango	377.12736	Rango	377.12736	Rango	377.12736	Rango	377.12736
Mínimo	152.002828	Mínimo	152.002828	Mínimo	152.002828	Mínimo	152.002828	Mínimo	152.002828	Mínimo	152.002828
Máximo	529.130188	Máximo	529.130188	Máximo	529.130188	Máximo	529.130188	Máximo	529.130188	Máximo	529.130188
Suma	25327.2213	Suma	25327.2213	Suma	25327.2213	Suma	25327.2213	Suma	25327.2213	Suma	25327.2213
Cuenta	58	Cuenta	58	Cuenta	58	Cuenta	58	Cuenta	58	Cuenta	58

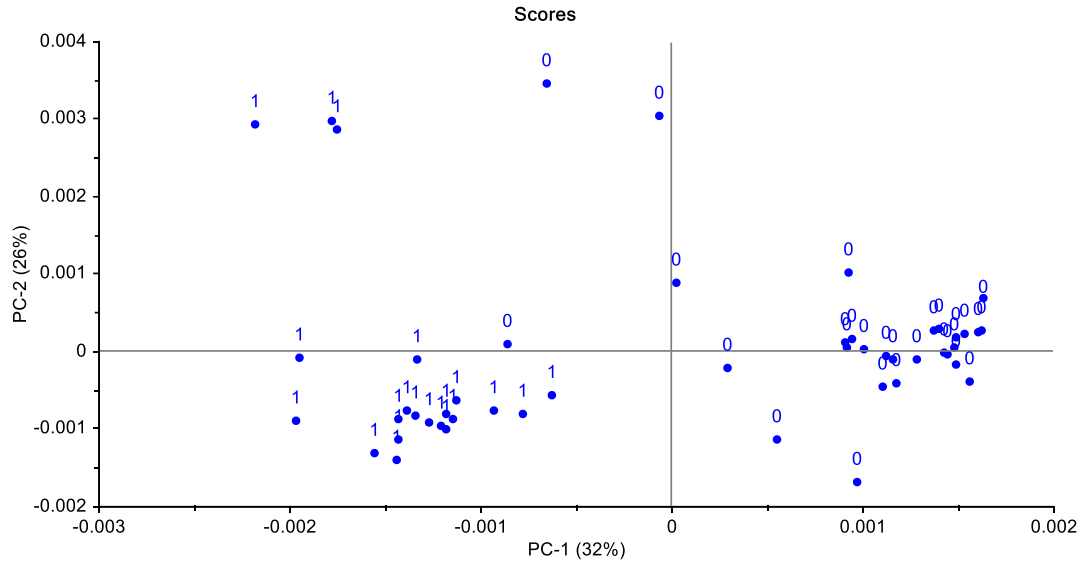
<i>cargue4</i>		<i>cargue4</i>		<i>cargue4</i>		<i>cargue4</i>		<i>cargue4</i>		<i>cargue4</i>	
Media	435.945637	Media	435.945637	Media	435.945637	Media	435.945637	Media	435.945637	Media	435.945637
Error típico	6.82463655	Error típico	6.82463655	Error típico	6.82463655	Error típico	6.82463655	Error típico	6.82463655	Error típico	6.82463655
Mediana	429.670183	Mediana	429.670183	Mediana	429.670183	Mediana	429.670183	Mediana	429.670183	Mediana	429.670183
Moda	#N/A	Moda	#N/A	Moda	#N/A	Moda	#N/A	Moda	#N/A	Moda	#N/A
Desviación estándar	51.9748835	Desviación estándar	51.9748835	Desviación estándar	51.9748835	Desviación estándar	51.9748835	Desviación estándar	51.9748835	Desviación estándar	51.9748835
Varianza de la muestra	2701.38851	Varianza de la muestra	2701.38851	Varianza de la muestra	2701.38851	Varianza de la muestra	2701.38851	Varianza de la muestra	2701.38851	Varianza de la muestra	2701.38851
Curtosis	14.9140246	Curtosis	14.9140246	Curtosis	14.9140246	Curtosis	14.9140246	Curtosis	14.9140246	Curtosis	14.9140246
Coefficiente de asimetría	-2.23892114	Coefficiente de asimetría	-2.23892114	Coefficiente de asimetría	-2.23892114	Coefficiente de asimetría	-2.23892114	Coefficiente de asimetría	-2.23892114	Coefficiente de asimetría	-2.23892114
Rango	388.537011	Rango	388.537011	Rango	388.537011	Rango	388.537011	Rango	388.537011	Rango	388.537011
Mínimo	156.045813	Mínimo	156.045813	Mínimo	156.045813	Mínimo	156.045813	Mínimo	156.045813	Mínimo	156.045813
Máximo	544.582825	Máximo	544.582825	Máximo	544.582825	Máximo	544.582825	Máximo	544.582825	Máximo	544.582825
Suma	25284.8469	Suma	25284.8469	Suma	25284.8469	Suma	25284.8469	Suma	25284.8469	Suma	25284.8469
Cuenta	58	Cuenta	58	Cuenta	58	Cuenta	58	Cuenta	58	Cuenta	58
<i>cargue5</i>		<i>cargue5</i>		<i>cargue5</i>		<i>cargue5</i>		<i>cargue5</i>		<i>cargue5</i>	
Media	425.876509	Media	425.876509	Media	425.876509	Media	425.876509	Media	425.876509	Media	425.876509
Error típico	4.98721092	Error típico	4.98721092	Error típico	4.98721092	Error típico	4.98721092	Error típico	4.98721092	Error típico	4.98721092
Mediana	429.082736	Mediana	429.082736	Mediana	429.082736	Mediana	429.082736	Mediana	429.082736	Mediana	429.082736
Moda	#N/A	Moda	#N/A	Moda	#N/A	Moda	#N/A	Moda	#N/A	Moda	#N/A
Desviación estándar	37.9814668	Desviación estándar	37.9814668	Desviación estándar	37.9814668	Desviación estándar	37.9814668	Desviación estándar	37.9814668	Desviación estándar	37.9814668
Varianza de la muestra	1442.59182	Varianza de la muestra	1442.59182	Varianza de la muestra	1442.59182	Varianza de la muestra	1442.59182	Varianza de la muestra	1442.59182	Varianza de la muestra	1442.59182
Curtosis	41.6277776	Curtosis	41.6277776	Curtosis	41.6277776	Curtosis	41.6277776	Curtosis	41.6277776	Curtosis	41.6277776
Coefficiente de asimetría	-6.12866855	Coefficiente de asimetría	-6.12866855	Coefficiente de asimetría	-6.12866855	Coefficiente de asimetría	-6.12866855	Coefficiente de asimetría	-6.12866855	Coefficiente de asimetría	-6.12866855
Rango	283.489689	Rango	283.489689	Rango	283.489689	Rango	283.489689	Rango	283.489689	Rango	283.489689
Mínimo	163.530422	Mínimo	163.530422	Mínimo	163.530422	Mínimo	163.530422	Mínimo	163.530422	Mínimo	163.530422
Máximo	447.020111	Máximo	447.020111	Máximo	447.020111	Máximo	447.020111	Máximo	447.020111	Máximo	447.020111
Suma	24700.8375	Suma	24700.8375	Suma	24700.8375	Suma	24700.8375	Suma	24700.8375	Suma	24700.8375
Cuenta	58	Cuenta	58	Cuenta	58	Cuenta	58	Cuenta	58	Cuenta	58
<i>cargue6</i>		<i>cargue6</i>		<i>cargue6</i>		<i>cargue6</i>		<i>cargue6</i>		<i>cargue6</i>	
Media	371.471324	Media	371.471324	Media	371.471324	Media	371.471324	Media	371.471324	Media	371.471324
Error típico	14.3639111	Error típico	14.3639111	Error típico	14.3639111	Error típico	14.3639111	Error típico	14.3639111	Error típico	14.3639111
Mediana	405.097473	Mediana	405.097473	Mediana	405.097473	Mediana	405.097473	Mediana	405.097473	Mediana	405.097473
Moda	#N/A	Moda	#N/A	Moda	#N/A	Moda	#N/A	Moda	#N/A	Moda	#N/A
Desviación estándar	109.392288	Desviación estándar	109.392288	Desviación estándar	109.392288	Desviación estándar	109.392288	Desviación estándar	109.392288	Desviación estándar	109.392288
Varianza de la muestra	11966.6727	Varianza de la muestra	11966.6727	Varianza de la muestra	11966.6727	Varianza de la muestra	11966.6727	Varianza de la muestra	11966.6727	Varianza de la muestra	11966.6727
Curtosis	2.75834288	Curtosis	2.75834288	Curtosis	2.75834288	Curtosis	2.75834288	Curtosis	2.75834288	Curtosis	2.75834288
Coefficiente de asimetría	-1.94756804	Coefficiente de asimetría	-1.94756804	Coefficiente de asimetría	-1.94756804	Coefficiente de asimetría	-1.94756804	Coefficiente de asimetría	-1.94756804	Coefficiente de asimetría	-1.94756804
Rango	432.282809	Rango	432.282809	Rango	432.282809	Rango	432.282809	Rango	432.282809	Rango	432.282809
Mínimo	24.3046549	Mínimo	24.3046549	Mínimo	24.3046549	Mínimo	24.3046549	Mínimo	24.3046549	Mínimo	24.3046549
Máximo	456.587464	Máximo	456.587464	Máximo	456.587464	Máximo	456.587464	Máximo	456.587464	Máximo	456.587464
Suma	21545.3368	Suma	21545.3368	Suma	21545.3368	Suma	21545.3368	Suma	21545.3368	Suma	21545.3368
Cuenta	58	Cuenta	58	Cuenta	58	Cuenta	58	Cuenta	58	Cuenta	58

Matriz de datos, el análisis estadístico completo de la matriz y la matriz con los pretratamientos aplicados para el análisis de componentes principales por su tamaño y extensión se adjuntan al documento de Excel llamado matriz.

De la evaluación de las pruebas estadísticas preliminares a la matriz de datos y las pruebas de validación de varianza arrojadas por el programa, en el cual se muestran independencia de los componentes, una variabilidad dentro de los rangos de aplicación se puede decir que el PCA se puede aplicar al sistema de llenado U-3160.

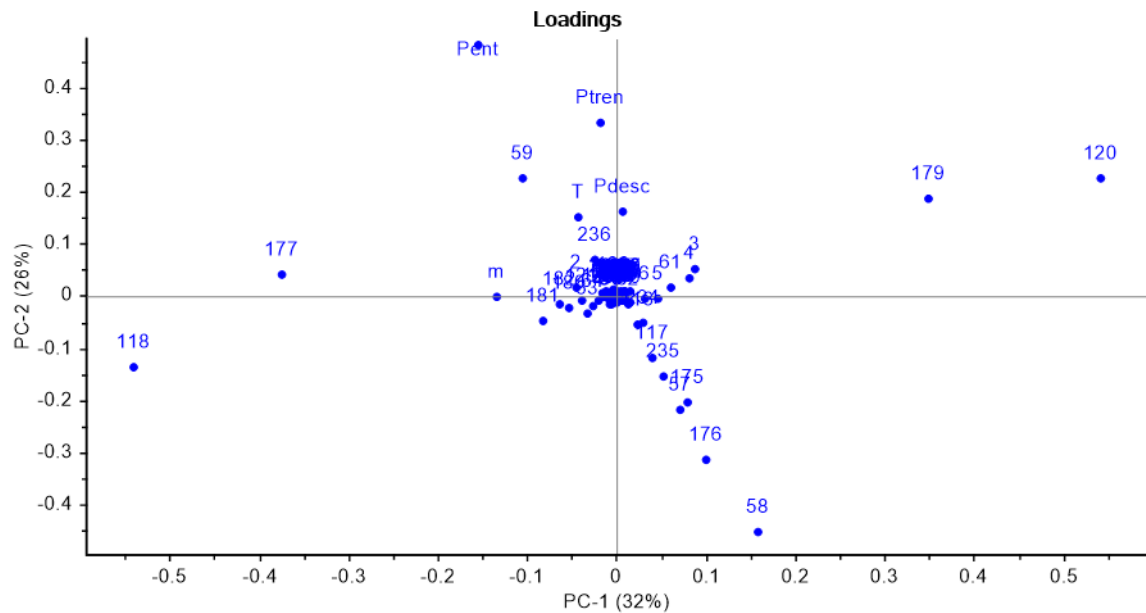
ANEXO I.

GRAFICA SCORES COMPONENTE 1 Y 2



Scores es un tipo de resultado arrojado por el programa The unscrambler X de CAMO, en el que se agrupan las tendencias por cargue.

GRAFICA DE LOADINGS COMPONENTES 1 Y 2



Loading son un tipo de resultado arrojado por el programa The unscrambler X de CAMO, en el que se arrojan la dispersión de datos de las diferentes variables.

TABLAS MOSTRADAS POR EL PROGRAMA PARA SCORES Y LOADING.

No se muestran las tablas completas puesto que los loadings son la cantidad de datos ingresados como matriz y los scores son 89 que representan cada cargue. Se adjunta archivo PCA llenadero U-3160 del programa *The Unscrambler® X*. para que pueda ser consultado.

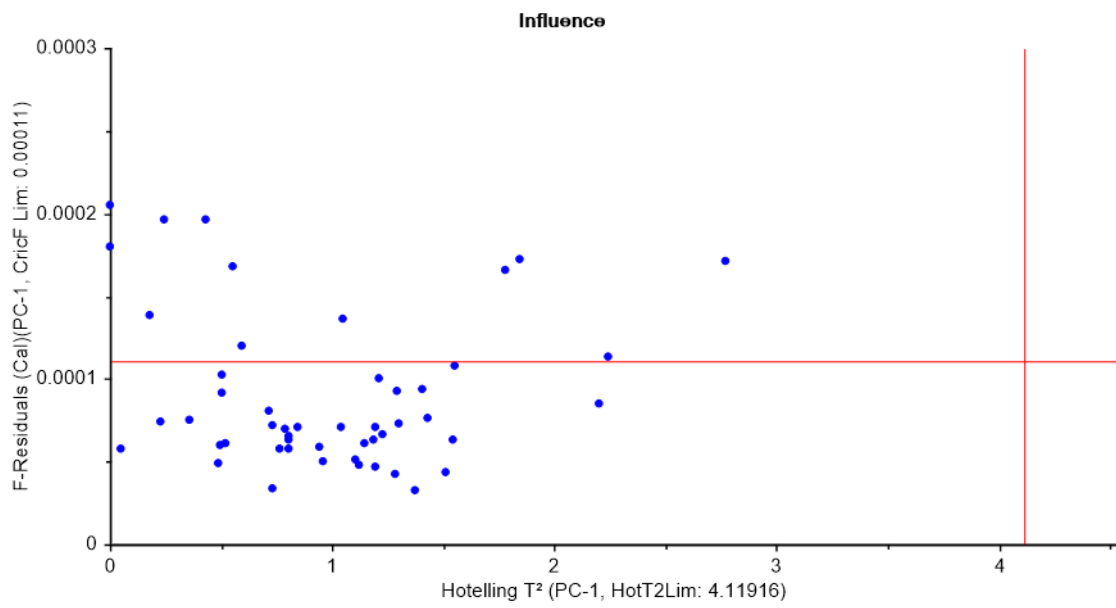
New Project [Clasificacion.unsb] - The Unscrambler X

File Edit View Insert Plot Tasks Tools Help

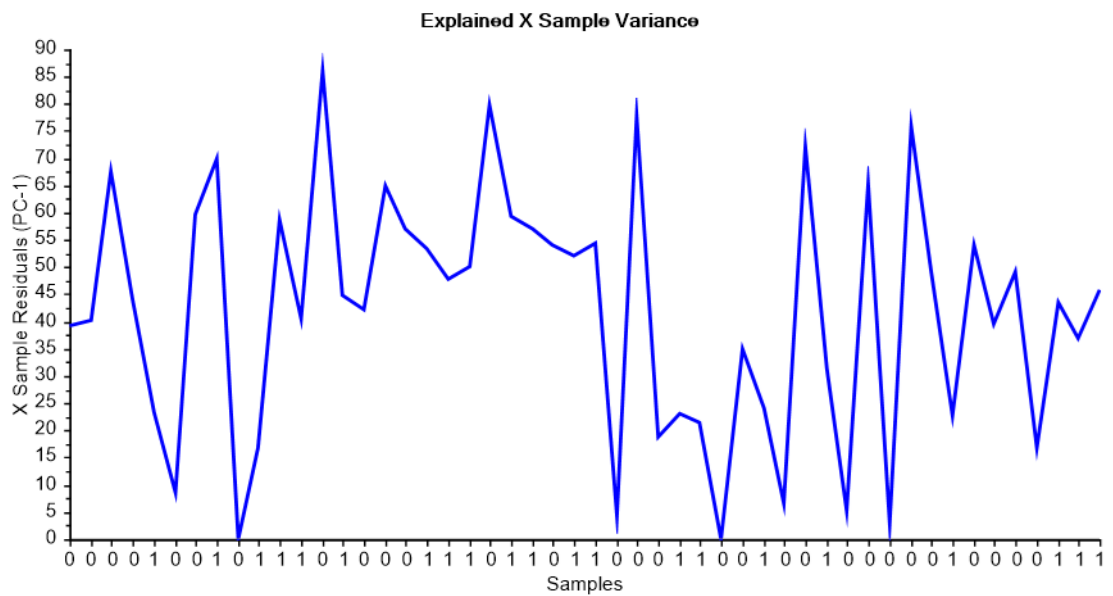
X: PC-1 Y: PC-2

		PC-1	PC-2		PC-1	PC-2	
		1	2	X Loadings	1	2	
0	1	0.0016	0.0007	m	1	-0.1344	0.0000
0	2	0.0009	0.0001		2	-0.0442	0.0151
0	3	0.0014	0.0003		3	0.0878	0.0521
0	4	0.0012	-0.0001		4	0.0812	0.0325
1	5	-0.0008	-0.0008		5	0.0471	-0.0045
0	6	0.0010	-0.0017		6	0.0325	-0.0039
0	7	0.0014	0.0003		7	0.0032	0.0014
1	8	-0.0014	-0.0008		8	-0.0125	-0.0029
0	9	0.0000	0.0009		9	0.0007	-0.0120
1	10	-0.0006	-0.0006		10	0.0156	-0.0122
1	11	-0.0014	-0.0009		11	0.0114	-0.0089
1	12	-0.0011	-0.0006		12	0.0049	-0.0082
0	13	0.0015	0.0002		13	0.0019	-0.0031
1	14	-0.0012	-0.0010		14	0.0039	0.0002
0	15	0.0015	-0.0002		15	0.0077	-0.0073
0	16	0.0016	0.0003		16	-0.0047	-0.0057
0	17	0.0015	0.0000		17	-0.0159	0.0041
1	18	-0.0014	-0.0011		18	-0.0104	0.0023
1	19	-0.0012	-0.0010		19	0.0015	-0.0064
1	20	-0.0013	-0.0001		20	0.0074	-0.0079
0	21	0.0016	0.0002		21	0.0056	-0.0009
1	22	-0.0019	-0.0001		22	-0.0018	-0.0012
1	23	-0.0013	-0.0009		23	-0.0129	-0.0032
0	24	0.0016	-0.0004		24	-0.0104	0.0011
1	25	-0.0011	-0.0009		25	0.0062	0.0046
1	26	-0.0012	-0.0008		26	0.0152	0.0073
0	27	0.0006	-0.0011		27	0.0042	0.0104
0	28	0.0015	0.0000		28	-0.0079	0.0031
0	29	0.0009	0.0010		29	-0.0075	-0.0099
1	30	-0.0018	0.0030		30	-0.0061	-0.0152
1	31	-0.0013	-0.0008		31	-0.0063	-0.0156
0	32	-0.0001	0.0030		32	-0.0042	-0.0081
0	33	0.0011	-0.0005		33	-0.0041	0.0044
1	34	-0.0017	0.0029		34	-0.0033	0.0072
0	35	0.0003	-0.0002		35	-0.0004	0.0041
0	36	0.0014	0.0000		36	-0.0008	0.0024
1	37	-0.0022	0.0029		37	-0.0014	-0.0013

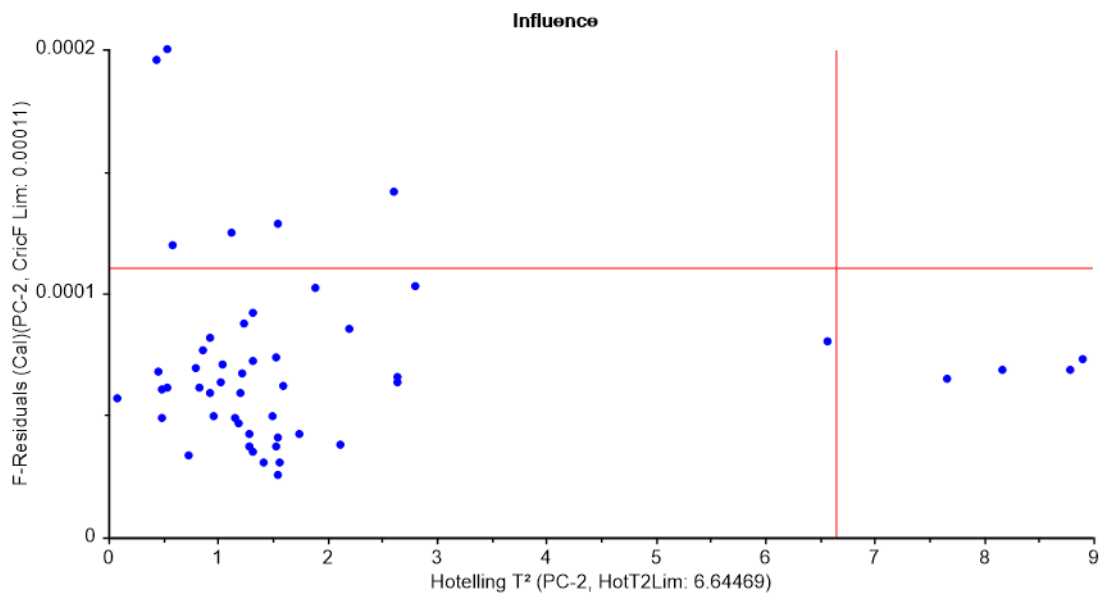
GRAFICA DE INFLUENCIA PUNTOS DE TENDENCIA COMPONENTE 1



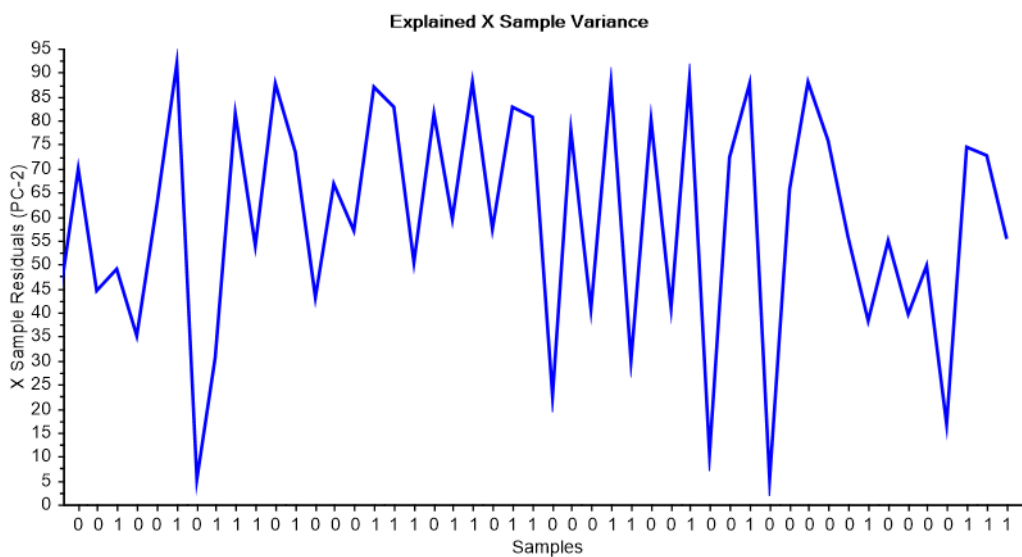
GRAFICA DE VARIANZA COMPONENTE 1



GRAFICA INFLUENCIA COMPONENTE 2



GRAFICA DE VARIANZA SEGUNDO COMPONENTE



Para corroborar la aplicación del PCA el programa genera las pruebas de validación se remite al adjunto del libro en los test de validación de varianza, archivo PCA llenadero U-3160 del programa *The Unscrambler*® X.