

PCA y Lógica Fuzzy para la detección de eventos anormales de operación, en una planta de Craqueo Catalítico Fluidizado FCC

Sandra Zambrano^{*}, Cristhiam Higuera^{**},

Rodolfo Villamizar^{***}, Carlos Agudelo^{*}.

Fecha de Recibido: 21/04/2011 Fecha de aprobación: 16/08/2011

Resumen

En el presente artículo se aplican y ajustan algunas técnicas y metodologías para la detección de fallos en una planta de Cracking Catalítico Fluidizado (FCC) modelo UOP, inicialmente se realiza un análisis de confiabilidad que permite definir el nivel de criticidad de cada uno de los equipos e identificar modos de fallo potenciales y su efecto sobre la operación de la planta; con la información entregada por este análisis se establecen casos de estudio, como requerimientos para un sistema de supervisión, detección y clasificación de situaciones anómalas, que pueden llevar al proceso a una condición de fallo. Para estudiar los casos se simulan las condiciones con un modelo de operación de la unidad FCC y se detectan los fallos con una herramienta basada en PCA (Análisis de componentes principales) y Lógica Fuzzy; finalmente se ajusta la herramienta utilizando datos históricos del proceso en presencia del fallo.

Palabras clave: *Análisis de componentes principales (PCA), Análisis de modos y efectos de falla (FMEA), Detección de Fallos, Craqueo Catalítico Fluidizado (FCC), Clustering basadas en Lógica Fuzzy (CLF), Knowledge Discovery in Databases (KDD), Técnicas Estadísticas Multivariadas (TEM).*

Abstract

In the present article we apply and adjust some techniques and methodologies for detecting faults in a Fluidized Catalytic Cracking refinery process (FCC) UOP model, initially performing a reliability analysis that defines the level of criticality of each equipment and identify potential fault modes and their effect on the operation of the process, with the information provided by this analysis, case studies are set as requirements for a monitoring, detection and classification of anomalous situations, which may lead to process a fault condition. To study cases, conditions are simulated with a model of the FCC unit operation and faults are detected with a tool based on PCA (Principal Component Analysis) and Fuzzy logic, and finally the tool is adjusted using historical process data in the presence of the fault.

Keywords: *Principal component analysis (PCA), failure modes and effects analysis (FMEA), fault detection, Fluid Catalytic Cracking (FCC), Fuzzy Logic based Clustering (FLC), Knowledge Discovery in Databases (KDD), multivariate statistical techniques (MST).*

* Instituto Colombiano del Petróleo, Centro de I+D, Kilómetro 7 Vía a Piedecuesta, Santander, Colombia., {sandra.zambrano, carlos.agudelo}@ecopetrol.com.co

** Ingeniero Electrónico, investigador de la supervisión experta y control de procesos, cristhiamhiguera@gmail.com

*** Universidad Industrial de Santander, Escuela de Ingenierías Eléctrica Electrónica y de Telecomunicaciones, kra 27 calle 9, Bucaramanga, Santander, Colombia., Profesor Asociado: rovillam@uis.edu.co

§Se concede autorización para copiar gratuitamente parte o todo el material publicado en la *Revista Colombiana de Computación* siempre y cuando las copias no sean usadas para fines comerciales, y que se especifique que la copia se realiza con el consentimiento de la *Revista Colombiana de Computación*.

1. Introducción

Actualmente diferentes áreas de ingeniería enfrentan un aumento en la demanda de seguridad y confiabilidad de los procesos. Contar con un sistema de identificación y detección de fallos anticipados proporciona beneficios orientados a corregir desviaciones del proceso productivo, evitar acontecimientos anormales, reducir pérdidas de productividad, disminuir las paradas no programadas y principalmente pronosticar situaciones que pongan en peligro a las personas, el proceso o el medio ambiente.

Existe literatura asociada al proceso de identificación y detección de fallos, algunas técnicas incluyen el uso de inteligencia artificial, estadísticos de fallos, entre otros; que relacionan los síntomas observados en los procesos reales con las fallas que los causaron. Los sistemas industriales automatizados son vulnerables a fallos en su instrumentación por sensores, actuadores, controladores y perturbaciones externas [1].

En la industria se ha documentado el impacto económico asociado al manejo inapropiado de situaciones anómalas haciendo énfasis en la necesidad de automatizar las tareas requeridas para la detección temprana y el diagnóstico de fallos, utilizando toda la información disponible en tiempo real. Existen muchas técnicas en la detección temprana y el diagnóstico de fallos, algunas de estas técnicas incluyen el uso de sistemas con bases de conocimiento que contienen las relaciones causales entre los síntomas observados en los procesos reales y las fallas que los causaron [1,2].

El objetivo de la detección y aislamiento de fallos es inferir sobre el comportamiento del proceso cuando ha ocurrido un fallo para generar hipótesis y acciones a partir de los síntomas observados. El proceso de detección y aislamiento de fallos cuenta con varios niveles: detección del fallo, indica que ha ocurrido un fallo en alguna parte del sistema, aislamiento del fallo, muestra específicamente la localización del fallo en el sistema, finalmente la identificación, cuantifica la magnitud y naturaleza del fallo [3]. El alcance del presente artículo es la detección del fallo en un caso de estudio con alto nivel de criticidad de una unidad FCC modelo UOP.

2. Proceso de Craqueo Catalítico Fluidizado (Fcc)

Mediante la ruptura catalítica, se llevan a cabo procesos de conversión primaria, donde por rompimiento molecular de fracciones del petróleo,

se obtienen productos de mayor valor comercial. Entre los productos valiosos de esta reacción se encuentran el gas combustible, Etano-Etileno, Gas Licuado del Petróleo (GLP), Gasolina (liviana y pesada), Aceite liviano de Ciclo (ALC) Aceite Pesado de ciclo (APC) y Slurry.

El Proceso de FCC emplea un catalizador que se comporta como un fluido cuando está aireado con vapor o aire. El catalizador fluidizado es continuamente circulado desde la zona de reacción (reactor), donde ocurre el rompimiento de partículas hasta la zona de regeneración (regenerador), donde el catalizador es reactivado. La acción del catalizador también consiste en servir de vehículo para transferir calor desde el regenerador a la zona de reacción. La carga precalentada y atomizada fluye por el elevador donde se encuentra con el catalizador regenerado. Finalmente se cuenta con una unidad fraccionadora donde se recuperan los productos valiosos de la reacción en corrientes de APC, ALC, nafta y gases.

El proceso FCC rompe moléculas de hidrocarburo de alto peso molecular (punto de destilación superior a los 360°C) y las transforma en productos más valiosos tales como etileno, propileno, gasolina y gasóleo (diesel). La reacción tiene lugar a temperaturas elevadas (superiores a los 500°C) en presencia de un catalizador. Es uno de los procesos de conversión más importantes en la industria de refinación de petróleo. [4].

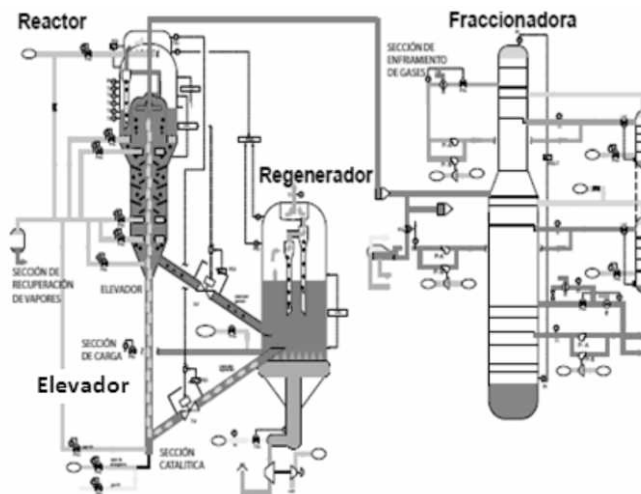


Figura 1. Esquem general de Proceso. (Fuente: Autor)

2.1 Control del proceso de la unidad FCC

Como en cualquier unidad de proceso, en las unidades de FCC existe interrelación entre las variables operativas. Es conveniente separar las condiciones de proceso que afectan el funcionamiento de la unidad en dos tipos de variables conocidas: independientes y dependientes [5].

Las *variables independientes*: Son aquellas variables manipuladas en la operación, con el objeto de estudiar cómo inciden sobre las variables dependientes. Para el caso de estudio seleccionado se describen las siguientes variables independientes, las cuales actúan sobre elementos finales de control:

- Caudal de alimentación (FC-1)
- Temperatura de precalentamiento (TC-3)
- Temperatura del reactor (TC-1)
- Diferencia de presión reactor-regenerador (PDC-1)
- Caudal de aire (FC-2)
- Presión en la fraccionadora (PC-3)

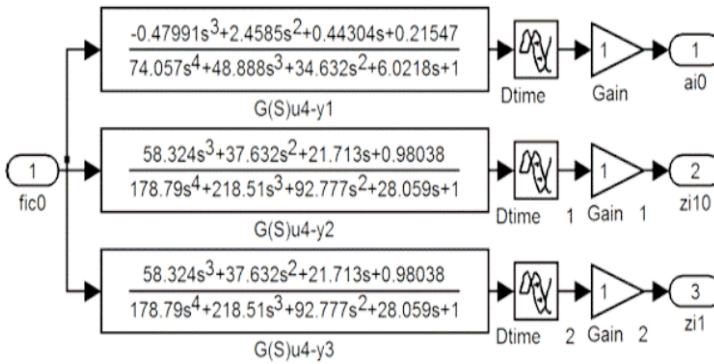
Las *variables dependientes*: De igual manera que para las variables independientes, son las variables de respuesta que se observan en el estudio y que podrían estar influenciadas por los valores de las variables independientes. Algunas de las variables pueden ser medidas directamente y otras obtenidas por medio de cálculos, en este caso de estudio son:

- Oxígeno en aire de combustión (AI-1)
- Temperatura del regenerador (TI-2)
- Presión en el reactor (PI-2)
- Temperatura de gases de combustión (TI-3)
- Posición de la válvula de catalizador gastado (ZI-1)
- Nivel de catalizador en el stripper (LI-1)
- Velocidad del compresor de gases del fraccionador
- Circulación de catalizador
- Severidad/Conversión

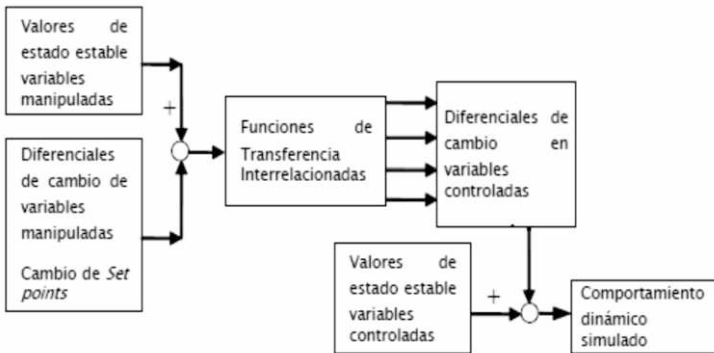
3. Modelo Dinámico

El fundamento primario del modelo de simulación es la representación de la unidad FCC tipo UOP en forma de bloques. Cada bloque representa el comportamiento de una variable dependiente o controlada, que la llamaremos CV, como resultado de efectuar movimientos sobre una o más variables independientes o manipuladas, que las llamaremos MV, debido a dicha relación, se elaboró una matriz

dinámica que establece la dependencia entre dichas variables, ver figura 2(a). La mejor manera de representar cambios en el tiempo consiste en sustituirla cada comportamiento dinámico interrelacionado con una función de transferencia. Esto permite aplicar los principios de linealidad y superposición, respuesta propia de estos modelos. El esquema de construcción se muestra en la figura 2(b). El alcance del modelo incluye el reactor, regenerador y la fraccionadora principal de la FCC y se desarrolló en el ICP como resultado de trabajos de investigación del grupo de automatización [6]. Para tal objeto las variables más relevantes dentro del proceso fueron clasificadas como manipuladas (MV), controladas (CV) y de perturbación (FF).



(a) Descripción bloques del Modelo de Proceso. (Fuente: Autor)



(b) Esquema de construcción.

Figura 2. Descripción del modelo de proceso FCC tipo UOP.

(Fuente: Autor)

4. Análisis de modos de fallo y confiabilidad

4.1 FMEA

El FMEA (Análisis de modos y efectos de falla), se entiende como una metodología sistemática que se emplea para analizar problemas potenciales de confiabilidad durante el ciclo de vida de un componente, equipo o sistema; este permite definir el nivel de criticidad e identificar modos de fallo potenciales y su efecto sobre la operación del sistema; facilita la toma de decisiones, permite anticiparse a las fallas y si estas ocurren puedan ser superadas rápidamente, todo lo anterior hace que esta metodología contribuya al aumento de la confiabilidad [1, 2]. El esquema general para un análisis de modos y efectos de fallo FMEA, se muestra en la figura 3.

4.2 Características dinámica e identificación del proceso

Una vez clasificados los equipos altamente críticos se determinan las características dinámicas tales como linealidad, variabilidad, retardo, ruido y se extraen de cada variable en estudio las distribuciones estadísticas, que por medio del comportamiento histórico permiten definir sus características en los diferentes modos de operación (normal o fallo), simular su comportamiento en el modelo y seleccionar una metodología de detección adecuada.

En este artículo, el caso de estudio consiste en un fallo de la turbina del compresor soplador. Una vez seleccionado el caso altamente crítico para el funcionamiento de la unidad FCC y teniendo en cuenta la relación que existe entre las variables se escogió un grupo puntual de variables a ser analizadas.

5. Análisis de los estados anormales de operación

El éxito de los modelos utilizados para identificación de fallos, depende del proceso de extracción de características. Debido a lo anterior, la disponibilidad de técnicas y herramientas computacionales para tal fin, ha motivado a los investigadores a explorar varias metodologías [1, 2, 3]. Es común encontrar análisis realizados en el dominio del tiempo o de la frecuencia. Se han implementado soluciones que utilizan descriptores de falla utilizando transformada de Fourier, modelos auto-regresivos y características temporales, entre otros [5].

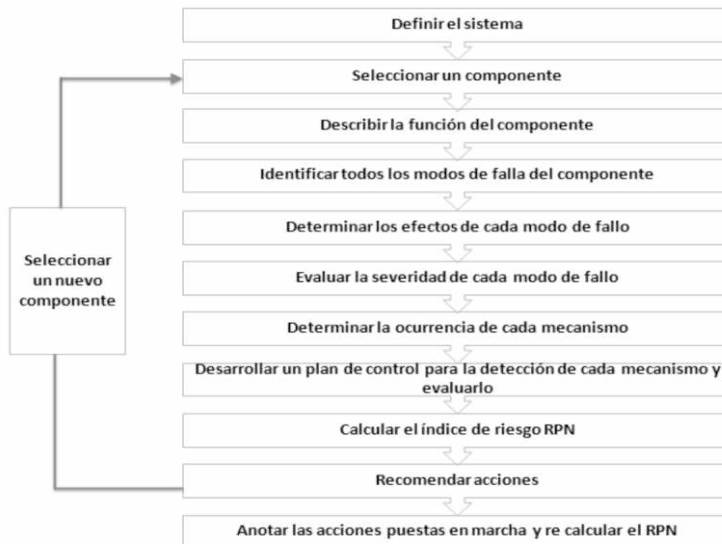


Figura 3. Diagrama de bloques para el FMEA. (Fuente: Autor)

Las estrategias Knowledge Discovery in Databases (KDD) y Minería de Datos (MD) han atraído la atención de investigadores e industriales como herramientas para explotar información en los datos de trabajo. El análisis clustering presenta la ventaja de asistir en la organización de los datos de un sistema e inferir información a partir de ello.

Las Técnicas Estadísticas Multivariantes (TEM) se han mostrado como herramientas potencialmente útiles para el análisis y supervisión de procesos. Dentro de las TEM, el Análisis de Componentes Principales (PCA) ha sido la técnica más estudiada en investigación y la que ha exhibido resultados prometedores en aplicaciones industriales [7,8].

En la última década, se han explorado diversas estrategias de análisis de datos para usar la información contenida en los datos de proceso. Una de estas propuestas es la que combina (TEM) con técnicas de Clustering basadas en Lógica Fuzzy (CLF) para el análisis y la supervisión de procesos sujetos a frecuentes cambios en las condiciones de operación (procesos multioperacionales).

5.1 El Análisis de Cluster o Clustering

El análisis Clustering se puede definir como el proceso de agrupar objetos o datos dentro de clases, grupos o clusters con el propósito de que los objetos dentro de un cluster sean similares entre sí y distintos a los otros clusters. Estas técnicas, con raíces en diversos campos (estadística, química analítica, biología molecular, aprendizaje con

maquinas, KDD y MD) han dado lugar a una amplia gama de investigaciones [9,10].

5.2 Utilización del análisis clustering en la industria petroquímica

En los últimos años, se han explorado las capacidades del análisis de cluster para asistir en problemas de supervisión de procesos. Así podemos encontrar en la literatura una serie de estrategias de supervisión que integran el análisis de cluster dentro de sus propuestas y que se pueden reagrupar como de la siguiente manera:

- Estrategias TEM-CLF: En estos casos se utilizan las TEM para obtener una representación reducida del proceso o de las tendencias de las variables del proceso y, luego, estas representaciones reducidas se analizan mediante técnicas CLF.
- Estrategias con S_{PCA} : En estas, se comienza por dividir los datos en grupos de tamaños que pueden responder a un ordenamiento natural (lotes de producción de procesos discontinuos), o según un criterio apropiado para el análisis actual (duración esperada de los fallos conocidos). Luego, se procesa cada grupo con PCA y, finalmente, se utiliza un índice S_{PCA} que mide la similaridad entre los modelos PCA de cada grupo [11].
- Estrategias con redes neuronales artificiales (RNA). Son similares a los TEM-CLF, dado que aplican una reducción inicial de los datos, usualmente con ACP, y sobre estos se aplica el análisis de cluster con RNA-SOM o con RNA-ART2.

Normalmente las estrategias de supervisión obedecen a un esquema tipo KDD (ver figura 4) con número de etapas variables, la primera etapa extracción de características (para reducir la dimensionalidad de los datos con PCA) y una segunda etapa de identificación de patrones con análisis de cluster. Ocasionalmente, se usa algún tipo de preprocesamiento inicial por ejemplo filtrado de los datos [12], o la división del conjunto original datos en ventanas sucesivas de datos [13,14].

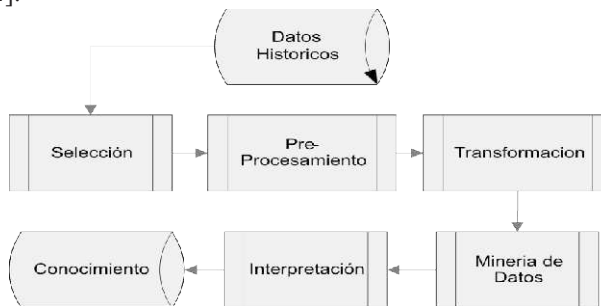


Figura 4. Arquitectura genérica del Proceso KDD. (Fuente: Autores)

5.3 Análisis de Componentes Principales PCA

La aplicación de componentes principales radica en la potencialidad de la técnica para realizar compresión de datos, correlacionar información y detectar patrones o tendencias. Uno de los objetivos al usar PCA es reducir la cantidad de datos registrados de varios datos de desviación, generados simultáneamente en un experimento; PCA es útil para identificar las variables responsables de los fallos [15].

La idea básica del PCA es extraer un nuevo conjunto de variables combinadas, llamadas componentes principales, que describen las variaciones y tendencias clave en los datos de operación en un espacio de dimensiones más pequeño que el original. Supóngase que se dispone de un conjunto de datos de operaciones sin fallos de proceso Y de dimensiones $m \times n$, ecuación 1, donde m es el número de muestras disponibles y n es el número de variables medidas. Además, se asume que los datos de las n variables en Y se han estandarizado previamente y que la distribución de los datos en Y se aproxima a una distribución normal.

$$Y = t_1 * P_1^T + t_2 * P_2^T + \dots + t_n * P_n^T = \sum_{i=1}^n t_i * P_i^T \quad (1)$$

$$y = T * P^T \quad (2)$$

T es la llamada matriz de scores, de dimensiones $m \times n$. Sus columnas, representadas por t_i de $m \times 1$, se conocen como las variables coordenadas o variables scores y representan las coordenadas de los datos al proyectarse en el nuevo espacio de los Componentes Principales (CP). P es la matriz de loadings, de dimensiones $n \times n$, cuyas columnas las forman los diferentes (P_i). Los (P_i), conocidos como cargas o loadings y de dimensión $n \times 1$, son los vectores propios que se derivan de la matriz de correlación de las variables en Y ; vectores unitarios que indican las direcciones de los CP y forman una base ortonormal en R^n . Luego, se cumple que $P^T \cdot P = P \cdot P^T = 1$ y $P^T = P^{-1}$. La matriz P y la descomposición aportada por el PCA, se suelen determinar mediante un método conocido como Descomposición en Valores Singulares o DVS [16]. Durante la resolución del PCA, también se obtiene una matriz diagonal Λ que contiene en su diagonal los valores propios λ_i , asociados a cada P_i , ordenados en orden decreciente $\lambda_1 > \lambda_2 > \dots > \lambda_n$.

El nuevo conjunto de variables obtenido con el PCA concentra la máxima varianza del proceso, en un número de CP que es significativamente menor al número de variables medidas, lo cual hace

que la información del proceso, útil para tareas de análisis y supervisión, sea mucho más fácil de manejar sobre todo si el proceso en estudio involucra muchas variables. El modelo en las ecuaciones (1) o (2) puede describirse como:

$$Y = T_a * P_a' + E = \tilde{Y} + E \rightarrow \tilde{Y} T_a * P_a' \quad (3)$$

En las ecuaciones anteriores, el sub-índice a corresponde al número de CP que retienen la máxima variabilidad, E es la matriz de residuos del modelo obtenido. La dimensión de las matrices Ta y Pa queda definida como mxa y nxa respectivamente. \hat{Y} representa la reconstrucción de los datos y se puede considerar como el modelo aproximado del proceso operando en condiciones normales. Su definición depende de la selección adecuada de los CP. Se cuenta con varios métodos bastante fiables para la selección de los CP, siendo los más usados el Porcentaje de Varianza Explicada y la selección basada en validación cruzada o cross-validation [17, 18].

5.4 Monitorización con PCA

Una vez establecido el modelo del proceso (ecuación 3), este se puede utilizar para monitorizar. Para hacer esto primero se calculan las siguientes estadísticas: La Estadística de Hotelling o T2: Si el PCA se obtiene a partir de datos históricos de operaciones sin fallo, el modelo resultante, según la ecuación 3, representará un Modelo del Proceso Bajo Control (MP-BC). Luego, se puede inferir el estado actual del proceso a través del cálculo de la T2, que no es más que la distancia mahalánobis, entre los nuevos datos, proyectados en el espacio de los CP, y el MP-BC. Para calcular esta desviación se toma en cada instante k en que se muestrea el proceso, el correspondiente vector de mediciones y(k), de dimensión 1xn, y se proyecta en el espacio de los CP como sigue:

$$t_{(k)} = y_{(k)} * P_a \quad (4)$$

En la ecuación anterior t(k) son los scores representando la proyección de y(k) en el instante k. Luego, T2 se determina como sigue:

$$T^2 = t_{(k)}^T * \Lambda_a^{-1} * t_{(k)} \quad (5)$$

Donde Λ_a es la matriz diagonal de valores propios λ_i asociados a los Pi de los a CP seleccionados.

El error de predicción al cuadrado o SPE: El SPE (Squared Prediction Error), también conocido como Q, se define como la suma del error al cuadrado entre la señal original y la señal reconstruida después de la reducción que produce el ACP. Se calcula como sigue:

$$e_{(k)} = (I - P_a * P_a^T) * y_{(k)} \rightarrow SPE = e_{(k)} * e_{(k)}^T \quad (6)$$

donde I es la matriz identidad de nxn y e(k) es el error entre los datos y el ajuste por el modelo. Luego, la monitorización con PCA se efectúa como sigue:

Con los nuevos datos y(k) que se recogen en cada nuevo instante k, se obtienen los correspondientes scores según la ecuación 4.

Se calcula el error entre los datos recibidos y(k) y el ajuste por el modelo según la ecuación 6.

Con la información anterior se calculan el T2 y el SPE, según las ecuaciones 5 y 6 respectivamente.

Se comparan el T2 y el SPE obtenidos con sus respectivos límites de control (T2lim y SPElim) tal que:

$$T^2 \leq T_{lim}^2 \text{ y } SPE \leq SPE_{lim} \quad (7)$$

Si alguna o ninguna de las 2 condiciones anteriores se cumple, entonces el sistema se encuentra fuera de sus condiciones normales de operación. Las comparaciones anteriores (ecuación 7) típicamente se aplican mediante gráficos de monitorización o control donde se observan las señales de T2 y el SPE con sus respectivos límites.

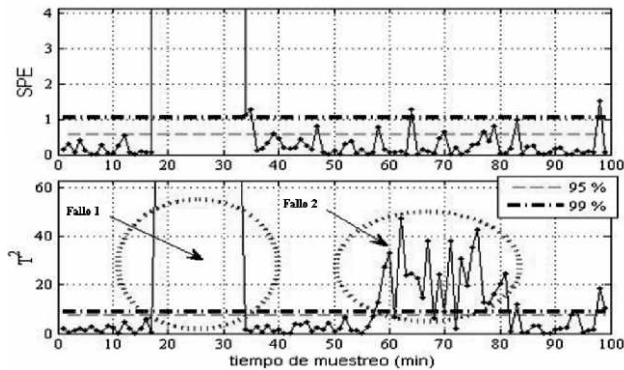


Figura 5. Detección con PCA clásico y límites.

En los trabajos realizados por [19, 20,21], se introduce el método para estimar estados fuera de la operación normal. La manera tradicional de detectar fallos con la monitorización basada en PCA (regla de detección) consiste en tomar en cada instante los nuevos valores calculados de SPE y T2, proyectarlos en los gráficos de control de dichos estadísticos y comprobar si superan los límites de control en dichos gráficos. Luego, si durante k periodos consecutivos (típicamente $k=3$) se produce una desviación se genera una alarma de fallo (figura 5). Aunque esta no es la regla universal se utiliza ampliamente, y por tanto es también la que se utilizará en este artículo para todos los métodos. [22].

5.5 Métodos de Clustering basados en Lógica Fuzzy (CLF)

Los métodos CLF caen dentro del grupo de métodos de partición. La idea de estos métodos es obtener una partición de los datos en c grupos y para ello se han propuesto 2 tipos de particiones posibles que condicionan a los diferentes métodos:

La Partición rígida: Sea $Y = \{y_1, y_2, \dots, y_m\}$ un conjunto finito de datos de proceso, donde cada muestra m es un vector en R^n y sea c un entero con valores $2 \leq c < m$. La matriz de pertenencias $U = [\mu_{ik}]$, representa la partición rígida de los datos en Y si y solo si sus elementos satisfacen las siguientes propiedades:

$$\begin{aligned} \mu_{ik} \in 0,1 \quad 1 \leq i \leq m \quad 1 \leq k \leq c \\ \sum_{k=1}^c \mu_{ik} = 1 \quad 1 \leq i \leq m \\ 0 < \sum_{i=1}^m \mu_{ik} < m \quad 1 \leq k \leq c \quad (8) \end{aligned}$$

Donde μ_{ik} es la pertenencia individual de un objeto o muestra i a un grupo k . Claramente, las pertenencias μ_{ik} asignan cada objeto a un solo grupo. La Partición Difusa (Fuzzy) es una generalización de la partición rígida, mediante un enfoque de Lógica Fuzzy, que permite valores en el intervalo $[0,1]$ para las μ_{ik} . Esto permite que los valores de pertenencia no solo indiquen a que grupo pertenece cada objeto sino también como se clasifica un objeto en cada grupo. Así, sea $Y = [y_1, y_2, \dots, y_m]$ un conjunto finito de datos de proceso y sea c un entero con valores $2 \leq c < m$, la matriz de pertenencias $U = [\mu_{ik}]$, representa la partición difusa de los datos en Y si y solo si sus elementos satisfacen las siguientes propiedades:

$$\begin{aligned} \mu_{ik} &\in [0,1] \quad 1 \leq i \leq m \quad 1 \leq k \leq c \\ \sum_{k=1}^c \mu_{ik} &= 1 \quad 1 \leq i \leq m \\ 0 < \sum_{i=1}^m \mu_{ik} &< m \quad 1 \leq k \leq c \end{aligned} \quad (9)$$

La k-ésima columna de U contiene los valores de la función de pertenencia del k-ésimo subconjunto difuso de Y.

5.5.1 Método Fuzzy C-Means (FCM)

Se basa en el concepto de particiones difusas (técnicas CLD). Sea la matriz de datos Y de dimensiones mxn. Se intenta determinar la pertenencia de cada objeto de Y a cada uno de los c clusters mediante la minimización de la siguiente función objetivo:

$$J(Y, U, V) = \sum_{k=1}^c \sum_{i=1}^m (\mu_{ik})^{\delta} d_{ik}^2 \quad d_{ik}^2 = |X_i - v_k|^2 \quad (10)$$

Donde δ es un índice de difusividad tal que $\delta \in [1, \infty)$ se recomienda el valor típico de la literatura $\delta = 2.5$ [23,24].

La minimización de la ecuación 10 involucra un problema de optimización no lineal que se podría resolver de distintas maneras el método más popular para resolver este problema se conoce como la iteración de Picard [25] en cada iteración los prototipos y las pertenencias de los clusters (centros de clusters) se recalculan como sigue:

$$v_k = \frac{\sum_{i=1}^m \mu_{ik}^{\delta} * y_i}{\sum_{i=1}^m \mu_{ik}^{\delta}} \quad (11)$$

5.6 Estimación del número de clusters

Todos los métodos anteriores requieren como entrada el número de clusters c en que se dividen los datos. En muchas situaciones este valor puede conocerse a-priori, en otras situaciones donde no se conozca esta información se asiste con herramientas para la determinación del número de grupos en que se dividen los datos.

6. Caso de estudio

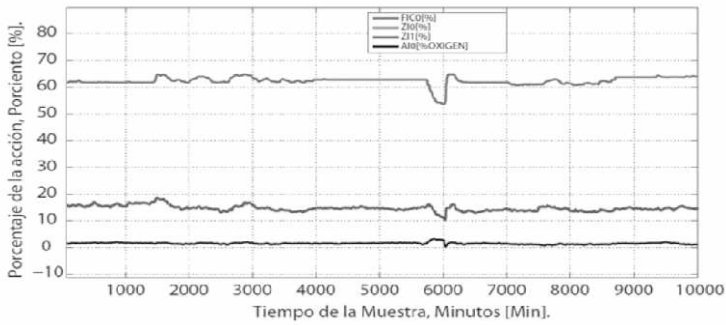
Se selecciona un caso de alta criticidad (fallo de la Turbina compresor soplador) que relaciona las variables controladas AIO% Analizador de O₂ de Gases de Combustión, ZIO Posición de la válvula corredera y ZI1 Posición de la válvula corredera (redundante), con las variables manipuladas FICO Controlador de eficiencia-Aire de Combustión como lo muestra la figura 2(a). Las características dinámicas estudiadas corresponden con las distribuciones estadísticas de los datos históricos de las variables relacionadas, en estado de operación normal y en estado de fallo, como se aprecia en la tabla 1.

Variable	Operación normal	Operación Anormal	Parada de Planta
AIO	$\sim N(2,0.06)$	$\sim N(4.26,0.31)$	$\sim N(0,0)$
FICO	$\sim N(63,0.046)$	$\sim N(31,5.1)$	$\sim N(2.21,0.8)$
ZIO	$\sim N(13.45,0.213)$	$\sim N(5.142,2.092)$	$\sim N(0.1,0.17)$
ZI1	$\sim N(13.52,0.198)$	$\sim N(5.139,2.457)$	$\sim N(0.4,0.24)$

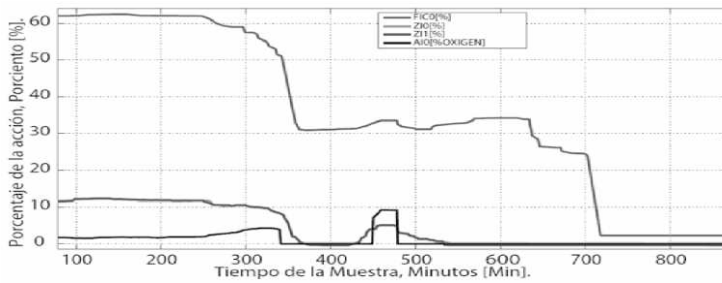
Tabla 1. Selección de variables y características dinámicas caso de estudio

La convención utilizada $\sim N(\mu, \sigma)$, para representar los datos en la tabla 1, corresponde a una distribución estadística normal, donde la variable μ representa la media poblacional y σ la desviación estándar, estos valores se calcularon para las variables asociadas al fallo estudiado.

Con estos datos se procede a simular el comportamiento de los eventos de operación normal y de fallo en el modelo de la FCC, obteniendo el comportamiento mostrado en las figuras 6(a) y 6(b).



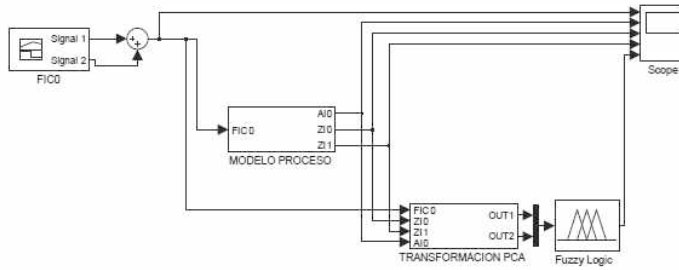
(a) Simulando Comportamiento Normal.



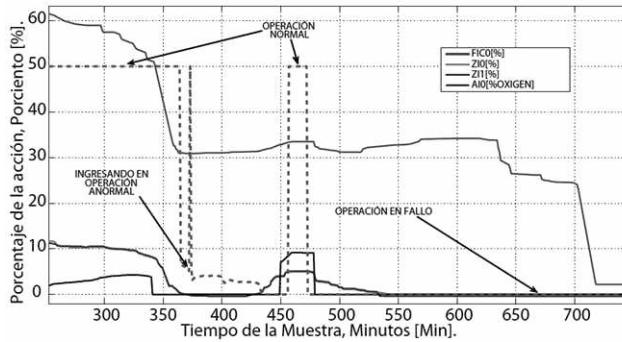
(b) Simulando Comportamiento Anormal.

Fig. 6. Simulando comportamiento del proceso FCC.

Después se ajustan los algoritmos basados en PCA y Lógica Fuzzy para detectar los dos tipos de operación del proceso y se procesa en simulink, como lo muestra el diagrama de la figura 7(a); para obtener los resultados que se muestran en la figura 7(b); donde se diferencian los estados de operación normal y operación anormal en el proceso.



(a) Diagrama Simulink



b) Resultado de la Detección.

Fig. 7. Detección de comportamientos en el proceso FCC.

7. Resultados

Clasificación de equipos de acuerdo con el nivel de criticidad, identificando modos de fallo potenciales; por ejemplo en el caso de estudio se presenta una falla altamente crítica asociada al fallo de la turbina del compresor soplador. Identificación de las características principales de las tendencias: tasas de cambio, abrupticidad e identificación paramétrica. Determinación de la existencia de señales de monitoreo de condición asociada a los equipos: En este caso la aplicación de PCA permite seleccionar las características principales de la respuesta y definir umbrales para diagnosticar condiciones anormales. Mediante la Clasificación obtenida con el método basado en Lógica Fuzzy se logró diferenciar los modos de operación normal, fallo y parada de planta.

8. Conclusiones

Se muestra que las técnicas utilizadas en la arquitectura híbrida propuesta para las tareas de detección y clasificación (PCA, Lógica Fuzzy) son herramientas efectivas en el reconocimiento de modos de operación. Con los resultados obtenidos en el ajuste de las técnicas es posible hacer una selección de las características más descriptivas de las variables del proceso asociadas a los equipos altamente críticos en una unidad FCC tipo UOP, para el caso de estudio presentado se logró diferenciar entre los diferentes estados de operación en un modelo del proceso de craqueo catalítico fluidizado. Para trabajos futuros se pueden aplicar variaciones en la arquitectura utilizando técnicas similares, por ejemplo, reemplazar PCA por FDA (Análisis por discriminante de Fischer) o KPCA (Análisis de componentes principales utilizando Kernel) o PLS (Regresión en mínimos cuadrados parciales), también se podría reemplazar el clasificador que utiliza Lógica Fuzzy por uno que utilice una red neuronal tipo SOM (Mapas con características auto-organizativas). En el futuro se espera usar esta herramienta en línea para la detección temprana de eventos que puedan desencadenar condiciones de fallo.

Referencias

- [1] V. Venkatasubramanian, R. Rengaswamy, and S. Kavuri, “A review of process fault detection and diagnosis Part II: Quantitative model and search strategies,” *COMPUTERS & CHEMICAL ENGINEERING*, vol. 27, pp. 313–326, MAR 15 2003.
- [2] V. Venkatasubramanian, R. Rengaswamy, S. Kavuri, and K. Yin, “A review of process fault detection and diagnosis Part III: Process history based methods,” *COMPUTERS & CHEMICAL ENGINEERING*, vol. 27, pp. 327–346, MAR 15 2003.
- [3] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. Kavuri, “A review of process fault detection and diagnosis Part I: Quantitative model-based methods,” *COMPUTERS & CHEMICAL ENGINEERING*, vol. 27, pp. 293–311, MAR 15 2003. M. Misra, H. Yue, S. Qin, and C. Ling, “Multivariate process monitoring and fault diagnosis by multi-scale PCA,” *COMPUTERS & CHEMICAL ENGINEERING*, vol. 26, pp. 1281–1293, SEP 15 2002V.
- [4] Sadeghbeigi, Reza. *Fluid Catalytic Cracking Handbook Design, Operation and Troubleshooting of FCC facilities*, 2Ed, Gulf Professional Publishing, Houston. 2000 2 Ed 234-275.

- [5] J. Acedo. Control de reactores. Repsol-YPF, S.A. Ingeniería Química., pages 168-187, octubre 2005.
- [6] Angulo, N. Aristizabal, L. Fase I Desarrollo de un modelo de simulación dinámico de la sección de reacción, regeneración y fraccionadora principal de una unidad de craqueo catalítico. Universidad Politécnica de Valencia-Universidad Santo Tomas, 2006.
- [7] MacGregor, J. F. Data-Based Latent Variable Methods for Process Analysis, Monitoring and Control, in European Symposium on Computer Aided Process Engineering - 14 (Barbosa Póvoa, A. y Matos, H. eds.). Elsevier, Lisbon, Portugal. 2004.
- [8] Han, J. y M. Kamber. Data mining: concepts and techniques. Morgan Kaufmann, San Francisco, CA, USA. 2001.
- [9] Jain, A. K., M. N. Murty y P. J. Flynn. Data Clustering: A Review. ACM Computing Surveys, 31(3), pp.264-323. 1999.
- [10] H. Aradhye, B. Bakshi, J. Davis, and S. Ahalt, "Clustering in wavelet domain: A multiresolution ART network for anomaly detection," AICHE JOURNAL, vol. 50, pp. 2455–2466, OCT 2004.
- [11] Krzanowski, W. J. Between-groups comparison of principal components. J. American Statistical Association, 74(367), pp.703-707. 1979.
- [12] Hwang, D. H. y C. Han. Real-time monitoring for a process with multiple operating modes. Control Engineering Practice., 7, pp.891-902. 1999.
- [13] Tona, R. V., Espuña, A., and Puigjaner, L. A Historical Data Based Methodology to identify longer variations and optimal process operating conditions, in 51st Canadian Chemical Engineering Conference Halifax, Canada. pp.375. 2001.
- [14] Johannesmeyer, M. C., A. Singhal y D. E. Seborg. Pattern matching in historical data. AIChE Journal, 48(9), pp.2022-2038. 2002.
- [15] Wenfu Ku, Robert H. Storer, and Christos Georgakis. Disturbance detection and isolation by dynamic principal component analysis. Chemometrics and intelligent laboratory systems, 30:179-196, November 1995.

- [16] Jackson, J. E. A User's Guide to Principal Components. John Wiley&Sons, New York. 1991.
- [17] Wise, B. M. y N. B. Gallagher. The process chemometrics approach to process monitoring and fault detection. *Journal of Process Control*, 6(6), pp.329-348. 1996.
- [18] Johannesmeyer, M. C., A. Singhal y D. E. Seborg. Pattern matching in historical data. *AIChE Journal*, 48(9), pp.2022-2038. 2002.
- [19] Z. Li, G. Liu, R. Zhang, and Z. Zhu, "Fault detection, Identification and reconstruction for gyroscope in satellite based on independent component analysis," *ACTA ASTRONAUTICA*, vol. 68, pp. 1015–1023, APR-MAY 2011.
- [20] V. H. Nguyen and J.-C. Golinval, "Fault detection based on Kernel Principal Component Analysis," *ENGINEERING STRUCTURES*, vol. 32, pp. 3683–3691, NOV 2010.
- [21] M. Misra, H. Yue, S. Qin, and C. Ling, "Multivariate process monitoring and fault diagnosis by multi-scale PCA," *COMPUTERS & CHEMICAL ENGINEERING*, vol. 26, pp. 1281–1293, SEP 15 2002.
- [22] J. Acedo. Control de reactores. Repsol-YPF, S.A. Ingeniería Química, pages 168-187, octubre 2005.
- [23] Teppola, P. y P. Minkkinen. Possibilistic and Fuzzy C-means clustering for process monitoring in an activated sludge wastewater treatment plant. *Journal of Chemometrics*, 13(3-4), pp.445-459. 1999.
- [24] Teppola, P., S. Mujunen y P. Minkkinen. Adaptive Fuzzy C-means Clustering in Process Monitoring. *Chemometrics and Intelligence Laboratory System*, 45, pp.23-38. 1999.
- [25] Bezdek, J. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Publishing Corporation, New York, N.Y. (1981).