



Universidad de Valladolid



ESCUELA DE INGENIERÍAS  
INDUSTRIALES

**UNIVERSIDAD DE VALLADOLID**  
**ESCUELA DE INGENIERIAS INDUSTRIALES**

**Grado en Ingeniería Electrónica Industrial y Automática**

**Detección y diagnóstico de fallos/anomalías  
en una planta industrial mediante técnicas  
de control estadístico multivariante**

**Autor:**

**Crespo García, Alejandro**

**Tutor:**

**De la Fuente Aparicio,**

**María Jesús**

**Valladolid, junio 2025**



**Universidad de Valladolid**



**ESCUELA DE INGENIERÍAS  
INDUSTRIALES**



# Índice

|   |    |
|---|----|
| Resumen.....  | 6  |
| Abstract .....  | 7  |
| Capítulo 1: Introducción y objetivos.....                   | 8  |
| 1.1 Introducción.....                                       | 10 |
| 1.2 Objetivos.....  | 10 |
| 1.3 Organización de la memoria.....                         | 11 |
| Capítulo 2: Control de calidad. Aspectos teóricos.....      | 13 |
| 2.1 Control de calidad .....                                | 15 |
| 2.2 Evolución del control de calidad.....                   | 16 |
| 2.3 Estadísticos.....                                       | 17 |
| 2.4 Análisis de componentes principales o PCA .....         | 17 |
| 2.4.1.- PCA y DPCA para detección de fallos .....           | 20 |
| 2.5 Análisis de factores lentos o SFA .....                 | 21 |
| 2.5.1.- SFA y DSFA para detección de fallos.....            | 22 |
| 2.6 Análisis estadístico de patrones o SPA.....             | 23 |
| 2.7 Análisis de Componentes Principales Sparse o SPCA ..... | 24 |
| Capítulo 3: Planta y datos.....                             | 26 |
| 3.1 introducción a la planta .....                          | 29 |
| 3.2 Datos históricos .....                                  | 31 |
| Capítulo 4: Aplicación.....                                 | 36 |
| 4.1 Introducción.....                                       | 38 |
| 4.2 PCA.....  | 39 |
| 4.2.1.- Detección de fallos usando PCA .....                | 41 |
| 4.2.2.- Detección de fallos usando DPCA.....                | 44 |
| 4.3 SFA .....   | 46 |
| 4.3.1.- Detección de fallos usando SFA.....                 | 48 |
| 4.3.2.- Detección de fallos usando DSFA .....               | 52 |
| 4.4 SPA .....   | 56 |
| 4.4.1.- Detección de fallos usando SPA.....                 | 57 |
| 4.5 SPCA.....   | 60 |
| 4.5.1.- Detección de Fallos usando SPCA .....               | 60 |
| 4.5.2.- Detección de fallos usando DSPCA.....               | 64 |
| Capítulo 5: Conclusiones y trabajo futuro .....             | 68 |



|                          |    |
|--------------------------|----|
| 5.1 Conclusiones .....   | 70 |
| 5.2 Trabajo futuro ..... | 71 |
| Bibliografía .....       | 73 |



Universidad de Valladolid



ESCUELA DE INGENIERÍAS  
INDUSTRIALES



## Resumen

Este TFG busca mejorar la detección y el diagnóstico de fallos/anomalías en una planta industrial mediante la aplicación de métodos de control estadístico multivariante. Actualmente la industria tiene infinidad de formas para medir y almacenar datos en cualquier planta de producción, el problema se encuentra a la hora de tratar estos datos para llegar a una conclusión.

Para hacer más fácil la interpretación de los datos, utilizamos las técnicas estadísticas de control multivariante, el objetivo de estas es controlar y analizar las variables deseadas para garantizar la mejor calidad posible, o, en nuestro caso concreto, aquellas variables que puedan llegar a causar fallos o anomalías en el proceso de producción. Usando métodos como el Análisis de Componentes Principales (PCA) o el Análisis de factores lentos (SFA) entre muchos otros, se va a reducir la dimensionalidad de los datos y evaluar correlaciones.

Los diferentes métodos de control estadístico implementados se van a comparar con los mismos datos de base, datos que pertenecen a una planta química, la planta Tennessee Eastman, estos datos son una referencia de la literatura científica para probar toda clase de métodos de detección de anomalías.

Una vez evaluados todos los métodos, se comparará cuáles de ellos responden mejor a la hora de detectar errores, no solo de forma gráfica, también de forma numérica, para ello se usarán diferentes estadísticos. De esta forma se podrá saber cuál de ellos es mejor detectando anomalías en una planta industrial o si hay uno que sea superior en todo frente a sus rivales.



## Abstract

This Final Degree Project aims to enhance fault detection and diagnosis in industrial plants by applying multivariate statistical process control models. In modern industry, vast amounts of data are continuously measured and stored in production plants. However, the challenge lies in processing and analyzing this data effectively to reach meaningful conclusions.

To facilitate data interpretation, we employ multivariate statistical process control techniques. These methods help monitor and analyze key variables to ensure the highest possible quality or, in our case, to identify variables that may lead to faults or anomalies in the production process. By using models such as Principal Component Analysis (PCA) and Slow Factor Analysis (SFA), among others, we aim to reduce data dimensionality and assess correlations.

The selected statistical methods will be tested using a reference dataset from the Tennessee Eastman chemical plant, a widely used benchmark for evaluating statistical methods. Once all models have been applied, their performance will be compared in terms of fault detection, both visually and numerically, using various statistical metrics. This analysis will determine which model performs best in specific scenarios or if one model consistently outperforms the others across all evaluation criteria.



# Capítulo 1: Introducción y objetivos



Universidad de Valladolid



ESCUELA DE INGENIERÍAS  
INDUSTRIALES



## 1.1 Introducción

En el contexto actual de la producción industrial, tanto la eficiencia como la calidad son cada vez recursos más importantes, necesarios para mantener un buen producto a un precio competitivo. El control constante de la producción, aparte de mejorar la calidad del producto final, reducir residuos y subir la eficiencia, también aumenta la seguridad ya que permite una mejora en la detección de errores que pueden causar pérdidas económicas o daños humanos. Saber cuáles son los fallos que se producen de la forma más rápida y acertada posible es la manera de evitarlos.

Detectar estos fallos parece sencillo sobre el papel, pero cuando intervienen un gran número de variables como pueden ser presión, temperatura o caudal, que a su vez se están relacionando entre ellas de manera diferente y todo esto a largo del tiempo, lo convierte en una tarea complicada, imposible de ser realizada por una persona o un grupo de personas.

Para resolver este problema, el control de procesos estadístico multivariante o MSPC por sus siglas en inglés, es la herramienta que más fuerza ha cogido en los últimos tiempos. Anteriormente se usaba el control estadístico univariante, que analizaba cada variable de forma independiente, pero en una planta industrial en la que todas las variables están relacionadas, MSPC puede detectar anomalías que conduzcan a un fallo.

Este TFG va a comparar diferentes métodos de control estadístico multivariante, en los que, dados unos datos de una planta, se va a ver cuál de estos modelos es el que mejor tiempo de respuesta tiene, menor número de falsas alarmas produce y más alarmas detectadas es capaz de acertar, para ello vamos a utilizar dos estadísticos, el de Hotelling o  $T^2$ , y el de SPE o Q. De esta manera compararemos como el MSPC en sus distintas formas va a ser una forma eficaz de detectar fallos y cuál de estas es la más útil.

## 1.2 Objetivos

Este TFG pretende implementar diferentes métodos de control estadístico multivariante para detectar fallos/anomalías en las plantas industriales. Y además se trata de comprobar de forma tanto gráfica, como numérica, cuál de los métodos implementados obtiene mejores resultados a la hora de detectar fallos o anomalías en una planta de producción, como es la planta Tennessee Eastman. Se van a comprobar el tiempo de fallo, las falsas alarmas y las alarmas detectadas de cada método estadístico, con el objetivo de comparar cuál de ellos es más eficaz en cada campo y cuál es el más eficaz en términos generales.



### 1.3 Organización de la memoria

Estos son los 5 capítulos que forman este TFG:

En el primer capítulo se verá de forma general un resumen del tema que se va a tratar y los diferentes puntos que se desean conseguir.

En el segundo capítulo vamos a ver en que se fundamentan desde el punto de vista teórico, los diferentes métodos de control estadístico que se aplicarán, además de ver de una forma más concreta en que consiste el control de calidad.

El tercer capítulo explica la planta Tennessee Eastman, que es y que vamos a obtener de la misma para comprobar los distintos modelos definidos en el capítulo anterior.

Una vez visto la teoría de los métodos estadísticos, en el cuarto capítulo se va a comentar como están programados, además de cuáles son los resultados que ofrece cada uno ante los mismos datos de la planta.

Sabiendo ya los resultados de cada método, en el quinto y último capítulo se van a comparar estos resultados, además de obtener unas conclusiones sobre qué método es el más deseable si hay uno que destaque sobre el resto. También se va a tratar como se puede seguir en futuros proyectos los resultados obtenidos en este.



Universidad de Valladolid



ESCUELA DE INGENIERÍAS  
INDUSTRIALES



## Capítulo 2: Control de calidad. Aspectos teóricos



Universidad de Valladolid



ESCUELA DE INGENIERÍAS  
INDUSTRIALES



## 2.1 Control de calidad

El control de calidad son un conjunto de actividades y técnicas cuyo objetivo es garantizar que tanto los productos como los servicios cumplan con la calidad que le prometen al cliente. El control de calidad es necesario para hacer una empresa más competitiva, bajar el número de residuos creados o hacer un proceso más eficiente entre otras. El control de calidad ha evolucionado mucho a lo largo de los años, se inició como un control visual y actualmente se utilizan sensores con una gran precisión, este tema se tratará más en profundidad en el punto siguiente.

El control de calidad aplicado a una planta industrial es cada vez más necesario debido a la creciente competitividad, un control de calidad inadecuado puede repercutir de forma muy negativa en la imagen percibida que el cliente tiene sobre una marca, haciendo que la próxima vez se decante por otra empresa, que no la recomiende o incluso que hable mal sobre ella a otras personas. Aparte de lo mencionado anteriormente, el control de calidad también es importante porque repercute de forma indirecta en la seguridad, tanto de la planta, el medio ambiente y los trabajadores, esto es así porque en caso de que una variable de un proceso en una planta de producción no esté dentro de los parámetros adecuados lo mínimo que va a hacer es que el producto no cumpla con los estándares marcados, pero como he dicho si la variable mide por ejemplo la cantidad de combustible en una reacción química y estas inyectando más de lo necesario se puede llegar a causar un incendio.

Para que el control de calidad sea efectivo, es necesario que se detecten los fallos de una manera rápida y eficaz, sin falsos positivos, tener un sistema de detección de fallos que marque falsos positivos es prácticamente como no tener nada. Para que la detección de fallos sea útil, es necesario recabar la mayor cantidad de información posible sobre las variables que se consideren importantes en los distintos procesos, para ello serán necesarios unos sensores adecuados para las variables que se desean medir, cuanto antes llegue la información recabada al punto donde se va a tratar más eficaz será la misma, una vez tengamos esa información es igual de importante tratarla correctamente para obtener unas conclusiones y en caso de que sea necesario actuar sobre la planta. Todas estas acciones tienen como objetivo mejorar el producto final y garantizar la salud y seguridad de los miembros de la planta.



## 2.2 Evolución del control de calidad

El ser humano ha realizado controles de calidad mucho tiempo antes de que llegase la revolución industrial, aunque no tenga nada que ver con los complejos procesos que se realizan en nuestro tiempo, en la prehistoria se inspeccionaba si una piedra podía ser adecuada para realizar un cierto trabajo, mucho después con la llegada de oficios como artesanos de cerámica o herreros, el control de calidad seguía estando presente con el objetivo de ofrecer un mejor producto.

En la revolución industrial, al estar todo más automatizado debido al uso de la producción en cadena, no se daba la importancia al control de calidad que sí se aplicaba en la producción más artesanal, esto produjo que el control de calidad se tuviese que ejecutar de forma externa a la producción en caso de que hubiese algún tipo de control de calidad, ya que se solían omitir.

En el control de calidad destaca la figura de Walter A. Shewhart, quién creó el concepto de control estadístico de procesos (SPC), con ello introdujo los gráficos en el mundo del control de calidad, con ello consiguió que fuese mucho más visual detectar fallos o elementos fuera del comportamiento normal en un proceso. En los siguientes años Joseph M. Juran hicieron de este sencillo concepto una teoría mucho más elaborada, se crearon métodos estadísticos para el control de calidad, esto fue el inicio del control de calidad tal y como lo conocemos actualmente. [1]

Según iba aumentando el uso y estudio de la detección de fallos, se crearon las normas ISO, estas son un tipo de normas que se deben aplicar en un proceso y que indican como o con que tolerancias debe crearse un producto. Como estas normas se usan a nivel global han ayudado a crear estándares internacionales en la fabricación.

Con la llegada de la industria 4.0 todos los elementos de la producción pasaron a estar conectados, esto permite saber todas las variables que se deseen de la planta de producción, el problema se encuentra a la hora de sacar conclusiones de toda esa cantidad de datos. Actualmente podemos hablar de control de calidad y relacionarlo con el internet de las cosas, el big data y la inteligencia artificial.

Esta evolución que ha tenido el control de calidad ha permitido que las empresas optimicen sus procesos para crear un mejor producto para el cliente, a la vez que ha dado más seguridad a sus trabajadores, solo las empresas que han incorporado las técnicas de control de calidad se han mantenido en el tiempo, toda aquella empresa que en el futuro no incorpore las nuevas técnicas que controlen la calidad de sus productos estará condenada a desaparecer, ya que el control de calidad será cada vez más importante.



## 2.3 Estadísticos

En el control de calidad se suelen usar dos estadísticas de monitorización, la primera es la estadística de Hotelling o  $T^2$ , que puede interpretarse como la distancia de una observación a la media del comportamiento normal, esta se compara con un umbral calculado y si lo sobrepasa es que hay un fallo. La segunda es la estadística SPE o Q, en este caso cuando el valor de Q es muy pequeño es que el sistema se encuentra en condiciones normales, al igual que en el anterior calculamos un umbral y si se sobrepasa es que se ha dado un error.

En las distintas muestras de datos, a la hora de detectar un error, vamos a tener en cuenta los siguientes parámetros, el primero de ellos es el tiempo de fallo, que viene dado por el primer instante en que cada estadística ( $T^2$  o Q) supera su umbral. Para evitar falsas alarmas, se decide que ha habido un fallo real cuando la estadística supera su umbral un número consecutivo de veces que en nuestro caso serán 10. El tiempo de fallo corresponde con la primera vez que se superó el umbral después de detectar que hay un fallo.

Las falsas alarmas, que es el porcentaje de veces que las estadísticas superan el umbral durante la operación normal de la planta. Para que un sistema de detección de anomalías sea útil, este índice debe ser lo más bajo posible.

Por último, las alarmas detectadas, que es aquel porcentaje en los que el valor obtenido supera el umbral cuando realmente hay un fallo.

En los métodos estadísticos que vamos a aplicar, vamos a intentar ver cuál de ellos se acerca a los siguientes valores para que sea ideal, un tiempo de fallo de 160 (que es donde sabemos que se produce el fallo en nuestros datos), las falsas alarmas con un porcentaje de 0 y las alarmas detectadas del 100, esto es ideal y prácticamente ningún método lo llega a conseguir.

## 2.4 Análisis de componentes principales o PCA

PCA o análisis de componentes principales es una técnica de control estadístico multivariante cuyo objetivo es la reducción de la dimensionalidad, es decir los datos se proyectan en un espacio de menor dimensión que el original, este contiene la mayor parte de la varianza de los datos originales, también considera la correlación entre las distintas variables, esto hace que al utilizar datos de dimensiones grandes con un alto ruido obtiene mejores resultados que otros métodos estadísticos univariantes. Al aplicar PCA a un conjunto de datos se mantiene la estructura de correlación de las variables del proceso, es más fácil obtener la máxima variabilidad de los datos y podemos dividir el espacio en dos subespacios diferentes, uno la tendencia y otro el ruido. Esto nos permite identificar las variables que crean fallos y las variables a las que los afectan. Lo único que

necesita son un conjunto de datos en los que el funcionamiento de la planta sea normal, esto es necesario para calibrar el modelo [2].

El funcionamiento de PCA es el siguiente, dado un conjunto de datos  $X \in \mathbb{R}^{n \times m}$  con un número  $n$  de observaciones y un número de variables  $m$ , lo primero que hace es calcular la matriz de covarianza de los datos, después obtiene los autovalores y autovectores de la matriz, los autovectores indican las componentes principales de la varianza máxima en la matriz de datos, los autovalores indican la magnitud de la varianza a lo largo de cada vector propio (Figura 1). Esta es la forma mediante la cual PCA identifica el componente principal más importante [3].

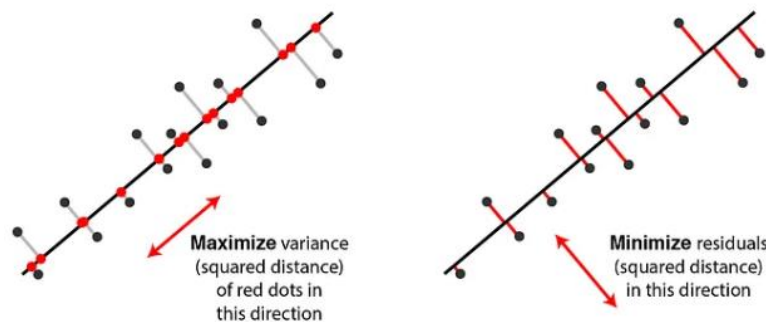


Figura 1.- Explicación gráfica modelo PCA

Inicialmente debemos usar en PCA unos datos que sean de funcionamiento normal de la planta, es decir, que ninguna de sus variables falle,  $X$ . Una vez tenemos los datos, realizaremos una normalización de estos a media cero y varianza uno, así lograremos que la escala sea la misma para todas las variables, para aplicar la normalización a cada una de las variables la restaremos la media de esta y a continuación en la ec.1 dividiremos entre su desviación estándar.

$$X_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j} \quad (1)$$

En la ec.2, haremos la matriz de covarianza, que describe como varían en su conjunto las variables.

$$\Sigma = \left( \frac{1}{(n-1)} \right) X^T X \quad (2)$$

Después, en la ec.3 necesitaremos calcular los valores y vectores propios (autovalores o autovectores), siendo  $\lambda$  el valor propio asociado, que indica la varianza definida por esa componente.

$$\Sigma v = \lambda v \quad \text{o} \quad R = V \Delta V^T \quad (3)$$



La matriz  $V$  está formada por los vectores propios. Ordenando los valores propios (cuya magnitud, como dijimos, nos dice cuánta varianza explica cada uno) en orden descendente, podemos seleccionar aquellos que contengan la mayor cantidad de información, descartando los demás. Se suelen escoger los  $k$  primeros componentes principales hasta llegar a un cierto porcentaje de varianza que queremos tener en el modelo simplificado. De la misma forma, se ordena la matriz  $V$  de vectores propios según la magnitud de sus valores propios asociados, y finalmente construimos la matriz de proyección con la  $k$  primeras columnas de  $V$  seleccionadas (los primeros  $k$  vectores propios), ec. 4.

$$V_k = [v_1, v_2, \dots, v_k] \quad (4)$$

La proyección de los datos en la ec.5 trata de representar los datos de dimensión reducida de dimensión  $n \times k$ .

$$X' = XV_k \quad (5)$$

En resumen, PCA hace más sencillos de entender los datos ya que reduce el número de variables, también elimina variables que son redundantes. Como puntos negativos puede llegar a eliminar variables necesarias y puede fallar si las variables no son lineales.

DPCA o análisis de componentes principales dinámico, es un modelo de detección de fallos dinámica [4], esto permite que pueda manejar datos con estructura temporal, al contrario que PCA, DPCA tiene en cuenta los datos de instantes de tiempo anteriores, este proceso, en teoría, le facilita una mejor detección de fallos respecto al modelo estadístico anterior. Como punto negativo podemos observar una mayor carga computacional, además de que necesita tener en cuenta una adecuada medida de los tiempos para que sea fiel con la realidad.

La única diferencia respecto a PCA es la llamada matriz de datos extendida, que como he mencionado antes tiene retrasos temporales, el resto de las operaciones será igual que en PCA.

$X(t)$  es la variable en el instante actual de tiempo y  $L$  es el número de retardos que se aplican, ec.6.

$$X_d = [X(t) X(t-1) X(t-2) \dots X(t-L)] \quad (6)$$

### 2.4.1.- PCA y DPCA para detección de fallos

Ya se ha explicado cómo funciona PCA, ahora el problema se encuentra en cómo hacer que se detecten los fallos, una vez aplicado este método. Para ello se van a aplicar dos estadísticas ( $T^2$  y  $Q$ ), junto con un umbral calculado para cada una de ellas, todo aquel valor por encima del umbral será considerado como un valor fuera de los límites establecidos y cuando se acumula un número consecutivo de valores fuera de su umbral se detecta un fallo.

En primer lugar, en la ec.7, se va a definir el estadístico de Hotelling (representado con  $T^2$ ), este se caracteriza por detectar patrones globales de anomalías, sirve para detectar características más generales, se obtiene de la siguiente forma:

$$T^2 = x s^{-1} x^t \quad (7)$$

En esta ecuación  $x$  son la proyección de las variables del proceso sobre los componentes principales y  $s$  representa la matriz covarianza de las componentes principales, o escrito de otra forma, y refiriéndonos a PCA, donde  $\Delta_k$  es la matriz diagonal formada por los  $k$  valores propios más grandes de la ec. 3:

$$T^2 = x V_k \Delta_k^{-1} V_k^t x^t \quad (8)$$

Como se ha dicho anteriormente el fallo se obtiene siempre y cuando se supere un umbral, el umbral se obtiene hallando el percentil 99 del valor calculado en la ecuación 7 u 8.

En cuanto al estadístico  $Q$  ec.9, también llamado residuo o error cuadrático de la reconstrucción, se centra más en las variables individuales, detecta fallos más locales, se calcula de la siguiente manera:

$$Q = r^t r \quad (9)$$

De forma que  $r$  es el vector residuos que resulta de la diferencia entre los valores originales y los valores originales proyectados por PCA, y nos queda en la ec.10 de la siguiente forma.

$$r = (I - V_k V_k^T) x^t \quad (10)$$

En este caso el umbral también se obtiene calculando el percentil 99 de la ec. (9).

Para el caso de DPCA la forma de calcular tanto  $T^2$  como  $Q$  será la misma explicada anteriormente, lo que cambia es la matriz original  $X \in \mathbb{R}^{n \times m}$ , ahora estará formada por la concatenación de la matriz actual más las tres matrices con datos anteriores, es decir  $X$  tendría el formato definido en la ec. (6),  $X_d \in \mathbb{R}^{n \times (4 \cdot m)}$ .

Una vez detectado el fallo para poder diagnosticarlo, vamos a calcular el diagrama de contribución de cada variable al fallo, es decir, como cada variable es responsable del fallo, y aquella variable con un valor más alto, es la variable que ha

fallado. Este diagrama de contribuciones, solo se calcula para la estadística Q y es el siguiente:

$$CON = [X(t_{fallo}, :) - X(t_{fallo}, :) * V_k V_k^T]^2 \quad (11)$$

Donde  $X(t_{fallo}, :)$  es el valor de todas las variables del proceso, en el instante en que la estadística Q detecta el fallo.

## 2.5 Análisis de factores lentos o SFA

Slow feature analysis es un modelo que está basado en la creencia de que los componentes más importantes en una serie temporal son los que varían más lentamente, por lo que dada una entrada vectorial aprende de señales cuya variación es mínima, para ello utiliza la expansión no lineal, cuyo propósito es hacer más notables las características complejas. Por ello es ideal para detectar pequeñas variaciones en vectores con un gran número de datos [5].

Normalizamos los datos de entrada de la planta,  $X \in \mathbb{R}^{n \times m}$ , para ello hacemos la resta respecto la media y dividimos entre la desviación típica, de forma análoga a lo que se hizo en PCA, en la ecuación (1). Después calculamos la matriz de covarianza de los datos que antes hemos normalizado, ec. (2), y nuevamente calculamos los autovalores y autovectores, ec. (3).

Calculamos la denominada matriz de blanqueo (ec. 12), y blanqueamos los datos (ec. 13), lo que permite eliminar las correlaciones cruzadas entre las variables de entrada.

$$Q = \Delta^{-1/2} V^t \quad (12)$$

$$Z = Q X^t \quad (13)$$

Se calcula la derivada primera de los datos blanqueados, Z, como los datos originales han sido medidos en cada instante de tiempo  $\Delta t$ , la derivada se puede aproximar empíricamente por la ec.14:

$$\dot{z}_j(t) = \frac{z_j(t) - z_j(t + \Delta t)}{\Delta t} \quad (14)$$

Si se considera que el periodo de muestreo donde se toman los datos,  $\Delta t=1$ , calcular la derivada primera es equivalente a diferenciar los datos ec.15.

$$Z_d[:, j] = Z[:, j] - Z[:, j - 1], \quad j \geq 1 \quad (15)$$

Después se calcula la matriz de covarianza de la diferencia temporal (ec. 16), y se descompone en sus valores singulares (ec. 17):

$$R_d = \frac{1}{n-1} Z_d Z_d^t \quad (16)$$

$$R_d = P \Omega P^t \quad (17)$$

Finalmente, la matriz de proyección  $W$  se calcula, como en la ec. 18 y ya se puede obtener los datos originales proyectados a través del método SFA (ec. 19):

$$W = P Q \quad (18)$$

$$S = W X^t \quad (19)$$

De igual forma que se hace en PCA, ahora se puede separar las componentes que varían rápidamente de las que lo hacen de forma más lenta, dividiendo  $W$  en dos matrices,  $W_a$  de componentes lentas y  $W_e$  de componente más rápidas y se obtienen dos nuevos espacios, la parte más dominante  $S_a = S^* W_a$  de componentes lentas y el residuo  $S_e = S^* W_e$ .

Como he dicho anteriormente SFA se basa en que los elementos en una serie temporal son también los que tienen cambios más lentos, esto puede suponer un problema debido a que solo tiene en cuenta el momento actual, por lo que los componentes lentos no pueden compararse con los vistos anteriormente, SFA no puede monitorear procesos dinámicos.

Para los procesos dinámicos se desarrolló DSFA [6], que aparte de tener en cuenta los componentes que cambian más lento, también tiene en cuenta las características dinámicas, este modelo tiene ventajas en una planta como la usada en este ejemplo ya que es un proceso dinámico. Como en PCA, la única diferencia entre SFA y DSFA es la matriz de datos de entrada, donde ahora en vez de tener sólo  $X$  en el instante actual, se introducen datos de instantes pasados (ec. 6).

### 2.5.1.- SFA y DSFA para detección de fallos

Para la detección de fallos con SFA primero hay que dividir el espacio  $S$  (ec. 18), en dos partes como en PCA, primero la parte dominante o más importante llamada  $S_a$  y el resto o residuo llamado  $S_e$ , como se indicó anteriormente ( $S_a = S^* W_a$  y  $S_e = S^* W_e$ ). A continuación, se calcula la estadística  $T^2$  para cada uno de esos espacios (ecs. 20 y 21). Donde  $T_a^2$  mide la variación estática dentro del subespacio definido por las componentes dominantes más lentas y  $T_e^2$  mide la variación dentro del subespacio definido por el resto de las componentes o del residuo.

$$T_a^2 = S_a^t S_a \quad (20)$$

$$T_e^2 = S_e^t S_e \quad (21)$$

Como también se ha trabajado con derivadas, se puede definir unos nuevos vectores como las derivadas primeras de  $S_a$  y  $S_e$ , es decir,  $\dot{S}_a$  y  $\dot{S}_e$  cómo se definió para la variable  $Z$ , y sus respectivas diferencias según las ecuaciones 14 y 15. Que en este caso quedaría:

$$S_{ad} = S_a(i, j) - S_a(i, j - 1) \quad (22)$$

$$S_{ed} = S_e(i, j) - S_e(i, j - 1) \quad (23)$$

Ahora se pueden definir dos estadísticas para estos nuevos espacios, de forma análoga a lo que se hizo para  $S_a$  y  $S_e$ , y que ahora llamaremos  $S^2$  como:

$$S_a^2 = S_{ad}^t \Omega_a^{-1} S_{ad} \quad (24)$$

$$S_e^2 = S_{ed}^t \Omega_e^{-1} S_{ed} \quad (25)$$

Donde  $\Omega_a$  y  $\Omega_e$  son la matriz de covarianza experimental de  $S_{ad}$  y  $S_{ed}$  respectivamente.

Como en PCA en SFA se calcularía el percentil 99 de  $S_a^2$ ,  $S_e^2$ ,  $T_a^2$  y  $T_e^2$  para obtener el umbral.

Una vez detectado el fallo, hay que diagnosticar el mismo, o lo que es lo mismo indicar la variable responsable del fallo, esto se hace calculando el diagrama de contribución de cada variable al fallo como se muestra en la siguiente ecuación:

$$CON = X(t_{falloT_a^2}, :) - \hat{X}(t_{falloT_a^2}, :) \text{ con } \hat{X}(t_{falloT_a^2}, :) = Q^{-1}P_a^T S_a \quad (26)$$

## 2.6 Análisis estadístico de patrones o SPA

Statistical Pattern Analysis es una metodología que usa varias estadísticas que analizan patrones de datos complejos. Esta metodología usa la relación entre distintas estadísticas de las variables del proceso, en vez de estudiar las propias variables, para estudiar el comportamiento de este. Es decir, SPA monitoriza las estadísticas del proceso, en nuestro caso, la media, la desviación estándar, la correlación, la skewness y la kurtosis de las variables, calculadas en una ventana de datos, en vez de las variables directamente [7, 8].

Para realizarlo debemos primero sacar las estadísticas de cada variable del proceso en una ventana de datos de magnitud  $k_v$ , como son:

La media:

$$\mu_j = \frac{1}{k_v} \sum_{i=1}^{k_v} x_{ij} \quad (27)$$

La varianza:



$$\sigma_j^2 = \frac{1}{k_v} \sum_{i=1}^{k_v} (x_{ij} - \mu_j)^2 \quad (28)$$

La Skewness:

$$skewness = \frac{1}{k_v} \sum_{i=1}^{k_v} \left( \frac{x_{ij} - \mu_j}{\sigma_j} \right)^3 \quad (29)$$

La Kurtosis:

$$kurt = \frac{1}{k_v} \sum_{i=1}^{k_v} \left( \frac{x_{ij} - \mu_j}{\sigma_j} \right)^4 - 3 \quad (30)$$

Y ahora con estas estadísticas de cada variable del proceso, en la ec.31 construimos la matriz de datos X, con dimensiones 4\*m, siendo m el número de variables del proceso:

$$X = [\mu_1, \mu_2, \dots, \mu_m, \sigma_1^2, \sigma_2^2, \dots, \sigma_m^2, Skew_1, \dots, Skew_m, Kurt_1, \dots, Kurt_m] \quad (31)$$

Y ahora se aplica la técnica PCA sobre esta matriz de datos, es decir, primero se normalizan los datos que nos dan, restando la media y dividiendo entre la varianza, como en la ec.(1), después se calcula la matriz de covarianza (ec.2), y luego se calculan los autovalores y autovectores de esta matriz de covarianza, o los valores singulares (ec.3).

En la ec.32 se elige un número a de elementos, es decir, un número a de componentes principales que conserven el 90% de la varianza total de los datos, este número marcará que datos cogemos y cual obviamos.

$$\sum_{i=1}^a \frac{S_i}{\sum S} \times 100 \geq 90\% \quad (32)$$

Una vez elegidos los componentes principales, podemos proyectar los datos originales sobre este nuevo espacio como en la (ec.4), y calculas las estadísticas T<sup>2</sup> y Q, como se obtuvieron en PCA (ecs. 7,8)

## 2.7 Análisis de Componentes Principales Sparse o SPCA

En PCA cada componente principal es una combinación lineal de todas las variables originales; esto puede llegar a dificultar la interpretación de los resultados. SPCA hace a grandes rasgos lo mismo que PCA, pero selecciona solo un subconjunto determinado de variables importantes, es decir, hace cero los



coeficientes de las variables que contribuyen poco a cada componente principal, de forma que ahora en SPCA, cada componente principal no es una combinación lineal de todos las variables del proceso, sino solo de unas cuantas, las más importantes para ese componente principal, esto, debería hacer que los resultados se interpretasen de una forma más clara, ya que es un modelo que esta más centrado en el objetivo.

Para ello se plantea PCA como un problema de regresión tipo lasso o red elástica, que, con el objetivo de crear un subconjunto de variables, da el valor de 0 a las que se quedan fuera del mismo, resolviéndolo como el siguiente problema de optimización, ec.33, donde  $P=V_k$ , es decir, la matriz de los componentes principales (ec.4):

$$\min \|X' - X' P P^T\|^2 + \alpha \|P\|_1 \quad (33)$$

El DSPCA, es decir, SPCA dinámico tiene en cuenta los datos dinámicos, es decir, busca el subconjunto de variables que describan mejor la naturaleza de los datos y también permite analizar componentes principales a lo largo del tiempo, como he mencionado anteriormente, esto debería resultar muy útil debido a la índole de nuestros datos. Además de lo mencionado anteriormente aplica la reducción de la dimensionalidad de PCA. Para resolver este problema simplemente se aplica SPCA sobre la matriz de datos aumentadas con los variables en instantes anteriores de tiempo (ec. 6).

Este método debería ser útil si los datos varían a lo largo del tiempo y si hay muchas variables, pero solo unas pocas son útiles para analizar.

Una vez calculados los componentes principales tanto con el método SPCA como DSPCA, para detectar anomalías se utilizan las estadísticas  $T^2$  y  $Q$ , definidas en la sección 2.4.1. de este capítulo.



Universidad de Valladolid



ESCUELA DE INGENIERÍAS  
INDUSTRIALES



Universidad de Valladolid



ESCUELA DE INGENIERÍAS  
INDUSTRIALES

## Capítulo 3: Planta y datos



Universidad de Valladolid



ESCUELA DE INGENIERÍAS  
INDUSTRIALES

### 3.1 introducción a la planta

Durante la posguerra, después de la primera guerra mundial, hubo un déficit en la adquisición de materias primas, y el fundador de Eastman kodak, la famosa empresa dedicada a la fotografía, necesitaba una serie de productos químicos para que la producción de su empresa fotográfica siguiese funcionando correctamente, para ello fundo la Eastman Chemical Company en Kingsport EEUU, en 1920, y en los años siguientes el éxito de la fábrica haría que expandiese su negocio hacia otras ramas diferentes a la fotografía, como polímeros o fibras sintéticas.

En 1994 la Eastman Chemical Company y la Eastman Kodak se separaron de forma definitiva dando de esta forma lugar a una empresa independiente.

El proceso Tennessee Eastman (Figura 2) fue desarrollado por Downs y Vogel, dos reputados químicos, en el año 1993, hoy en día, este modelo se sigue usando, es muy útil porque lo crearon en base a un modelo real, esto dio lugar a que fuese muy complejo y exacto [9].

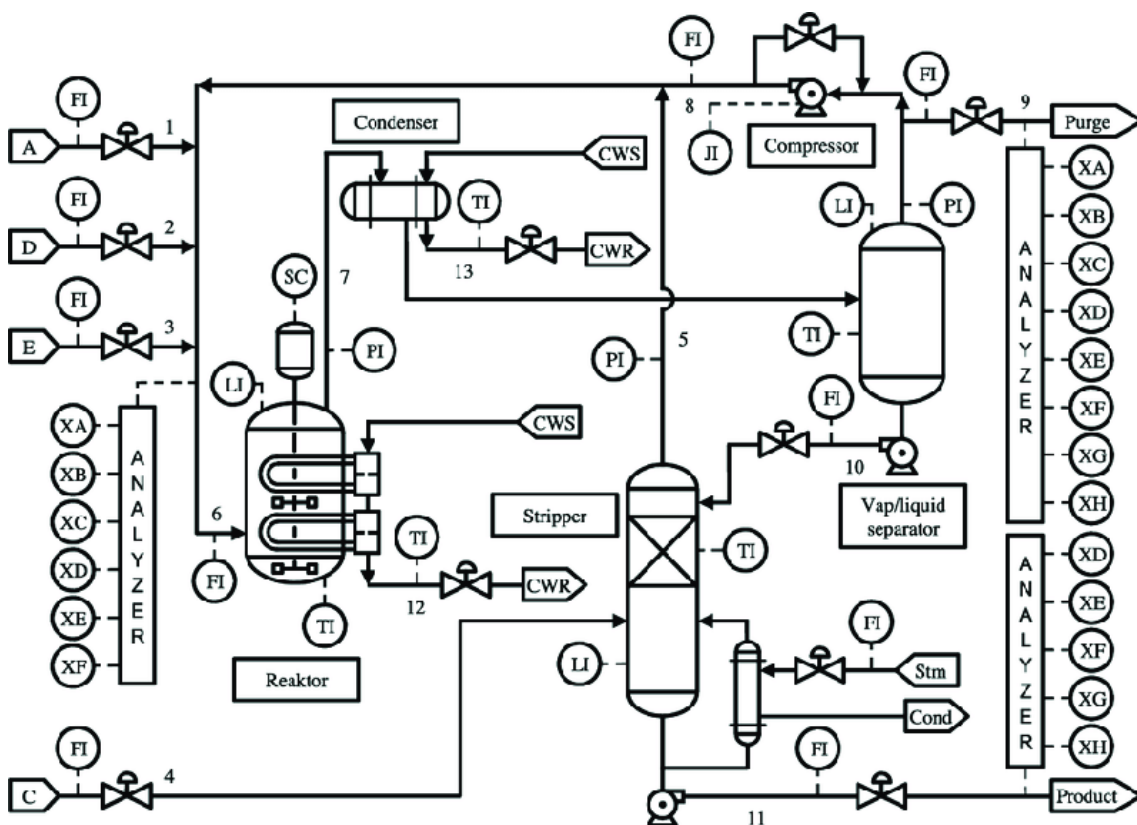


Figura 2.- Esquema de la planta TE



En este proceso químico se obtienen cuatro reactivos (Figura 3) a partir de dos productos iniciales, estos son los elementos deseados, pero en la reacción también se obtienen un componente inerte y un subproducto.

Las cuatro reacciones son irreversibles, es decir que una vez que obtienes el producto no puedes volver a obtener el original, y exotérmicas, esto quiere decir que el proceso de obtener el producto se desprende calor de la reacción. Las velocidades de reacción siguen una expresión de Arrhenius, lo que implica que la velocidad de reacción se ve afectada por la temperatura, cuanto más velocidad de reacción más temperatura, esto es peligroso por las grandísimas temperaturas que se pueden llegar a alcanzar, de forma que es necesario controlar de forma precisa la velocidad de reacción para que no se descontrole la temperatura y se pueda causar un accidente.

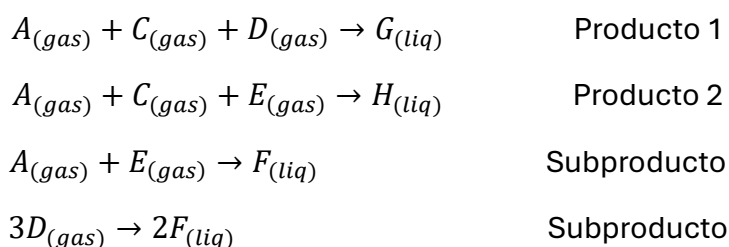


Figura 3.- Reacciones en la planta TE .

Este modelo es útil porque permite evaluar diferentes técnicas o estrategias para obtener un control de la planta lo más preciso posible. En el caso que nos vamos a centrar permite estudiar fallos y situaciones anómalas, esto hace más sencillo probar diferentes sistemas de detección y diagnóstico de fallos, en el siguiente apartado es lo que se va a comprobar, como se comportan los diferentes modelos estadísticos que hemos explicado de forma teórica ante los datos tomados del modelo de Tennessee Eastman. También permite optimizar la eficacia de un proceso, tanto de materias primas como de energía, y sirve además para mejorar los sistemas de seguridad de una planta, ya que permite probar su eficacia en condiciones de funcionamiento muy difíciles dentro del modelo, por último, también se usa para crear nuevos modelos estadísticos de control [9].

### 3.2 Datos históricos

El proceso Tennessee Eastman se creó usando un modelo de la planta simulada durante 2880 minutos, en este tiempo se crearon 960 medidas, de todas estas muestras las 160 primeras se dan en condiciones normales, y 800 contienen algún tipo de fallo. Se midieron un total de 52 variables, 41 medidas y 11 manipuladas, esto da aproximadamente 49920 medidas únicas durante la duración del experimento. En la base de datos se dan 21 condiciones operativas diferentes (ensayos) en los que se encuentra algún tipo de fallo, aparte del ensayo en el que la planta se encuentra funcionando en condiciones normales, necesario para saber cuáles son los valores buscados para el correcto funcionamiento.

En la figura 4 se muestran los 21 tipos de fallos diferentes que se pueden dar en la planta, en la figura se ve una breve descripción del fallo acompañada de una indicación del tipo de fallo que nos indica si es rápida (step), o sticking que serían válvulas atascadas que no responden a las señales.

| No. | Fault Description                                      | Fault Type        |
|-----|--|-------------------|
| 1   | A/C feed ratio, B composition constant (stream 4)      | Step              |
| 2   | B composition, A/C ratio constant (stream 4)           | Step              |
| 3   | D feed temperature (stream 2)                          | Step              |
| 4   | Reactor cooling water inlet temperature                | Step              |
| 5   | Condenser cooling water inlet temperature              | Step              |
| 6   | A feed loss (stream 1)                                 | Step              |
| 7   | C header pressure loss-reduced availability (stream 4) | Step              |
| 8   | A, B, C feed composition (stream 4)                    | Random variation  |
| 9   | D feed temperature (stream 2)                          | Random variation  |
| 10  | C feed temperature (stream 4)                          | Random variation  |
| 11  | Reactor cooling water inlet temperature                | Random variation  |
| 12  | Condenser cooling water inlet temperature              | Random variation  |
| 13  | Reaction kinetics                                      | Slow drift        |
| 14  | Reactor cooling water valve                            | Sticking          |
| 15  | Condenser cooling water valve                          | Sticking          |
| 16  | Unknown  | -                 |
| 17  | Unknown  | -                 |
| 18  | Unknown  | -                 |
| 19  | Unknown  | -                 |
| 20  | Unknown  | -                 |
| 21  | The valve for stream 4                                 | Constant position |

Figura 4.- Fallos del proceso Tennessee Eastman

En la figura 5 se muestran las variables medidas en el Tennessee Eastman Process, las variables XMEAS son las variables que pueden ser medidas usando los distintos sensores que tiene la planta, por ejemplo, el flujo de alimentación que tiene el producto A o el flujo del condensador.



Las variables XMV son aquellas variables que pueden ser manipuladas, un ejemplo se da en la manipulación del caudal de A, que anteriormente hemos dicho que también podía ser una variable medida, otro ejemplo es el flujo de vapor de la torre de destilación.

| #  | Variable  | Descripción  |
|----|-----------|--|
| 1  | XMEAS(1)  | Flujo de A (alimentación, línea 1)                           |
| 2  | XMEAS(2)  | Flujo de D (alimentación, línea 2)                           |
| 3  | XMEAS(3)  | Flujo de E (alimentación, línea 3)                           |
| 4  | XMEAS(4)  | Flujo de A + C (alimentación, línea 4)                       |
| 5  | XMEAS(5)  | Flujo de reciclaje   |
| 6  | XMEAS(6)  | Flujo a la entrada del reactor (línea 6)                     |
| 7  | XMEAS(7)  | Presión del reactor  |
| 8  | XMEAS(8)  | Nivel del reactor  |
| 9  | XMEAS(9)  | Temperatura del reactor                                      |
| 10 | XMEAS(10) | Flujo de purga de gas  |
| 11 | XMEAS(11) | Temperatura del separador                                    |
| 12 | XMEAS(12) | Nivel del separador  |
| 13 | XMEAS(13) | Presión del separador  |
| 14 | XMEAS(14) | Flujo de salida del separador (línea 10)                     |
| 15 | XMEAS(15) | Nivel del stripper   |
| 16 | XMEAS(16) | Presión del stripper   |
| 17 | XMEAS(17) | Flujo de salida del stripper (línea 11)                      |
| 18 | XMEAS(18) | Temperatura del stripper                                     |
| 19 | XMEAS(19) | Flujo de vapor del stripper                                  |
| 20 | XMEAS(20) | Energía del compresor  |
| 21 | XMEAS(21) | Temperatura de salida de agua de enfriamiento del reactor    |
| 22 | XMEAS(22) | Temperatura de salida del agua de enfriamiento del separador |



|    |           |   |
|----|-----------|---|
| 23 | XMEAS(23) | Concentración de componente A a la entrada del reactor (línea 6)  |
| 24 | XMEAS(24) | Concentración de componente B a la entrada del reactor (línea 6)  |
| 25 | XMEAS(25) | Concentración de componente C a la entrada del reactor (línea 6)  |
| 26 | XMEAS(26) | Concentración de componente D a la entrada del reactor (línea 6)  |
| 27 | XMEAS(27) | Concentración de componente E a la entrada del reactor (línea 6)  |
| 28 | XMEAS(28) | Concentración de componente F a la entrada del reactor (línea 6)  |
| 29 | XMEAS(29) | Concentración de componente A en el gas de purga (línea 9)        |
| 30 | XMEAS(30) | Concentración de componente B en el gas de purga (línea 9)        |
| 31 | XMEAS(31) | Concentración de componente C en el gas de purga (línea 9)        |
| 32 | XMEAS(32) | Concentración de componente D en el gas de purga (línea 9)        |
| 33 | XMEAS(33) | Concentración de componente E en el gas de purga (línea 9)        |
| 34 | XMEAS(34) | Concentración de componente F en el gas de purga (línea 9)        |
| 35 | XMEAS(35) | Concentración de componente G en el gas de purga (línea 9)        |
| 36 | XMEAS(36) | Concentración de componente H en el gas de purga (línea 9)        |
| 37 | XMEAS(37) | Concentración de componente D a la salida de la planta (línea 11) |
| 38 | XMEAS(38) | Concentración de componente E a la salida de la planta (línea 11) |
| 39 | XMEAS(39) | Concentración de componente F a la salida de la planta (línea 11) |

|    |           |   |
|----|-----------|---|
| 40 | XMEAS(40) | Concentración de componente G a la salida de la planta (línea 11) |
| 41 | XMEAS(41) | Concentración de componente H a la salida de la planta (línea 11) |
| 42 | XMV(1)    | Flujo de alimentación de D (línea 2)                              |
| 43 | XMV(2)    | Flujo de alimentación de E (línea 3)                              |
| 44 | XMV(3)    | Flujo de alimentación de A (línea 1)                              |
| 45 | XMV(4)    | Flujo de alimentación de A y C (línea 4)                          |
| 46 | XMV(5)    | Válvula de recirculación del compresor                            |
| 47 | XMV(6)    | Válvula de la purga (línea 9)                                     |
| 48 | XMV(7)    | Flujo de líquido del separador (línea 10)                         |
| 49 | XMV(8)    | Flujo de producto del stripper (línea 11)                         |
| 50 | XMV(9)    | Válvula de vapor del stripper                                     |
| 51 | XMV(10)   | Flujo de agua de enfriamiento del reactor                         |
| 52 | XMV(11)   | Flujo de agua de enfriamiento del condensador                     |

Figura 5.- Variables medidas de la TEP

Los datos utilizados para este trabajo son públicos y se pueden descargar en el siguiente enlace (<http://web.mit.edu/braatzgroup/links.html>). Sobre estos, se aplicarán los diferentes métodos de análisis estadístico multivariante, explicados en el capítulo anterior, y de esta manera, monitorizar el proceso.

Además, estos datos del proceso Tennessee Eastman han servido para muchas aplicaciones, entre ellas destacan para nuestro caso las de desarrollo de métodos para distinguir distintos errores o casos anómalos en una planta de producción [8], para ello se utilizó la técnica estadística PCA, además de los estadísticos Q y T2, en este trabajo además de PCA vamos a utilizar otras técnicas estadísticas como SFA o SPA entre otras. Con PCA se logró una precisión a la hora de encontrar errores de un 54%, con otras técnicas estadísticas esto podría ser más preciso.



Universidad de Valladolid



ESCUELA DE INGENIERÍAS  
INDUSTRIALES



## Capítulo 4: Aplicación



Universidad de Valladolid



ESCUELA DE INGENIERÍAS  
INDUSTRIALES



## 4.1 Introducción

Vamos a comparar las distintas técnicas de control estadístico multivariante explicadas en el capítulo 2, para ello se van a usar los datos de la planta Eastman Chemical Company de la que se ha hablado en el capítulo anterior, es decir, ante la misma entrada se va a observar cuál de las salidas es la que mejor funciona y así saber cuál de todos los métodos es el más adecuado.

El tamaño de los datos utilizados es de cincuenta y dos variables por novecientas sesenta observaciones, se dispone de unos datos en los que el funcionamiento de la planta es normal, este se usará para entrenar los distintos modelos estadísticos de forma que así sepan cómo debe funcionar la planta en condiciones normales, además del archivo de funcionamiento normal hay otros veintiún archivos en los que la planta sufre distintos tipos de errores, de todos ellos el siete y el doce son los más representativos a la hora de representarlos en una gráfica.

Los fallos tres, nueve y quince, son los más difíciles de detectar, los dos primeros están relacionados con la variable D, puede ser que se dé el caso en el que algún modelo estadístico no los llegue a detectar, esto también es un gran problema debido a que el fallo está ahí, pero si no se detecta la fabricación es posible que salga defectuosa o que se sufra algún tipo de accidente en la planta.

Cada uno de estos ficheros de datos de las novecientas sesenta observaciones, tiene ciento sesenta en las que no hay errores, en caso de que alguna técnica de control estadístico multivariante diese error en este periodo se marcará un error de funcionamiento muy claro de ese modelo, esto es así porque es igual de importante detectar los fallos de forma correcta como no dar falsos positivos, los falsos positivos provocan una bajada en la producción y en caso de que se repitan de forma excesiva pueden llegar a que el operario tome por un falso positivo un error verdadero.

Se van a tener en cuenta una serie de parámetros que nos indicaran como de buenos son los distintos modelos estadísticos que vamos a estudiar.

El primer parámetro es el porcentaje de falsas alarmas, que es importante por lo que se acaba de mencionar. Este se obtiene del número de veces que la correspondiente estadística ( $T^2$  o  $Q$ ) ha superado su umbral en la zona en la que no puede haber errores, es decir antes del instante ciento sesenta, el umbral se obtiene haciendo el percentil noventa y nueve de cada estadístico en condiciones normales de operación.

Se tendrán en cuenta también el número de alarmas que detecten los distintos métodos estadísticos, una alarma se detecta cuando las estadísticas superan el umbral. Solo se medirá en el periodo de ciento sesenta a novecientos sesenta, la zona en la que se pueden dar fallos.



Por último, se observará el tiempo de detección, este es el intervalo de tiempo desde que ocurre un fallo hasta que el modelo estadístico lo detecta.

Para que las distintas técnicas estadísticas multivariantes sean lo más efectivas lo deseable es que el porcentaje de alarmas detectadas sea el más cercano a cien posible, el porcentaje de falsas alarmas lo más próximo a cero que se pueda, y en cuanto al tiempo de detección lo ideal es que cuanto más pequeño sea mejor, en la realidad nunca va a ser cero.

Todas las técnicas estadísticas están programadas en el lenguaje Python, en la versión 3.11, usando el entorno de desarrollo Spyder

## 4.2 PCA

Para aplicar el modelo de PCA lo primero que hay que hacer es entrenar el modelo utilizando unos datos en los que el funcionamiento de la planta está en condiciones normales de operación, el primer paso consiste en preparar estos datos aplicando un pretratamiento de estos, este consiste en eliminar variables inapropiadas como las que tienen errores muy grandes, escalar las variables para asegurar que todas tienen el mismo peso y eliminar datos espurios (outliers), para ello vamos a normalizarlos a media 0 y varianza 1 para después aplicar el método de PCA.

Una vez obtenida la matriz de datos normalizados, calculamos la matriz covarianza, a continuación vamos a aplicar la descomposición en valores singulares SVD para sacar los valores propios y vectores propios de la matriz, después vamos a multiplicar los datos normalizados por los vectores singulares para tener las componentes principales, a continuación hay que definir la variabilidad, esta es la cantidad de varianza total en los datos originales que es capturada por los componentes principales, en nuestro caso será del 90%. La detección de fallos es la última parte del programa, en ella sacamos los estadísticos  $T^2$ , y  $Q$  y guardamos los umbrales obtenidos de los datos con funcionamiento normal, también guardaremos otras variables como la media o la desviación estándar.

Con los datos guardados anteriormente volvemos a aplicar PCA, esta vez sobre unos datos en los que el funcionamiento de la planta no es el deseado para ver como este método detecta los posibles fallos que pueden ocurrir en la planta.

En primer lugar, se van a mostrar el funcionamiento normal de la planta, sin errores, esto se hace para que el programa sepa que valores serían los adecuados, de esta forma si están fuera de los mismos el modelo estadístico lo detectará como fallo.

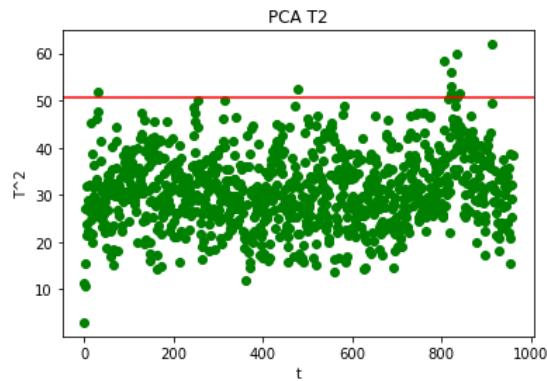


Figura 6.- Estadística  $T^2$  y su umbral en condiciones normales de operación

En las figuras 6 y 7, en verde se muestra el estadístico  $T^2$  y en azul el  $Q$ , las líneas rojas representan el umbral, este servirá para saber si se detecta fallo.

El estadístico  $T^2$  o de Hotelling se va a aplicar con una significancia ( $\alpha$ ) con valor 0,01; al aplicarla la matriz que se obtiene posee una dimensión  $1 \times 960$ , o lo que es lo mismo, el valor del estadístico  $T^2$  para cada una de las 960 observaciones, el valor del umbral obtenido es 50,63.

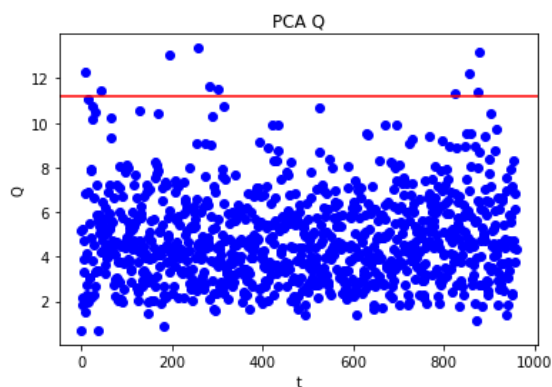


Figura 7.- Estadística  $Q$  y su umbral en condiciones normales de operación

Para el estadístico  $Q$  o también llamado SPE, vamos a usar el mismo  $\alpha$  con valor 0,01; en este caso el valor del umbral será de 11,18.

### 4.2.1.- Detección de fallos usando PCA

Se van a estudiar los 21 ficheros con cada uno de los fallos pertenecientes a la figura 4 mostrada en el capítulo anterior, de todos los posibles fallos solo se van a mostrar dos por ser los que mejor representan el buen o mal funcionamiento del modelo, estos son el fallo siete y el fallo doce, el fallo siete indica la pérdida de presión en un cabezal, mientras que el doce es la variación aleatoria en la temperatura de entrada del agua de enfriamiento del condensador.

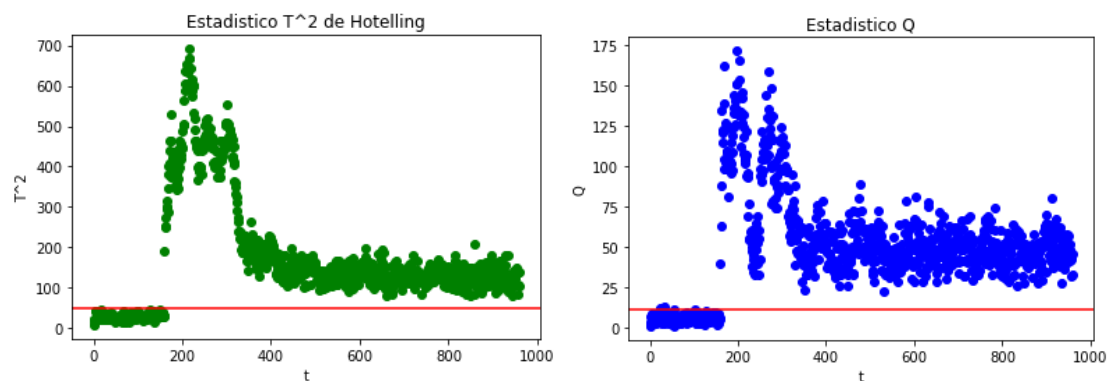


Figura 8.- Estadísticas  $T^2$  y  $Q$  para el fallo 7 usando el método PCA

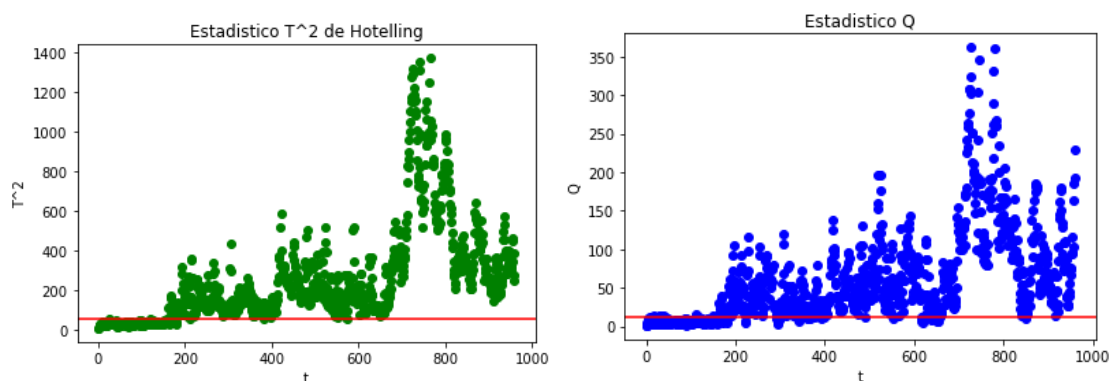


Figura 9.- Estadísticas  $T^2$  y  $Q$  para el fallo 12 usando el método PCA

En las figuras 8 y 9, se pueden ver las estadísticas  $T^2$  y  $Q$ , para los dos fallos comentados. Inicialmente hasta el instante de tiempo 160, el funcionamiento de la planta es el correcto, ningún valor supera el umbral, (indicado por la línea roja), a partir de ese instante es cuando pueden producirse fallos.

En las mismas figuras también, podemos ver tanto en el estadístico  $T^2$  como en el  $Q$  y como de forma clara salta un fallo y se sobrepasa el umbral en el instante 160 en ambos casos. Este es el ejemplo en el que el salto es más acusado y en el que mejor se ve el antes y después del momento 160.

En la Figura 10 se muestra el diagrama de contribución de las 52 variables que nos dan los datos y en cual de esas variables de la planta se encuentra el error, se puede ver como en el fallo 7 el error viene dado en la variable 4, perteneciente al flujo de alimentación de C (en la línea 4), lo que tiene sentido ya que este fallo es una caída de la presión en el stripper, por lo que funciona mal y como consecuencia la entrada de flujo en ese elemento se modifica (línea 4), además se ve también que la siguiente variable más importante es la 45, que es la variable de control que trata de controlar este flujo de alimentación al stripper. En el fallo 12 además de la variable 4 también se encuentra con un valor bastante alto la variable 11, correspondiente a la temperatura del separador, lo que también es correcto ya que el fallo 12 es una modificación de la temperatura del agua de entrada al condensador, lo que modifica su temperatura de salida que es el agua de refrigeración que va al separador.

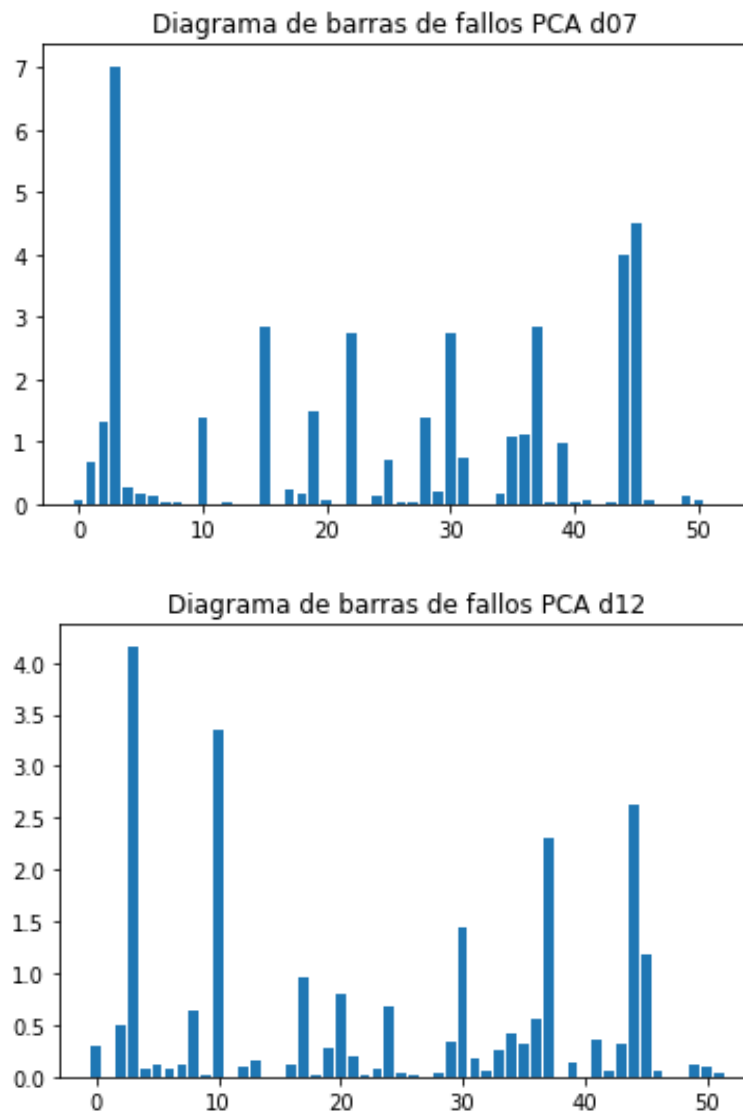


Figura 10.- Variables que fallan PCA



Para los distintos modelos estadísticos que vamos a utilizar el más deseable es aquel que muestre de una forma más clara donde se encuentra el error o los errores.

La Tabla 1 muestra los valores exactos que detecta PCA para los veintiún archivos de fallos que tiene la planta Tennessee Eastman, se puede observar cómo no aparece ningún fallo antes del instante de tiempo 160, en caso de aparecer habría problemas con el modelo estadístico puesto que antes del instante 160 el funcionamiento de la planta es el deseado. Las falsas alarmas sufridas tienen un valor bajo para ambos estadísticos, aunque para el Q es un poco mayor, este entra dentro de los parámetros deseados. Las alarmas detectadas no son todo lo altas que deberían, se encuentran en torno al 70 por ciento del total para la estadística T<sup>2</sup> y 80% para la estadística Q.

En conclusión, podemos decir que este método tiene un funcionamiento adecuado, pero no muy bueno, no sufre un excesivo número de falsas alarmas, pero no funciona correctamente a la hora detectar alarmas, siendo este su principal objetivo.

| PCA   | T2      |              |              | Q       |              |              |
|-------|---------|--------------|--------------|---------|--------------|--------------|
|       | T fallo | falsa alarma | alarma detec | T fallo | falsa alarma | alarma detec |
| 1     | 6       | 1,250        | 99,250       | 2       | 3,750        | 99,750       |
| 2     | 11      | 1,875        | 98,625       | 14      | 4,375        | 98,625       |
| 3     | nd      | 1,250        | 1,875        | nd      | 2,500        | 7,500        |
| 4     | 504     | 0,625        | 41,875       | 0       | 2,500        | 100,000      |
| 5     | 10      | 0,625        | 25,750       | 0       | 2,500        | 33,875       |
| 6     | 9       | 0,625        | 98,875       | 0       | 2,500        | 100,000      |
| 7     | 0       | 0,000        | 100,000      | 0       | 2,500        | 100,000      |
| 8     | 22      | 0,000        | 97,250       | 17      | 4,375        | 96,875       |
| 9     | nd      | 3,125        | 3,375        | nd      | 4,375        | 4,500        |
| 10    | 103     | 1,875        | 30,750       | 47      | 0,000        | 46,000       |
| 11    | 50      | 0,000        | 51,250       | 6       | 5,000        | 69,375       |
| 12    | 21      | 1,250        | 98,625       | 22      | 6,250        | 95,125       |
| 13    | 45      | 1,875        | 94,625       | 40      | 1,250        | 95,125       |
| 14    | 0       | 0,625        | 99,500       | 1       | 5,000        | 99,875       |
| 15    | nd      | 1,250        | 2,500        | nd      | 3,125        | 6,125        |
| 16    | 310     | 3,750        | 15,125       | 193     | 3,125        | 43,500       |
| 17    | 28      | 0,625        | 78,625       | 21      | 3,125        | 95,625       |
| 18    | 92      | 1,250        | 89,375       | 83      | 5,625        | 90,125       |
| 19    | nd      | 0,625        | 12,250       | nd      | 4,375        | 21,875       |
| 20    | 86      | 2,500        | 31,250       | 84      | 3,750        | 55,875       |
| 21    | 504     | 1,250        | 41,375       | 249     | 3,750        | 50,125       |
| MEDIA | 105,941 | 1,176        | 70,125       | 45,824  | 3,493        | 80,581       |

Tabla 1: Tabla de resultados con PCA

En la tabla 1 se pueden ver los datos obtenidos para cada uno de los 21 fallos, en caso de que el tiempo de fallo sea nd significa que ese fallo no ha sido detectado, y por lo tanto no se tendrá en cuenta para la media.

En la figura 3 se pueden observar cuales son las distintas variables, la tabla 2 analiza cuales de esas variables son las causantes de cada fallo. Para detectar los fallos se ha usado la siguiente ecuación, siendo P una matriz con un numero de columnas k, es decir,  $P=V_k$ , como se vio en la ec. 11 del capítulo 2.

$$fallo = (Xn[t_{fallo}] - Xn[t_{fallo}] p p^t)^2 \quad (34)$$

En la tabla 2 se muestran las variables que provocan cada fallo, el espacio vacío de la tabla se debe a que el error no se llega a detectar, por tanto, la ec. 34 no se puede calcular y eso produce que el valor del fallo obtenido no sea relevante.

| Nº fallo | Variable |  | Nº fallo | Variable |
|----------|----------|--|----------|----------|
| 1        | 21       |  | 12       | 4, 11    |
| 2        | 31       |  | 13       | 29       |
| 3        | -        |  | 14       | 21       |
| 4        | 35       |  | 15       | -        |
| 5        | 22       |  | 16       | 19       |
| 6        | 44       |  | 17       | 21       |
| 7        | 4        |  | 18       | 11       |
| 8        | 21       |  | 19       | -        |
| 9        | -        |  | 20       | 46       |
| 10       | 18       |  | 21       | 42       |
| 11       | 51       |  |          |          |

Tabla 2: Variables que fallan en PCA

#### 4.2.2.- Detección de fallos usando DPCA

Para utilizar el modelo estadístico DPCA vamos a seguir los pasos usados anteriormente, lo único que cambiará es la forma en la que leemos los datos para hacerlos dinámicos, para ello dada la matriz de datos iremos creando copias de está desplazando la columna de datos una posición a la derecha, en nuestro caso haremos esta operación 4 veces, después concatenaremos las matrices nuevas obteniendo una matriz de tamaño 960x208.

$$X(t) = [X(t) X(t - 1) X(t - 2)X(t - 3)X(t - 4)] \quad (35)$$

Para DPCA los resultados en los estadísticos de Hotelling y Q (Figuras 11 y 12), son muy similares a los de PCA, en la Figura 7, se observa que los errores se repiten cuatro veces, esto es así porque al hacer dinámico el modelo creamos cuatro matrices desplazadas, los errores que detecta siguen teniendo los mismos valores que PCA.

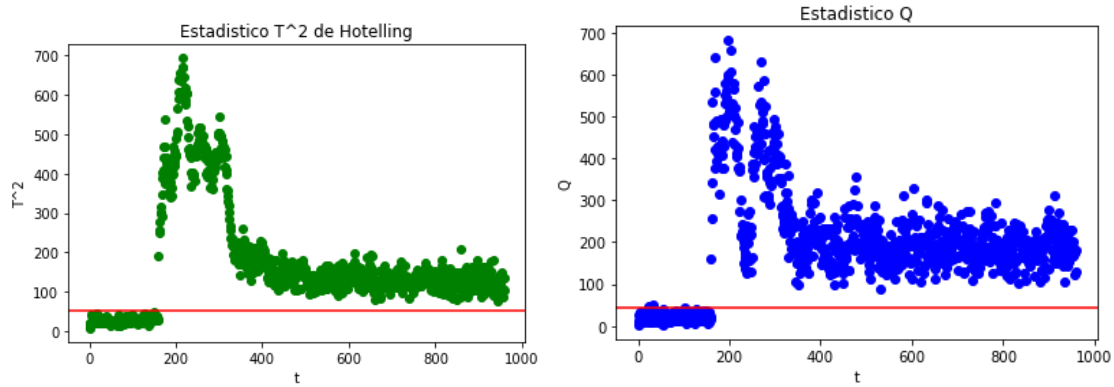


Figura 11: Detección de fallos usando DPCA para el fallo 7

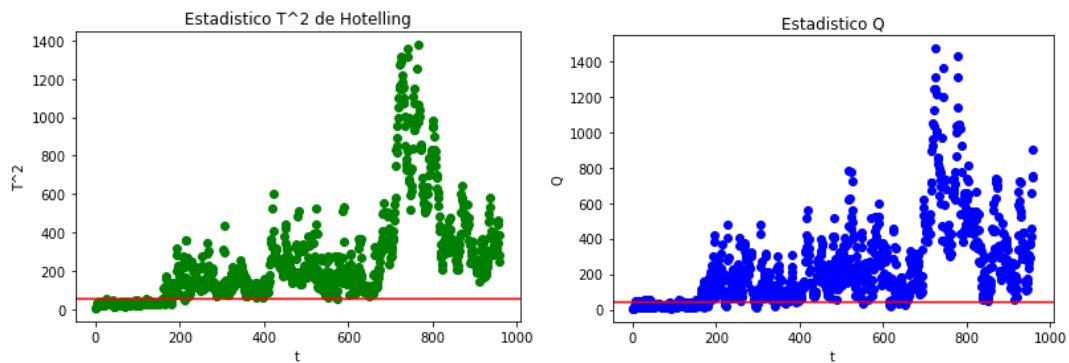


Figura 12: Detección de fallo usando DPCA para el fallo 12

En la figura 13, se representa el valor de las variables cuando se detectan los fallos (7 a la izquierda y 12 a la derecha). Como DPCA es un modelo dinámico, este capta relaciones dinámicas a lo largo del tiempo, en cambio PCA solo tiene en cuenta la estructura estática de los datos, este es el motivo de que aplicando los mismos datos los dos modelos salgan tan diferentes, los picos que aparecen de igual valor y distribuidos de igual forma en el espacio temporal se deben al modelo dinámico. Y se puede ver como en el fallo 7 el error viene dado en la variable 4, y la misma variable retardada una, dos y tres veces, mientras que en el fallo 12 se encuentra en la variable 11, y sus respectivos retardos perteneciente a la temperatura del separador.

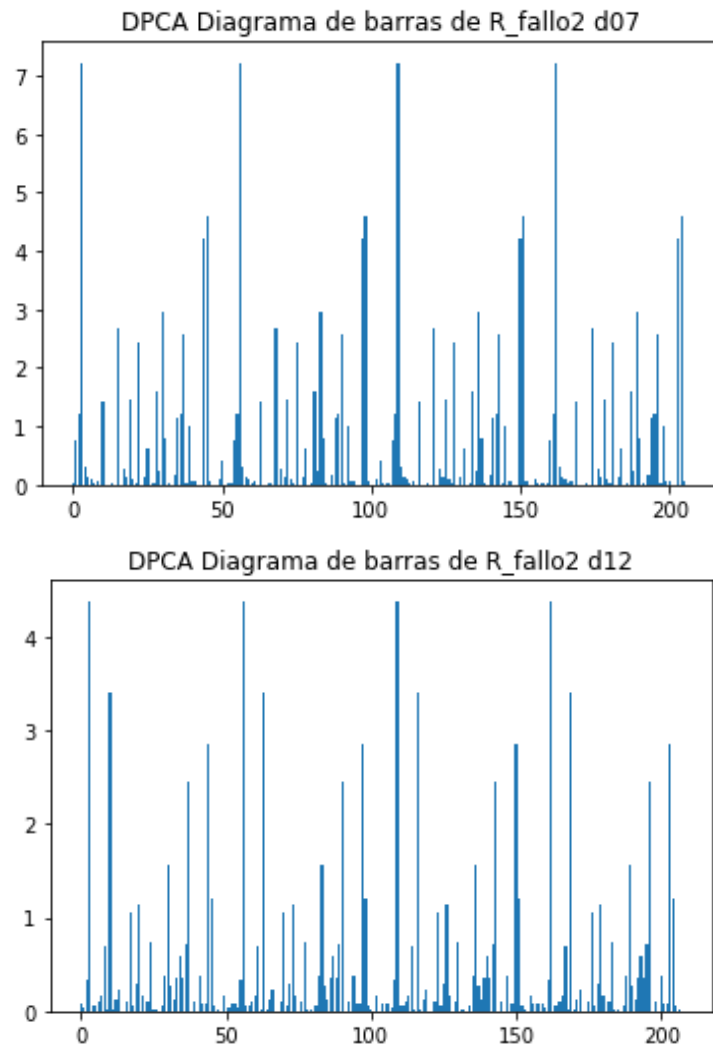


Figura 13.- Variables que fallan usando DPCA

Mirando la tabla de resultados de este método DPCA (Tabla 3), respecto a la de PCA (Tabla 1), podemos ver un ligero aumento en el número de alarmas detectadas, pero a su vez también han subido las falsas alarmas, por tanto, podemos concluir que en este caso el modelo dinámico no aporta grandes mejoras respecto al convencional, siendo los resultados obtenidos en el primero peores de lo deseado.

### 4.3 SFA

Se van a comprobar los mismos datos usados anteriormente ahora aplicando la técnica estadística de SFA, estos datos son los que reflejan el comportamiento normal de la planta.

Las diferencias respecto a PCA son que SFA usa una matriz de covarianza modificada con diferencia temporal y los estadísticos usados, que en este caso son  $T_a^2$ ,  $T_e^2$ ,  $S_a^2$  y  $S_e^2$  explicados en el capítulo 2 de este TFG, siguiendo las ecuaciones



(ecs. 20,25). De igual forma el umbral para cada estadística se calcula usando el percentil 99 de su valor en condiciones normales de operación.

La figura 14 representa lo mismo que las del apartado anterior, es decir el resultado de SFA en condiciones normales de operación, la diferencia es que en SFA hay cuatro en lugar de dos, las gráficas se dividen  $T^2_a$ ,  $T^2_e$ ,  $S^2_a$  y  $S^2_e$ , las a representan las 31 primeras posiciones, es decir los valores que me quedo al reducir la dimensión del espacio original con SFA mientras las e son el resto, es decir si en total son 52 variables la e representa las 21 últimas. Comentar, que la reducción del espacio con SFA se ha hecho con los mismos valores que en PCA, para poder comparar resultados.

| DPCA  | T2      |              | Q            |         |              |              |
|-------|---------|--------------|--------------|---------|--------------|--------------|
|       | T fallo | falsa alarma | alarma detec | T fallo | falsa alarma | alarma detec |
| 1     | 6       | 1,250        | 99,250       | 2       | 4,375        | 99,750       |
| 2     | 11      | 1,250        | 98,625       | 14      | 4,375        | 98,500       |
| 3     | nd      | 0,625        | 1,750        | nd      | 0,625        | 6,250        |
| 4     | 0       | 0,625        | 7,500        | 0       | 3,125        | 100,000      |
| 5     | 10      | 0,625        | 25,000       | 0       | 3,125        | 28,750       |
| 6     | 5       | 0,000        | 99,625       | 0       | 1,875        | 100,000      |
| 7     | 0       | 0,000        | 100,000      | 0       | 1,875        | 100,000      |
| 8     | 22      | 0,000        | 97,250       | 17      | 3,750        | 96,750       |
| 9     | nd      | 3,125        | 3,000        | nd      | 3,750        | 4,250        |
| 10    | 103     | 0,625        | 30,125       | 47      | 0,000        | 44,875       |
| 11    | 382     | 0,000        | 38,375       | 6       | 4,375        | 69,875       |
| 12    | 21      | 1,250        | 98,500       | 22      | 5,625        | 95,375       |
| 13    | 45      | 1,875        | 94,625       | 40      | 1,250        | 95,000       |
| 14    | 0       | 0,625        | 99,000       | 1       | 4,375        | 99,875       |
| 15    | nd      | 1,250        | 2,000        | nd      | 1,875        | 5,250        |
| 16    | 312     | 3,750        | 14,625       | 194     | 3,125        | 41,125       |
| 17    | 28      | 0,625        | 79,750       | 21      | 2,500        | 95,250       |
| 18    | 92      | 1,250        | 89,375       | 83      | 4,375        | 90,125       |
| 19    | nd      | 0,625        | 11,250       | nd      | 4,375        | 21,500       |
| 20    | 86      | 0,000        | 32,000       | 84      | 3,750        | 54,500       |
| 21    | 504     | 0,625        | 41,125       | 249     | 4,375        | 49,875       |
| MEDIA | 101,688 | 0,859        | 71,078       | 48,750  | 3,320        | 78,727       |

Tabla 3: Tabla de resultados de DPCA

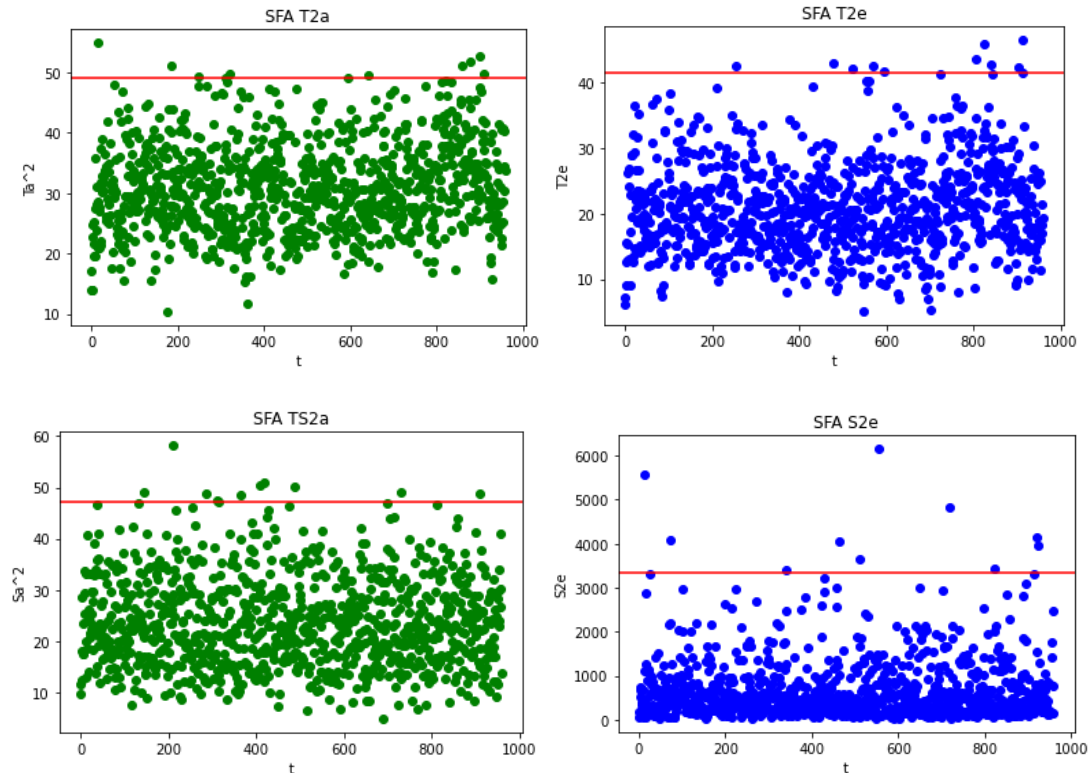


Figura 14.- Estadística  $T^2$  y su umbral en condiciones normales de operación

#### 4.3.1.- Detección de fallos usando SFA

Al igual que los modelos anteriores lo primero que hacemos es cargar y normalizar los datos, después aplicamos la reducción de dimensiones cogiendo los 31 primeros valores de la matriz, tendrán el sufijo a y el resto el sufijo e. Después calculamos los estadísticos de todos los elementos.

En las figuras siguientes, 15 y 16, se muestran los resultados de aplicar el método SFA a los fallos 7 y 12 respectivamente. Se muestran las 4 estadísticas junto a sus umbrales para cada fallo. Se ve como con los estadísticos  $T^2_a$  y  $T^2_e$  ambos fallos se detectan perfectamente, ya que los estadísticos superan su umbral, nada más detectarse el fallo, en cambio los estadísticos  $S^2_a$  y  $S^2_e$  no son capaces de detectar el fallo, en ninguno de los dos casos (fallo 7, figura 15 y fallo 12, figura 16).

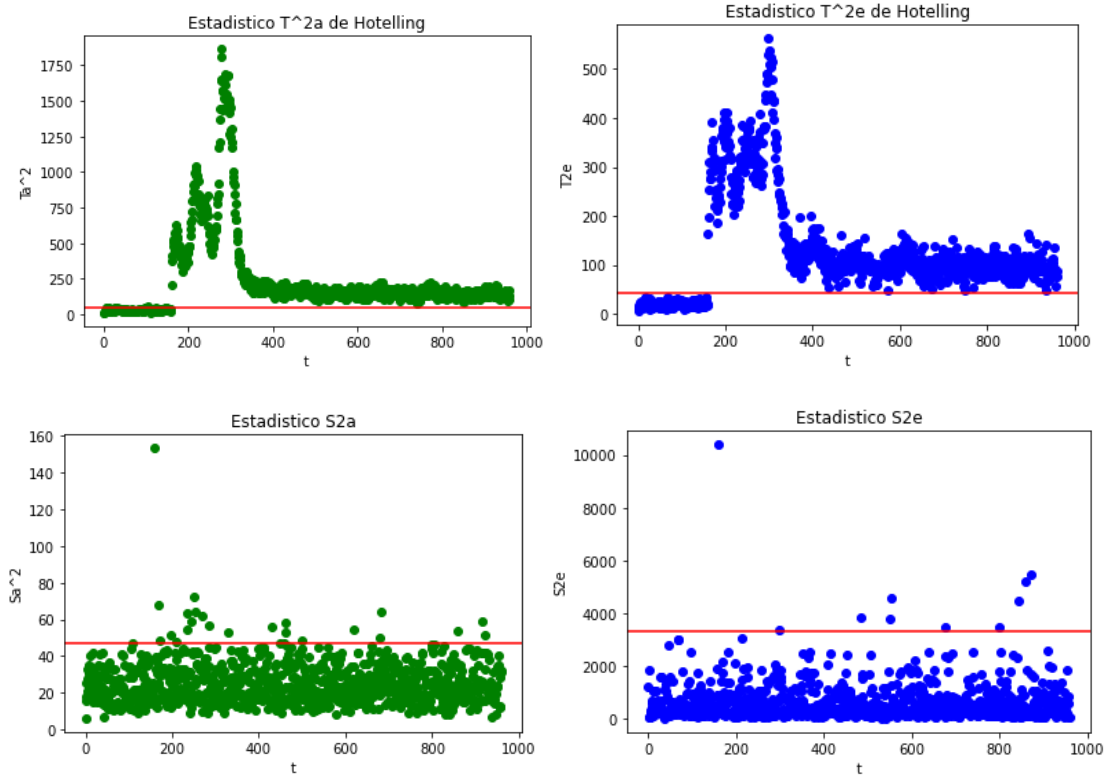


Figura 15: Detección de fallo usando SFA para el fallo 7

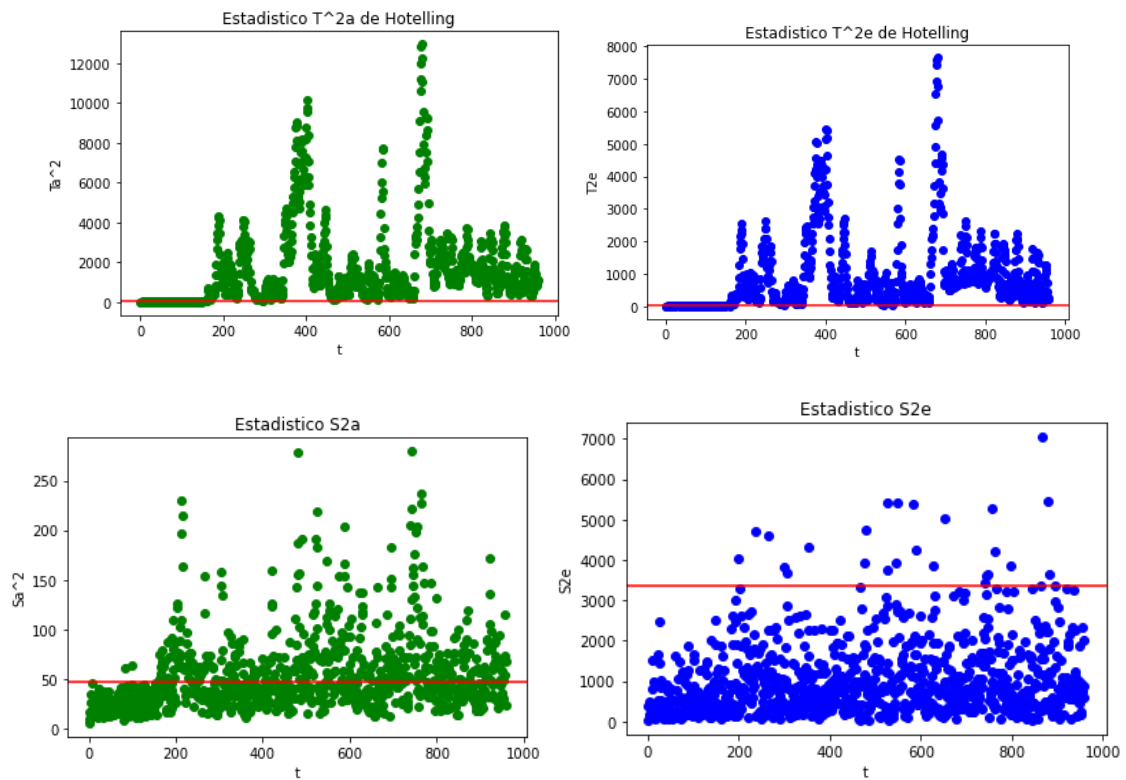


Figura 16: Detección de fallo usando SFA para el fallo 12



De la misma forma que antes, en la Figura 17 se muestra el valor de las 52 variables cuando SFA detecta el fallo, para indicarnos cual es la variable responsable del fallo, que en este caso el resultado es diferente de PCA. Para el fallo 7, la variable con valor más grande es la variable 15, el nivel del stripper, para el fallo 12, SFA también nos dice que la responsable del fallo es la variable 11, que pertenece a la temperatura del separador.

Los resultados totales se muestran en las Tablas 4 y 5 (Tabla 4 con  $T^2_e$  y  $T^2_a$ , y la Tabla 5 con las estadísticas  $S^2$ ), las cuales muestran que los resultados obtenidos usando el estadístico  $T^2$  son bastante mejores que los modelos anteriores (PCA y DPCA), los datos que llevan el subíndice a son mejores a la hora de detectar alarmas, llegan a detectare en media un 93% y el estadístico  $T^2_e$  un 84%, aunque por el contrario el número de falsas alarmas también aumenta. El estadístico  $S^2$  tiene unos resultados bastante malos, el único caso en el que es útil es si se busca reducir al mínimo el número de falsas alarmas, pero si no detecta alarmas no sirve de mucho, al igual que ocurría en el estadístico  $T^2$ , el número de alarmas detectadas se reduce en los últimos valores, representados por la e.

| SFA   | T2a     |              |              | T2e     |              |              |
|-------|---------|--------------|--------------|---------|--------------|--------------|
|       | T fallo | falsa alarma | alarma detec | T fallo | falsa alarma | alarma detec |
| 1     | 2       | 5,625        | 99,875       | 4       | 0,000        | 99,500       |
| 2     | 12      | 4,375        | 98,625       | 14      | 0,625        | 98,250       |
| 3     | nd      | 11,875       | 10,125       | nd      | 0,000        | 1,625        |
| 4     | 0       | 6,875        | 100,000      | 38      | 1,250        | 67,250       |
| 5     | 0       | 6,875        | 100,000      | 0       | 1,250        | 100,000      |
| 6     | 0       | 1,250        | 100,000      | 0       | 0,000        | 100,000      |
| 7     | 0       | 2,500        | 100,000      | 0       | 0,000        | 100,000      |
| 8     | 16      | 3,750        | 98,125       | 21      | 0,625        | 97,375       |
| 9     | nd      | 10,000       | 5,750        | nd      | 1,250        | 1,500        |
| 10    | 22      | 2,500        | 87,875       | 24      | 0,625        | 80,750       |
| 11    | 6       | 0,625        | 80,500       | 24      | 2,500        | 54,500       |
| 12    | 0       | 3,125        | 100,000      | 2       | 0,625        | 99,625       |
| 13    | 35      | 1,250        | 95,875       | 42      | 0,625        | 94,750       |
| 14    | 0       | 3,750        | 100,000      | 1       | 0,000        | 99,875       |
| 15    | nd      | 5,625        | 13,000       | nd      | 1,250        | 2,250        |
| 16    | 6       | 5,000        | 93,125       | 10      | 2,500        | 74,250       |
| 17    | 21      | 4,375        | 96,625       | 23      | 0,000        | 89,000       |
| 18    | 78      | 3,750        | 91,125       | 83      | 0,000        | 90,000       |
| 19    | 9       | 2,500        | 91,750       | 112     | 0,000        | 58,250       |
| 20    | 64      | 2,500        | 90,125       | 69      | 0,000        | 88,125       |
| 21    | 262     | 14,375       | 61,125       | 565     | 1,250        | 33,500       |
| Media | 25,381  | 4,881        | 81,601       | 26,286  | 0,685        | 72,875       |

Tabla 4.- Resultados de SFA con las estadísticas  $T^2_a$  y  $T^2_e$



|         | S2a          |              |         | S2e          |              |
|---------|--------------|--------------|---------|--------------|--------------|
| T fallo | falsa alarma | alarma detec | T fallo | falsa alarma | alarma detec |
| nd      | 2,500        | 1,375        | nd      | 0,000        | 0,750        |
| nd      | 3,125        | 1,875        | nd      | 0,625        | 0,625        |
| nd      | 1,875        | 3,375        | nd      | 0,000        | 0,375        |
| nd      | 0,625        | 2,375        | nd      | 1,250        | 0,875        |
| nd      | 0,625        | 3,125        | nd      | 1,250        | 0,750        |
| nd      | 2,500        | 21,375       | nd      | 0,000        | 1,250        |
| nd      | 0,000        | 2,875        | nd      | 0,000        | 1,250        |
| nd      | 3,750        | 6,875        | nd      | 0,625        | 1,375        |
| nd      | 1,875        | 3,250        | nd      | 1,250        | 1,125        |
| nd      | 1,875        | 4,500        | nd      | 0,625        | 1,375        |
| nd      | 0,625        | 9,375        | nd      | 2,500        | 1,750        |
| 42      | 1,250        | 53,000       | nd      | 0,625        | 3,375        |
| nd      | 1,250        | 10,000       | nd      | 0,625        | 2,000        |
| nd      | 3,125        | 100,000      | 32      | 0,000        | 88,375       |
| nd      | 4,375        | 2,625        | nd      | 1,250        | 0,750        |
| nd      | 1,250        | 9,250        | nd      | 2,500        | 0,750        |
| 46      | 3,125        | 29,625       | nd      | 0,000        | 5,375        |
| 218     | 3,125        | 24,125       | nd      | 0,000        | 2,000        |
| 36      | 0,000        | 70,250       | nd      | 0,000        | 11,625       |
| nd      | 0,625        | 11,750       | nd      | 0,000        | 4,500        |
| nd      | 2,500        | 0,000        | nd      | 0,625        | 70,38        |
| 68,400  | 2,125        | 55,400       | 32      | 0,000        | 88,375       |

Tabla 5.- Resultados de SFA con las estadísticas  $S_a^2$  y  $S_e^2$

| SFA      |          |  |          |          |
|----------|----------|--|----------|----------|
| Nº fallo | Variable |  | Nº fallo | Variable |
| 1        | 23       |  | 12       | 22, 44   |
| 2        | 3        |  | 13       | 51       |
| 3        | 52       |  | 14       | 9        |
| 4        | 51       |  | 15       | 8        |
| 5        | 22       |  | 16       | 29       |
| 6        | 2        |  | 17       | 21       |
| 7        | 4, 9     |  | 18       | 40       |
| 8        | 39       |  | 19       | 9        |
| 9        | 43       |  | 20       | 40       |
| 10       | 39       |  | 21       | 2        |
| 11       | 9        |  |          |          |

Tabla 6: Variables que fallan en SFA

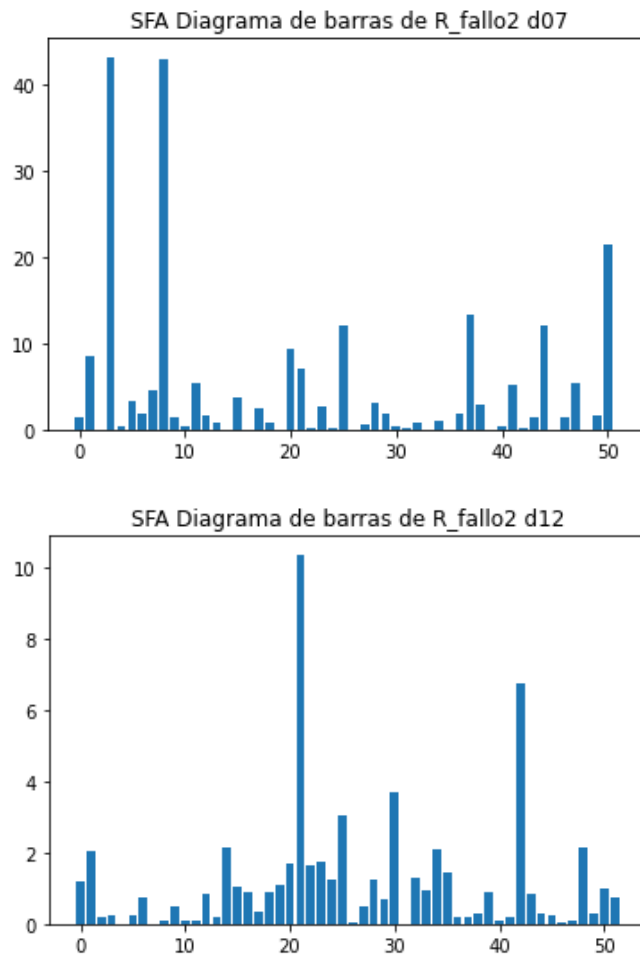


Figura 17.- Variables que fallan SFA

De la misma forma, en la tabla 6 se muestran las variables que provocan cada fallo, calculadas con la ec. 26 del capítulo 2, el espacio vacío de la tabla se debe a que el error no se llega a detectar, por tanto, la ec. 26 no se puede calcular y eso produce que el valor del fallo obtenido no sea relevante.

#### 4.3.2.- Detección de fallos usando DSFA

Al igual que en PCA en SFA se puede hacer un modelo dinámico que está basado en este mismo, la diferencia es que cambia la forma de la matriz de datos, esta pasará a tener 960x208, ahora tendrá 208 variables, lo que permite tener en cuenta instantes de tiempo anteriores, es decir ser dinámica.

Para DSFA los resultados en los estadísticos (Figuras 18 y 19), son muy similares a los de SFA, también se observa que los estadísticos  $T^2$  detectan los fallos perfectamente, mientras que los estadísticos  $S^2$ , apenas superan el umbral, por lo

que su comportamiento no es bueno, tanto para el fallo 7 como para el fallo 12, y como veremos más adelante para todos los demás fallos.

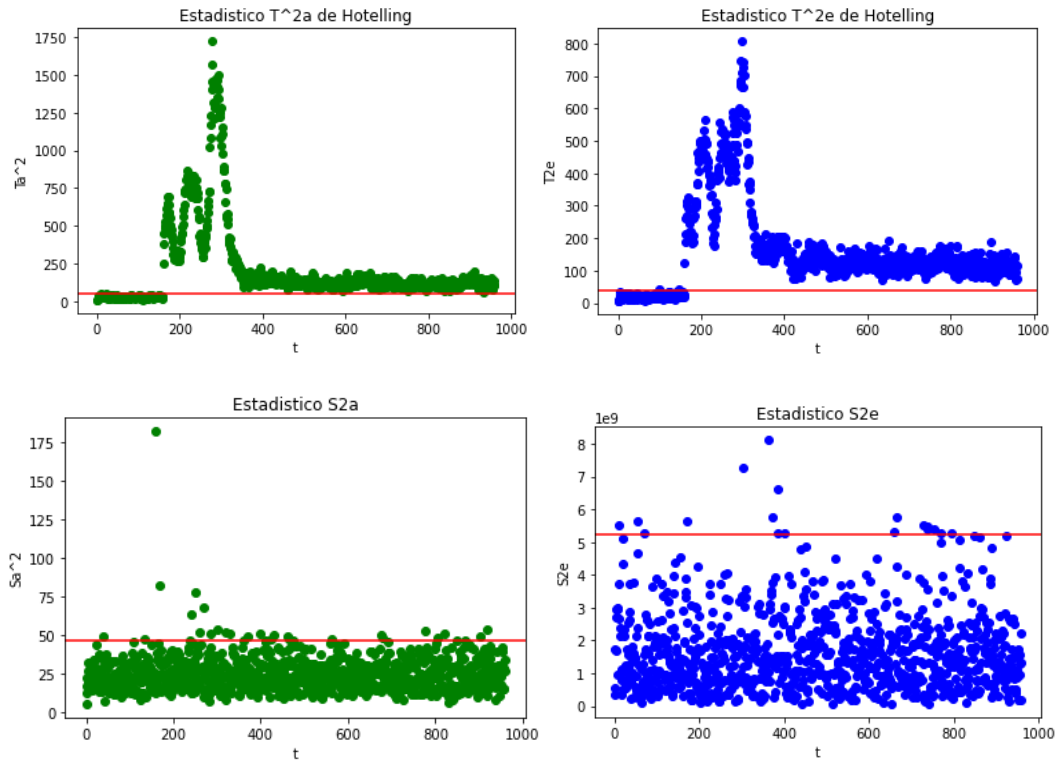


Figura 18: Detección de fallo usando DSFA para el fallo 7

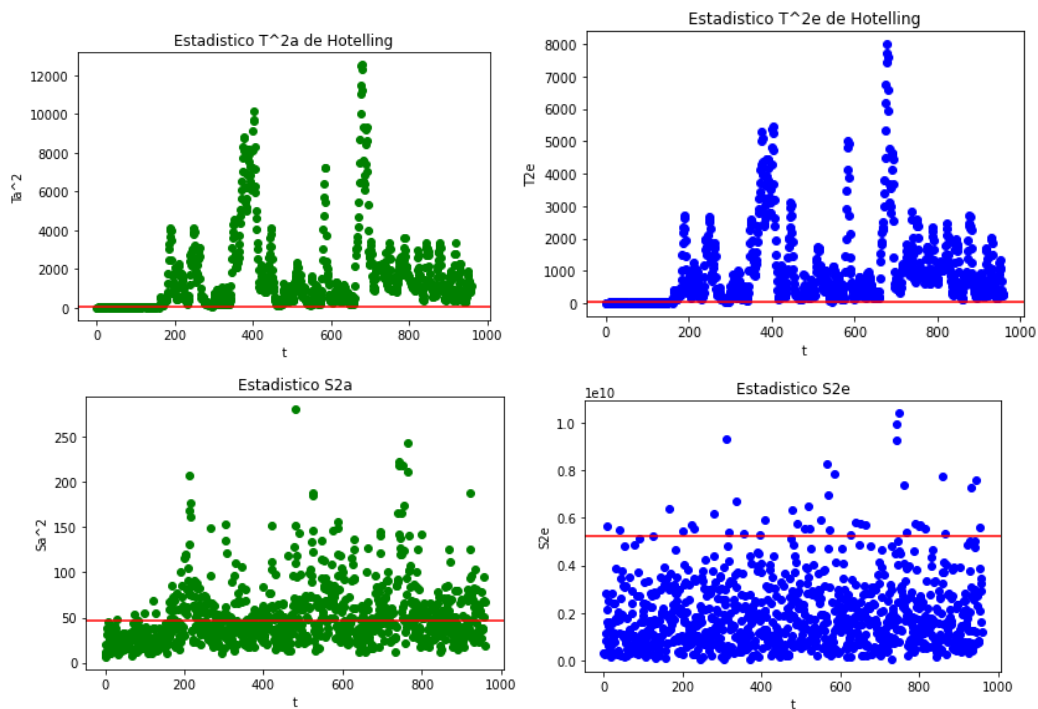


Figura 19: Detección de fallo usando DSFA para el fallo 12

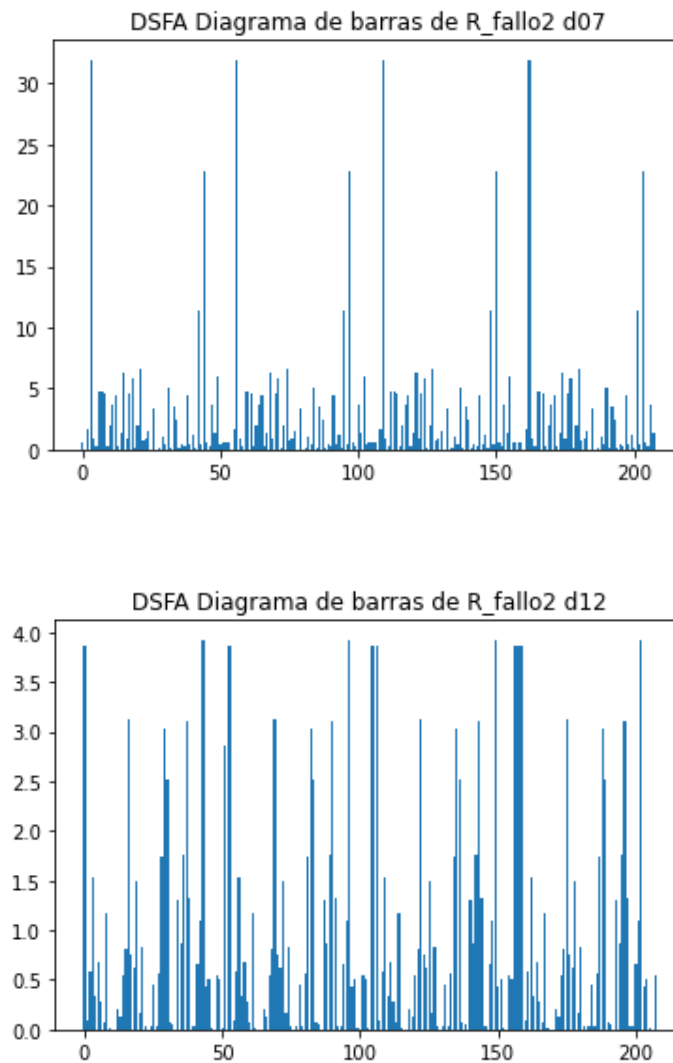


Figura 20: Variables que fallan con el método DSFA

En la Figura 20 se muestra el valor de todas las variables cuando DSFA detecta el fallo, para indicarnos cual es la variable responsable de este, que como en el caso de DPCA se muestran unos pocos repetidos que son las variables responsables, pero con sus respectivos retardos. El resultado es similar a SFA, el fallo 7 las responsables son las variables 4 y 9, para el fallo 12, la responsable es la variable 11 con sus respectivos retardos.

Los resultados totales se muestran en las Tablas 7 y 8, donde se puede ver que el estadístico  $T_a^2$  es el que mejor detecta las alarmas seguido de  $T_e^2$ ,  $T^2$  por tanto detecta muchos más fallos que  $S^2$  pero también genera más falsas alarmas, los estadísticos  $S_a^2$  y  $S_e^2$  no son correctos a la hora de eliminar las falsas alarmas ni las alarmas detectadas, solo son superiores en la detección de falsas alarmas, ya que logran un porcentaje menor.



| DSFA  | T <sub>2a</sub> |              |              | T <sub>2e</sub> |              |              |
|-------|-----------------|--------------|--------------|-----------------|--------------|--------------|
|       | T fallo         | falsa alarma | alarma detec | T fallo         | falsa alarma | alarma detec |
| 1     | 0               | 3,750        | 100          | 4               | 1,875        | 99,5         |
| 2     | 12              | 3,125        | 98,875       | 11              | 1,250        | 98,625       |
| 3     | nd              | 6,875        | 7,375        | nd              | 2,500        | 2,875        |
| 4     | 0               | 2,500        | 100,000      | 309             | 1,875        | 51,625       |
| 5     | 0               | 2,500        | 100,000      | 0               | 1,875        | 100,000      |
| 6     | 0               | 2,500        | 100,000      | 0               | 0,000        | 100,000      |
| 7     | 0               | 0,625        | 100,000      | 0               | 1,250        | 100,000      |
| 8     | 16              | 4,375        | 98,125       | 21              | 0,000        | 97,750       |
| 9     | nd              | 8,125        | 6,750        | nd              | 4,375        | 3,375        |
| 10    | 21              | 1,875        | 90,250       | 21              | 0,000        | 82,375       |
| 11    | 6               | 5,000        | 80,375       | 24              | 0,625        | 51,250       |
| 12    | 1               | 4,375        | 99,875       | 1               | 3,125        | 99,750       |
| 13    | 36              | 0,000        | 95,625       | 46              | 0,000        | 94,750       |
| 14    | 0               | 3,125        | 100,000      | 1               | 1,875        | 99,875       |
| 15    | nd              | 5,000        | 13,125       | nd              | 3,125        | 4,250        |
| 16    | 6               | 11,250       | 92,375       | 10              | 5,625        | 79,625       |
| 17    | 21              | 4,375        | 96,000       | 21              | 1,250        | 95,000       |
| 18    | 78              | 1,250        | 91,000       | 84              | 1,875        | 89,750       |
| 19    | 109             | 3,125        | 91,250       | 168             | 2,500        | 58,750       |
| 20    | 69              | 1,875        | 90,500       | 63              | 1,875        | 89,500       |
| 21    | 471             | 6,875        | 56,000       | 462             | 3,125        | 46,250       |
| MEDIA | 47              | 3,472        | 93,347       | 69,222          | 1,667        | 85,243       |

Tabla 7.- Resultados de DSFA con las estadísticas T<sub>a</sub><sup>2</sup> y T<sub>e</sub><sup>2</sup>.



| DSFA  | S2a     |              |              | S2e     |              |              |
|-------|---------|--------------|--------------|---------|--------------|--------------|
|       | T fallo | falsa alarma | alarma detec | T fallo | falsa alarma | alarma detec |
| 1     | nd      | 3,75         | 1,75         | nd      | 1,875        | 1,25         |
| 2     | nd      | 3,750        | 1,750        | nd      | 1,250        | 1,000        |
| 3     | nd      | 1,875        | 2,500        | nd      | 2,500        | 0,875        |
| 4     | nd      | 3,125        | 2,875        | nd      | 1,875        | 1,125        |
| 5     | nd      | 3,125        | 2,875        | nd      | 1,875        | 1,125        |
| 6     | nd      | 0,000        | 20,750       | nd      | 0,000        | 1,375        |
| 7     | nd      | 1,250        | 3,625        | nd      | 1,250        | 1,875        |
| 8     | nd      | 3,750        | 6,875        | nd      | 0,000        | 2,125        |
| 9     | nd      | 1,875        | 3,625        | nd      | 4,375        | 0,750        |
| 10    | nd      | 1,875        | 5,875        | nd      | 0,000        | 1,250        |
| 11    | nd      | 1,250        | 12,750       | nd      | 0,625        | 1,500        |
| 12    | 42      | 3,125        | 49,875       | nd      | 3,125        | 4,875        |
| 13    | nd      | 1,875        | 10,875       | nd      | 0,000        | 1,125        |
| 14    | 0       | 1,250        | 100,000      | 3       | 1,875        | 93,875       |
| 15    | nd      | 1,250        | 4,125        | nd      | 3,125        | 1,250        |
| 16    | nd      | 3,125        | 9,750        | nd      | 5,625        | 1,250        |
| 17    | 47      | 1,875        | 25,875       | nd      | 1,250        | 5,750        |
| 18    | 230     | 2,500        | 25,250       | nd      | 1,875        | 0,625        |
| 19    | 37      | 1,250        | 70,375       | nd      | 2,500        | 2,375        |
| 20    | nd      | 2,500        | 11,875       | nd      | 1,875        | 3,000        |
| 21    | nd      | 2.5          | 2,000        | nd      | 3,125        | 1,000        |
| MEDIA | 71,2    | 2            | 54,275       | 3       | 2,125        | 21,500       |

Tabla 8.- Resultados de DSFA con las estadísticas  $S_a^2$  y  $S_e^2$

#### 4.4 SPA

Ahora se va a aplicar el método SPA a los datos de la planta TE, recordar que este método no trabaja con las variables directamente, sino que primero se calcula en una ventana de datos deslizante (en nuestro caso de 50 muestras), diferentes medidas estadísticas de las variables, como la media, la varianza, la skewness y la kurtosis, y con estos valores se construye el matiz inicial,  $X$ , de dimensión (960, 208) sobre la que luego se usa PCA.

$$X = [\mu_1, \mu_2, \dots, \mu_n, \sigma_1^2, \sigma_2^2, \dots, \sigma_n^2, skew_1, skew_n, kurt_1, kurt_n] \quad (36)$$

A continuación, en la Figura 21, se muestra el funcionamiento sin fallos usando el modelo estadístico SPA, y calculando los estadísticos  $T^2$  y  $Q$  y sus umbrales con el percentil 99, de sus valores en condiciones normales de operación, y se puede ver prácticamente ningún punto supera el umbral que indicaría un fallo.

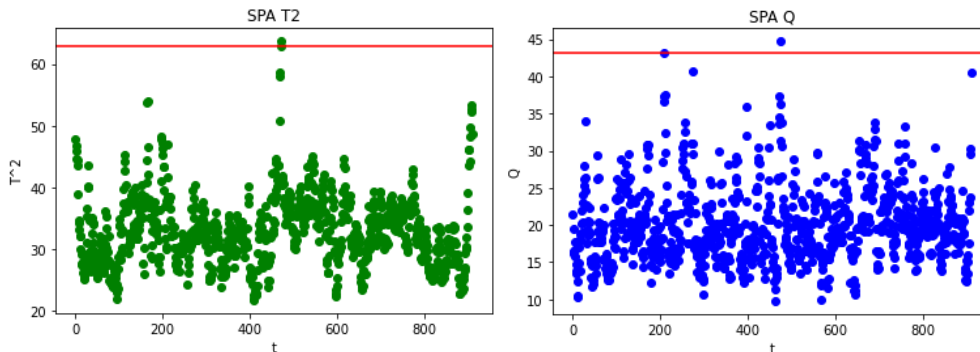


Figura 21: Estadísticos  $T^2$  y  $Q$  para el método SPA con datos normales de funcionamiento

#### 4.4.1.- Detección de fallos usando SPA

Ahora vamos a aplicar SPA a dos archivos de datos diferentes que contienen fallos. Se va a realizar el mismo procedimiento usado anteriormente, en el método anterior, cargamos y normalizamos los datos y cogemos los primeros 31 valores. Para finalizar se calculan los estadísticos.

En las figuras 22 y 23 se observa como ahora si hay valores que sobrepasan el umbral, como se ha mencionado anteriormente, esto se debe a que estos datos se obtuvieron cuando se producía un fallo en la planta.

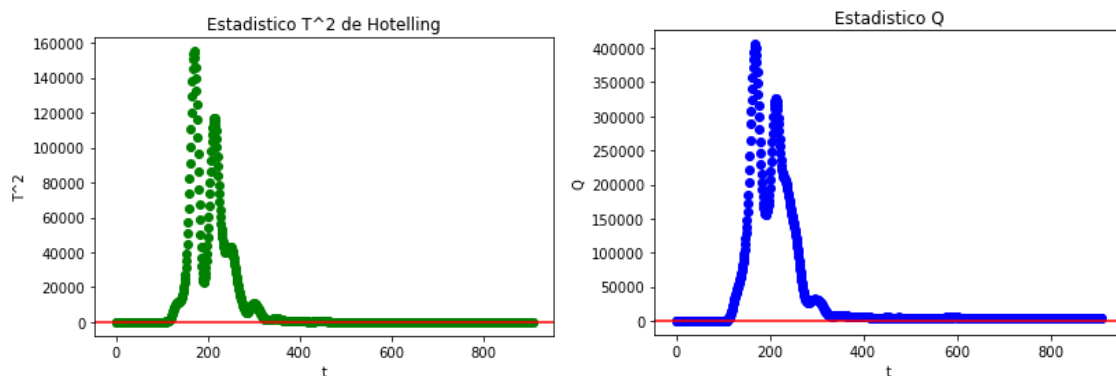


Figura 22: Detección de fallo usando SPA para el fallo 7

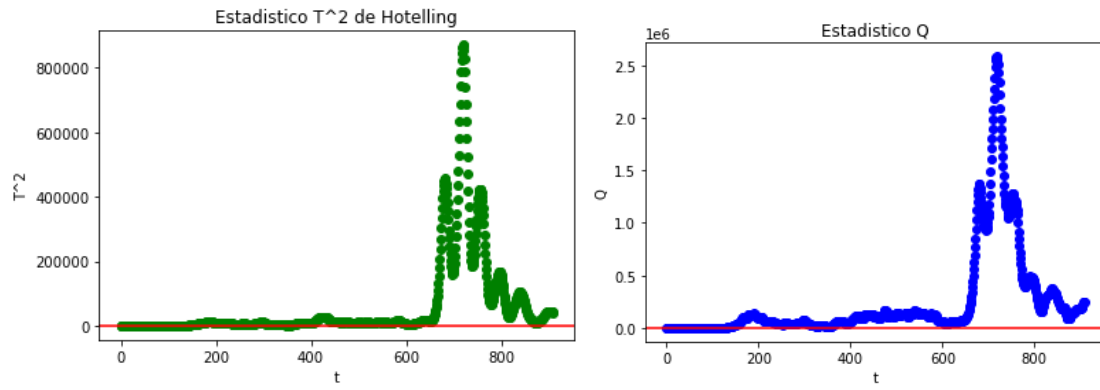


Figura 23: Detección de fallo usando SPA datos de fallo 12

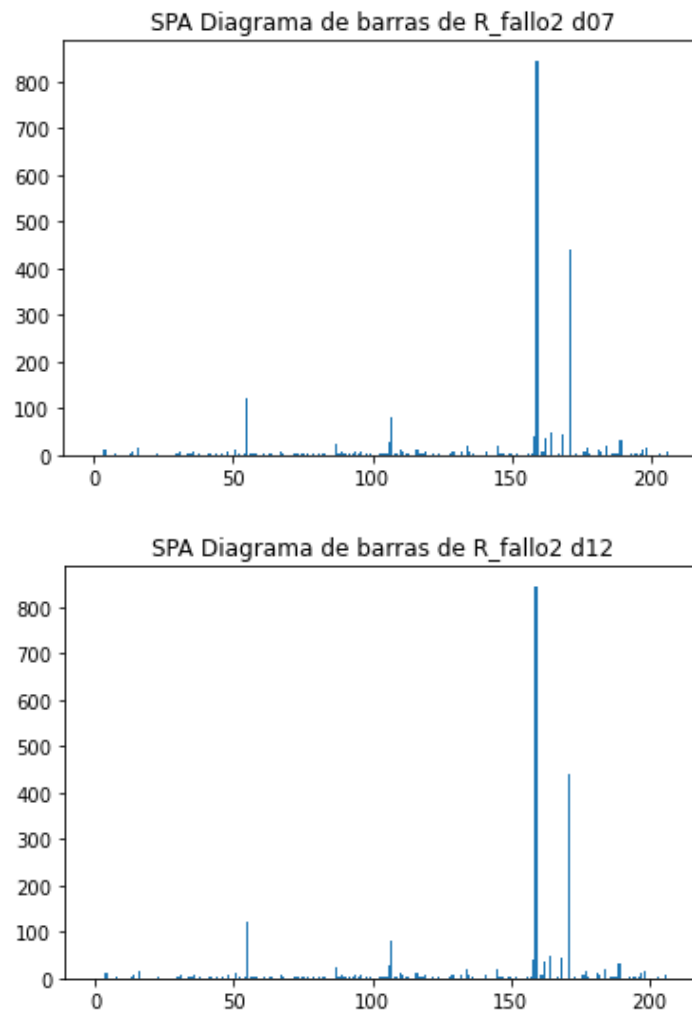


Figura 24: Variables que fallan SPA

En la figura 24 se observa como las variables que fallan lo hacen de una forma muy clara, hay muy poco ruido, esto se ve reflejado en la tabla 9, donde el número de falsas alarmas es casi inexistente, para  $T^2$  es 0 y para Q un 1,11%. Como contraparte tenemos el número de alarmas detectadas que no entra dentro de unos valores buenos, además de que el tiempo de fallo tiempo destaca en sus valores.

Para SPA el tiempo de fallo es de 110, esto se debe a como están representados los datos en el programa, lo que pasa es que los 5 estadísticos modifican la ventana de datos. Como tenemos una ventana de datos de 50, el fallo se detecta en el 110 no en el 160 como en ejemplos anteriores, esto también influye a la hora de detectar fallos, para saber a qué variable pertenecen habría que dividir entre 5.

| SPA   | T2      |              |              | Q       |              |              |
|-------|---------|--------------|--------------|---------|--------------|--------------|
|       | T fallo | falsa alarma | alarma detec | T fallo | falsa alarma | alarma detec |
| 1     | 6       | 0            | 99,25        | 0       | 0            | 99,75        |
| 2     | 0       | 18,330       | 97,750       | 0       | 45,000       | 98,375       |
| 3     | 399     | 0,000        | 1,750        | 0       | 10,900       | 3,500        |
| 4     | 0       | 0,000        | 91,500       | 0       | 0,900        | 100,000      |
| 5     | 1       | 0,000        | 91,500       | 0       | 1,660        | 100,000      |
| 6     | 0       | 0,000        | 20,875       | 0       | 0,000        | 20,875       |
| 7     | 0       | 0,000        | 100,000      | 0       | 0,000        | 100,000      |
| 8     | 21      | 0,000        | 97,375       | 22      | 0,000        | 97,250       |
| 9     | 693     | 0,000        | 2,625        | nd      | 16,360       | 8,125        |
| 10    | 54      | 0,000        | 48,500       | 57      | 0,000        | 64,625       |
| 11    | 10      | 0,000        | 98,875       | 6       | 0,000        | 99,250       |
| 12    | 2       | 0,000        | 99,750       | 2       | 0,000        | 99,750       |
| 13    | 51      | 0,000        | 93,625       | 0       | 5,000        | 94,000       |
| 14    | 1       | 0,000        | 99,875       | 1       | 0,000        | 99,875       |
| 15    | 622     | 0,000        | 6,875        | 281     | 0,000        | 19,375       |
| 16    | 0       | 6,660        | 28,625       | 48      | 0,000        | 54,870       |
| 17    | 27      | 0,000        | 97,125       | 14      | 0,000        | 98,250       |
| 18    | 88      | 0,000        | 23,750       | 86      | 0,000        | 24,000       |
| 19    | 109     | 0,000        | 67,000       | 22      | 0,000        | 97,500       |
| 20    | 91      | 0,000        | 85,370       | 77      | 0,000        | 90,375       |
| 21    | nd      | 0,000        | 0,500        | 35      | 0,000        | 1,625        |
| MEDIA | 120,83  | 0,000        | 68,090       | 38,294  | 1,113        | 69,161       |

Tabla 9: Tabla errores SPA

| SPA      |          |  |          |          |
|----------|----------|--|----------|----------|
| Nº fallo | Variable |  | Nº fallo | Variable |
| 1        | 199      |  | 12       | 178      |
| 2        | 159      |  | 13       | 190      |
| 3        | 195      |  | 14       | 207      |
| 4        | 207      |  | 15       | 158      |
| 5        | 178      |  | 16       | 71       |
| 6        | 157      |  | 17       | 179      |
| 7        | 160      |  | 18       | 74       |
| 8        | 187      |  | 19       | 57       |
| 9        | 71       |  | 20       | 56       |
| 10       | 56       |  | 21       |          |
| 11       | 207      |  |          |          |

Tabla 10: Variables que fallan en SPA

## 4.5 SPCA

En la figura 25 se muestran las gráficas en las que el funcionamiento de la planta es normal aplicando SPCA, hay momentos en los que se supera el umbral, esta situación no es para nada deseable pero la mayor parte de los puntos se encuentran por debajo del mismo lo que nos indica que no se detectan fallos.

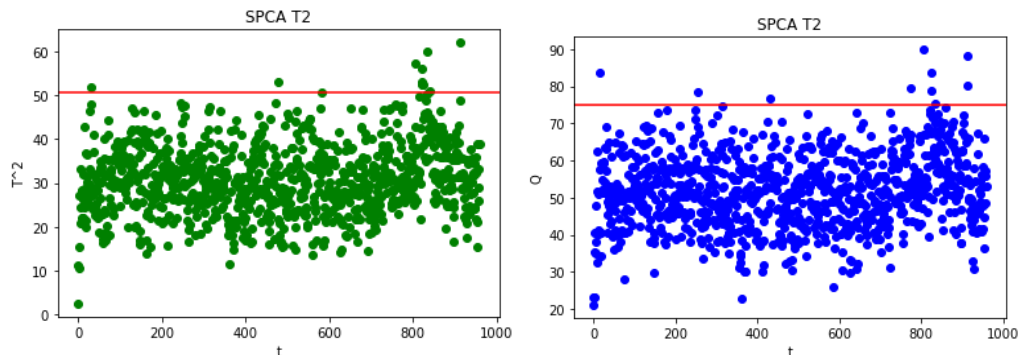


Figura 25: Estadísticos  $T^2$  y  $Q$  para el método SPCA con datos normales de funcionamiento

### 4.5.1.- Detección de fallos usando SPCA

El funcionamiento es muy similar al tradicional de PCA, solo que a mayores se aplicara una librería que realiza la reducción a 0 de los valores menos importantes. Es importante destacar que este modelo es uno de los que son más complejos, lo que lleva a que sea uno de los que tiene un tiempo de computación más elevado.

Las figuras 26 y 27 ya contienen fallos, en ellas se ve como antes del instante 160 no salta ningún error, pero a partir de este ya empiezan a ocurrir (se sobrepasa el umbral). En las figuras se ve un funcionamiento prácticamente ideal, en el que una vez que salta el fallo después del instante 160 lo detecta de una forma muy clara.

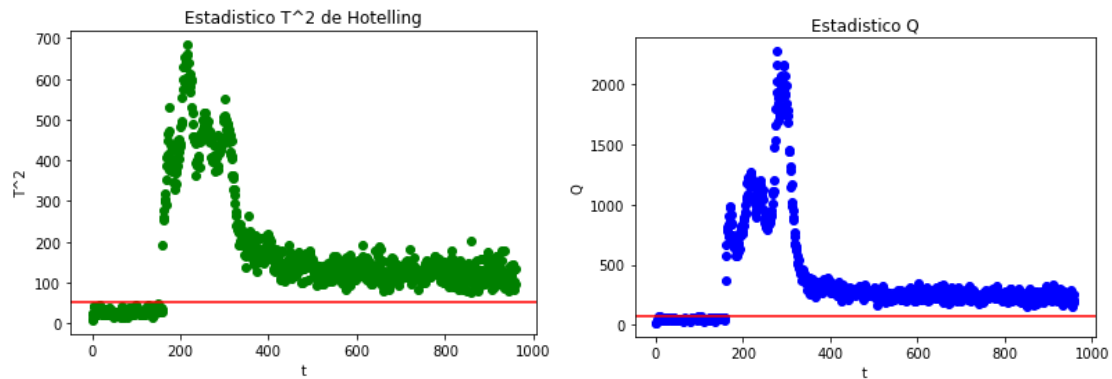


Figura 26: Detección de fallo usando SPCA datos de fallo 7

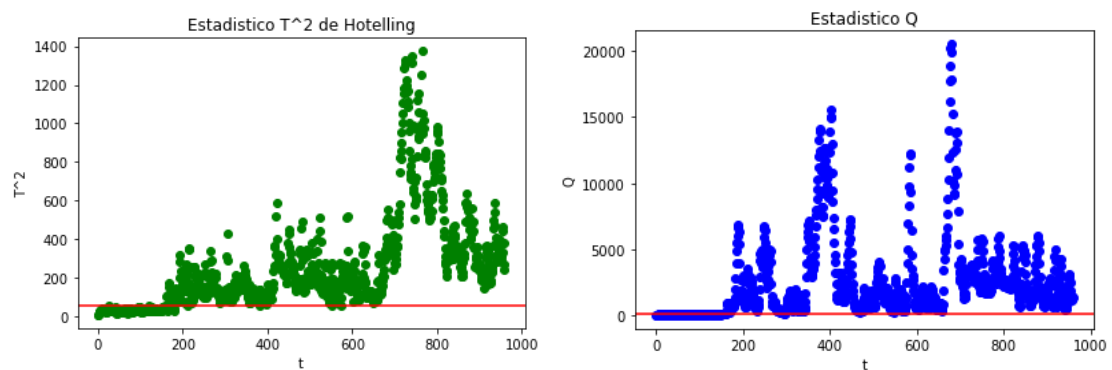


Figura 27: Detección de fallo usando SPCA datos del fallo 12

En la Figura 28 se muestra el valor de las 52 variables cuando SPCA detecta el fallo, para indicarnos cual es la variable responsable del fallo, que en este caso el resultado es diferente de PCA. Para el fallo 7, la variable con valor más grande es la variable 4 como siempre, el flujo de alimentación de A y C, para el fallo 12, SPCA nos dice que la variable responsable del fallo es la variable 33 perteneciente a la purga del gas, en concreto a la concentración de E en dicha purga.

Los resultados arrojados en la tabla 11 por este método son bastante buenos, mejora a los obtenidos usando el método convencional de PCA, cabe resaltar el buen tiempo de detección de fallo y las alarmas detectadas utilizando el estadístico Q. En la tabla 12 se ven las variables que fallan en cada archivo de datos.

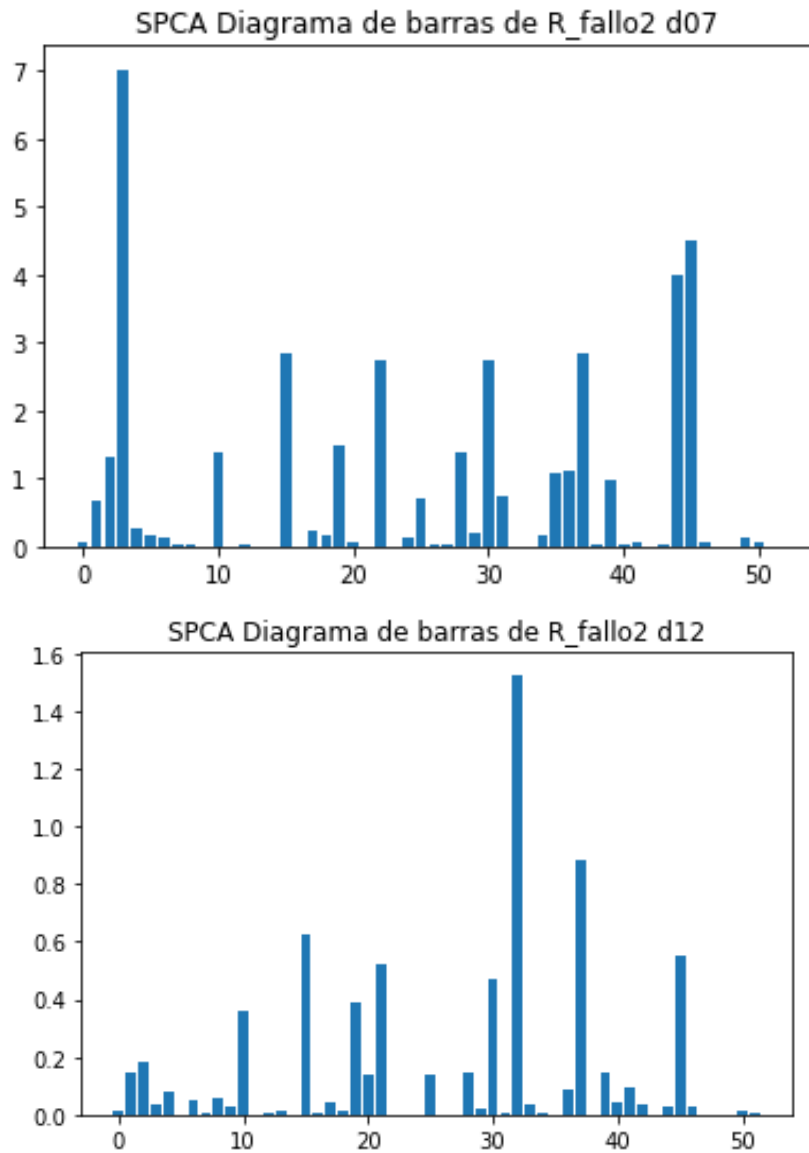


Figura 28: Variables que fallan SPCA



| SPCA  | T2      |              |              | Q       |              |              |
|-------|---------|--------------|--------------|---------|--------------|--------------|
|       | T fallo | falsa alarma | alarma detec | T fallo | falsa alarma | alarma detec |
| 1     | 6       | 1,875        | 99,25        | 0       | 3,125        | 100          |
| 2     | 11      | 1,250        | 98,625       | 10      | 3,125        | 98,750       |
| 3     | nd      | 0,625        | 2,000        | nd      | 10,625       | 11,250       |
| 4     | 0       | 0,625        | 43,250       | 0       | 5,000        | 100,000      |
| 5     | 10      | 0,625        | 26,000       | 0       | 5,000        | 100,000      |
| 6     | 8       | 0,625        | 99,000       | 0       | 1,250        | 100,000      |
| 7     | 0       | 0,000        | 100,000      | 0       | 0,625        | 100,000      |
| 8     | 22      | 0,000        | 97,250       | 15      | 2,500        | 98,125       |
| 9     | nd      | 3,750        | 3,620        | nd      | 11,875       | 8,250        |
| 10    | 103     | 1,250        | 30,875       | 21      | 0,000        | 92,250       |
| 11    | 50      | 0,000        | 51,500       | 5       | 3,750        | 82,250       |
| 12    | 21      | 0,625        | 98,500       | 1       | 6,875        | 99,875       |
| 13    | 46      | 1,875        | 94,500       | 36      | 1,250        | 95,500       |
| 14    | 0       | 0,625        | 99,500       | 0       | 3,750        | 100,000      |
| 15    | nd      | 1,250        | 2,750        | nd      | 5,000        | 15,250       |
| 16    | 310     | 5,000        | 15,625       | 6       | 8,125        | 94,000       |
| 17    | 28      | 1,250        | 79,125       | 21      | 3,750        | 97,000       |
| 18    | 92      | 1,875        | 89,375       | 77      | 3,750        | 91,125       |
| 19    | nd      | 1,250        | 12,000       | 9       | 4,375        | 94,250       |
| 20    | 86      | 2,500        | 31,000       | 64      | 2,500        | 91,125       |
| 21    | 504     | 1,250        | 41,875       | 394     | 10,625       | 59,875       |
| MEDIA | 81,063  | 1,289        | 72,000       | 36,611  | 3,854        | 94,118       |

Tabla 11: Tabla errores SPCA

| SPCA     |          |          |          |
|----------|----------|----------|----------|
| Nº fallo | Variable | Nº fallo | Variable |
| 1        | 21       | 12       | 33       |
| 2        | 38       | 13       | 31       |
| 3        | 42       | 14       | 31       |
| 4        | 35       | 15       | 42       |
| 5        | 22       | 16       | 23       |
| 6        | 44       | 17       | 21       |
| 7        | 4        | 18       | 22       |
| 8        | 31       | 19       | 38       |
| 9        | 22       | 20       | 25       |
| 10       | 29       | 21       | 16       |
| 11       | 51       |          |          |

Tabla 12: Variables que fallan es SPCA

### 4.5.2.- Detección de fallos usando DSPCA

A continuación, se va a estudiar el modelo dinámico de SPCA, sus siglas siguen la dinámica de los modelos anteriores, estas son DSPCA.

Las figuras 29 y 30 nos indican los fallos en los archivos 7 y 12 aplicando DSPCA, se puede observar cómo se sobrepasa el umbral indicando que se ha producido un fallo, en este caso en la figura 30 el error no está tan marcado como en la figura 29, pero aun así se detecta de una forma adecuada.

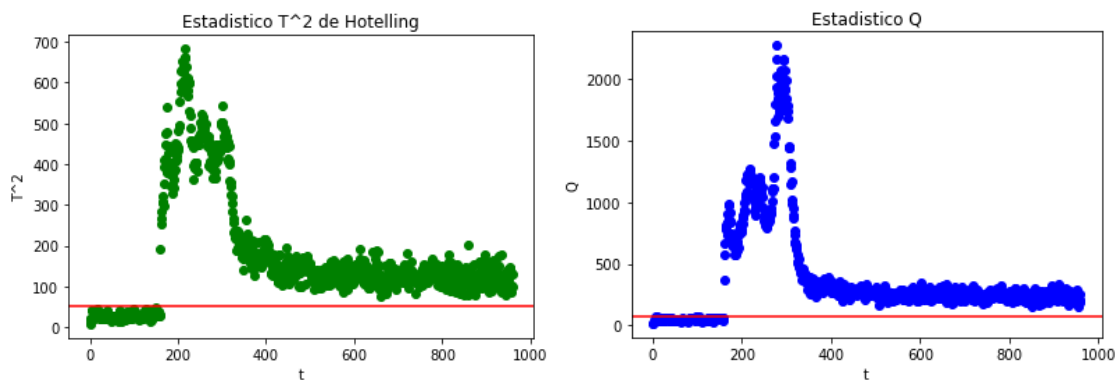


Figura 29: Detección de fallo usando DSPCA datos de fallo 7

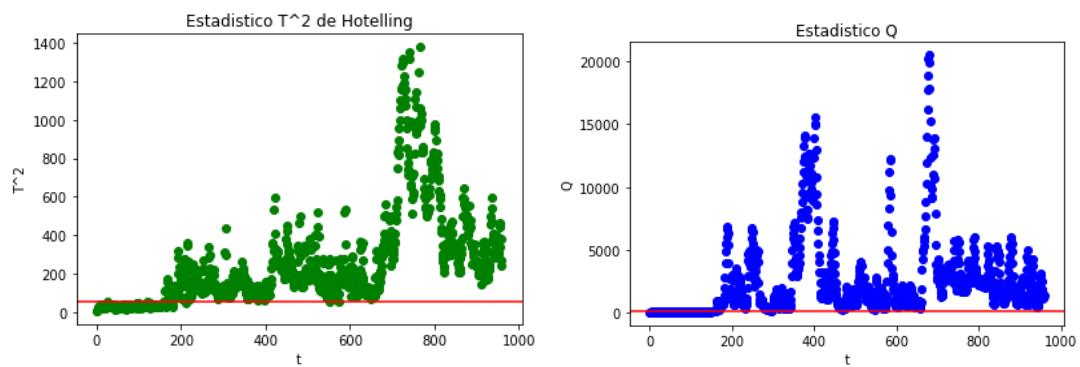


Figura 30: Detección de fallo usando DSPCA datos de fallo 12

En la Figura 31 se muestra el valor de las 52 variables cuando DSPCA detecta los fallos 7 y 12, para indicarnos cual es la variable responsable del fallo, que en este caso el resultado es muy similar a SPCA.

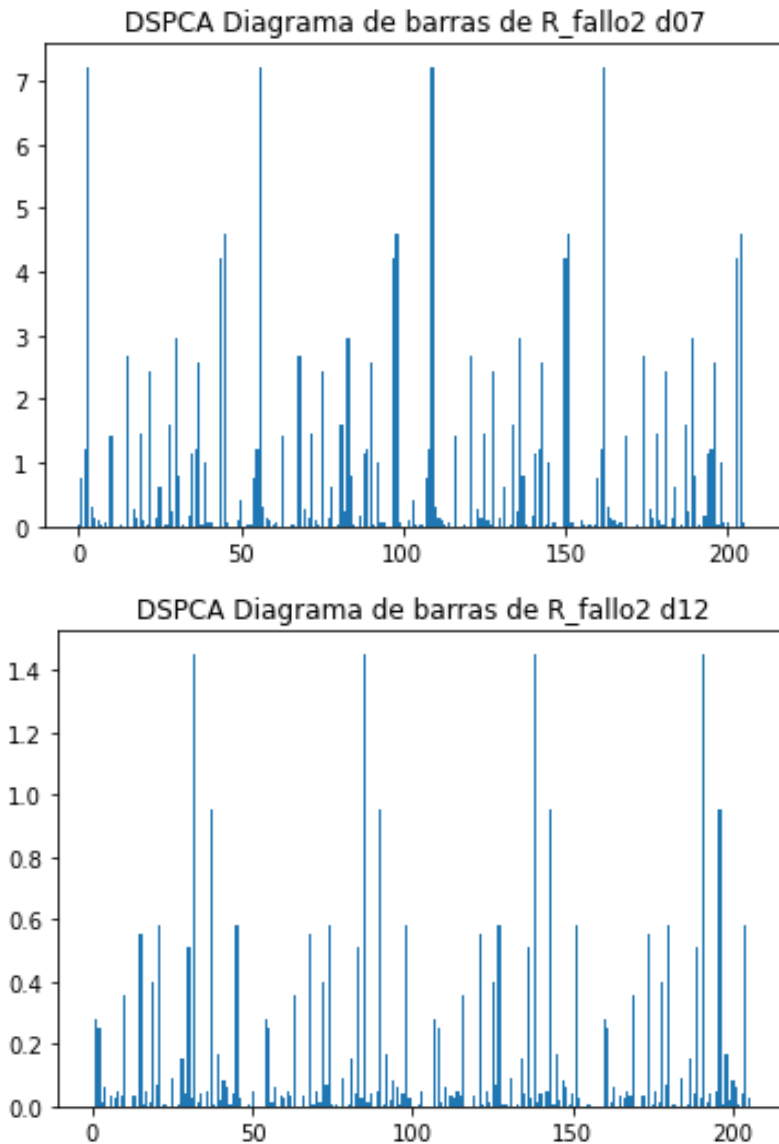


Figura 31: Variables que fallan DSPCA

En la tabla 13 se observa que este método es el que mejores resultados debería haber obtenido, pero en cambio es muy similar a SPCA, los tiempos de fallo son superiores y las falsas alarmas y alarmas detectadas prácticamente iguales, esto podría deberse a los fallos por acumulados por el mayor número de operaciones que realiza, operaciones que se notan también en el tiempo de ejecución.



| DSPCA | T2      |              |              |         | Q            |              |
|-------|---------|--------------|--------------|---------|--------------|--------------|
|       | T fallo | falsa alarma | alarma detec | T fallo | falsa alarma | alarma detec |
| 1     | 6       | 1,875        | 99,25        | 0       | 3,125        | 100          |
| 2     | 11      | 1,250        | 98,625       | 10      | 3,125        | 98,750       |
| 3     | nd      | 0,625        | 1,750        | nd      | 10,625       | 11,250       |
| 4     | 0       | 0,625        | 8,875        | 0       | 5,000        | 100,000      |
| 5     | 10      | 0,625        | 25,375       | 0       | 5,000        | 100,000      |
| 6     | 5       | 0,625        | 99,625       | 9       | 1,250        | 100,000      |
| 7     | 0       | 0,000        | 100,000      | 0       | 0,625        | 100,000      |
| 8     | 22      | 0,000        | 97,250       | 15      | 2,500        | 98,125       |
| 9     | nd      | 3,750        | 3,500        | nd      | 11,875       | 8,250        |
| 10    | 103     | 1,875        | 30,625       | 21      | 0,000        | 92,250       |
| 11    | 374     | 0,000        | 40,125       | 5       | 3,750        | 82,250       |
| 12    | 21      | 1,250        | 98,500       | 1       | 6,875        | 99,875       |
| 13    | 45      | 1,875        | 94,625       | 36      | 1,250        | 95,500       |
| 14    | 0       | 0,625        | 99,250       | 0       | 3,750        | 100,000      |
| 15    | nd      | 1,250        | 2,750        | nd      | 5,000        | 15,250       |
| 16    | 310     | 3,750        | 16,250       | 6       | 8,125        | 94,000       |
| 17    | 28      | 0,625        | 80,625       | 21      | 3,750        | 97,000       |
| 18    | 92      | 1,250        | 89,375       | 77      | 3,750        | 91,125       |
| 19    | nd      | 0,625        | 12,250       | 9       | 4,375        | 94,250       |
| 20    | 86      | 2,500        | 32,500       | 64      | 2,500        | 91,125       |
| 21    | 504     | 0,625        | 42,500       | 394     | 10,625       | 59,875       |
| MEDIA | 101,063 | 1,172        | 71,531       | 37,111  | 3,854        | 94,118       |

Tabla 13: Tabla errores DSPCA



Universidad de Valladolid



ESCUELA DE INGENIERÍAS  
INDUSTRIALES



## Capítulo 5: Conclusiones y trabajo futuro



Universidad de Valladolid



ESCUELA DE INGENIERÍAS  
INDUSTRIALES

## 5.1 Conclusiones

En la actualidad con la gran cantidad de sensores que existen y la posibilidad de conectarlos todos ellos con una latencia muy baja se pueden cuantificar todas las variables de una planta industrial de una forma muy precisa, el cuello de botella se encuentra a la hora de tratar esa información, hay tal cantidad de datos que se hace muy difícil obtener una conclusión de todos ellos. Las diferentes técnicas estadísticas que se han probado intentan obtener a partir de unos datos de la planta Tennessee Eastman si el funcionamiento de esta es el correcto o alguna de las variables ha sufrido alguna desviación que pueda derivar en un fallo o en una pérdida de eficiencia.

Se han comprobado técnicas estadísticas como PCA, SFA, SPA o SPCA además de sus respectivos modelos dinámicos, como actúan ante los mismos datos, cuál de ellos es más eficaz para detectar mayor número de fallos manteniendo bajo el número de falsas alarmas. En cada uno de ellos primero se ha evaluado el funcionamiento normal de la planta para saber cuál es el umbral de funcionamiento en el que las variables no tienen fallos, para después comprobar cómo reaccionan ante valores de variables que indican fallos en el desarrollo del proceso.

En la tabla 14 se muestra un resumen de los resultados de cada uno de los métodos, indicando únicamente el valor medio obtenido de cada uno de los parámetros indicados, tiempo de fallo, falsas alarmas y alarmas detectadas.

|       | T <sup>2</sup> |              |              | Q       |              |              |
|-------|----------------|--------------|--------------|---------|--------------|--------------|
|       | T fallo        | falsa alarma | alarma detec | T fallo | falsa alarma | alarma detec |
| PCA   | 105,941        | 1,176        | 70,125       | 45,824  | 3,493        | 80,581       |
| DPCA  | 101,688        | 0,859        | 71,078       | 48,75   | 3,320        | 78,727       |
| SPA   | 120,83         | 0            | 68,09        | 38,24   | 1,11294118   | 69,16147059  |
| SPCA  | 81,06          | 1,289        | 72,000       | 36,11   | 3,854        | 94,118       |
| DSPCA | 101,06         | 1,172        | 71,531       | 37,11   | 3,854        | 94,118       |

|      | T2a     |              |              | T2e     |              |              |
|------|---------|--------------|--------------|---------|--------------|--------------|
|      | T fallo | falsa alarma | alarma detec | T fallo | falsa alarma | alarma detec |
| SFA  | 25,38   | 4,167        | 93,597       | 26,28   | 0,660        | 84,722       |
| DSFA | 47      | 3,472        | 93,347       | 69,22   | 1,667        | 85,243       |
|      | S2a     |              |              | S2e     |              |              |
|      | T fallo | falsa alarma | alarma detec | T fallo | falsa alarma | alarma detec |
| SFA  | -       | 2,125        | 55,4         | -       | 0            | 88,375       |
| DSFA | -       | 2            | 54,275       | -       | 2,125        | 21,5         |

Tabla 14: Comparación de los diferentes métodos de MSPC implementados



PCA ha sido el primer método estudiado, este es el más conocido y usado a lo largo de los años, por ser el más popular, es el que se va a usar como referencia para los demás. Para el estadístico Q (el que mejores resultados obtiene) consigue un número de alarmas detectadas en torno al 80% y de falsas alrededor del 3%, el tiempo de fallo es 45,82. Su modelo dinámico con siglas DPCA tiene unas estadísticas muy parecidas, incluso un poco inferiores, lo que no lo haría recomendable debido a que necesita más recursos para ejecutarse.

En SPA destaca en el número de falsas alarmas, en cambio el tiempo de fallo y el número de alarmas detectadas es bastante mejorable, el umbral que se obtiene no es lo suficientemente ajustado.

SFA mejora el número de alarmas detectadas respecto a PCA logrando un 93% con el estadístico  $T^2$ , las falsas alarmas se mantienen bajas, alrededor del 4% con un tiempo de fallo de 25,38. La mejora en el número de alarmas detectadas y tiempo de fallo hacen que este modelo sea mejor que PCA. En este caso el modelo dinámico no solo no obtiene unos resultados parecidos, sino que estos son inferiores al modelo estándar.

SPCA obtiene unos magníficos resultados con un 94% de alarmas detectadas, un 4% de falsas alarmas y un tiempo de fallo de 36,11, por ahora es el mejor modelo que se ha analizado, posee un gran número de alarmas detectadas y mantiene bajo el número de falsas alarmas. Por último, DSPCA obtiene prácticamente los mismos resultados que el modelo estándar, sumado a su complejidad a la hora de ejecutarse, este método es menos aconsejable que el modelo estándar.

Se puede concluir que el mejor método es SPCA, ya que obtiene un buen compromiso entre su alto porcentaje de alarmas detectadas y su porcentaje de falsas alarmas, sobre todo considerando el estadístico Q.

## 5.2 Trabajo futuro

Como se ha visto en el punto anterior el método más eficiente a la hora de detectar errores es SPCA, en el futuro este trabajo puede seguirse aplicando la inteligencia artificial, ya que las redes neuronales pueden llegar a conocer el funcionamiento normal de una planta una vez la hayas alimentado con suficientes datos, esto conseguiría una forma más eficiente y eficaz sobre la teoría a la hora de detectar situaciones anormales.

Se pueden crear diferentes umbrales que definan el funcionamiento normal de la planta, ya que actualmente se está trabajando con un único archivo de funcionamiento normal, pero hay muchas variables del proceso que cambian dependiendo de las diferentes condiciones térmicas exteriores o de la pureza de las materias usadas, crear más de un archivo de funcionamiento normal de la planta y



poder aplicarlos dependiendo de las condiciones haría más eficientes todos los métodos para detectar fallos.

Desarrollar un sistema capaz de recomendar mejoras en el proceso estudiando los datos históricos de la planta, por ejemplo, si en el último año la pureza de una materia prima ha decaído respecto a la pureza usada inicialmente para entrenar al método, este sugiera un cambio que compense esa disminución en la pureza.



## Bibliografía

- [1].- Shewhart, W. A. (1931). *Economic control of quality of manufactured products*. Van Nostrand.
- [2].- De la Fuente, M. (2024). *Control estadístico de procesos* [Apuntes de clase]. Universidad de Valladolid.
- [3].- Omarzai, F. (n.d.). *Principal Component Analysis (PCA) in depth*. Medium. Recuperado el 16 de abril de 2025, de <https://medium.com/@fraidoonomarzai99/principal-component-analysis-pca-in-depth-93c871f25dfa> (Último intento de acceso: 2 noviembre de 2024)
- [4].- Rodríguez Campos, A., & Ramos González, J. (2019). *Aplicación del control estadístico de procesos en una línea de producción de la empresa Apis*. (Disponible en: [https://dehesa.unex.es/bitstream/10662/8439/1/978-84-09-04460-3\\_554.pdf](https://dehesa.unex.es/bitstream/10662/8439/1/978-84-09-04460-3_554.pdf)) (Último intento de acceso: 16 de noviembre de 2024)
- [5].- Wiskott, L., & Sejnowski, T. (2022). Slow feature analysis: Unsupervised learning of invariances. *Neural Computation*, Vol. 42, páginas 715–770. [https://www.cnbcmu.edu/~tai/readings/learning/wiskott\\_sejnowski\\_2002.pdf](https://www.cnbcmu.edu/~tai/readings/learning/wiskott_sejnowski_2002.pdf) (Último intento de acceso: 18 de noviembre de 2024)
- [6].- Ma, X., Si, Y., Yuan, Z., Qin, Y., & Wang, Y. (2020). Multistep dynamic slow feature analysis for industrial process monitoring, *IEEE Transactions on Instrumentation and Measurement*, Vol. 69(12), páginas 9535-9548. <https://ieeexplore.ieee.org/document/9123944> (Último intento de acceso: 18 de noviembre de 2024)
- [7] Wang, J., & He, Q. P. (2010). Multivariate statistical process monitoring based on statistics pattern analysis, *Industrial & Engineering Chemistry Research*, Vol. 49, páginas: 7858–7869 [https://pubs.acs.org/doi/epdf/10.1021/ie901911p?ref=article\\_openPDF](https://pubs.acs.org/doi/epdf/10.1021/ie901911p?ref=article_openPDF) (Último intento de acceso: 3 de octubre de 2024)
- [8] Zhang, Y., & Zhao, H. (2015). A novel fault diagnosis method based on statistics pattern analysis and KPCA. *Expert Systems with Applications*, Vol. 38(6), paginas 6895-6905 <https://www.sciencedirect.com/science/article/pii/S0957417410013977> (Último intento de acceso: 4 de octubre de 2024)
- [9]. - Downs, J. J., & Vogel, E. F. (1993). A plant-wide industrial process control problem. *Computers & Chemical Engineering*, Vol 17(3) paginas 245-255 <https://es.scribd.com/document/182323726/Downs-and-Vogel-1993> (Último intento de acceso: 18 de enero de 2025)