UNIVERSITAT POLITÈCNICA DE VALÈNCIA

ESCUELA TÉCNICA SUPERIOR INGENIEROS INDUSTRIALES VALENCIA

TRABAJO FIN DE GRADO EN INGENIERÍA EN TECNOLOGÍAS INDUSTRIALES

DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA PARA LA DETECCIÓN AUTOMÁTICA DE CIBERATAQUES EN INSTALACIONES CRÍTICAS

AUTOR: ALEJANDRO VALENZUELA NAVARRO

TUTOR: JAVIER SANCHÍS SAEZ

COTUTORES: SUSANA BARCELÓ CERDÁ
PEDRO LUIS IGLESIAS REY

Curso Académico: 2017-18

Agradecimientos

A mis amigos, especialmente a Luis.

Abstract

The project tries to design a computer application with connectivity to SCADA systems through the industrial communication protocol OPC with the aim of detecting cyber-attacks on the devices that make up the control system (programmable automata, drives, sensors, the SCADA itself). Multivariate statistical monitoring techniques have been used, specifically statistical models based on Principal Component Analysis (PCA) and control charts T^2 and SPE. With this, it has been possible to carry out an online monitoring of the state of the installation, generating warnings, alarms and corresponding recommendations when anomalous behaviors are detected. The development has been applied to two practical cases. The main case is the detection of cyber-attacks in a water distribution network proposed in the BATADAL competition (BATtle of the Attack Detection ALgorithms). This case study was raised in the area of hydraulic engineering within the activities of the Environmental & Water Resources Institute (EWRI) belonging to the American Society of Civil Engineers (ASCE). The second case is the detection of the different faults in the multiphase flow installation of the University of Cranfield. The data of this case have been published in the MATLAB Central site for later use in benchmarks related to statistical process control. The MATLAB and LabVIEW integrated development environments have been used for the design and implementation of the application.

Key Words: PCA, cyber-attack, multivariate, T^2, SPE, MATLAB, LabVIEW, OPC

Resumen

El proyecto trata de diseñar una aplicación informática con conectividad a sistemas SCADA a través del protocolo de comunicación industrial OPC con el objetivo de detectar ciberataques en los dispositivos que conforman el sistema de control (autómatas programables, accionamientos, sensores, el propio SCADA). Se han utilizado técnicas de monitorización estadística multivariante, concretamente modelos estadísticos basados en el Análisis de Componentes Principales (PCA) y en los gráficos de control T^2 y SPE. Con ello se ha conseguido realizar una monitorización en línea del estado de la instalación, generando los avisos, las alarmas y las recomendaciones correspondientes cuando se detecten comportamientos anómalos. El desarrollo se ha aplicado a dos casos prácticos. El caso principal es la detección de ciberataques en una red de distribución de agua propuesta en la competición BATADAL (BATtle of the Attack Detection ALgorithms). Este caso práctico se planteó en el área de ingeniería hidráulica dentro de las actividades de la Environmental & Water Resources Institute (EWRI) perteneciente a la American Society of Civil Engineers (ASCE). El segundo caso es la detección de los diferentes fallos en la instalación de flujo multifase de la Universidad de Cranfield. Los datos de este caso han sido publicados en el sitio MATLAB Central para su posterior uso en benchmarks relacionados con el control estadístico de procesos. Para el diseño y la implementación de la aplicación se han utilizado los entornos de desarrollo integrado MATLAB y LabVIEW.

Palabras Clave: PCA, ciberataque, multivariante, T^2, SPE, MATLAB, LabVIEW, OPC

Resum

El projecte tracta de dissenyar una aplicació informàtica amb connectivitat a sistemes SCADA a través del protocol de comunicació industrial OPC amb l'objectiu de detectar ciberatacs en els dispositius que conformen el sistema de control (autòmats programables, accionaments, sensors, el propi SCADA). S'han utilitzat tècniques de monitoratge estadístic multivariant, concretament models estadístics basats en l'Anàlisi de Components Principals (PCA) i en els gràfics de control T^2 i SPE. Amb això s'ha aconseguit realitzar un monitoratge en línia de l'estat de la instal·lació, generant els avisos, les alarmes i les recomanacions corresponents quan es detecten comportaments anòmals. El desenvolupament s'ha aplicat a dos casos pràctics. El cas principal és la detecció de ciberatacs en una xarxa de distribució d'aigua proposada en la competició BATADAL (BATtle of the Attack Detection ALgorithms). Aquest cas pràctic es va plantejar en l'àrea d'enginyeria hidràulica dins de les activitats de l'Environmental & Water Resources Institute (EWRI) pertanyent a l'American Society of Civil Engineers (ASCE). El segon case és la detecció de les diferents fallades en la instal·lació de flux multifase de la Universitat de Cranfield. Les dades d'aquest cas han sigut publicats en el lloc MATLAB Central per al seu posterior ús en benchmarks relacionats amb el control estadístic de processos. Per al disseny i la implementació de l'aplicació s'han utilitzat els entorns de desenvolupament integrat MATLAB i LabVIEW.

Paraules Clau: PCA, ciberatac, multivariant, T^2, SPE, MATLAB, LabVIEW, OPC

Índice general

Índice de figuras

Índice de tablas

Lista de símbolos

Letras latinas

A	Número de componentes principales retenidas
c	Factor de corrección
$Cont()$	Contribuciones de T_A^2 y SPE
e	Vector residuo
E	Matriz residuo
F	Distribución F de Snedecor
K	Número de variables del modelo de referencia
m	Número de observaciones del modelo de referencia
p	Vector de loadings, componente principal
P	Matriz de loadings
$\hat{P}$	Matriz P reducida
$\hat{P}(:)$	Vector p con mayor valor t_{norm}
Q	Hace referencia al SPE
s_0	Desviación típica
t	Vector de scores
t_{norm}	Valor normalizado del vector t
T	Matriz de scores
T^2	Hotelling's T-Squared
T_A^2	Hotelling's T-Squared con A componentes principales retenidas
$UCL()_\alpha$	Límite de control de T_A^2 y SPE
$\vec{v}$	Autovector, vector propio o eigenvector
$\vec{v_\lambda}$	Autovector asociado a λ
y	Vector observación de datos originales tipificados
$\hat{y}$	Vector observación reducido
Y	Matriz de datos originales tipificados
$\hat{Y}$	Matriz Y reducida
z	Variable normalizada

Letras griegas

α	Nivel de significación. Riesgo Tipo I
λ	Autovalor, valor propio o eigenvalue
σ	Media
μ	Desviación típica

Siglas

ANN	Artificial Neural Networks
ANSI	American National Standards Institute
ASCE	American Society of Civil Engineers
BATADAL	BATtle of the Attack Detection ALgorithms
BBDD	Base de Datos
CPU	Central Processing Unit
CSV	Comma-Separated Values
CVA	Canonical Variate Analysis
DCS	Distributed Control System
DDoS	Distributed denial-of-service
DPCA	Dynamic Principal Component Analysis
EPANET	Software para el análisis de sistemas de distribución de agua
EWRI	Environmental and Water Resources Institute
IDE	Integrated Development Environment
ISA	International Society of Automation
LabVIEW	Laboratory Virtual Instrument Engineering Workbench
LAN	Local Area Network
MATLAB	MATrix LABoratory
ML	Machine Learning
OPC	Open Platform Communications
OPC UA	OPC Unified Architecture
PDF	Portable Document Format
PLC	Programmable Logic Controller
PLS	Partial Least Squares Regression
RAM	Random Access Memory
RTU	Remote Terminal Unit
SCADA	Supervisory Control And Data Acquisition
SPC	Statistical Process Control
SPE	Squared Prediction Error
SVD	Singular-Value Decomposition
SVM	Support Vector Machines
WDS	Water Distribution System.
Z-Score	Unidad tipificada, variable estandarizada o normalizada

Parte I

Memoria

Capítulo 1

Introducción

1.1 Objeto del proyecto

El presente proyecto tiene como propósito el diseño e implementación de una aplicación que permita la detección de anomalías durante la operación de un proceso cualquiera. A partir de un histórico de datos, dicho software ha de ser capaz de generar un modelo matemático que represente el funcionamiento en condiciones normales del proceso, de monitorizar su actividad en línea, de realizar una correcta gestión de sus alarmas y, si fuese el caso, indicar cuál es el origen del problema que las ha activado.

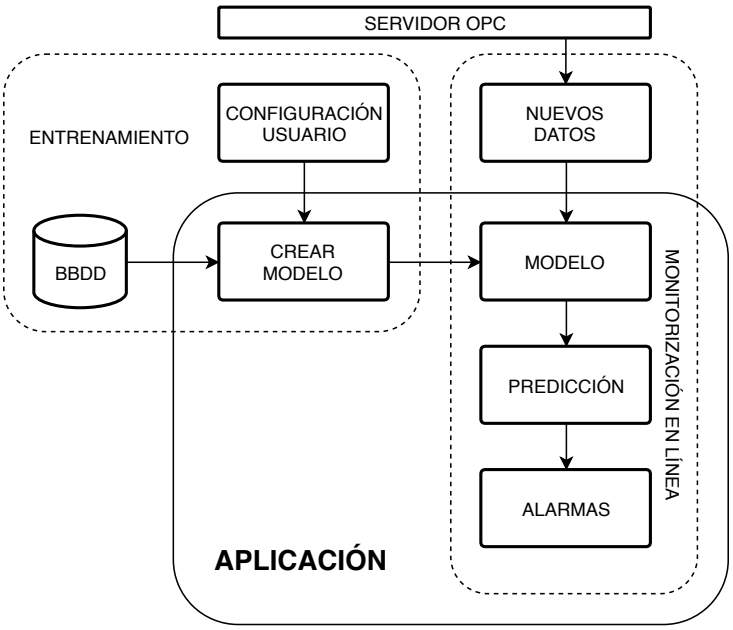

Figura 1.1: Esquema general del proyecto

1.2 Objetivos

La plataforma debe aproximarse a la realidad tanto como sea posible. Desde el proceso de adquisición de los datos, hasta su representación en una interfaz gráfica de usuario. Para ello es necesario cumplir los objetivos del proyecto que han sido agrupados de la siguiente forma:

Análisis previo

- Estudio de los dos casos considerados. La red de distribución de agua propuesta en el concurso BATADAL y la instalación de flujo multifase de la Universidad de Cranfield.
- Análisis previo de los conjuntos de datos en bruto proporcionados. Visualización del conjunto de variables originales. Eliminación de los datos anómalos o valores atípicos del conjunto de entrenamiento.
- Identificación de las distribuciones de probabilidad. Búsqueda de posibles correlaciones entre variables. Correlación temporal.

Desarrollo del algoritmo matemático

- Estandarización de las variables mediante su centrado y escalado. Cálculo de medias y desviaciones típicas. Cálculo del Z-Score o valor normalizado.
- Reducción de la dimensionalidad del conjunto de datos mediante el Análisis de Componentes Principales. Reducción del número de variables siguiendo los diferentes criterios. Criterio del valor propio unitario. Criterio del porcentaje de variabilidad explicada. Criterio del gráfico Scree Plot.
- Cálculo del modelo estadístico que representa el comportamiento normal del proceso.
- Cálculo de los gráficos de control Hotelling's T-Squared (T^2) y Squared Prediction Error (SPE, Q). Cálculo de los límites de control superiores de dichos gráficos.
- Implementación del análisis de contribuciones. Cálculo de los diagramas de contribuciones sobre las variables originales.

Implementación de la aplicación

- Elección justificada de las tecnologías a utilizar. El lenguaje de programación y el entorno de desarrollo.
- Elección de una arquitectura modular y escalable adecuada para su implementación.
- Diseño de la interfaz de usuario. Número de pantallas, elementos interactivos, información que se mostrará, etc.
- Fase de entrenamiento. Configuración manual de los diferentes parámetros para realizar la correcta carga de un fichero CSV. Tipo de delimitador, lectura de la cabecera, criterio para la elección del número de componentes, etc. Carga del fichero CSV que contenga el conjunto de datos. Ejecución del algoritmo previamente definido y generación del modelo matemático.

- Modo en línea. Elección del servidor OPC del que se leerán los datos en el modo en línea. Utilización del modelo generado durante la fase de entrenamiento. Ejecución del programa. Visualización en tiempo real de los datos y los gráficos de control. Gestión de alarmas.

Simulación de una fuente de datos

- Creación de un servidor OPC que permita la distribución de datos en este estándar de comunicación para el control y supervisión de procesos industriales.
- Creación de una interfaz que permita leer los datos de un archivo CSV con una determinada frecuencia de muestreo, simulando así la lectura de las variables de un proceso cualquiera.
- La interfaz también debe escribir los datos en el servidor OPC anterior para que este los pueda distribuir a sus clientes, tal y como sería en un caso real.

Resultados

- Descripción de los resultados del caso BATADAL. Comparación con los publicados en la web de la competición. Elección del criterio más adecuado, gráficos de control, diagramas de contribuciones.
- Descripción de los resultados del caso Cranfield. Elección del criterio más adecuado, gráficos de control, diagramas de contribuciones.

Capítulo 2

Antecedentes

2.1 Ciberataques

Es sabido que la seguridad es uno de los principales puntos a tener en cuenta durante el desarrollo de cualquier actividad industrial. Con la implantación de nuevas tecnologías en la industria han aparecido nuevos retos en materia de seguridad que deben ser superados, especialmente en lo que a su vertiente digital se refiere. Cada día más dispositivos industriales se encuentran conectados a Internet, y con ello mayor es el número de riesgos a los que se ven sometidos. Uno de los mayores peligros a los que está expuesto el sector industrial es el de los ciberataques.

Un ataque informático o ciberataque no es más que una acción intencionada y dirigida para causar daño a un sistema informático u obtener un beneficio de éste. Para poder llevar a cabo estos ataques se explotan las vulnerabilidades de seguridad que existen en los sistemas de información. Estas vulnerabilidades se pueden encontrar tanto en el software como en el hardware (ver Figura 2.1).

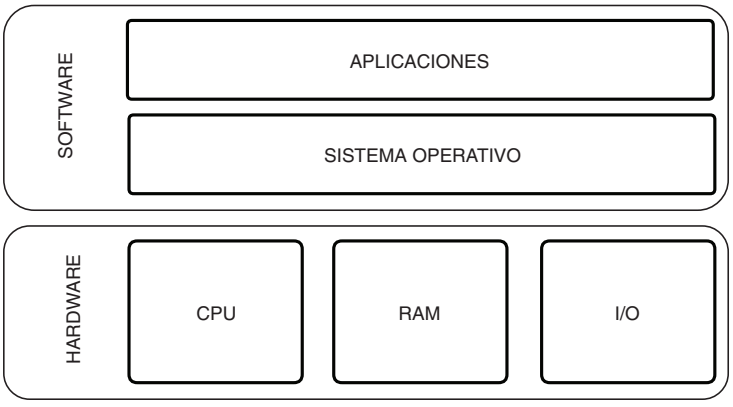

Figura 2.1: Diagrama simplificado de las capas de un sistema informático

Cuando se habla de vulnerabilidades en el software se hace referencia a aquellas que se presentan en las diferentes capas que conforman un sistema operativo, en los drivers que controlan los dispositivos físicos o en los propios programas de usuario que hayan sido instalados. Estos agujeros de seguridad pueden permitir que terceros realicen acciones maliciosas como instalar software sin consentimiento expreso, acceder a bases de datos, realizar modificaciones en ellas sin tener los permisos necesarios e interrumpir procesos del propio sistema, entre otras. Normalmente, son de estas debilidades de las que suelen aprovecharse los diferentes tipos de malware. Los también llamados programas maliciosos son la amenaza más extendida. Los hay de diferentes clases: virus, gusanos, troyanos y spyware son los más comunes, y cada uno tiene sus propias peculiaridades. Los virus y los gusanos tienen como objetivo propagarse e infectar todos los sistemas posibles para luego tomar el control de éstos. Los troyanos incluyen una gran variedad de programas que efectúan acciones sin el consentimiento del usuario. Recolectan datos y los envían; los destruyen o los alteran. En cambio, el spyware son programas que recogen información de forma no autorizada y pasan totalmente inadvertidos para el usuario.

	MELTDOWN	SPECTRE
CPU afectadas	Intel, Apple	Intel, Apple, ARM, AMD
Vector de ataque	Ejecución de código en el sistema	
Método	Escalado de privilegios	Predictor de saltos
Método 2	Ejecución especulativa	
Corrección	Actualización de software	

Tabla 2.1: Resumen de Meltdown y Spectre

Cuando se habla de vulnerabilidades en el hardware se hace referencia a aquellas que se presentan directamente en los dispositivos físicos. Esto puede ser la memoria RAM, los procesadores, unidades de almacenamiento o incluso la arquitectura de los dispositivos de una red. Meltdown [9] y Spectre [8] (Tabla 2.1) son dos vulnerabilidades que han aparecido recientemente y que afectan a multitud de sistemas, entre ellos los procesadores Intel x86, muy utilizados en todo el mundo. Mientras que Meltdown es una vulnerabilidad que permite que cualquier proceso pueda leer de la memoria virtual aún sin contar con los permisos para hacerlo, Spectre permite que terceros tengan acceso a información privada. Ambas son vulnerabilidades extremadamente graves, y aunque se han publicado actualizaciones de software para mitigarlas, no serán totalmente eliminadas hasta que no se modifique la arquitectura interna de los procesadores afectados.

Normalmente los ataques informáticos suelen aprovecharse al mismo tiempo de diferentes debilidades para cumplir su propósito, siendo el factor humano una de las más críticas. Son muchas las amenazas que se aprovechan de los descuidos de los usuarios. Estos descuidos pueden ser producto de la falta de información o de una confianza excesiva por parte de las personas. Así se generan un sin fin de oportunidades con las que se puede vulnerar la seguridad de un sistema.

Retomando el tema principal de esta sección, los ciberataques se clasifican en dos grandes grupos dependiendo del tipo de objetivo al que estén dirigidos: ataques destructivos y ataques indiscriminados o indistintos.

Por un lado, los primeros son ataques que se dirigen hacia un objetivo específico, habitualmente contra organizaciones y estados. Como ejemplo reciente tenemos el ataque que sufrió la plataforma GitHub el pasado día 28 de febrero de 2018. Fue un ataque de denegación de servicio

distribuido o DDoS. Este tipo de ataque, mediante la saturación y la sobrecarga de los recursos del sistema, causa que el servicio objetivo sea inaccesible por sus usuarios legítimos. En el caso de GitHub, gracias a la rápida actuación de su proveedor de servicios Akamai el ataque fue mitigado en tan solo 8 minutos (ver Figura 2.2) [7].

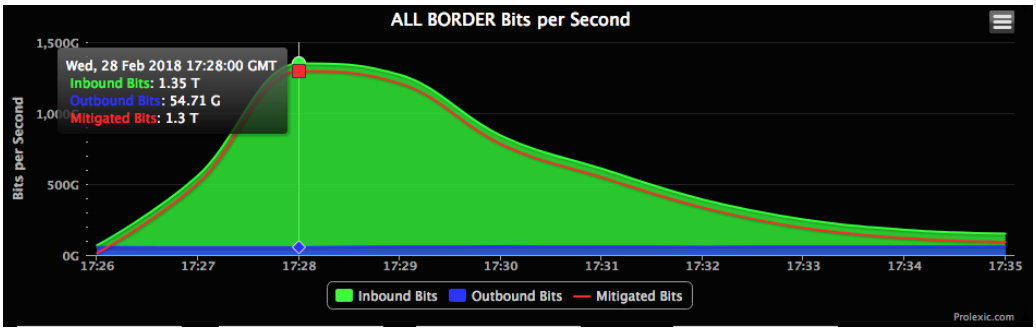

Figura 2.2: Ancho de banda utilizado durante el ciberataque a GitHub [7]

Por otro lado, los segundos son aquellos que no se dirigen hacia un objetivo concreto. Suelen ser globales y no distinguen de gobiernos, empresas o usuarios domésticos. Así, un ejemplo reciente sería el ciberataque WannaCry (ver Figura 2.3) [18], que se produjo el 12 de mayo de 2017. Un ataque a escala global que afectó a más de 200.000 sistemas en todo el mundo. En este caso, se trata de un Ransomware, un tipo de programa capaz de restringir el acceso a ciertas partes de los ordenadores mediante su cifrado, y después pedir un rescate por ellas a sus usuarios legítimos. Al contrario que en el caso GitHub, este ataque tuvo un gran impacto en todo el mundo viéndose afectados los diferentes sistemas durante horas e incluso días. Grandes empresas españolas se vieron comprometidas, siendo algunas de ellas gestoras de infraestructuras críticas.

Figura 2.3: Captura de pantalla de Wannacry [18]

2.1.1 Ciberataques a instalaciones críticas

Por su propia definición las infraestructuras críticas de un país son un objetivo muy atractivo para los ataques informáticos. Especialmente los relacionados con ciberterrorismo o con ciberespionaje. Un ataque puede causar desde pequeñas incidencias en el funcionamiento habitual de una planta industrial hasta desestabilizar toda una región. Otros, simplemente se producen con el fin de obtener información y tratan de ser indetectables.

El 23 de diciembre de 2015 sucedió el primer ciberataque con éxito a una red de distribución eléctrica [3] (ver Figura 2.4). Este hecho tuvo lugar en Ucrania. Los atacantes tuvieron acceso a los sistemas de información de al menos tres compañías de distribución de energía de dicho país, e interrumpieron temporalmente el suministro de electricidad a los consumidores finales.

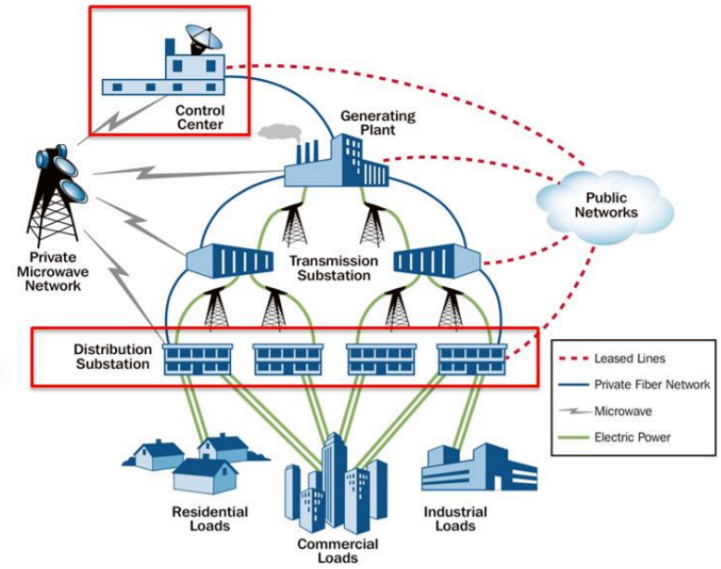

Figura 2.4: Diagrama del ataque a la red eléctrica de Ucrania [3]

El ataque tuvo tres objetivos principales. Tomar bajo control los sistemas SCADA para desconectar subestaciones remotamente, la destrucción de archivos almacenados en servidores y estaciones de trabajo, y deshabilitar componentes clave de la infraestructura como sistemas de alimentación ininterrumpida, conmutadores, modems, RTUs, etc.

2.1.2 Ciberataques a instalaciones hidráulicas

Debido al origen del caso principal que nos ocupa cabe destacar como las instalaciones hidráulicas también se han visto afectadas por los ciberataques. A continuación, se muestran dos ejemplos:

En 2002, en una planta de fabricación de componentes electrónicos de Estados Unidos, un PLC del sistema de osmosis inversa que se utilizaba para la purificación de agua en la fabricación de semiconductores fue atacado y deshabilitado. Esto ocurrió debido a que dicho PLC estaba inadecuadamente protegido y directamente conectado a una red de área local (LAN) que era accesible desde Internet. Así, personas no autorizadas y ajenas a la instalación tuvieron acceso al PLC. Esto se tradujo en un paro total del sistema de osmosis inversa durante varias horas,

aunque el coste económico fue mínimo gracias a las reservas de agua que había acumuladas en tanques de la instalación. La solución para resolver este problema fue sencilla, simplemente se movió el PLC a una LAN protegida por un cortafuegos (Figura 2.5).

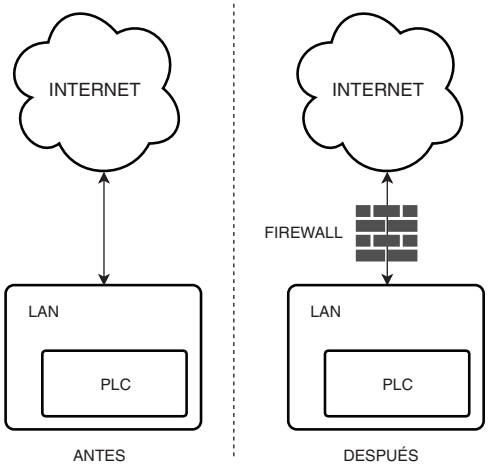

Figura 2.5: Solución adoptada en la planta de componentes electrónicos

El otro caso ocurrió durante 2003, en Australia. Mientras se realizaba una auditoría a un sistema de control de agua se tuvieron problemas con el funcionamiento de la instalación. En un principio, se sospechó de posibles errores en una de las bases de datos, pero no fue así. Durante unas pruebas de funcionamiento el sistema se bloqueó completamente. Tras investigar lo sucedido, se descubrió que el virus Blaster se había instalado en el sistema de control. Este virus deja abierta una puerta trasera que permite la intrusión a terceros, haciendo que la máquina infectada pueda ser fácilmente atacada por otros virus, o accesos remotos no autorizados. El software para el control de la instalación no había sido actualizado con los parches de seguridad correspondientes y por ello se vio afectado. En este caso, la solución que se llevó a cabo fue corregir la vulnerabilidad mediante dichas actualizaciones y se propuso la separación del sistema SCADA llevándolo una LAN protegida para prevenir futuras incidencias.

2.2 Control Estadístico de Procesos Multivariante, MSPC

El desarrollo de los sistemas de control distribuido ha permitido que cada vez sea mayor la cantidad de datos que se recogen y almacenan de los procesos industriales. Generalmente, los datos se pueden utilizar para modelar, monitorizar y controlar un proceso cualquiera. El desarrollo de nuevas técnicas matemáticas ha permitido que incluso los procesos más complejos puedan gestionarse desde esta perspectiva. Esto se puede llevar a cabo extrayendo del bruto de información la parte más importante que nos permita generar modelos basados en esos datos donde se vea reflejado el comportamiento de dicho proceso. En lo que se refiere a la monitorización de procesos, el denominado Control Estadístico de Procesos Multivariante (MSPC) [6] juega un papel fundamental en la industria. El MSPC es la generalización del Control Estadístico de Procesos (SPC) dirigida a facilitar el manejo de conjuntos de datos con un alto número de variables correlacionadas entre sí. Esto es así porque uno de los pasos a la hora de aplicar técnicas de MSPC es la reducción de dimensionalidad del conjunto de datos original. Dicha reducción se lleva a cabo mediante la descomposición de la información teniendo en cuenta las correlaciones entre las diferentes variables.

El Análisis de Componentes Principales (PCA) junto con la Regresión de Mínimos Cuadrados Parciales (PLS) son los dos métodos más utilizados en la actualidad para aplicar la reducción de dimensionalidad en monitorización de procesos. Extraer las componentes principales de los datos de proceso y calcular los estadísticos T-Squared (T^2) y Squared Prediction Error (SPE) permite, de manera sencilla, la monitorización de procesos, la detección de fallos y su diagnóstico identificando las variables originales vinculadas.

2.2.1 Análisis de Componentes Principales

El PCA es un procedimiento estadístico que, haciendo uso de una transformación ortogonal (Fórmula 2.1), convierte un conjunto de observaciones originales Y en las que posiblemente existan correlaciones entre las variables que lo componen, en un nuevo conjunto de observaciones T en el que sus variables son linealmente independientes. A este nuevo grupo de variables se le denomina componentes principales P. Es decir, se proyecta el espacio que contiene las variables originales en un nuevo subespacio en el que los vectores directores resultan linealmente independientes u ortogonales.

$$Y = TP^T \tag{2.1}$$

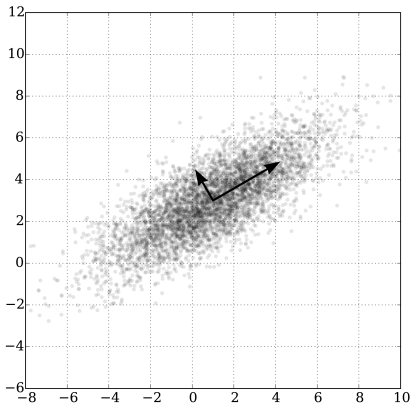

Figura 2.6: Direcciones principales de un set de datos. [21]

Estos nuevos vectores directores ($\vec{v_\lambda}$) también llamados autovectores, vectores propios o eigenvectors, tienen asociado un autovalor, valor propio o eigenvalue (λ). Dicho valor propio representa la varianza del conjunto de datos según la dirección de su vector propio. Cuando se obtienen las componentes principales de un determinado set de datos, éstas se ordenan de mayor a menor según el valor absoluto de su valor propio. Así, la primera componente principal será la que mayor variabilidad del conjunto de datos original contenga y la última la que represente en menor medida los datos de partida.

Existen diferentes métodos para llevar a cabo el PCA. Los principales son la Descomposición en Valores Singulares (SVD) [22] y la obtención de los autovectores y sus autovalores a partir de la matriz de covarianza del conjunto de datos. Ambos métodos suelen aplicarse tras realizar un centrado de los datos mediante la sustracción de la media o de calcular sus valores tipificados.

Después del cálculo de las nuevas direcciones principales se procede a elegir el número adecuado de ellas que serán suficientes para representar el comportamiento del modelo original. Este modelo reducido viene expresado por la Fórmula 2.2 [10].

$$\hat{Y} = T\hat{P}^T \tag{2.2}$$

Aunque existen numerosos estudios sobre como realizar correctamente esta elección, en la mayoría de los casos se opta por un criterio arbitrario, siempre teniendo en cuenta el porcentaje de variabilidad retenido. La información que se pierde al reducir la dimensionalidad del conjunto se llama residuo E y viene expresado por la Fórmula 2.3:

$$E = Y - \hat{Y} \tag{2.3}$$

Una vez se ha realizado el volcado de los datos sobre el conjunto reducido de nuevas direcciones principales se puede proceder al cálculo de los estadísticos que se requieran.

2.2.2 *Gráficos de control*

Los gráficos de control no son más que la representación gráfica de determinadas variables estadísticas. Aunque existen diversos tipos de gráficos de control, una elección habitual suele ser el Hotelling's T-Squared (T^2) junto con el Squared Prediction Error (SPE) [10] [5]. El primero es un estadístico que representa las variaciones que sufren las variables dentro del modelo matemático generado, mientras que el SPE representa las variaciones fuera de dicho modelo. Esto se puede ver gráficamente en la Figura 2.7. En ella el modelo matemático se representa mediante una elipse, el estadístico T^2 se encuentra en el plano generado por la elipse (valores inusuales dentro del modelo) y el estadístico SPE, representado como Q en la figura, se encuentra en otro plano totalmente diferente (valores inusuales fuera del modelo). Ambos tienen significados matemáticos distintos y se complementan mutuamente haciendo de su uso una buena elección.

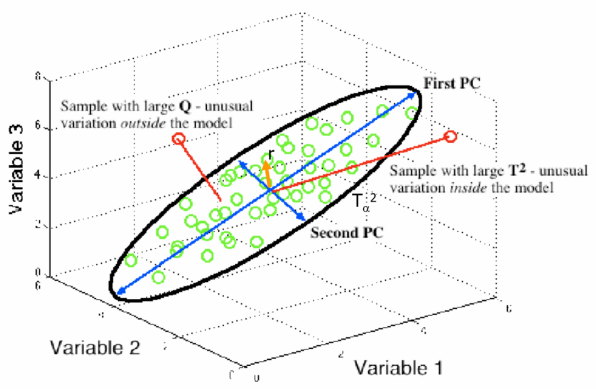

Figura 2.7: Interpretación de T^2 y Q (SPE). [11]

Hotelling's T-Squared, T^2

El estadístico Hotelling's T-Squared se calcula a partir de los *scores* (t_i), que no son más que los valores que toman las variables en el nuevo subespacio tras la transformación realizada durante el PCA. Siendo A el número de componentes principales retenidas, T_A^2 se calcula según de la Fórmula 2.4:

$$T_A^2 = \sum_{i=1}^{A} t_i \lambda_i^{-1} t_i^T \tag{2.4}$$

Luego por cada observación se tendrá un valor de dicho estadístico. Al dibujar los valores obtenidos para cada observación se genera el gráfico de control Hotelling's T-Squared.

Para completar este gráfico hay que añadirle el límite de control, que no es más que un valor límite superior a partir del cual se dice que los valores que están tomando los datos son atípicos, y que en el caso del T_A^2, como ya se ha mencionado antes, representan variaciones dentro del modelo matemático generado. Dicho límite se calcula mediante la Fórmula 2.5. Los parámetros se especifican en la Tabla 2.2.

$$UCL(T_A^2)_\alpha = \frac{A(m^2-1)}{m(m-A)} F_{(A,(m-A)),\alpha} \tag{2.5}$$

En la Figura 2.8 podemos ver un ejemplo de un gráfico de control Hotelling's T-Squared.

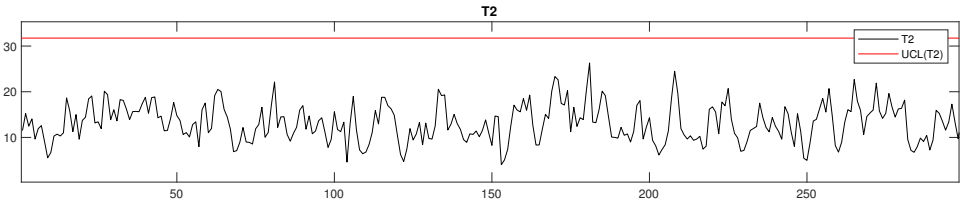

Figura 2.8: Ejemplo de gráfico de control Hotelling's T-Squared

Como se muestra, los valores de T_A^2 dibujados en negro siempre están por debajo de la línea roja. Dicha línea representa el límite $UCL(T_A^2)_\alpha$, límite a partir del cual se considera que el proceso representado por T_A^2 se encuentra fuera de control. Luego, en el caso de la figura, el proceso se encuentra bajo control.

Squared Prediction Error, SPE

El estadístico Squared Prediction Error se calcula a partir de la Fórmula 2.6

$$SPE = e_i^T e_i = (y_i - \hat{y_i})^T (y_i - \hat{y_i}) \tag{2.6}$$

Tal y como su nombre indica, el SPE es el error de predicción al cuadrado. Así, según la fórmula anterior, y_i es el valor original de la variables e $\hat{y_i}$ el valor predicho por el modelo matemático generado a partir del PCA. A esta diferencia entre el valor original y el valor predicho se le denomina residuo. En este caso, a cada nueva observación le corresponde un valor SPE, que dibujado en una gráfica forma el gráfico de control de SPE. Este gráfico de control también tiene su respectivo límite de control, que viene dado por la Fórmula 2.7:

$$UCL(SPE)_\alpha = \frac{K-A}{c^2} s_0^2 F_{(K-A,(m-A-1)(K-A)),\alpha} \tag{2.7}$$

En el cálculo de ambos límites de control (UCL, Upper Control Limit) se utiliza el valor F que hace referencia a la distribución F de Snedecor. Para su cálculo se necesitan indicar los parámetros de la Tabla 2.2.

A	Nº de componentes retenidas en el modelo PCA
m	Nº de observaciones del modelo de referencia
K	Nº de variables originales
s_0	Desviación típica
c	Factor de corrección
α	Nivel de significación (Riesgo Tipo I)

Tabla 2.2: Parámetros para el cálculo de los UCL

En la Figura 2.9 podemos ver un ejemplo de un gráfico de control Squared Prediction Error. Al igual que en la Figura 2.8, el proceso se encuentra bajo control debido a que todos los valores representados en negro (SPE) se encuentran por debajo de su respectivo límite representado de color rojo.

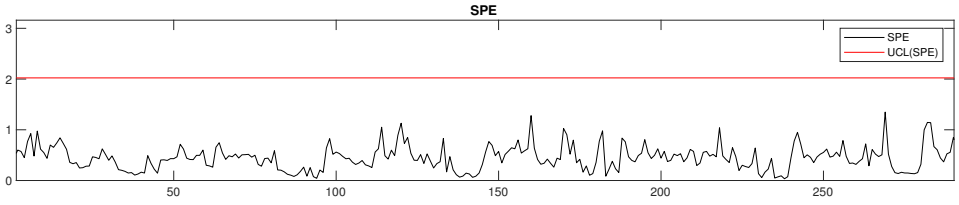

Figura 2.9: Ejemplo de gráfico de control Squared Prediction Error

2.2.3 Diagramas de Contribuciones

Hasta el momento se ha descrito una manera de detectar fácilmente comportamientos atípicos mediante dos simples gráficos. ¿Hay alguna forma de averiguar que variables son las causantes de estos comportamientos atípicos? Sí, este diagnóstico se puede realizar a través de los Diagramas de Contribuciones. Estos diagramas indican como han contribuido las diferentes variables a generar los valores anómalos en los gráficos de control. Se trata de un procedimiento que permite revertir los cálculos realizados hasta el momento y obtener en qué proporción las diferentes variables están afectando al comportamiento del sistema. Como se muestra en la Figura 2.10, los diagramas de contribuciones son diagramas de columnas con tantas barras como variables originales haya, y de ellos no es tan importante el valor absoluto que se le atribuya a cada barra como su valor relativo respecto al resto. Es decir, hay que fijarse en si una o varias barras destacan sobre el resto en términos relativos.

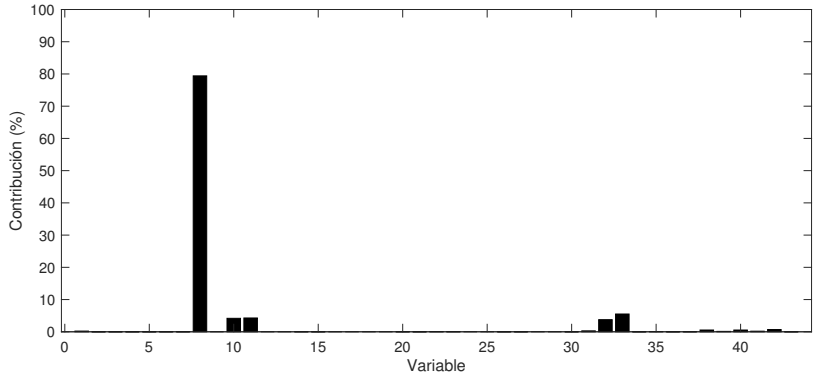

Figura 2.10: Ejemplo de diagrama de contribuciones

Contribuciones Hotelling's T-Squared, T^2

El cálculo de las Contribuciones de T_A^2 se realiza mediante el siguiente procedimiento. Primero se calculan los *scores* t de la nueva muestra y con la Fórmula 2.8.

$$t = \hat{P}y \tag{2.8}$$

Luego se calcula el valor normalizado de los *scores* t. Para ello hay que dividir cada *score* t entre su respectivo autovalor λ y calcular su cuadrado. Como indica la Fórmula 2.9:

$$t_{norm} = \left(\frac{t}{\lambda}\right)^2 \tag{2.9}$$

De la matriz P reducida $\hat{P}$ se elegirá el vector correspondiente al mayor valor normalizado t_{norm}. Este vector se le llamará $\hat{P}(:)$ y tendrá tantos elementos como variables originales se tengan. Así, las contribuciones se pueden calcular según la Fórmula 2.10:

$$Cont(T_A^2) = \hat{P}(:) \circ y \tag{2.10}$$

Esta última fórmula expresa la multiplicación elemento a elemento [20] de dos vectores con las mismas dimensiones.

Contribuciones Squared Prediction Error, SPE

El cálculo de las contribuciones al SPE se realiza mediante la Fórmula 2.11:

$$Cont(SPE) = e_i^2 = (y_i - \hat{y_i})^2 \tag{2.11}$$

Como se puede apreciar, la contribución al SPE es equivalente al valor del residuo al cuadrado e_i^2. A diferencia del cálculo de las contribuciones del T_A^2, al realizar la operación de la Fórmula

2.11 se obtienen automáticamente las contribuciones referidas a las variables del espacio original, las variables reales.

Capítulo 3

Descripción del problema

En los siguientes apartados se ha procedido a realizar una explicación de los dos casos que nos atañen: caso BATADAL y caso Cranfield. Se ha realizado una breve introducción del contexto en el que se sitúan ambos problemas, seguida de una detallada descripción de los recursos de los que se ha partido para poder resolverlos. Como ya se verá en el capítulo correspondiente, ambos casos han sido afrontados desde la misma perspectiva. Se ha tratado de mantener un enfoque generalista que, utilizando métodos estadísticos, permita abstraerse del contexto del problema y así poder aplicar la solución propuesta a otros casos diferentes.

3.1 Caso BATADAL

Como ya se ha visto en el capítulo anterior, en los últimos años los riegos a los que se expone la industria han crecido considerablemente. Cada vez son más las redes de distribución de agua que son supervisadas por complejos sistemas conectados a Internet. Estas soluciones permiten tener un mejor control de estas redes a costes más bajos. Si bien esto es cierto, la implementación de estas nuevas tecnologías expone a los sistemas de distribución de agua a nuevas vulnerabilidades de seguridad[12] con respecto a las alternativas cableadas que se usan tradicionalmente en el sector [14][16]. Al igual que para otras instalaciones críticas, la posibilidad de que se produzcan ciberataques supone una gran amenaza.

En la industria, normalmente los ciberataques están dirigidos hacia los sistemas de Supervisión, Control y Adquisición de Datos (SCADA, Supervisory Control and Data Acquisition) o hacia los Controladores Lógicos Programables (PLC, Programmable Logic Controllers) que suelen gobernar actuadores [2]. Esto se corresponde con los niveles 1 y 2 de la Jerarquía de los Procesos Industriales descrita en la norma ANSI/ISA-95 (Figura 3.1).

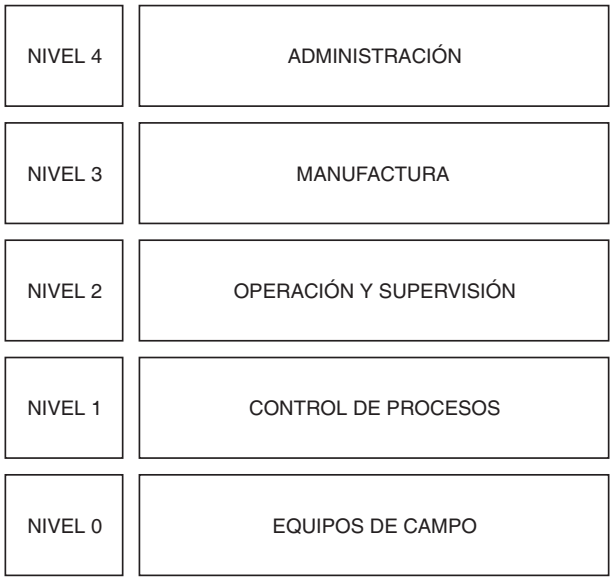

Figura 3.1: Niveles establecidos en el estándar ANSI/ISA-95

El riesgo que conlleva este problema se puede minimizar añadiendo medidas adicionales de seguridad a la red, o un mayor número de capas en el SCADA. Lo que se conoce con el término defensa en profundidad [19] (Figura 3.2). Aunque esto es un primer paso para garantizar la seguridad no es una metodología que resulte infalible. Con el paso del tiempo aparecen nuevos agujeros de seguridad resultando poco efectivas las medidas anteriormente mencionadas.

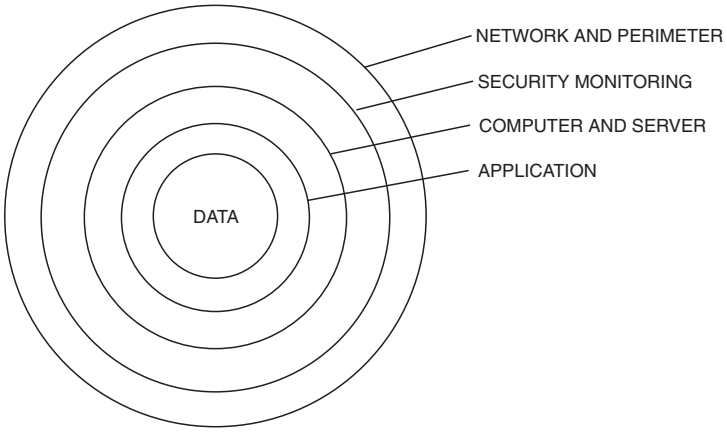

Figura 3.2: Diagrama del concepto de defensa en profundidad

Por otra parte, sigue existiendo la posibilidad de que haya una manipulación directa de las medidas de los sensores. Un ataque directo a la capa física del sistema, nivel 0 según el estándar ANSI/ISA-95. Así se considerarán como ataques ciber-físicos aquellos que combinen ambos orígenes. Aunque el problema que propone el Environmental and Water Resources Institute a través de su concurso BATADAL [4] hace referencia a este último concepto, de aquí en adelante también se llamarán simplemente ciberataques a aquellos que también involucren el citado nivel 0.

Tal y como dice la web del concurso, el objetivo es detectar de manera fiable la presencia de lecturas anómalas en el SCADA y hacerlo en el menor tiempo posible desde que comienzan a aparecer. Además, el algoritmo debe evitar falsas alarmas y reconocer cuando una amenaza ha desaparecido. Debido a la naturaleza distribuida del WDS, el algoritmo ideal también debería ser capaz de identificar qué componentes de la red física son atacados para facilitar la resolución de la incidencia. La interdependencia inherente de los elementos en la red de agua teóricamente debería permitir la detección de anomalías incluso cuando el adversario intente ocultar sus acciones alterando las lecturas de SCADA de un pequeño grupo de los sensores desplegados.

3.1.1 Descripción de la red de distribución de agua C-Town

C-Town Public Utility (CPU) es el principal operador del sistema de distribución de agua de una ciudad llamada C-Town [15]. En la Figura 3.3 se muestra la representación de dicha red.

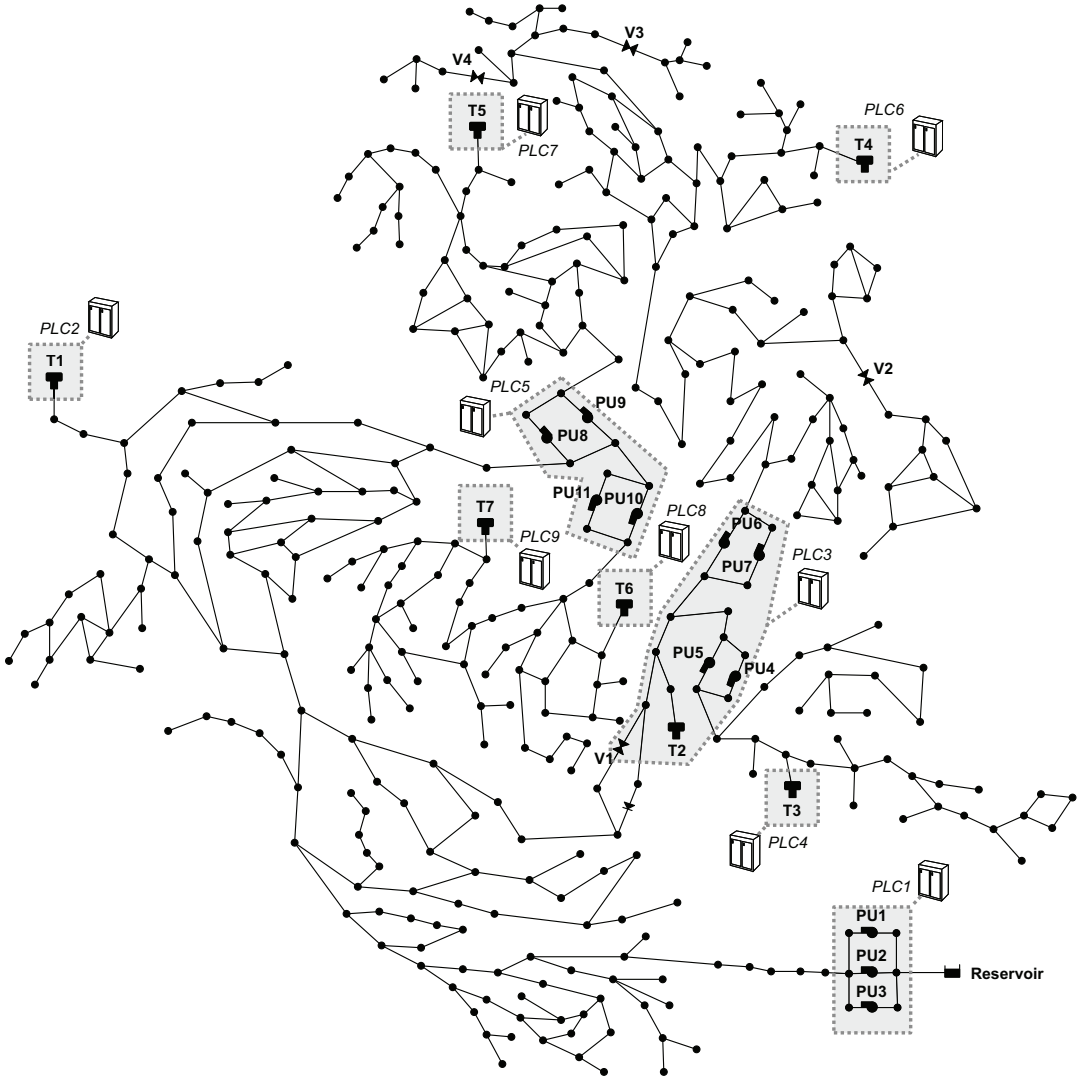

Figura 3.3: Red de distribución de agua de la ciudad C-Town. [16]

En los últimos años, CPU ha operado lo que se denomina una tipología de red estática, sin estar conectada a Internet. El año pasado, el operador CPU implementó una novedosa tecnología que permitía la recolección de datos a distancia de los sensores de la red de agua y también el control de los actuadores. Poco después de que esta nueva tecnología fuese desplegada se observaron niveles anormalmente bajos en el Tanque T5 y anormalmente altos en el Tanque T1. Un mes más tarde, inesperadamente se produjo un desbordamiento en este último tanque. Este incidente se produjo bajo extrañas circunstancias. Mientras sucedía, los operarios observaron que las lecturas que se recibían del nivel de dicho tanque siempre estaban por debajo de los límites de alarma y el comportamiento de las bombas parecían ser normales. Buscando las causas que habían provocado estos últimos episodios, se contempló la posibilidad de que, potencialmente, se tratase de un ciberataque. En concreto se consideró que los atacantes fuesen capaces de activar y desactivar los actuadores de la red a voluntad, así como también que pudiesen alterar las medidas de los sensores desplegados e interferir en las conexiones entre los diferentes componentes tal y como se muestra en el esquema de la Figura 3.4.

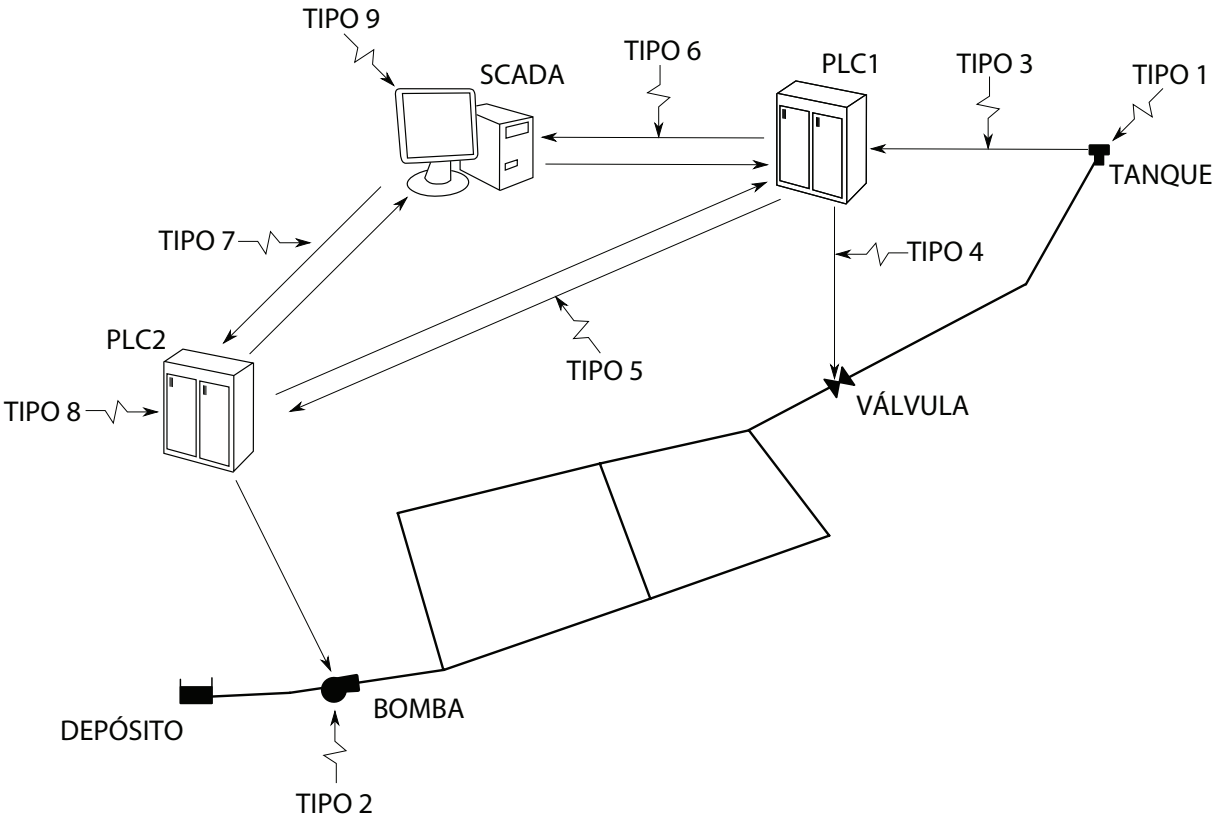

Figura 3.4: Tipos de ataques [16]

En la figura anterior se puede observar los diferentes tipos de incidentes que se pueden producir en la red de distribución de agua [16]. Estos han sido descritos a continuación:

Tipo 1. Ataque físico a un sensor

Para poder realizar este ataque se parte de la premisa de que el atacante tiene acceso físico al sensor. Así dicho sensor puede ser manipulado, dañado o incluso reemplazado. Como consecuencia las lecturas de los PLC asociados a ese sensor pueden ser erróneas, comprometiendo así el control de parte de la instalación.

Tipo 2. Ataque físico a un actuador

De la misma manera que en el ataque Tipo 1, en este caso se presupone que se tiene acceso físico directo al componente en cuestión. El actuador podría ser activado o desactivado a voluntad por el atacante o su punto de operación (apertura de una válvula, velocidad de una bomba, etc) cambiado afectando a la integridad del sistema.

Tipo 3. Ataque a la conexión entre sensores y PLC

La conexión entre los sensores y el PLC se puede realizar tanto de manera cableada como mediante tecnología wireless. Dependiendo de si la conexión es de un tipo u otro, el atacante necesitará tener o no acceso físico al sistema para llevar a cabo sus acciones. Éstas normalmente están orientadas a interrumpir la conexión (denegación de servicio), manipular los paquetes de datos u obtener información del estado del sistema.

Tipo 4. Ataque a la conexión entre actuadores y PLC

Teniendo las mismas consideraciones que el Tipo 3 en cuanto a los diferentes modos de conexión, este tipo de ataque puede afectar al control de los equipos de campo. La pérdida total del control de los equipos o alteración de las órdenes comandadas por los PLC pueden afectar a los elementos finales como válvulas y bombas provocando graves incidencias en la operación de la instalación.

Tipo 5. Ataque a la conexión entre dos PLC

Generalmente los PLC están conectados entre ellos a través de una red privada para intercambiar información del estado del sistema. Una casuística que se puede dar es cuando un PLC controla una bomba de agua que propulsa agua a un determinado depósito cuyo nivel está medido por sensores conectados a otros PLC. En este caso puede existir una comunicación directa entre ambos PLC y que el atacante se centre en ella con el objetivo de manipular la información que circula, obtenerla o directamente, cortarla.

Tipo 6. Ataque a la conexión entre PLC y SCADA

En este ataque, la comunicación de PLC a SCADA (generalmente establecida a través de una red privada o Internet) se manipula, escucha a escondidas o se interrumpe temporalmente saturando el canal de comunicación. Como resultado, información incompleta o incorrecta sobre el estado del sistema alcanza al SCADA. El atacante podría recurrir a este ataque para ocultar otras acciones de operadores humanos o algoritmos de detección de eventos implementados a nivel SCADA.

Tipo 7. Ataque a la conexión entre SCADA y PLC

Este ataque es similar al Tipo 6. En este ataque, las señales enviadas por el SCADA a un PLC son bloqueadas (denegación de servicio), manipuladas (en un ataque de engaño) o interceptadas por el atacante.

Tipo 8. Ataque a un PLC

Con este ataque, el atacante tiene el control directo de un PLC de la red. Dependiendo del nivel de control obtenido, el atacante puede detener por completo las operaciones normales del proceso controlado, manipular la lógica de control en el PLC o informar deliberadamente con datos incorrectos al SCADA. Aunque este tipo de ataque está relacionado con algunos de los ataques descritos previamente (tipos 6 y 7), este ataque en particular es generalmente más persistente. Se debe asumir que un PLC comprometido está bajo el control del atacante hasta que el ataque sea detectado y restaurado. Por el contrario, los otros ataques generalmente se caracterizan por un comportamiento intermitente que requiere la interacción constante del atacante.

Tipo 9. Ataque a un SCADA

Este ataque representa una situación en la que el atacante ha comprometido el sistema SCADA, ya sea a través de un ataque local o remoto. Así, es capaz de cambiar arbitrariamente cualquier configuración del sistema y obtener todos los datos medidos por los sensores. Esta familia de ataques se ha incluido en el listado para completar todos los tipos, pero no se considera en el resto del trabajo porque, en última instancia, no se puede detectar un sistema SCADA comprometido con los métodos que se han utilizado en este trabajo. Las técnicas eficaces contra este tipo de intrusiones quedan fuera del alcance de este proyecto.

Una vez descrito lo anterior se procede a detallar los archivos de partida que se proporcionaron desde la organización del concurso BATADAL.

3.1.2 Descripción de los datos suministrados

Desde la organización de BATADAL se han proporcionado a los concursantes los archivos numerados en la Tabla 3.1.

ID	Nombre del archivo	Formato	Descripción
1	rules.pdf	PDF	Enunciado y reglas del concurso
2	CTOWN.INP	EPANET INP	Modelo de la red en EPANET
3	BATADAL_dataset01.csv	CSV	Datos sin ataques*
4	BATADAL_dataset02.csv	CSV	Datos con ataques identificados
5	BATADAL_dataset03.csv	CSV	Datos sin ataques
6	BATADAL_dataset04.csv	CSV	Datos con ataques identificados
7	BATADAL_test_dataset.csv	CSV	Datos con ataques sin identificar

Tabla 3.1: Archivos proporcionados por BATADAL

El documento *rules.pdf* [15] contiene una breve introducción al tema, el calendario con las fechas clave del concurso, una descripción del problema, las reglas, los objetivos y los criterios de evaluación de las soluciones propuestas por los participantes. El archivo *CTOWN.INP* es un fichero de EPANET que contiene el modelo de la red de agua sobre la que se ha trabajado (Figura 3.3). Los ficheros *BATADAL_ datasetXX.csv* contienen los datos de la red generados en diferentes periodos de tiempo. En su momento, los archivos 3 y 4 fueron proporcionados por los organizadores con el propósito de que los participantes pudiesen ver la estructura que tendrían los archivos finales (5 y 6) y así poder comenzar a diseñar sus algoritmos. Finalmente, estos conjuntos de datos fueron marcados como obsoletos por la organización cuando los archivos nuevos 5 y 6 se publicaron. Esto fue debido a que fueron generados con patrones de demanda diferentes y, por lo tanto, se dio la recomendación de no utilizarlos durante el desarrollo de la competición. En este proyecto se ha seguido esta recomendación y solo se han utilizado los tres últimos ficheros de datos, aunque como ya se verá más adelante, la solución propuesta funciona relativamente bien con ellos.

Así pues, los archivos de partida quedan tal y como indica la Tabla 3.2.

ID	Nombre del archivo	Formato	Descripción
1	rules.pdf	PDF	Enunciado y reglas del concurso
2	CTOWN.INP	EPANET INP	Modelo de la red en EPANET
3	BATADAL_dataset03.csv	CSV	Datos sin ataques
4	BATADAL_dataset04.csv	CSV	Datos con ataques identificados
5	BATADAL_dataset05.csv	CSV	Datos con ataques sin identificar

Tabla 3.2: Archivos de partida para el caso BATADAL

Además, debido a que el concurso ya ha finalizado, también se ha tenido acceso a los dos archivos que contienen la información de los ataques que existen en los archivos 6 y 7. En estos archivos se detallan las fechas de los diferentes ataques y se realiza una breve descripción de cada uno de los incidentes.

Respecto a los archivos CSV, entrando más en detalle, incluyen las mediciones realizadas en la ya mencionada red de EPANET. En particular, se han incluido los niveles de los tanques, las presiones en diferentes nudos y los caudales que circulan por las diferentes bombas, entre otros. Los ficheros siguen una estructura claramente definida. La primera fila se denomina cabecera y contiene los nombres o etiquetas de cada una de las variables. El resto de filas son mediciones. Cada columna representa una variable. Así tenemos un total de 45 columnas que se corresponden con las 45 variables diferentes. Cabe destacar la primera columna *DATETIME*, donde se muestra la fecha y hora en la que se ha realizado cada medición mediante un formato dd/mm/aaaa HH, y la última *ATT_FLAG* donde se indica si en una determinada muestra se está o no bajo un incidente. El listado completo de las variables junto con sus correspondientes descripciones se muestra en la Tabla 3.3.

Variable	Etiqueta	Descripción	Unidades	Estación	Dato
1	DATETIME	Fecha de muestra	dd/mm/aaaa hh	-	String
2	L_T1	Nivel del tanque T1	m	PLC2	Float
3	L_T2	Nivel del tanque T2	m	PLC3	Float
4	L_T3	Nivel del tanque T3	m	PLC4	Float
5	L_T4	Nivel del tanque T4	m	PLC6	Float
6	L_T5	Nivel del tanque T5	m	PLC7	Float
7	L_T6	Nivel del tanque T6	m	PLC8	Float
8	L_T7	Nivel del tanque T7	m	PLC9	Float
9	F_PU1	Caudal en bomba PU1	L/s	PLC1	Float
10	S_PU1	Estado de bomba PU1	-	PLC1	Boolean
11	F_PU2	Caudal en bomba PU2	L/s	PLC1	Float
12	S_PU2	Estado de bomba PU2	-	PLC1	Boolean
13	F_PU3	Caudal en bomba PU3	L/s	PLC1	Float
14	S_PU3	Estado de bomba PU3	-	PLC1	Boolean
15	F_PU4	Caudal en bomba PU4	L/s	PLC3	Float
16	S_PU4	Estado de bomba PU4	-	PLC3	Boolean
17	F_PU5	Caudal en bomba PU5	L/s	PLC3	Float
18	S_PU5	Estado de bomba PU5	-	PLC3	Boolean
19	F_PU6	Caudal en bomba PU6	L/s	PLC3	Float
20	S_PU6	Estado de bomba PU6	-	PLC3	Boolean
21	F_PU7	Caudal en bomba PU7	L/s	PLC3	Float
22	S_PU7	Estado de bomba PU7	-	PLC3	Boolean
23	F_PU8	Caudal en bomba PU8	L/s	PLC5	Float
24	S_PU8	Estado de bomba PU8	-	PLC5	Boolean
25	F_PU9	Caudal en bomba PU9	L/s	PLC5	Float
26	S_PU9	Estado de bomba PU9	-	PLC5	Boolean
27	F_PU10	Caudal en bomba PU10	L/s	PLC5	Float
28	S_PU10	Estado de bomba PU10	-	PLC5	Boolean
29	F_PU11	Caudal en bomba PU11	L/s	PLC5	Float
30	S_PU11	Estado de bomba PU11	-	PLC5	Boolean
31	F_V2	Caudal en válvula V2	L/s	PLC3	Float
32	S_V2	Estado de válvula V2	-	PLC3	Boolean
33	P_J280	Presión en nudo J280	m	PLC1	Float
34	P_J269	Presión en nudo J269	m	PLC1	Float
35	P_J300	Presión en nudo J300	m	PLC3	Float
36	P_J256	Presión en nudo J256	m	PLC3	Float
37	P_J289	Presión en nudo J289	m	PLC3	Float
38	P_J415	Presión en nudo J415	m	PLC3	Float
39	P_J302	Presión en nudo J302	m	PLC5	Float
40	P_J306	Presión en nudo J306	m	PLC5	Float
41	P_J307	Presión en nudo J307	m	PLC5	Float
42	P_J317	Presión en nudo J317	m	PLC5	Float
43	P_J14	Presión en nudo J14	m	PLC3	Float
44	P_J422	Presión en nudo J422	m	PLC3	Float
45	ATT_FLAG	Aviso de ciberataque	-	-	Boolean

Tabla 3.3: Variables de la red de distribución de agua

En la Figura 3.5 se puede ver el formato que tienen estos ficheros.

	A	B	C	D	E	F	G	H	I	J	K
	DATETIME	**L_T1**	**L_T2**	**L_T3**	**L_T4**	**L_T5**	**L_T6**	**L_T7**	**F_PU1**	**S_PU1**	**F_PU2**
	Text	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number
1	DATETIME	L_T1	L_T2	L_T3	L_T4	L_T5	L_T6	L_T7	F_PU1	S_PU1	F_PU2
2	06/01/14 00	0.509729922	2.049002886	3.191145182	2.792634249	2.656090975	5.316830635	1.562320828	98.9984436	1	99.01815033
3	06/01/14 01	0.41258049	2.009071827	3.642565489	2.831672668	3.126387358	5.494855404	1.852042913	99.09590149	1	99.11563873
4	06/01/14 02	0.320111841	1.986092925	4.140191555	3.256733179	3.574600697	5.5	2.246126175	98.42095947	1	98.44049835
5	06/01/14 03	0.332878858	2.009203434	4.673478127	3.744497061	3.952379227	5.5	3.203572989	97.57517242	1	97.59445953
6	06/01/14 04	0.483495772	2.089049101	5.237936974	4.409456253	3.504675627	5.5	4.439714432	97.35105896	1	97.3702774
7	06/01/14 05	0.791114032	2.77317667	5.15580225	3.937262058	3.19152832	5.322743416	3.988905907	94.13546753	1	94.15374756
8	06/01/14 06	1.186589003	3.536068201	4.983953476	3.018010616	2.859591007	5.066728115	2.977463007	95.25800323	1	95.27661133
9	06/01/14 07	1.420448899	3.872925758	4.747457504	3.581882238	2.359944105	5.152646065	2.95374155	96.94745636	1	96.96656036
10	06/01/14 08	1.534827471	4.138433933	4.417932034	3.959265471	1.748312831	5.395834923	3.228595972	96.97029114	1	96.98940277
11	06/01/14 09	1.576541185	4.500040054	4.130156517	4.232002258	1.666736722	5.5	3.628677845	97.15647125	1	97.17563629
12	06/01/14 10	1.558549762	4.962010384	3.66521287	2.962581635	2.107415676	5.5	3.445806503	97.8139801	1	97.83333588
13	06/01/14 11	1.480653882	5.09930563	3.267485142	2.984019995	2.51137495	5.5	3.180091143	97.17651367	1	97.19568634
14	06/01/14 12	1.464823008	5.165903568	3.014807701	3.339660406	2.993499279	5.5	3.408365726	96.73049164	1	96.74953461
15	06/01/14 13	1.483956218	4.995484352	3.487367868	3.579058409	3.326382637	5.5	3.785472631	98.11657715	1	98.13602448
16	06/01/14 14	1.383690715	4.80074358	4.016099453	3.867702961	3.713591337	5.5	3.423741341	97.75424194	1	97.77357483
17	06/01/14 15	1.351577759	4.622988701	4.467820644	3.870457172	3.834753752	5.5	3.306835175	97.2080307	1	97.227211

Figura 3.5: Visualización del archivo BATADAL_dataset03.csv

Finalmente, tanto en la Figura 3.3 como en Tabla 3.3 se puede observar que las variables de las que se han obtenido los datos están principalmente asociadas a tres estaciones de control gobernadas por su respectivos PLC; PLC1, PLC3 y PLC5. Esto es un dato a tener en cuenta cuando se realice la identificación de los comportamientos anómalos debido a que posiblemente las causas se encuentren en los PLC o en sus conexiones.

En la sección siguiente se ha procedido a realizar la descripción del caso secundario, Caso Cranfield.

3.2 Caso Cranfield

En *Ruiz-Cárcel* [13] se expuso un estudio sobre la eficacia de una metodología basada en datos llamada Análisis Canónico de Variables (CVA, Canonical Variate Analysis) para la monitorización de procesos con condiciones de operación variables. Mientras que otros estudios se basaban en datos de procesos simulados por ordenador como por ejemplo el benchmark de Tennessee Eastman Process Plant [17], el trabajo de Ruiz-Cárcel ha tenido como objetivo proporcionar una serie de conjuntos de datos de una instalación real que sirvan como benchmark para la puesta a prueba de diferentes técnicas de monitorización de procesos, estando especialmente indicado para la detección y diagnóstico de fallos. En dicha instalación, que se encuentra en la Universidad de Cranfield, se ha llevado a cabo un proceso multifase donde se mezclan distintos fluidos para luego proceder a su separación. Los datos de los diferentes ensayos realizados en el estudio han sido publicados para su uso en el sitio web MATLAB Central[1]. Se procede pues a realizar una detallada explicación de la instalación y de los diferentes datasets obtenidos en los ensayos.

3.2.1 Descripción de la instalación

La instalación de flujo trifásico situada en la Universidad de Cranfield está diseñada para suministrar bajo control un determinado caudal de una mezcla de agua, aire y aceite a un sistema presurizado. En cierta medida trata de simular el proceso de extracción de petróleo. La Figura 3.6 muestra un esquema de la instalación. La instalación tiene dos partes claramente diferenciadas: zona de generación de la mezcla y zona de separación.

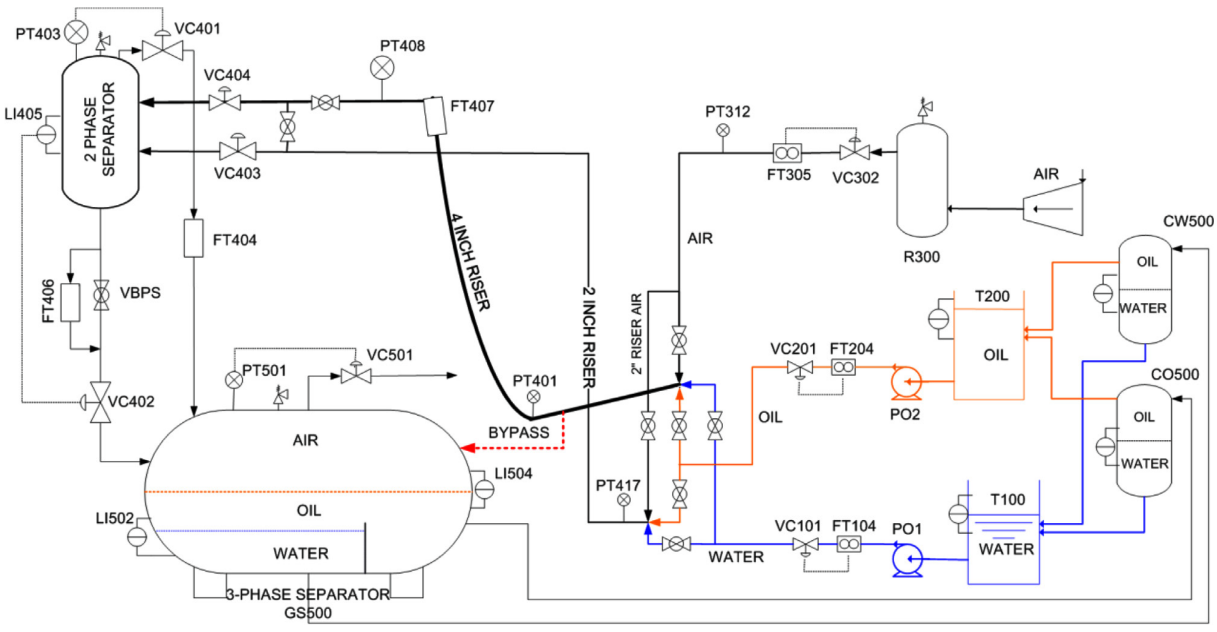

Figura 3.6: Instalación de Cranfield [13]

La zona de generación tiene todo lo necesario para medir y controlar todas las propiedades de la mezcla que se genera. La admisión de aire se realiza directamente de la atmósfera. Aunque en la figura solo se muestra uno, el aire es captado a través de dos compresores y es introducido en un

depósito presurizado (R300) para poder moderar las variaciones de presión generada por éstos. A la salida del depósito, tras ser filtrado para garantizar la completa eliminación de las partículas en suspensión, el aire pasa por una válvula neumática (VC302) y un medidor de caudal (FT305) con los que se puede regular fácilmente su flujo. En cambio, el agua y el aceite se mueven por un circuito cerrado. Ambos siguen el mismo esquema, son almacenados en sus tanques (T100 y T200) y se suministran independientemente a través sus respectivas bombas (PO1 y PO2). El caudal de agua es medido por el sensor FT104, el de aceite por el FT204 y sus respectivos caudales son controlados mediante las válvulas VC101 y VC201.

En el riser, una tubería de mayor diámetro que asciende hasta la altura del primer separador, es donde se realiza la mezcla de los tres fluidos. A partir de aquí ya nos encontramos en la zona de separación. Ésta consta de dos separadores, uno bifásico y otro trifásico en donde se consigue obtener los tres fluidos iniciales. El aire es emitido a la atmósfera mediante la válvula VC501 mientras que el aceite y el agua retornan a los tanques iniciales previo paso por dos decantadores que aseguren la total separación de los fluidos (CW500 y CO500).

3.2.2 Descripción de los datos suministrados

Como ya se ha mencionado antes, los datos de los ensayos se encuentran alojados en Internet. Tras descargar los archivos de MATLAB Central solo se han mantenido aquellos que resultan de interés para este trabajo, las bases de datos. Se parte de un total de 7 bases de datos (Tabla 3.4). Cada una de ellas contiene los diferentes ensayos realizados en la instalación.

ID	Nombre del archivo	Formato	Descripción
1	FaultyCase1.mat	MAT	Tablas con datos anómalos
2	FaultyCase2.mat	MAT	Tablas con datos anómalos
3	FaultyCase3.mat	MAT	Tablas con datos anómalos
4	FaultyCase4.mat	MAT	Tablas con datos anómalos
5	FaultyCase5.mat	MAT	Tablas con datos anómalos
6	FaultyCase6.mat	MAT	Tablas con datos anómalos
7	Training.mat	MAT	Datos de entrenamiento

Tabla 3.4: Archivos de partida para el caso Cranfield

En todos los ensayos, los datos han sido tomados con una frecuencia de muestreo de 1Hz. Las variables de la instalación que han sido medidas suman un total de 24 y pueden verse en la Tabla 3.5. En total se han realizado tres pruebas en condiciones nominales y 16 donde, deliberadamente, se han producido fallos con la introducción de perturbaciones en el sistema. Aunque las perturbaciones se han producido manualmente, responden a comportamientos que podrían darse en caso de que la instalación fuese controlada remotamente y sufriera un ciberataque como ocurre en el caso BATADAL. Para la realización de dichas pruebas solamente se ha utilizado aire y agua.

ID	Localización	Descripción	Unidades
1	PT312	Presión de aire suministrado	MPa
2	PT401	Presión abajo del riser	MPa
3	PT408	Presión arriba del riser	MPa
4	PT403	Presión arriba del separador	MPa
5	PT501	Presión del separador trifásico	MPa
6	PT408	Presión diferencial (PT401-PT408)	MPa
7	PT403	Presión diferencial en VC404	MPa
8	FT305	Caudal de entrada de aire	Sm3/s
9	FT104	Caudal másico de entrada de agua	kg/s
10	FT407	Caudal másico arriba del riser	kg/s
11	LI405	Nivel del separador bifásico	m
12	FT406	Caudal de salida del sep. bifásico	kg/s
13	FT407	Densidad arriba del riser	kg/m3
14	FT406	Densidad salida separador bifásico	kg/m3
15	FT104	Densidad de entrada del agua	kg/m3
16	FT407	Temperatura arriba del riser	ºC
17	FT406	Temperatura salida separador bifásico	ºC
18	FT104	Temperatura de entrada de agua	ºC
19	LI504	Nivel separador trifásico	%
20	VC501	Posición de válvula VC501	%
21	VC302	Posición de válvula VC302	%
22	VC101	Posición de válvula VC101	%
23	PO1	Intensidad bomba de agua	A
24	PT417	Presión en la zona de mezcla 2"	MPa

Tabla 3.5: Variables medidas de la instalación

Entrenamiento

Siguiendo el orden del caso BATADAL comenzaremos la explicación por el archivo *Training.mat*. En dicho fichero se tienen tres conjuntos de datos (T1, T2 y T3) que representan el funcionamiento del sistema en condiciones normales en tres ensayos diferentes. Conjuntos que, para abreviar, llamaremos de entrenamiento de ahora en adelante. Durante las pruebas en las que se han tomado las mediciones, los puntos de funcionamiento del sistema han sido deliberadamente alterados para obtener datos del proceso trabajando bajo diferentes condiciones de operación. El objetivo que se persigue con ello es asegurar que las condiciones han sido representativas de un funcionamiento nominal. Se han probado en cada uno de estos ensayos hasta 20 combinaciones diferentes de entrada de aire y agua. También se han realizado incrementos y decrementos a diferentes velocidades de sus caudales másicos de entrada (Figura 3.7) con el objetivo de tener una mayor variedad de puntos de operación y así, poder asegurar que la dinámica del sistema queda reflejada en los datos. Los puntos de operación han sido seleccionados con el fin de abarcar todo el rango de funcionamiento de la instalación, aunque siempre tratando de evitar los caudales excesivamente bajos que puedan causar "slugging" en la tubería por la que asciende la mezcla de fluidos y que lleva al primer separador.

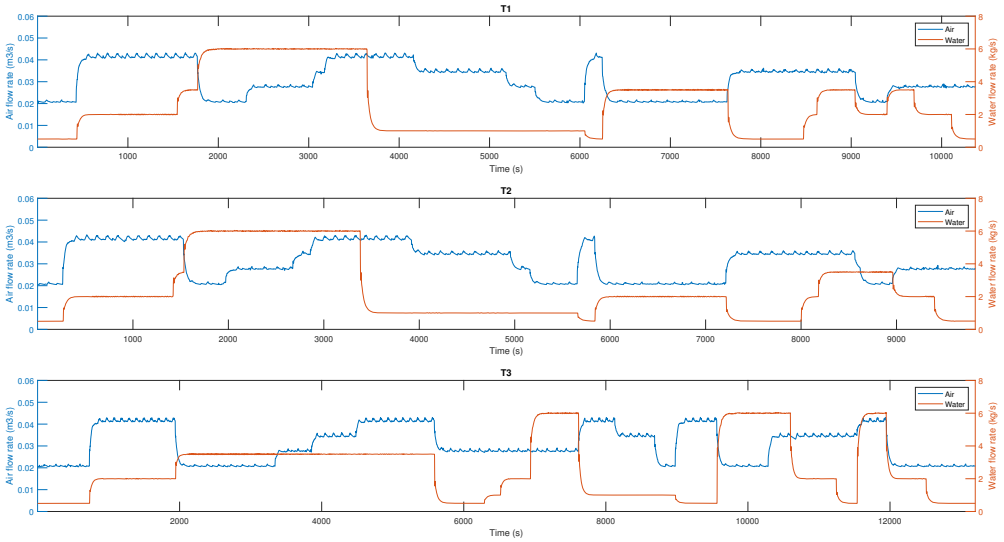

Figura 3.7: Caudales de entrada

A continuación se han descrito los 6 tipos de fallos que se han ensayado en la instalación y que se corresponden con cada uno de los datasets enumerados en la Tabla 3.4

Fallo 1: bloqueo entrada de aire

Estrangulamiento parcial de la válvula manual que controla la cantidad de aire que se incluirá en la mezcla de fluidos. Para este caso se han incluido tres conjuntos de datos. Uno con condiciones de operación cambiantes, y los otros dos con un suministro de aire y agua constante.

Fallo 2: bloqueo entrada de agua

Estrangulamiento parcial de la válvula manual que controla la cantidad de agua que se incluirá en la mezcla de fluidos. Aún siendo similar al Fallo 1 las consecuencias son diferentes debido a las diferentes propiedades del agua respecto a las del aire. Para este caso se han incluido tres conjuntos de datos. Al igual que en el anterior caso, uno con condiciones de operación cambiantes, y los otros dos con un suministro de aire y agua constante.

Fallo 3: bloqueo entrada de separador bifásico

Bloqueo en la válvula VC404, en la entrada del primer separador. Para este caso se han incluido tres conjuntos de datos. Uno con condiciones de operación cambiantes, y los otros dos con un suministro de aire y agua constante.

Fallo 4: conducto bypass abierto

Durante el cuarto ensayo se ha simulado una fuga en la parte inferior del riser mediante la apertura de una válvula que extrae parte del fluido en esa zona enviándolo directamente al segundo separador. Para este caso se han incluido tres conjuntos de datos. Uno con condiciones de operación cambiantes, y los otros dos con un suministro de aire y agua constante.

Fallo 5: condiciones de slugging

El Fallo 5 se da cuando con las condiciones apropiadas el gas de una mezcla asciende mientras el líquido de la mezcla se acumula en la parte baja en una tubería de ascenso. Para ello se ha reducido la velocidad de circulación de los fluidos mediante la estrangulación de las válvulas pertinentes. Para el primer set el fallo por slugging fue introducido y quitado dos veces. En cambio, en el tercer set el slugging se produjo tres veces.

Fallo 6: presurización del riser 2

En los casos anteriores la tubería que se ha utilizado para llevar el fluido a la entrada del primer separador ha sido la de 4". La de 2' estaba totalmente aislada del sistema. En el Fallo 6, dicha tubería ya no se encuentra aislada del sistema debido a la apertura en la válvula que conecta los dos riser, justo antes de las válvulas VC403 y VC403. Así dicha tubería queda presurizada causando el fallo de la instalación. En este caso se han generado dos datasets, ambos bajo condiciones de operación variables.

Una vez explicados los dos casos que se han tratado en este proyecto se puede pasar al siguiente capítulo donde se ha detallado la solución que se ha implementado.

Capítulo 4

Solución propuesta

Como se ha detallado en el capítulo anterior, tanto en el caso BATADAL como en el caso de la instalación de flujo multifase de la Universidad de Cranfield se han proporcionado distintos conjuntos de datos en los que se ve reflejado el comportamiento de cada instalación. Ambos casos proporcionan datos del funcionamiento normal de la instalación y otros datos que incluyen comportamientos anómalos, ya sean los ciberataques en el caso BATADAL o la introducción de perturbaciones en el caso de la Universidad de Cranfield. De ahí que se haya propuesto como solución el desarrollo de una aplicación que, partiendo de un histórico de datos obtenidos durante su funcionamiento normal de una instalación cualquiera, sea capaz de realizar una monitorización activa de la propia y una detección en línea de los diferentes fallos, comportamientos anómalos y ciberataques que ocurran.

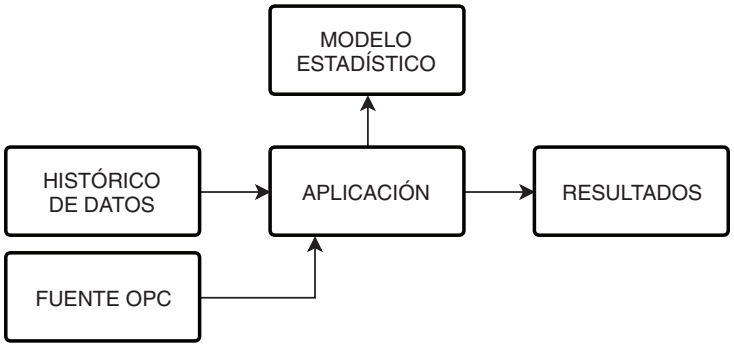

Figura 4.1: Diagrama general

Para ello se han utilizado diferentes técnicas matemáticas relacionadas con el Control Estadístico de Procesos (SPC). Concretamente, se han utilizado las técnicas de monitorización estadística multivariante (MSPC) basadas en el Análisis de Componentes Principales (PCA) y los gráficos de control Hotelling's (T^2) Control Chart y Squared Prediction Error (SPE) Chart.

Para el diseño de los algoritmos se ha utilizado el software MATLAB ya que su lenguaje permite al usuario una rápida implementación de los cálculos matriciales y su interfaz de usuario es amigable facilitando la visualización de los resultados. En cuanto a la implementación final de la aplicación, se ha sido utilizado el software LabVIEW ya que favorece el desarrollo de la interfaz de usuario y la conexión con fuentes externas de información.

Por último, la simulación de una fuente de datos se ha realizado utilizando el estándar de comunicación OPC (OLE for Process Control) basado en el modelo Cliente-Servidor, siendo el cliente la aplicación desarrollada en LabVIEW y el servidor el generado mediante el software MatrikonOPC Server for Simulation.

En los apartados siguientes se ha procedido a explicar detalladamente cómo llevar a cabo cada uno de los elementos que componen la solución propuesta. La explicación se ha dividido en varias partes:

1. Análisis previo
2. Diseño del algoritmo
3. Implementación de la aplicación
4. Simulación de fuente de datos con OPC

En la primera, se ha realizado un análisis de los datos de partida para su posterior uso. En la segunda parte se ha explicado paso por paso el funcionamiento del algoritmo, tanto durante la generación del modelo matemático (etapa de entrenamiento) como en la ejecución del modo online (etapa de monitorización en línea). En la tercera, se ha explicado cómo se ha llevado a cabo la implementación final de la aplicación, la interfaz de usuario y cómo funciona internamente ésta. En la última se ha detallado cómo se ha llevado a cabo la implementación de una fuente de datos OPC para simular la obtención de los datos de una instalación.

4.1 Análisis previo

Antes de comenzar con el diseño e implementación de la aplicación se debe realizar un análisis previo de los diferentes conjuntos de datos que se tienen. Se ha realizado una representación gráfica de cada variable para detectar en caso de que existan puntos anómalos que posteriormente puedan afectar al entrenamiento de los algoritmos y correlaciones temporales.

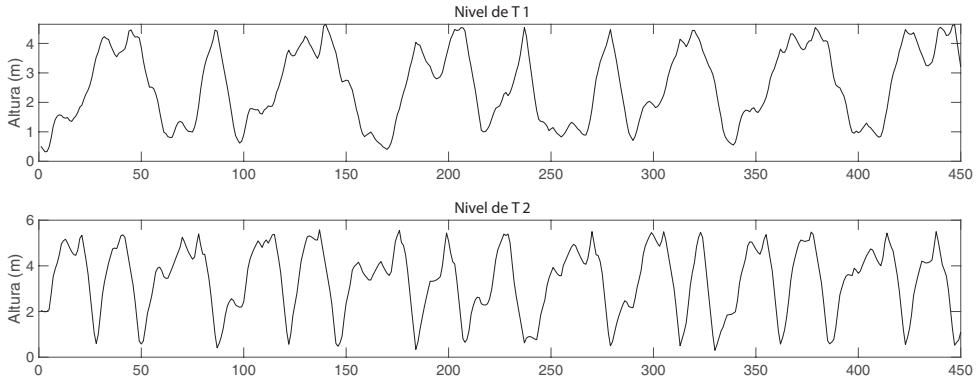

Figura 4.2: Patrones temporales

En la Figura 4.2 se puede observar que algunas variables tienen correlación temporal. Esto significa que en este caso sería más adecuado realizar un Análisis de Componentes Principales Dinámico (DPCA) aunque queda fuera del alcance del proyecto y se realizará un PCA normal.

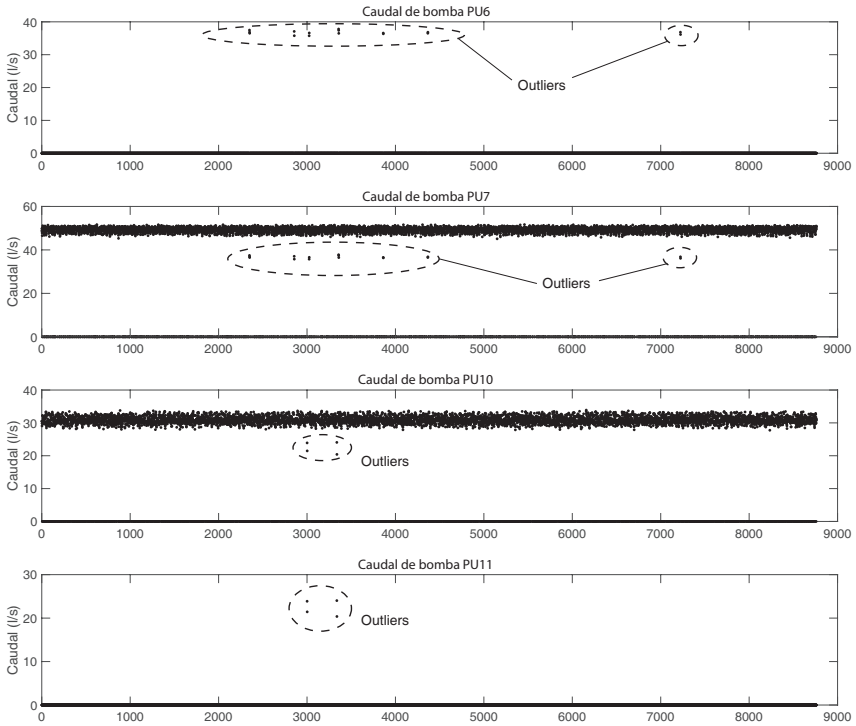

Figura 4.3: Outliers

En la Figura 4.3 se puede observar algunas variables del set de entrenamiento presentan puntos anómalos (outliers). Estos puntos deberán ser eliminados antes de la utilización de dicho conjunto de datos en el entrenamiento del modelo matemático. También se han graficado variables derivadas de las variables originales para facilitar la localización de otros puntos anómalos. En la

Figura 4.4 se puede observar como algunos puntos caen fuera de lo que sería el rango de operación habitual del conjunto de las dos bombas en paralelo.

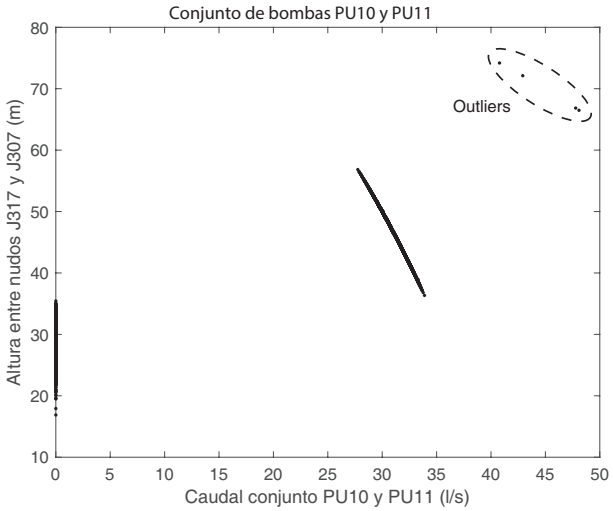

Figura 4.4: Outliers 2

Se ha procedido a representar los histogramas de cada una de las variables para verificar si las distribuciones probabilidad que siguen las variables son distribuciones normales. Tras realizar su representación se ha comprobado que la distribución que más se aproxima a las variables es la distribución normal. Esto significa que es posible utilizar los gráficos de control T^2 y SPE ya que una de las condiciones es que el conjunto de datos siga una distribución normal multivariante, es decir, que cada variable del conjunto siga la distribución de Gauss.

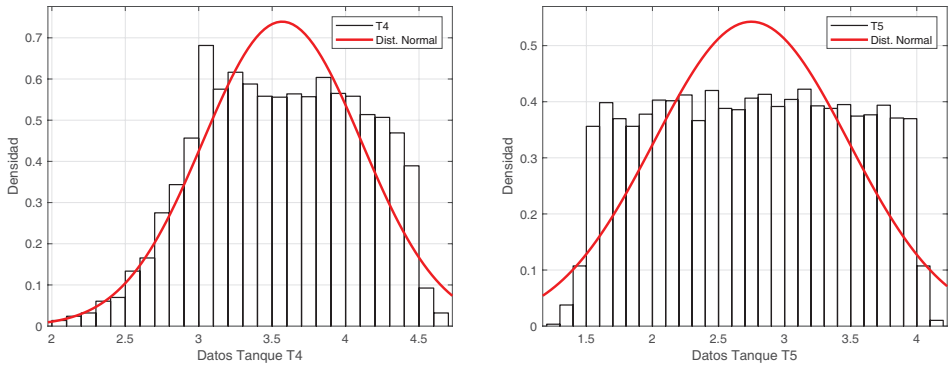

Figura 4.5: Distribuciones de las variables T4 y T5

Finalmente, el único archivo que ha sido modificado para su posterior utilización es el archivo de entrenamiento BATADAL_dataset03.csv, del cual se han eliminado los puntos marcados en las Figuras 4.3 y 4.4. Al conjunto que queda tras realizar esta operación se le ha llamado MOD_dataset03.csv y ya se encuentra totalmente libre de medidas anómalas. Con ello ya se puede proceder a su utilización en el algoritmo de entrenamiento.

4.2 Diseño del algoritmo

Tal y como se ha comentado anteriormente, para poder llevar a cabo de manera correcta la detección de los fallos y ciberataques se ha tenido que dividir la ejecución del algoritmo principal en dos etapas: la etapa de entrenamiento y la de monitorización en linea. En la Figura 4.6 se puede observar un diagrama en el que se ha resumido de manera concisa los pasos que el algoritmo sigue.

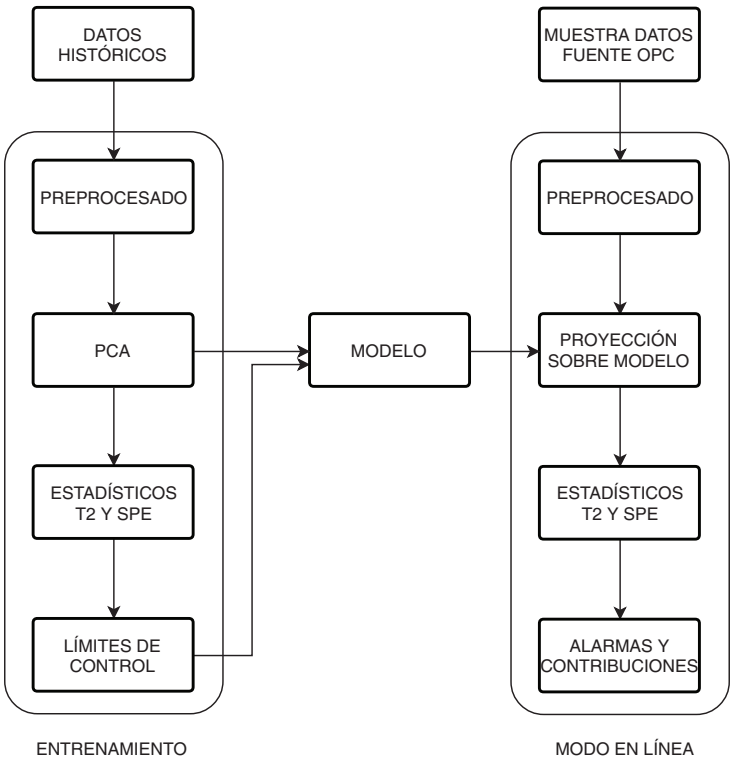

Figura 4.6: Diagrama del algoritmo

Seguidamente, se ha detallado cada uno de los bloques del diagrama anterior.

4.2.1 Entrenamiento

Durante la etapa de entrenamiento se genera un modelo matemático a partir de un conjunto histórico de datos. El modelo matemático representa el comportamiento del sistema y se utilizará en el Modo en Línea para saber si el sistema está o no bajo normales de funcionamiento, si hay o no ciberataques.

Como se puede observar en el diagrama anterior el punto clave del algoritmo radica en el cálculo de los estadísticos $T2$ y SPE y de sus respectivos límites. Para poder proceder con el cálculo de dichos estadísticos, primero se deben realizar una serie de operaciones previas sobre los datos proporcionados. Estas operaciones son dos: el preprocesado y el Análisis de Componentes Principales.

Preprocesado

Los ficheros de partida con formato CSV contienen los datos históricos en bruto del funcionamiento del sistema. Habitualmente, la información contenida en los ficheros en bruto no puede ser utilizada directamente en los pasos posteriores. Hay que realizar un tratamiento sobre ella. Mediante el preprocesado se puede determinar la estructura que siguen los datos y así mantener la parte de estos que resulte interesante para el trabajo. Así se pueden eliminar variables irrelevantes (columnas) o muestras (filas) que contengan valores anómalos. En el caso BATADAL la columna DATETIME y la columna ATT_FLAG no resultan de interés y por lo tanto son eliminadas, en cambio en el caso Cranfield no hay ninguna variable que resulte irrelevante y por lo tanto no se ha eliminado ninguna de ellas.

Durante el preprocesado también podemos realizar lo que en estadística se conoce como el proceso de estandarización. Éste no es más que un centrado y escalado de cada una de las variables. Otros nombres por los que se le conoce a esta técnica es normalización o Z-Score. Para llevar a cabo su cálculo simplemente hay que aplicar a cada columna del set de datos la siguiente fórmula:

$$z = \frac{x - \mu}{\sigma} \tag{4.1}$$

Donde x es el valor de cada variable, μ es la media de dicha variable, y σ su desviación típica. Al realizar esta operación se consigue que todas las variables tengan media nula y desviación típica unitaria y, por lo tanto, también varianza unitaria. Con ello se evita que posteriormente unas variables tengan mayor peso respecto de otras del que deberían, por ejemplo al estar expresadas diferentes unidades.

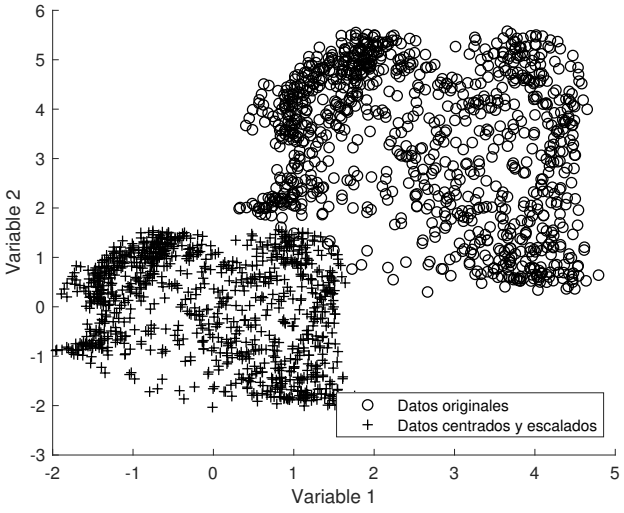

Figura 4.7: Datos centrados y escalados frente a los originales

En la Figura 4.7 se puede observar el antes y el después de aplicar el centrado y escalado a un pequeño set de datos. Una vez centrados, todos los datos se encuentran entorno al origen de coordenadas y se muestran con una escala diferente.

Una vez hecho el preprocesado podemos pasar a la siguiente etapa: el PCA.

Análisis de Componentes Principales

El objetivo principal del PCA es reducir la dimensionalidad del set de datos. Es decir, reducir el número de variables con el que se está trabajando perdiendo la mínima información por el camino. Así, si partimos de un número cualquiera de variables, tras ejecutar el PCA podemos reducir dicho número conservando un porcentaje de la información original próximo al 100 %. Con ello se persigue reducir el coste computacional del algoritmo en los pasos posteriores.

Si se ha partido de una matriz de datos centrados y escalados Y, el PCA hace una descomposición tal que:

$$Y = TP^T \tag{4.2}$$

Siendo P una matriz de cambio de espacio que contiene los vectores propios que definen las direcciones principales de Y, y T el valor que resulta de proyectar las variables del espacio original sobre las direcciones principales. Hay que tener en cuenta que los autovectores de la matriz P están ordenados de mayor a menor según su valor propio asociado, y por lo tanto, se pueden eliminar aquellas componentes que tengan asociados los valores propios más pequeños. Si se realiza esta operación tenemos lo siguiente:

$$\hat{Y} = T\hat{P}^T \tag{4.3}$$

Donde $\hat{Y}$ se calcula con la matriz P reducida. Dicho cálculo conlleva una pérdida de información que se ve reflejada en el residuo. Finalmente, podemos obtener la matriz del residuo E a partir de las anteriores:

$$E = Y - \hat{Y} \tag{4.4}$$

El número de componentes principales que se elijan para crear $\hat{P}$ determinará el número de variables que tendrá nuestro modelo matemático. Existen diferentes criterios para una correcta elección de dicho número. En este proyecto se hace referencia a los siguientes:

1. Valores propios mayores a la unidad. Este criterio se basa en elegir aquellas direcciones principales en las que su autovalor asociado sea mayor que 1.
2. Porcentaje de variabilidad explicada. Se basa en elegir tantas componentes como sea necesario para que el porcentaje de variabilidad explicada acumulado sea mayor que un determinado valor.
3. Basado en el gráfico de sedimentación. Este método se basa en realizar una elección a estima del número de componentes mediante la visualización del diagrama Scree Plot, donde se muestra como varía el porcentaje resultante de variabilidad explicada en función del número de componentes que se eligen. En la Figura 4.8 se muestra dicho gráfico, pero esta vez representando los valores propios en vez de sus respectivos porcentajes. Un forma igualmente válida.

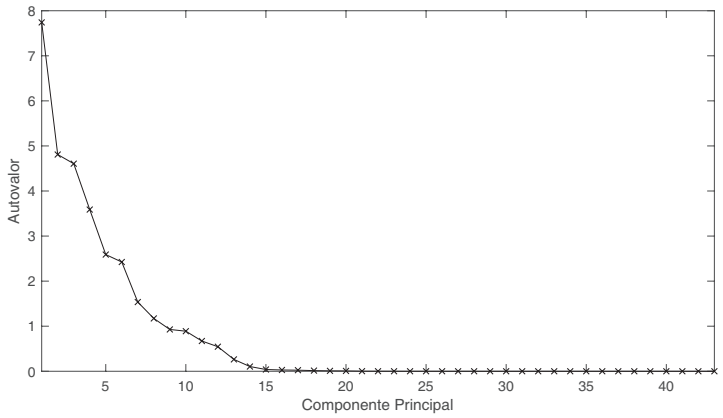

Figura 4.8: Gráfico de sedimentación o Scree Plot

Como se puede observar los tres criterios son arbitrarios. Existen otros que no lo son. Las técnicas basadas en la validación cruzada son una buena opción a considerar, pero debido a la complejidad de su implementación quedan fuera del alcance de este trabajo.

Cálculo de T_A^2 y SPE

Una vez ha sido realizada la reducción de la dimensionalidad mediante el PCA se puede proceder al cálculo de los estadísticos T_A^2 y SPE. Se calculan con los datos históricos ya preprocesados y proyectados sobre el nuevo subespacio que ha sido generado a partir de las componentes principales elegidas por uno de los criterios anteriores.

El T^2 se calcula mediante la Fórmula 4.5:

$$T_A^2 = \sum_{i=1}^{A} t_i \lambda_i^{-1} t_i^T \tag{4.5}$$

Con λ_i los autovalores y t_i los valores de las variables en el espacio proyectado. Así el sumatorio llega hasta A componentes principales. Por otro lado, el SPE se calcula a partir de los residuos con la Fórmula 4.6:

$$SPE = e_i^T e_i = (y_i - \hat{y_i})^T (y_i - \hat{y_i}) \tag{4.6}$$

Siendo y_i el valor original de la variable e $\hat{y_i}$ el valor de la variable predicho mediante el modelo PCA.

Cálculo de los límites de control $UCL(T_A^2)_\alpha$ y $UCL(SPE)_\alpha$

Los límites de control no son más que los valores máximos o mínimos que pueden tomar los gráficos de control para considerar que el proceso en cuestión está dentro de un comportamiento normal. En este caso se realiza el cálculo de los límites superiores $UCL(T_A^2)_\alpha$ y $UCL(SPE)_\alpha$. No se calculan los límites inferiores ya que no tiene sentido cuando ambos estadísticos resultan de sumar valores que se han elevado al cuadrado previamente. Así, las formulas con las que se calculan ambos son las siguientes:

$$UCL(T_A^2)_\alpha = \frac{A(m^2-1)}{m(m-A)} F_{(A,(m-A)),\alpha} \tag{4.7}$$

$$UCL(SPE)_\alpha = \frac{K-A}{c^2} s_0^2 F_{(K-A,(m-A-1)(K-A)),\alpha} \tag{4.8}$$

Una vez se han calculado los límites de control $UCL(T_A^2)_\alpha$ y $UCL(SPE)_\alpha$ junto con la anterior matriz $\hat{P}$, ya tenemos el modelo estadístico generado para los datos del problema. Adicionalmente, resulta esencial guardar las medias y las desviaciones típicas calculadas durante el preprocesado para poder utilizarlas en el Modo en Línea. Con este modelo se podrán detectar anomalías en futuros datos, entre ellos los ciberataques.

4.2.2 Modo Online

Una vez se ha generado el modelo matemático a partir de los datos del entrenamiento se puede ejecutar el Modo en Línea de la aplicación. En este modo se simula la lectura de los datos a través de un servidor OPC, tal y como se realizaría en un proceso real. Cada cierto tiempo, los nuevos datos son leídos del servidor e introducidos en la aplicación como si de una nueva muestra se tratase. Esta nueva observación es computada por el modelo estadístico y arroja unos resultados. En la Figura 4.9 se muestra el diagrama de flujo que se sigue en la ejecución de este modo.

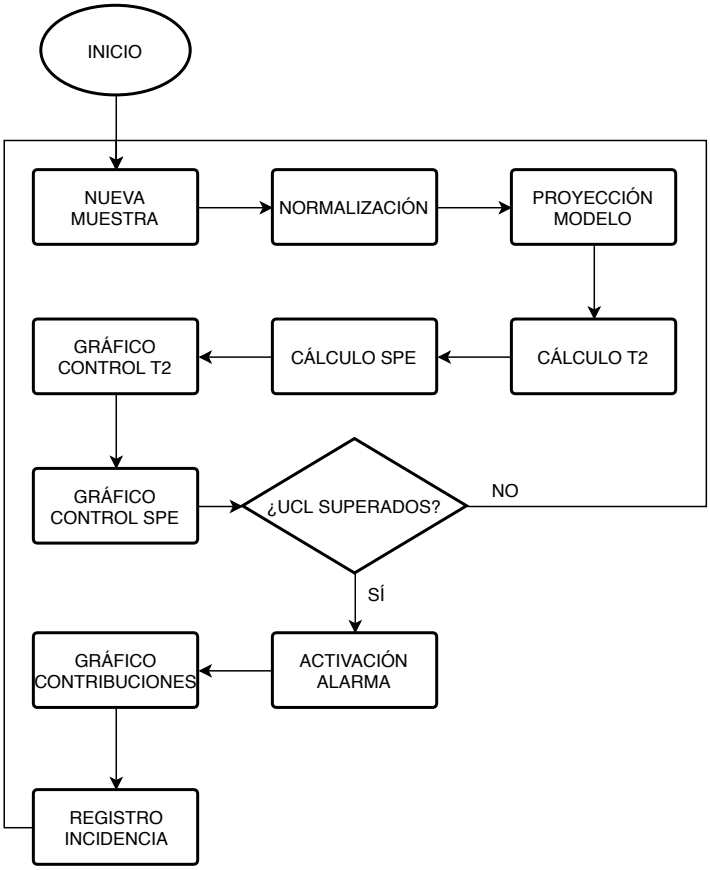

Figura 4.9: Diagrama de flujo del Modo Online

A continuación se ha procedido a una detallada explicación de cada uno de los pasos que se producen durante la ejecución del Modo Online.

Preprocesado

Los datos directamente proporcionados a través del servidor OPC son servidos en bruto, es decir, sin preprocesar. Hay que realizar dicho tratamiento. Por lo tanto, se calcula el valor Z-Score de la muestra que acaba de ser servida. Ello se realiza a partir de la media μ y de la desviación típica σ calculadas para el escalado y centrado de los datos en el preprocesado de la etapa de entrenamiento aplicando la Fórmula 4.1.

PCA

Una vez se ha realizado el preprocesado de la nueva observación se procede a su proyección sobre el modelo matemático PCA mediante la Fórmula 4.9.

$$t_i = y_i \hat{P} \tag{4.9}$$

Con ello se reduce la dimensionalidad de la observación y_i obteniendo los *scores* t_i que, como en apartados anteriores se ha mencionado, representan los valores de los datos originales en el nuevo subespacio generado por las direcciones principales.

Cálculo de T_A^2 y SPE

Con los anteriores *scores* t_i se puede realizar el cálculo de los estadísticos de control Hotelling's T-Squared (T^2) y Squared Prediction Error (SPE) mediante la Fórmula 4.5 y la Fórmula 4.6. Nótese que para proceder a su cálculo serán necesarios los valores propios de los autovectores que conforman la matriz $\hat{P}$.

Activación de alarmas

Una vez calculados T_A^2 y SPE se procede a comparar ambos valores con sus respectivos límites de control $UCL(T_A^2)_\alpha$ y $UCL(SPE)_\alpha$. En caso de que alguno de los dos supere a su respectivo límite de control se procede a la activación de una alarma (Figura 4.10).

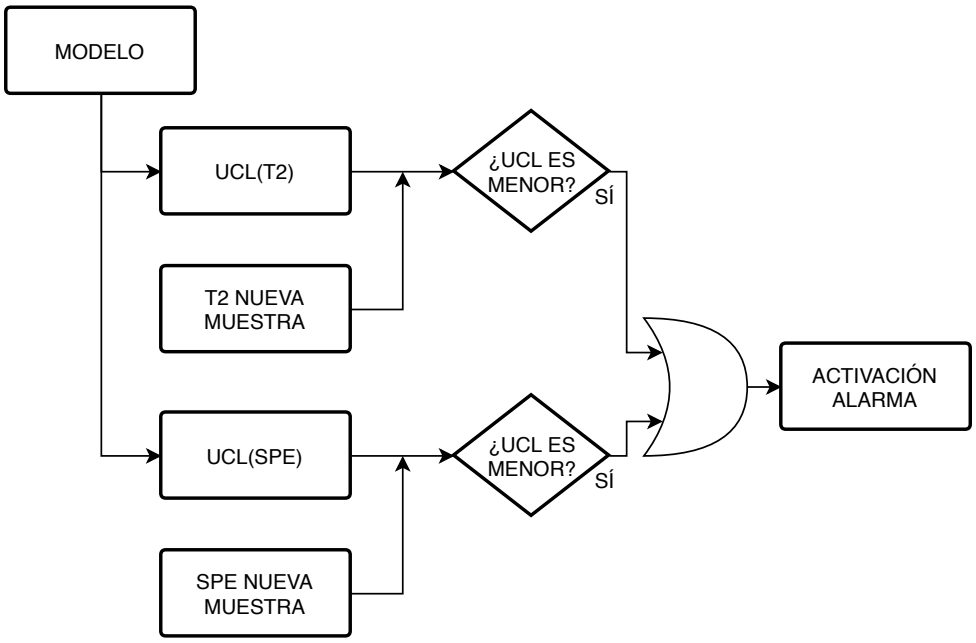

Figura 4.10: Lógica de la activación de alarmas

Cálculo del diagrama de contribuciones

Cuando se activa la alarma se activa un mecanismo de diagnóstico del fallo llamado diagrama de contribuciones. Se trata de un diagrama de barras donde se pueden ver las aportaciones de las variables del espacio original a los estadísticos T_A^2 y SPE. Luego a partir de estos dos gráficos de contribuciones (uno por cada estadístico) se puede deducir que variables son las que están relacionadas con el fallo que se ha producido. Por ejemplo, en la Figura 4.11 se puede observar como el valor de la variable número 8 destaca considerablemente sobre el resto en el diagrama de contribuciones del SPE. La muestra de datos utilizados para generar esta figura previamente han activado la alarma por superar el límite de control $UCL(SPE)_\alpha$. Luego se puede deducir que dicha variable está directamente relacionada con la causa que ha activado dicha alarma.

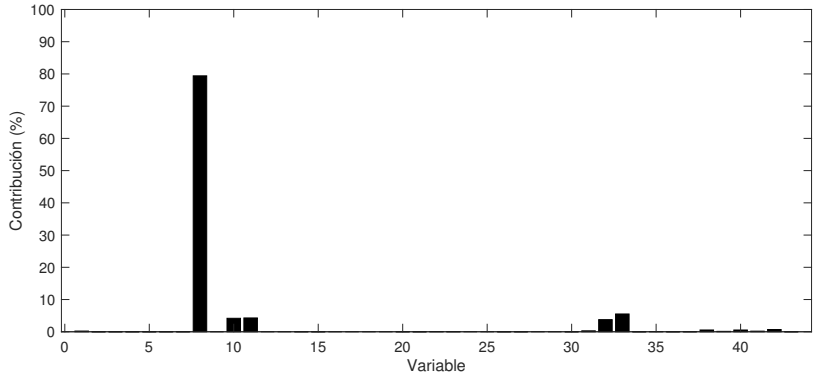

Figura 4.11: Contribuciones SPE

4.3 Implementación de la aplicación

La aplicación final se ha desarrollado utilizando el lenguaje visual gráfico G, lenguaje propietario de la empresa National Instruments y que es utilizado en el entorno de desarrollo integrado LabVIEW, también propiedad de la misma. Dicha aplicación sigue una arquitectura llamada Queued State Machine with event-driven Producer-Consumer. Una combinación que aúna los beneficios de las máquinas de estados genéricas y los de la arquitectura productor-consumidor, esta última derivada del patrón de diseño maestro-esclavo.

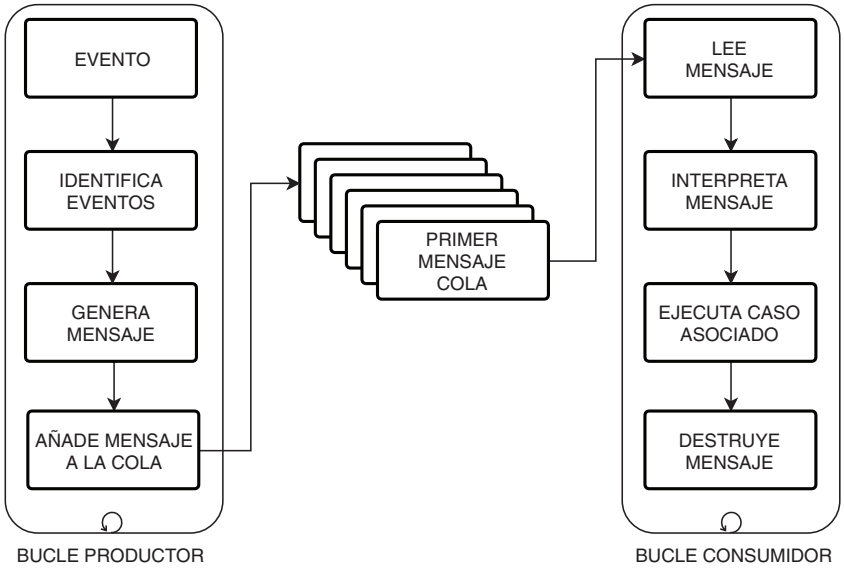

Figura 4.12: Diagrama de la arquitectura Productor - Consumidor

Tal y como muestra la Figura 4.12, por un lado tendremos un bucle principal que controlará los eventos que sucedan en la interfaz de usuario y las excepciones que se produzcan (bucle productor), y por otro, tendremos un bucle secundario (bucle consumidor) que se encargará de ejecutar los diferentes estados dependiendo de las condiciones que haya en ese momento. Esta comunicación entre ambos bucles se realiza mediante una cola de mensajes. Una cola no es más que un buffer donde se almacenan temporalmente los mensajes que se creen en el bucle productor. Estos mensajes se eliminan cuando el bucle consumidor los interpreta y ejecuta. En este trabajo, para facilitar la implementación de la aplicación se han utilizado dos bucles consumidores que dependen de un mismo bucle productor.

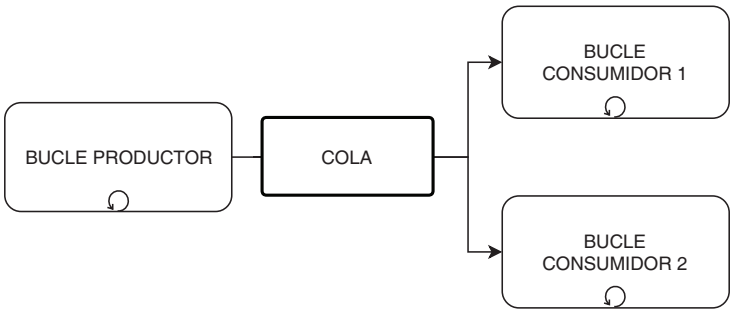

Figura 4.13: Diagrama de la arquitectura Productor - Doble Consumidor

En la Figura 4.13 se observan ambos bucles. Se ha decidido implementar este diseño porque de esta manera se consigue separar en dos hilos de ejecución diferentes la etapa de entrenamiento del modelo matemático y la etapa de ejecución del Modo Online. Así, se aumenta la legibilidad y la escalabilidad del código facilitando su depuración y la implementación de futuras funcionalidades.

En lo que se refiere a la interfaz de usuario se ha seguido el esquema de la Figura 4.14:

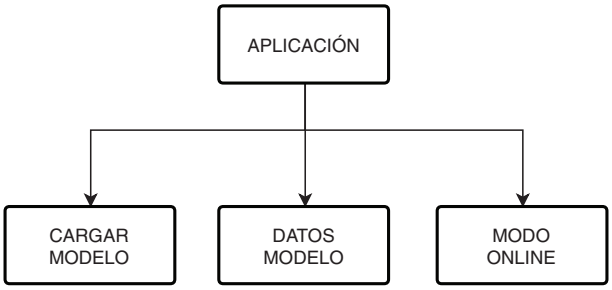

Figura 4.14: Diagrama interfaz de usuario

Cada uno de los bloques de la figura anterior se corresponde con una pestaña de la aplicación. A continuación se ha procedido a realizar una descripción de cada uno de ellos.

4.3.1 Cargar Modelo

Esta parte de la aplicación es la que ve el usuario al iniciarla y consta de dos subapartados: Configuración Inicial y Datos Cargados. En la Figura 4.15 se muestra una captura de pantalla de dicha pestaña.

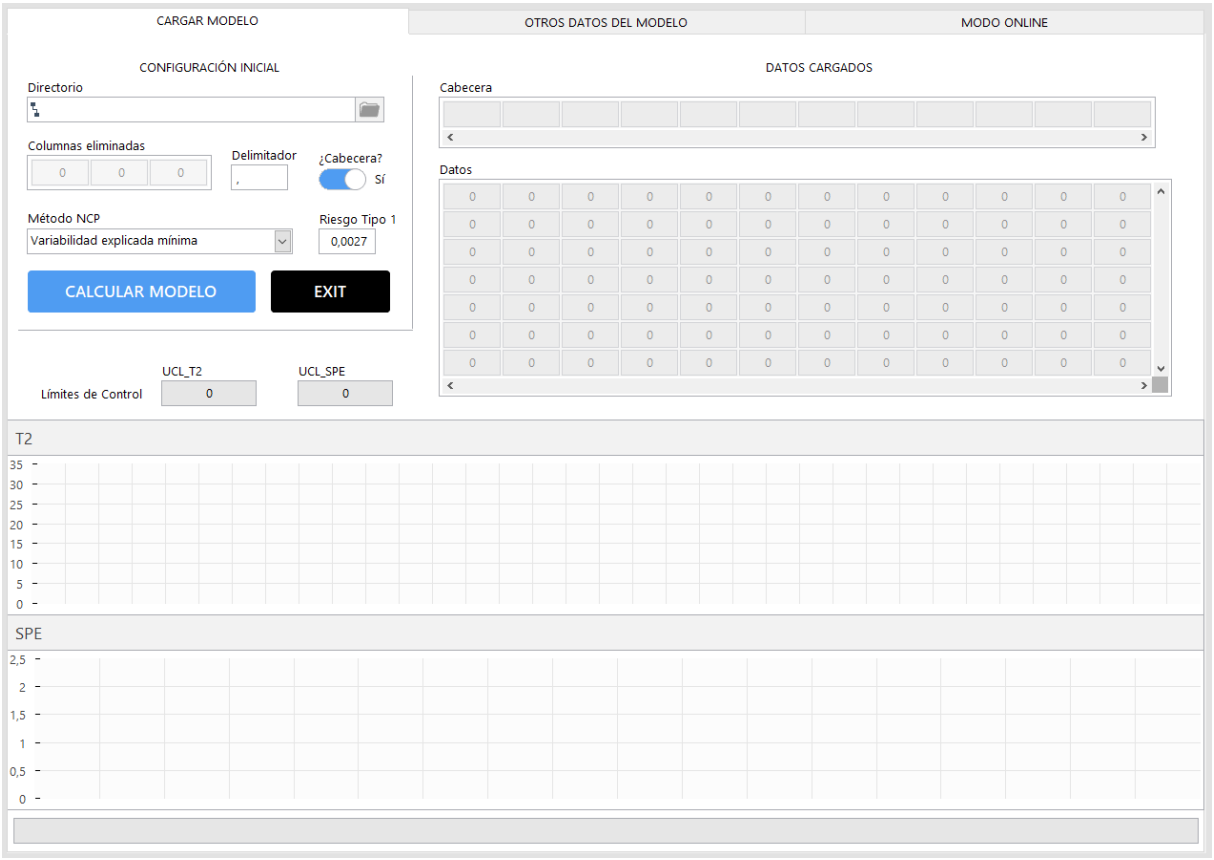

Figura 4.15: Captura de la pestaña Cargar Modelo

Configuración Inicial

Este subapartado no es más que el lugar donde se deben indicar todos los parámetros necesarios para realizar una correcta carga de los datos que la aplicación utilizará posteriormente para generar el modelo matemático. El usuario deberá de indicar los siguientes parámetros:

1. Directorio donde se encuentra el archivo de los datos.
2. Delimitador que se utiliza en dicho archivo para separar las columnas.
3. Si el archivo tiene o no cabecera.
4. Columnas a eliminar.
5. Método de elección del número de componentes que se utilizará en la generación del modelo.
6. Nivel de significación (Riesgo Tipo I).

Además, se incluyen dos botones que permiten realizar la carga del modelo (botón Calcular Modelo) o parar por completo la ejecución de la aplicación (botón Exit).

Datos Cargados

Aquí encontramos una visualización de cómo han sido cargados los datos del archivo CSV. Arriba tenemos una fila de datos que indica los valores de la cabecera en caso de que se haya marcado esta opción en la configuración anterior. En caso negativo aparecerá en la primera posición de esa fila un mensaje indicándolo.

Debajo de la cabecera tenemos la matriz principal que contiene los datos que se mostrará de una u otra manera dependiendo de la configuración que se haya indicado anteriormente. El objetivo que se persigue con este subapartado es que el usuario sea capaz de identificar si los datos han sido cargados o no correctamente por la aplicación y que se hagan las correcciones oportunas.

Por último, se dibujarán los gráficos de control T_A^2 y SPE, así como de los límites de control correspondientes. Estos se crearán a partir de las muestras que contenga el set de datos de entrenamiento utilizado, es decir, de manera *offline*. Con ello se pretende que el usuario pueda detectar errores relacionados con dicho set. Por ejemplo, el uso de conjuntos de datos que no hayan sido previamente limpiados (que contengan datos anómalos) y que, por lo tanto, no sean válidos para realizar la generación del modelo matemático.

Una vez hemos cargado los datos se dibujarán los gráficos de control T_A^2 y SPE junto con sus límites $UCL(T_A^2)_\alpha$ y $UCL(SPE)_\alpha$. El objetivo de mostrar estos gráficos, que han sido calculados con los datos de un funcionamiento normal, no es otro que el de visibilizar los posibles errores que se hayan cometido en los pasos previos. Como muestra la Figura 4.16 los datos han sido correctamente cargados ya que la cabecera aparece correctamente visualizada.

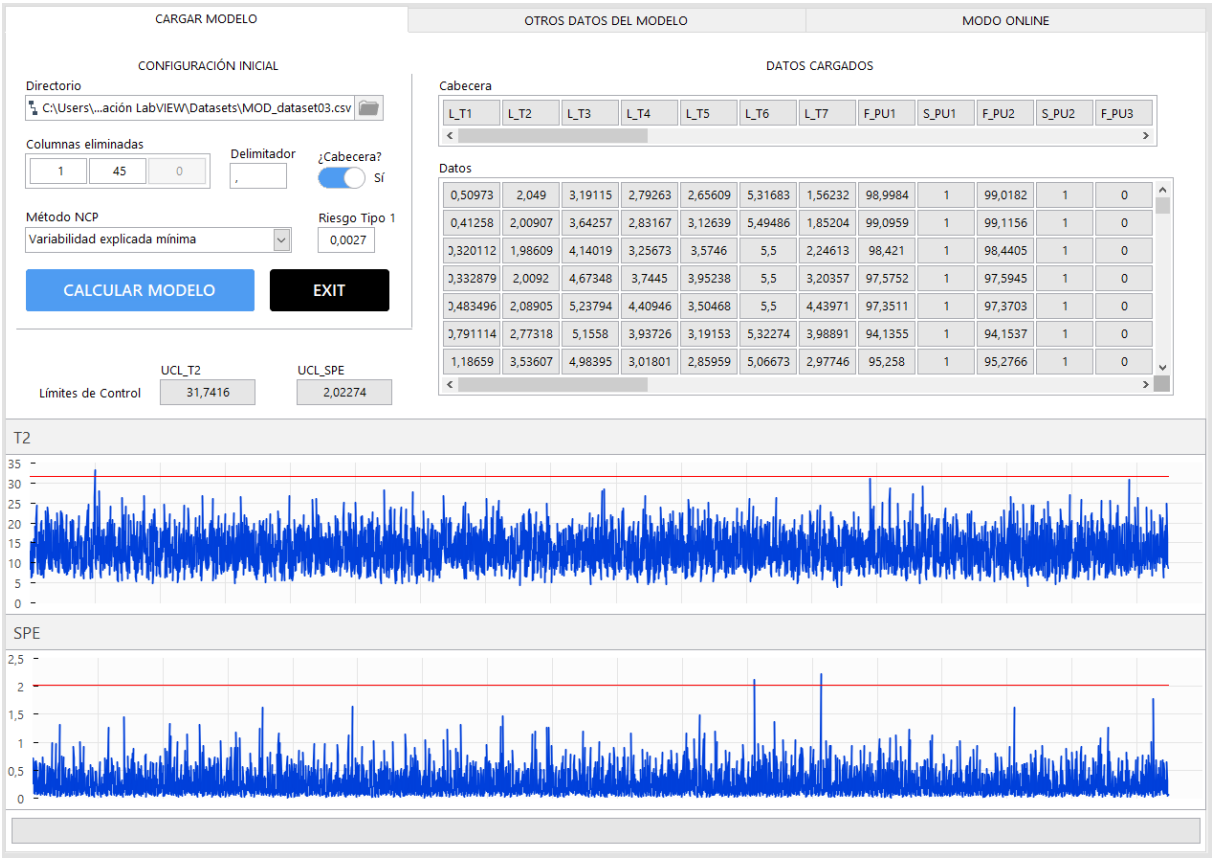

Figura 4.16: Datos cargados correctamente

También, se ha de recordar que al haber utilizado los datos de un funcionamiento normal de la instalación para generar los gráficos de control, éstos deberían estar siempre por debajo de sus límites correspondientes (líneas rojas). De esta forma, el usuario podrá ser capaz de visualizar directamente en pantalla si los datos para generar el modelo estadístico han sido elegidos correctamente o no.

En cambio en la Figura 4.17 los datos han sido cargados incorrectamente ya que en muchos puntos los límites de control son superados por la gráfica T_A^2.

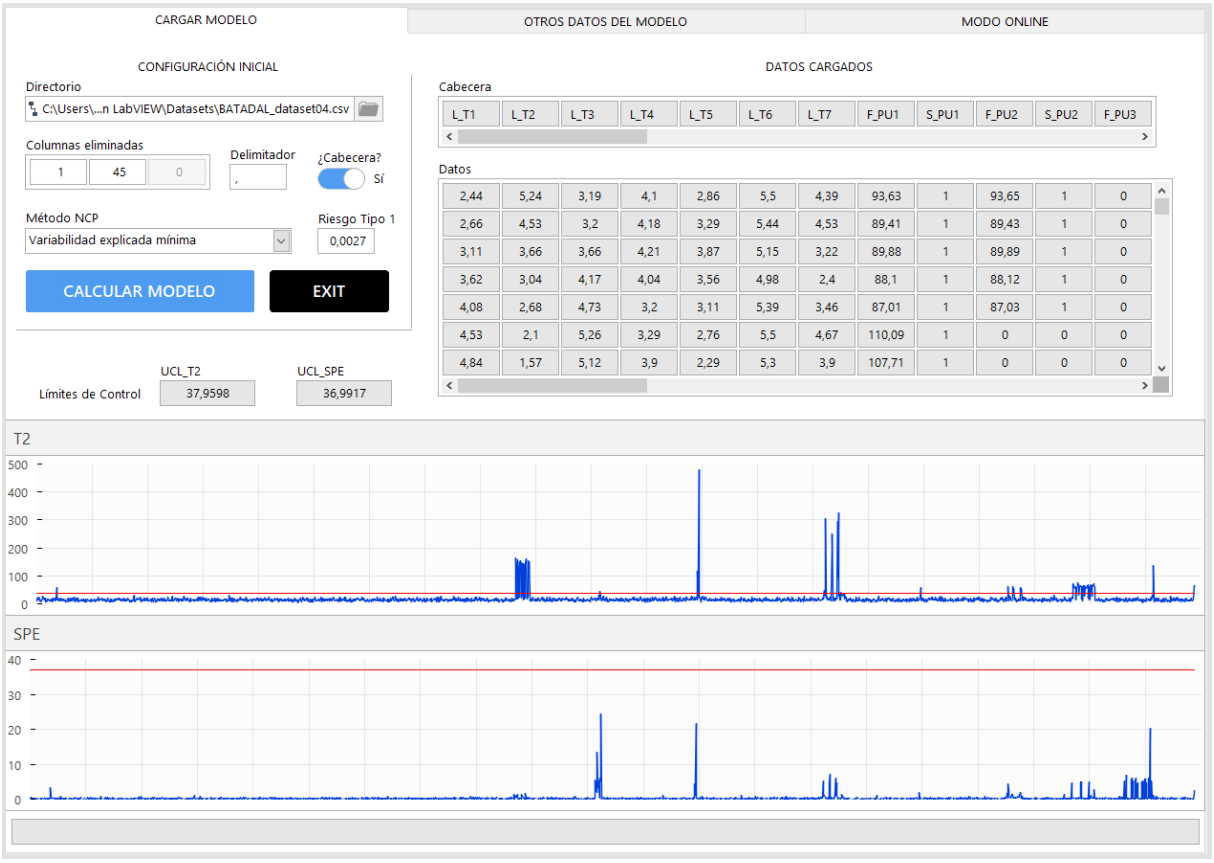

Figura 4.17: Datos cargados incorrectamente

4.3.2 Datos del Modelo

En esta segunda parte de la aplicación se muestran otros datos del modelo estadístico generado. En concreto, la matriz $\hat{P}$ que permite el cambio de subespacio, los valores propios *latent*, el porcentaje de variabilidad explicada acumulada según el número de componentes elegidas, las medias y las desviaciones típicas de las variables en los datos utilizados para calcular el modelo. Ver Figura 4.18.

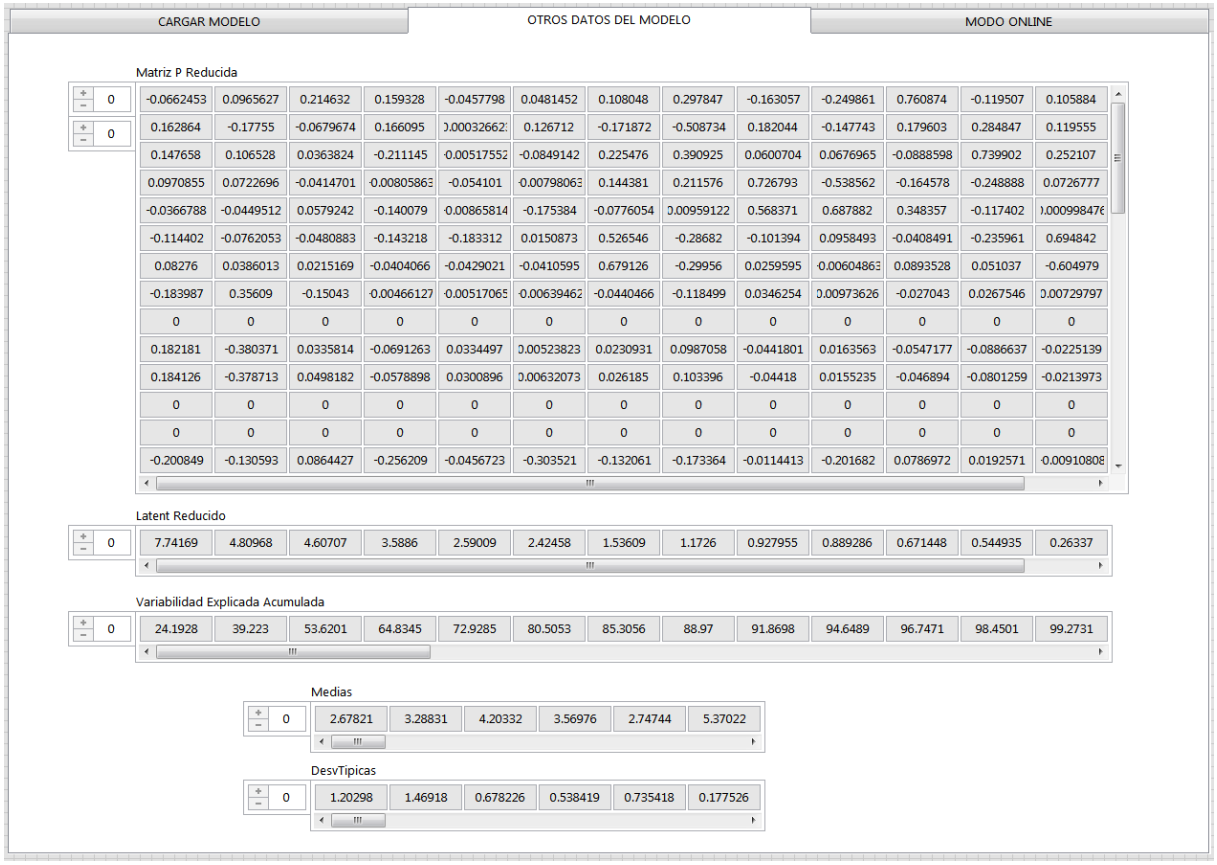

Figura 4.18: Captura de la pestaña Otros Datos del Modelo

4.3.3 Modo Online

El Modo Online o Modo en Línea es la última parte de la aplicación pero no por ello resulta menos importante. Como se puede ver en la Figura 4.21, se compone de 2 secciones. La primera permite configurar al usuario los datos del servidor OPC al que quiere conectarse. Para ello el usuario deberá hacer clic en el botón Abrir OPC (ver Figura 4.19).

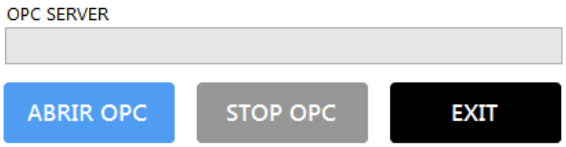

Figura 4.19: Botones de la pestaña Modo Online

Una vez pulsado dicho botón se abrirá una ventana emergente en la que aparecerá el listado de servidores OPC disponibles que haya detectado la aplicación en ese momento. En nuestro caso seleccionamos el generado para la realización de este trabajo, MatrikonOPC (ver Figura 4.20).

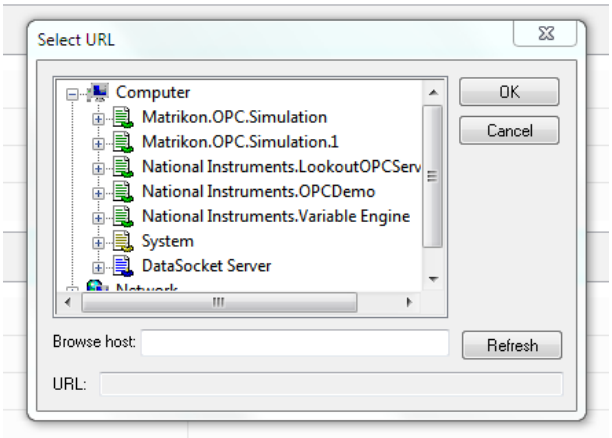

Figura 4.20: Ventana de selección del servidor OPC

Adicionalmente también se han creado los botones Stop OPC y Exit que permiten parar la conexión OPC o salir de la aplicación.

Abajo de la configuración se han añadido los gráficos de control T_A^2 y SPE. Dichos gráficos se dibujan en azul y en tiempo real tras procesar los datos recibidos del servidor OPC tal y como se muestra en la Figura 4.21. En rojo se muestran los límites de control correspondientes a cada gráfico de control. Cuando se sobrepasa uno de estos límites, tal y como se ha explicado anteriormente, la alarma se activa iluminándose un vistoso recuadro rojo. Además, se activa la función de diagnóstico de fallos, los gráficos de contribuciones. En ellos se muestra como han contribuido las diferentes variables a la activación de la alarma. Con ello se puede intuir el origen de la incidencia.

Por último, cada vez que se activa una alarma se guarda registro de ésta. A la derecha de la pantalla se muestra el Listado de Incidencias. Este muestra todo el registro de las anomalías que hayan sucedido, mostrando la fecha y hora de cuando se produjo la incidencia y las variables a las que apuntaban los gráficos de contribuciones.

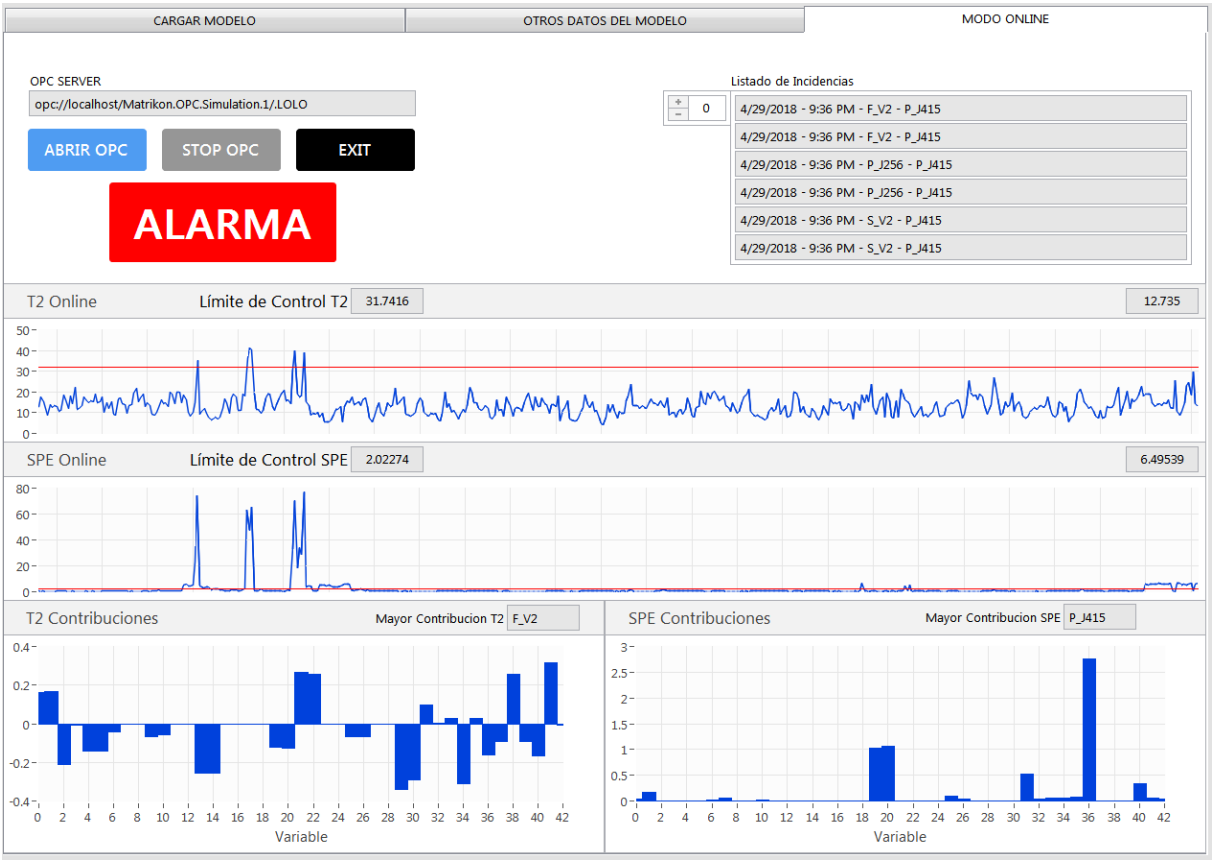

Figura 4.21: Captura del Modo Online

4.4 Configuración del servidor OPC

OPC es un estándar para las telecomunicaciones industriales desarrollado y mantenido por el consorcio OPC Foundation. Aunque a día de hoy el significado que han adoptado las siglas OPC es Open Platform Communications, en sus inicios era Object Linking and Embedding for Process Control y como ellas mismas indican este estándar está especialmente diseñado para llevar a cabo las comunicaciones relacionadas con el control de procesos industriales y para facilitar el intercambio de datos en tiempo real entre los diversos dispositivos de control de los diferentes fabricantes que nos podemos encontrar en una planta industrial. El concepto de interoperabilidad hace referencia a esto último y resulta fundamental en el planteamiento del estándar OPC. La Figura 4.22 muestra como la sencillez aumenta significativamente si se aplica la tecnología OPC.

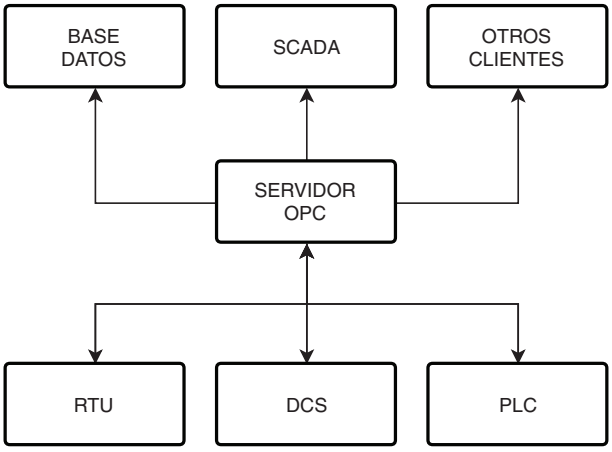

Figura 4.22: Diagrama de las comunicaciones mediante OPC

Debido a la alta interoperabilidad que presenta y a que, a diferencia de otros estándares de comunicación industriales, OPC es un estándar abierto ha sido elegido para realizar la simulación de una fuente de datos lo más realista posible. OPC ha evolucionado con el paso del tiempo y sobre él se han creado nuevas arquitecturas especialmente optimizadas para ciertas situaciones que están basadas en el modelo Cliente-Servidor. Algunas de ellas son:

- Intercambio de Datos OPC (OPC DX)
- Acceso de Datos XML (OPC XML DA)
- Arquitectura Unificada OPC (OPC UA)

No obstante, el análisis y elección de una arquitectura concreta en lo que a comunicaciones industriales se refiere queda fuera del alcance de este proyecto y por lo tanto no se ha llevado a cabo. Se ha realizado una conexión Cliente-Servidor básica. Donde se tiene un cliente principal que obtiene los datos del servidor OPC y un cliente secundario que proporciona los datos al servidor con una determinada frecuencia. El flujo de información sigue el esquema de la Figura 4.23

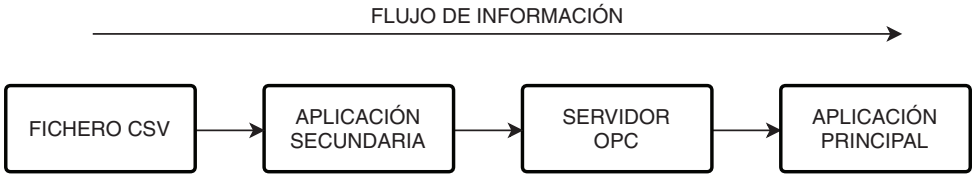

Figura 4.23: Diagrama del flujo de información

El servidor OPC se ha creado con el software MatrikonOPC Server for Simulation (Figura 4.24 que permite servir de manera sencilla y eficazmente cualquier tipo de dato con dicho protocolo. Como la finalidad que se persigue es que la aplicación principal (Cliente OPC) sea capaz de recibir periódicamente los datos como si de muestras de un proceso real se tratase y tras ello

ejecutar su algoritmo para detectar los comportamientos anómalos, el servidor OPC emitirá un array de números reales con doble precisión (64 bits) cada 100 milisegundos.

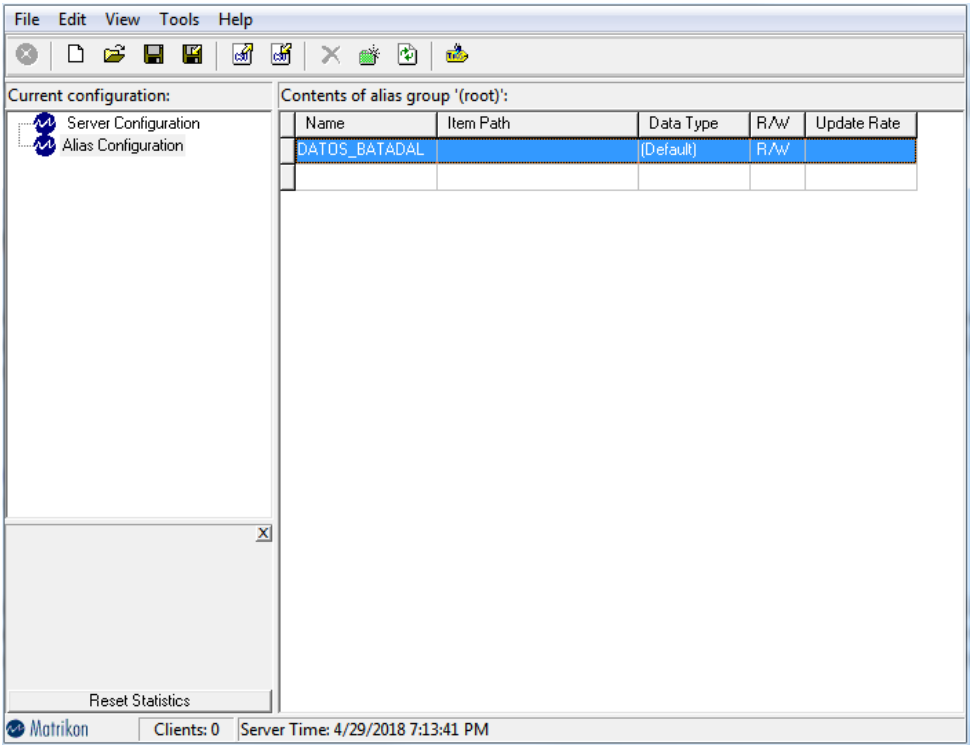

Figura 4.24: Listado de variables servidas en MatrikonOPC

En cuanto a los datos que deben ser servidos periódicamente a través del servidor, estos se encuentran en un archivo CSV y el servidor OPC no puede servir muestra tras muestra directamente. Para ello se necesita crear una pequeña aplicación secundaria que realice esta función. Esta interfaz también se ha diseñado e implementado mediante LabVIEW.

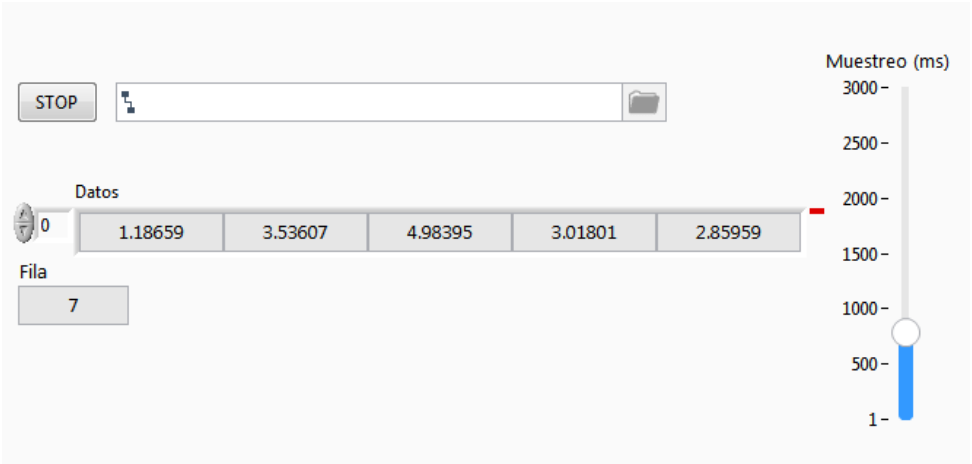

Figura 4.25: Captura aplicación secundaria

Como se puede observar en la Figura 4.25, el usuario puede seleccionar el archivo CSV que se quiere servir, así como el periodo en milisegundos al que se escribirá cada dato en el servidor OPC. Es importante señalar que la frecuencia a la que la aplicación principal lea los datos del servidor debe de ser la misma a la que la interfaz que permite escribir en el servidor esté funcionando. Es decir, que se encuentren sincronizados. En el caso de que esto no se cumpla se pueden dar dos casos:

- La frecuencia de lectura del cliente OPC es mayor que la frecuencia con la que cambian los datos del servidor OPC. Si esto es así, no se perderán datos pero se realizarán un mayor número de lecturas de las necesarias pudiendo verse sobrecargado el sistema.
- La frecuencia de lectura del cliente OPC es menor que la frecuencia con la que cambian los datos del servidor OPC. Si se da esta situación es posible que se pierdan datos ya que el servidor actualizará las variables más rápidamente de lo que el cliente será capaz de leerlas.

Para sincronizar ambas aplicaciones, la aplicación auxiliar cuenta con un dial que permite ajustar con precisión la velocidad a la que se leen y se sirven las filas del archivo CSV del que se parte. La aplicación lee el archivo CSV que le haya indicado el usuario y lo carga en memoria. Después, con un determinado periodo especificado en milisegundos por el usuario mediante el dial, es capaz de escribir fila por fila los datos del archivo en cuestión en una variable de tipo array en el servidor OPC para que este los distribuya a los clientes que haya conectados.

De esta sencilla manera se ha simulado la recogida de datos como si de una instalación en la que un SCADA o un DCS estuviesen funcionando se tratase.

Capítulo 5

Resultados

Una vez ha sido explicada la solución que se ha propuesto en este trabajo se puede proceder a exponer los resultados que se han obtenido. Para ello, se ha dividido dicha exposición en dos partes: una dedicada al caso BATADAL y otra al caso Cranfield. Antes de profundizar en ello se debe aclarar que los resultados expuestos en este capítulo, y en concreto las gráficas mostradas, han sido los generados en su mayoría mediante MATLAB y no mediante la aplicación de LabVIEW, aunque ambos resultados son iguales ya que utilizan exactamente el mismo algoritmo.

5.1 Resultados BATADAL

Tras haber realizado el preprocesado de los datos del caso BATADAL tal y como se ha especificado en el capítulo anterior, se ha procedido a realizar el Análisis de Componentes Principales a partir del fichero MOD_dataset03, que recordemos es el fichero de datos de entrenamiento una vez eliminados los ouliers y otros datos anómalos conforme lo especificado en el apartado 4.1 de este trabajo.

5.1.1 Elección del número de componentes principales a retener

Como ya se ha mencionado anteriormente, se han implementado dos criterios diferentes para la elección del número de componentes principales retenidas por el modelo matemático: el criterio de variabilidad explicada y el criterio de autovalores mayores a la unidad. De aquí en adelante a estos criterios se les ha llamado Criterio 1 y Criterio 2 para abreviar. Dependiendo del criterio elegido los resultados varían considerablemente, como se puede apreciar en el resumen de la Tabla 5.1. El número de componentes principales retenidas es un parámetro fundamental para el cálculo de los gráficos de control, que afecta tanto a los estadísticos T^2 y SPE como a sus límites de control correspondientes.

	Criterio 1	Criterio 2
Número de Componentes Principales retenidas	13	8
Variabilidad Explicada Acumulada	99.2731 %	88.9700 %
Límite UCL(T^2)	31.7416	23.6197
Límite UCL(SPE)	2.0227	255.0183

Tabla 5.1: Caso BATADAL. Cuadro resumen del Criterio 1 y Criterio 2

Criterio 1. Modelo basado en 13 componentes

En el Criterio 1 se retienen tantas componentes principales como sean necesarias para garantizar una variabilidad explicada acumulada mínima del 99 %. Con este criterio el número de componentes principales con el que se genera el modelo es 13, es decir se retienen las 13 primeras componentes ya que son las que más variabilidad tienen asociada. Así, los límites de control calculados a partir de las Fórmulas 4.7 y 4.8 tienen un valor de 31.7416 y 2.0227 respectivamente, tal y como indica en el resumen de la Tabla 5.1. Estos límites se visualizan de color rojo en las figuras siguientes.

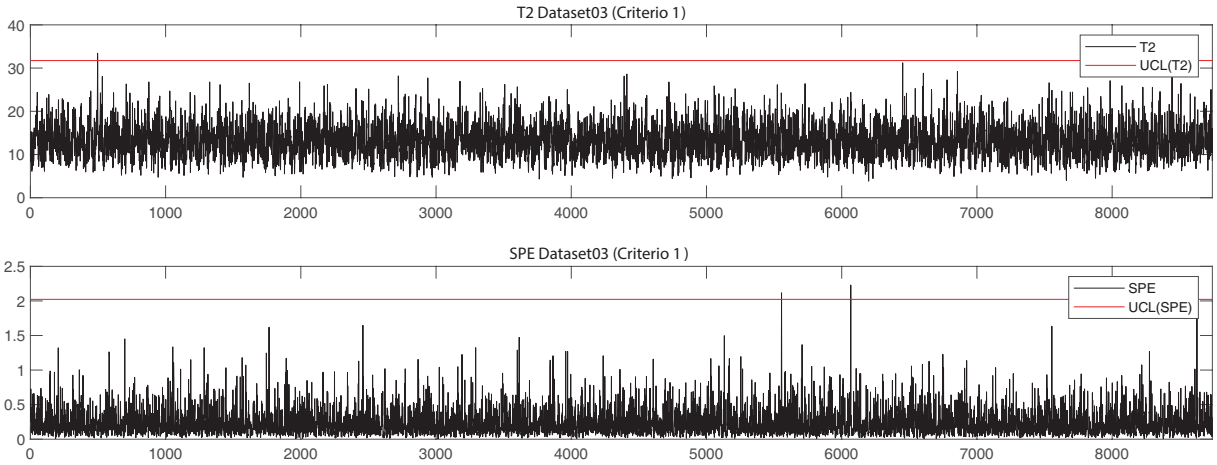

Figura 5.1: Gráficos de Control T^2 y SPE del fichero MOD_dataset03.

En la Figura 5.1 se muestran los gráficos de control del set de entrenamiento MOD_dataset03. Como se puede apreciar, la mayoría de puntos quedan por debajo de los límites de control.

La Figura 5.2 y en la Figura 5.3 se corresponden con los gráficos de control para los conjuntos de datos BATADAL_dataset04 y BATADAL_dataset05. En cambio, en ellos existen algunos puntos que superan los límites de control mientras que el resto no llega a ello. Esto se debe a que en estos sets existen medidas atípicas, pequeños periodos de tiempo en los que la red de agua presenta un comportamiento fuera de lo normal. Como se verá más adelante, estos periodos son los diferentes ciberataques que ha recibido la red.

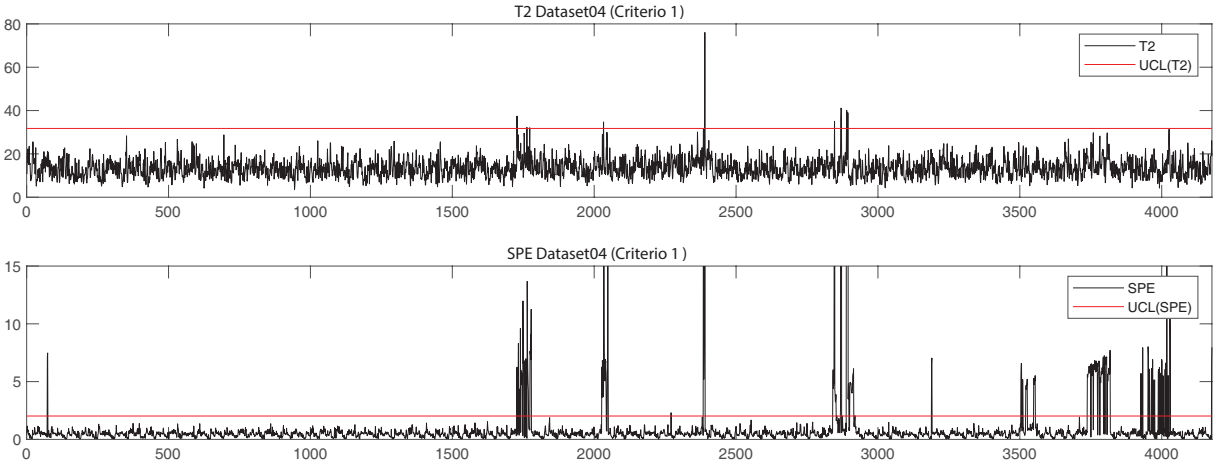

Figura 5.2: Gráficos de Control T^2 y SPE del fichero BATADAL_dataset04.

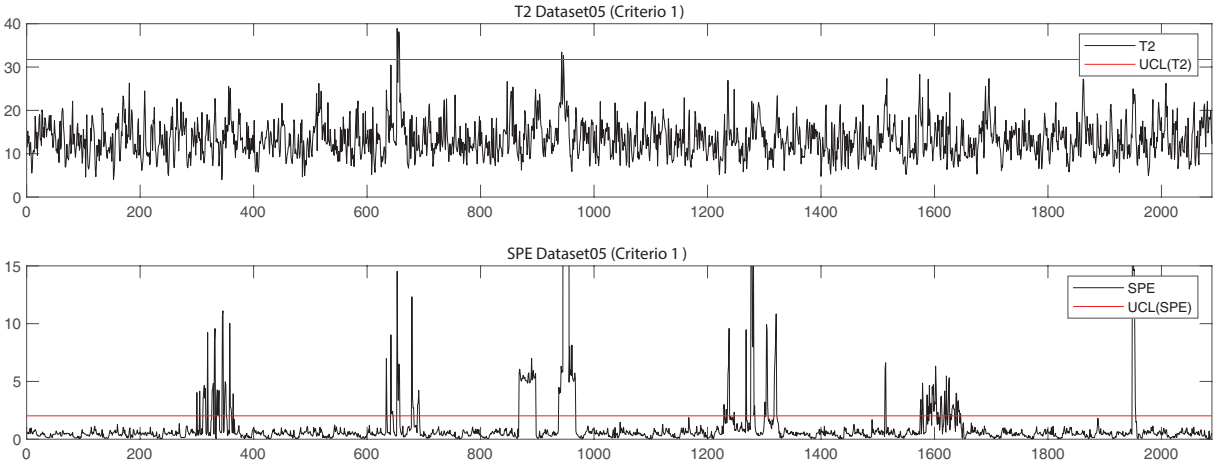

Figura 5.3: Gráficos de Control T^2 y SPE del fichero BATADAL_dataset05.

Criterio 2. Modelo basado en 8 componentes

En el Criterio 2 se retienen tantas componentes principales como valores propios mayores a la unidad haya. En este caso son 8 componentes las que cumplen esta condición dando lugar a un porcentaje de variabilidad explicada acumulada del 88.9700 %. En cuanto a los límites de control UCL(T2) y UCL(SPE), siendo calculados de la misma manera que en el Criterio 1, tienen un valor de 23.6197 y 255.0183 respectivamente. En la Figura 5.4 se aprecia como los gráficos de control siguen quedando por debajo de sus límites para los datos de entrenamiento.

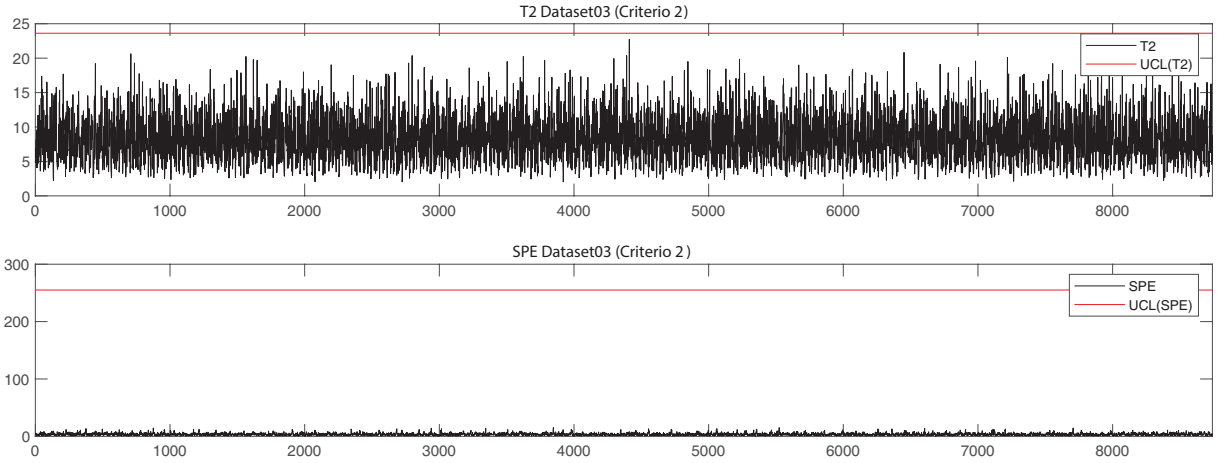

Figura 5.4: Gráficos de Control T^2 y SPE del fichero MOD_dataset03.

En cambio, en la Figura 5.5 y la Figura 5.6 vemos que a diferencia del Criterio 1 los gráficos de control no consiguen superar en prácticamente ningún caso los límites marcados aunque sí que se aprecian variaciones sustanciales en sus valores, es decir se detectan los comportamientos anómalos de los sets de datos 4 y 5 pero no llega a producirse la alerta generada al sobrepasar los límites de control.

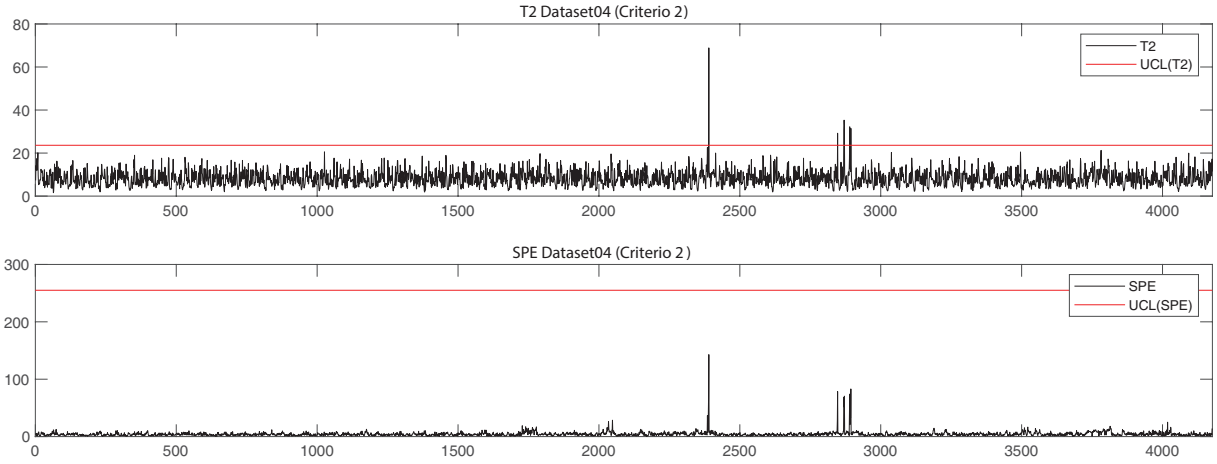

Figura 5.5: Gráficos de Control T^2 y SPE del fichero BATADAL_dataset04.

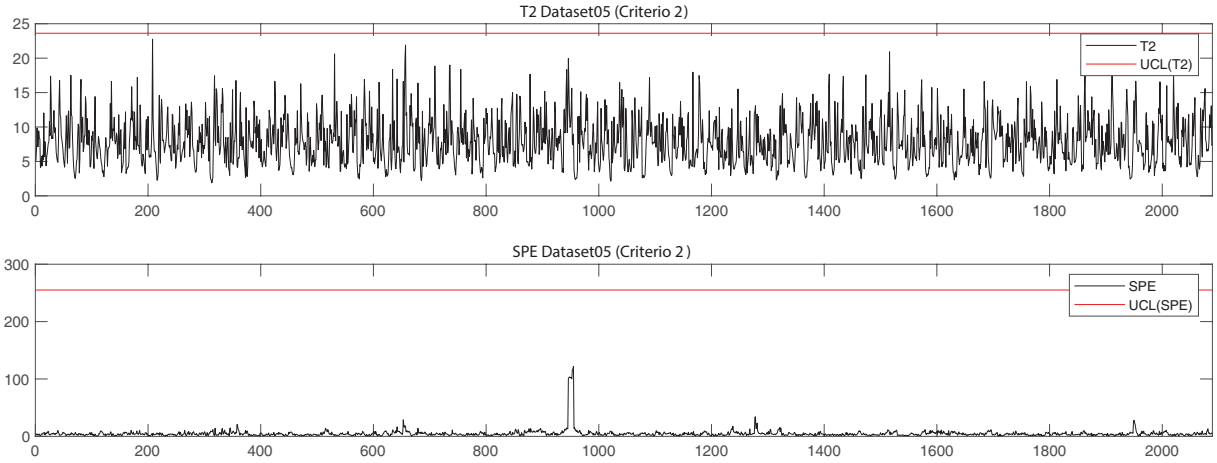

Figura 5.6: Gráficos de Control T^2 y SPE del fichero BATADAL_dataset05.

Por lo tanto, y a vista de los resultados anteriores, el primer criterio es mucho mejor para el caso que nos ocupa que el segundo criterio. Luego para proceder en los siguientes apartados se tomará como referencia el Criterio 1.

5.1.2 Diagnóstico del fallo. Comparación de resultados

Como se ha mencionado en apartados anteriores se puede identificar que variables del espacio original son las responsables de que los gráficos de control hayan superado sus límites correspondientes. Esto se realiza mediante el llamado Diagrama de Contribuciones. Teniendo los resultados de la Figura 5.2 y la Figura 5.3 se ha procedido a realizar un breve análisis de cada una de las incidencias detectadas.

Comenzando por la primera gráfica, en la Figura 5.7 se pueden observar que los comportamientos atípicos detectados por el algoritmo han sido sombreados con un determinado color. Sobre un total de 9 incidencias 7 de ellas, las sombreadas de color rojo, han sido detectadas correctamente mientras que las sombreadas de color amarillo son falsos positivos arrojados por la aplicación.

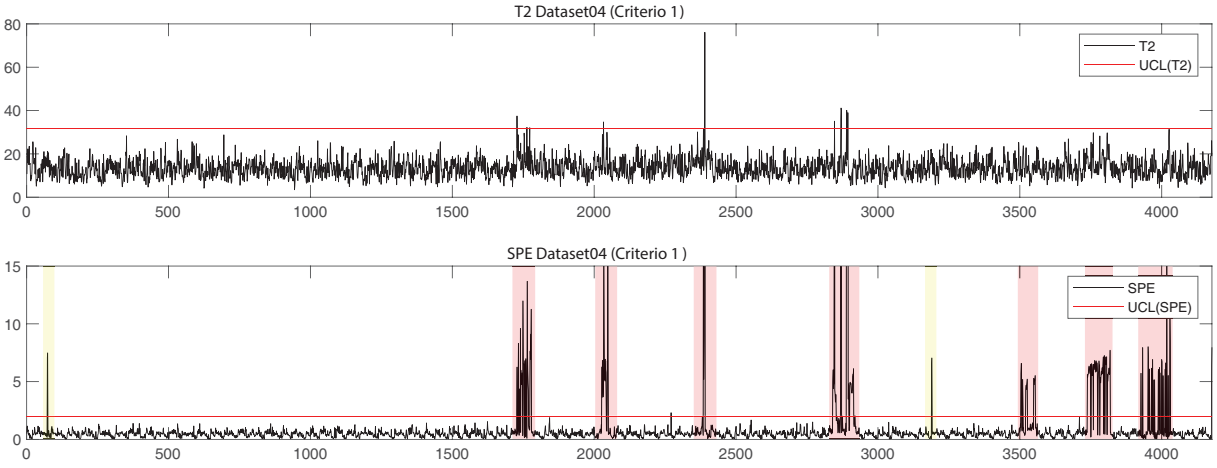

Figura 5.7: Anomalías detectadas en los datos del fichero BATADAL_dataset04.

En cuanto a segunda gráfica, en la Figura 5.8 se muestran un total de 8 incidencias de las cuales, 7 han sido detectadas correctamente mientras que solo ha habido un falso positivo.

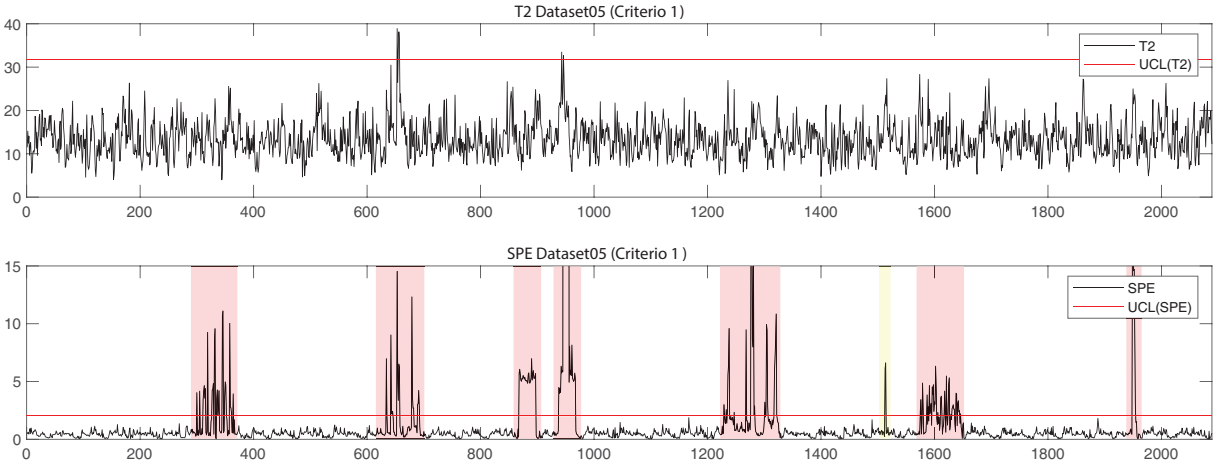

Figura 5.8: Anomalías detectadas en los datos del fichero BATADAL_dataset05.

Volviendo a la Figura 5.7 las anomalías que se identifican en color rojo ocurren en los periodos de tiempo que especifica la Tabla 5.2.

ID	Muestra	Fecha Inicio	Fecha Fin
1	1728 - 1778	13/09/16 23H	16/09/16 01H
2	2027 - 2049	26/09/16 10H	27/09/16 08H
3	2385 - 2390	11/10/16 08H	11/10/16 13H
4	2841 - 2919	30/10/16 08H	02/11/16 14H
5	3504 - 3554	26/11/16 23H	29/11/16 01H
6	3738 - 3818	06/12/16 17H	10/12/16 01H
7	3927 - 4030	14/12/16 14H	18/12/16 21H

Tabla 5.2: Incidencias en el dataset04

Mientras que en la Tabla 5.3 se han identificado los resultados de la Figura 5.8.

ID	Muestra	Fecha Inicio	Fecha Fin
1	300 - 364	16/01/17 11H	19/01/17 03H
2	634 - 691	30/01/17 09H	01/02/17 18H
3	868 - 897	09/02/17 03H	10/02/17 08H
4	938 - 967	12/02/17 01H	13/02/17 06H
5	1229 - 1322	24/02/17 04H	28/02/17 01H
6	1575 - 1646	06/12/16 17H	10/12/16 01H
7	1949 - 1954	26/03/17 04H	26/03/17 09H

Tabla 5.3: Incidencias en el dataset05

Seguidamente se muestran los diagramas de contribuciones referentes a algunos de los ataques identificados.

Dataset04. Primer ataque

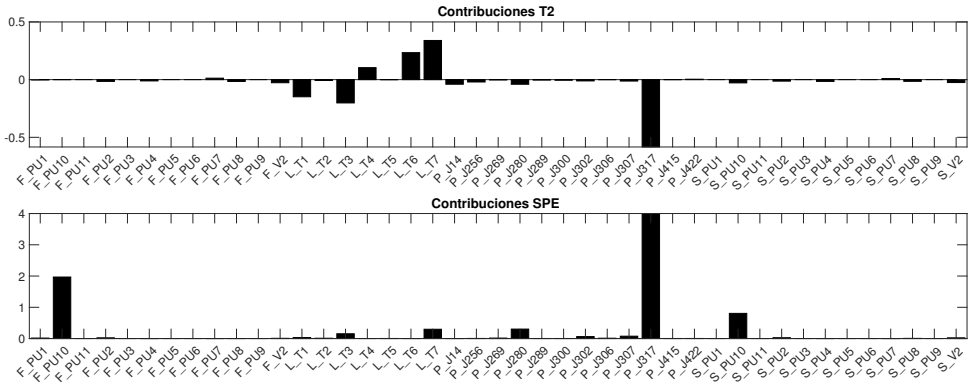

Figura 5.9: Diagrama de contribuciones del ataque 1 en el set 04

Como se observa en la Figura 5.9 las variables P_J317 y F_PU10 son las que más aportan a la generación del valor SPE anómalo. Los resultados publicados en la web del concurso BATADAL

indican que este ataque ha tenido como objetivo el cambio de los límites del tanque T7 que es controlado por las bombas PU10 y PU11.

Dataset04. Cuarto ataque

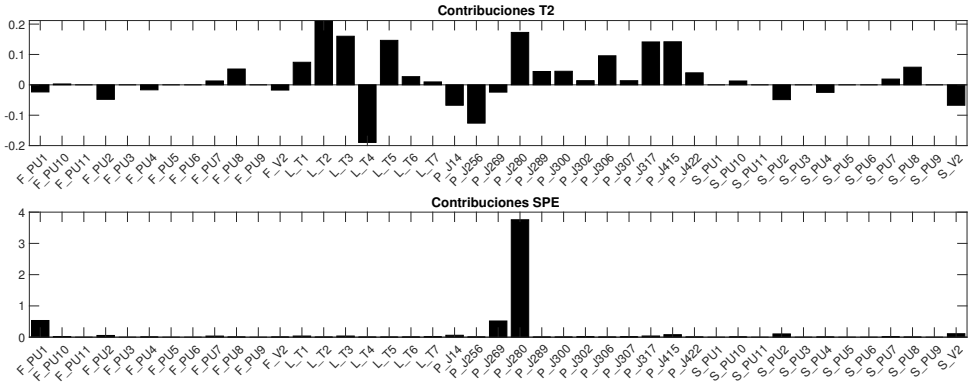

Figura 5.10: Diagrama de contribuciones del ataque 4 en el set 04

En la Figura 5.10 se puede observar que una de las variables destaca considerablemente sobre el resto, la J_280. En plano de la red de agua dicha variable se encuentra justo a la entrada del conjunto de bombas en paralelo formado por PU1, PU2 y PU3. En este caso, los resultados oficiales de BATADAL indican que este ataque ha provocado que las bombas PU1 y PU2 se mantengan siempre encendidas.

Dataset05. Primer ataque

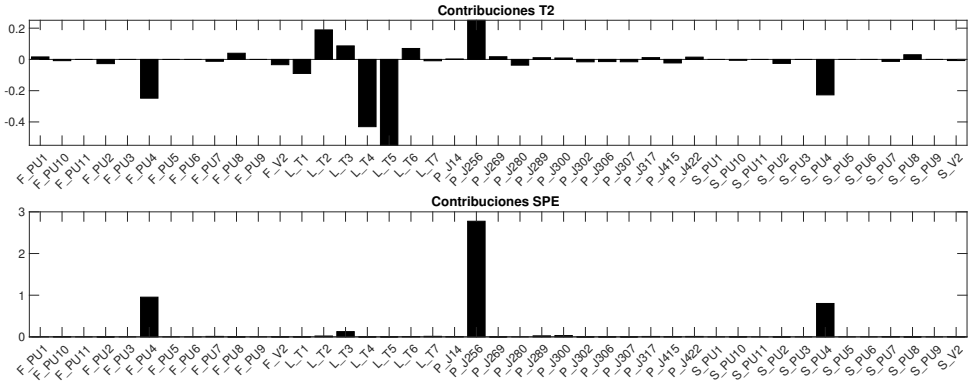

Figura 5.11: Diagrama de contribuciones del ataque 1 en el set 05

En el primer ataque del set 05 las variables J_256, F_PU4 y S_PU4 son señaladas por el gráfico de contribuciones. Según los resultados oficiales este ataque ha consistido en el cambio del límite del tanque T3. Dicho límite controla la activación de las bombas PU4 y PU5 que se encuentran

en paralelo. Luego el diagrama de contribuciones señala correctamente parte de las variables implicadas.

Dataset05. Segundo ataque

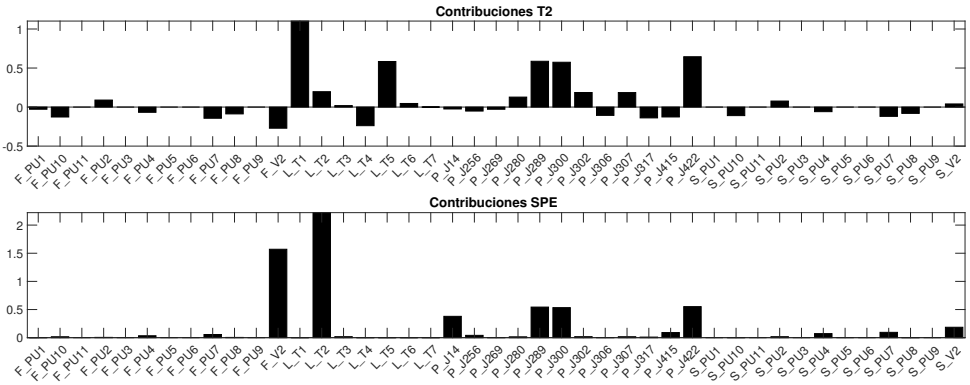

Figura 5.12: Diagrama de contribuciones del ataque 2 en el set 05

Por último, en el segundo ataque del set 05 de BATADAL el diagrama de contribuciones señala las variables L_T2 y F_V2 que se corresponden con el ataque descrito en los resultados de la web. El ataque consiste en una alteración de las lecturas del tanque T2 lo que hace que la válvula de control V2 se mantenga abierta cuando no debería estarlo, causando finalmente un desbordamiento en el propio tanque T2 por un mal control.

Como conclusión del caso BATADAL se puede decir que se han identificado correctamente los ataques en la línea de tiempo, y en lo que refiere a sus causas, se consiguen identificar variables que guardan relación con el origen del problema aunque en la mayoría de casos no se consigue señalar exactamente dicho origen del problema.

5.2 Resultados Cranfield

De la misma manera que en el caso BATADAL, después de realizar el preprocesado de los datos se ha procedido a realizar el PCA. Dado que en este caso se tienen tres sets de datos diferentes para llevar a cabo el entrenamiento se han concatenado los tres para formar uno solo. Con éste se ha realizado el entrenamiento al igual que en el caso anterior.

5.2.1 Elección del número de componentes principales a retener

Siguiendo el mismo procedimiento, hay dos criterios para elegir el número de componentes principales retenidas por el modelo matemático. En la Tabla 5.4 aparece un resumen con los parámetros que arrojan ambos entrenamientos.

	Criterio 1	Criterio 2
Número de Componentes Principales retenidas	13	5
Variabilidad Explicada Acumulada	99.1446 %	84.7495 %
Límite UCL(T^2)	31.6790	18.2120
Límite UCL(SPE)	2.3204	244.8361

Tabla 5.4: Caso Cranfield.Cuadro resumen del Criterio 1 y Criterio 2

En la Figura 5.13 y en la Figura 5.14 se muestran los resultados tras realizar el entrenamiento con ambos criterios.

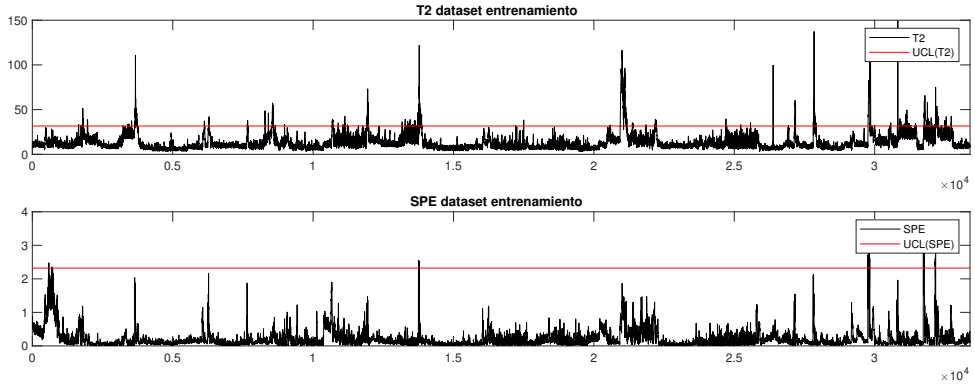

Figura 5.13: Gráficos de control T^2 y SPE del set de entrenamiento. Criterio 1.

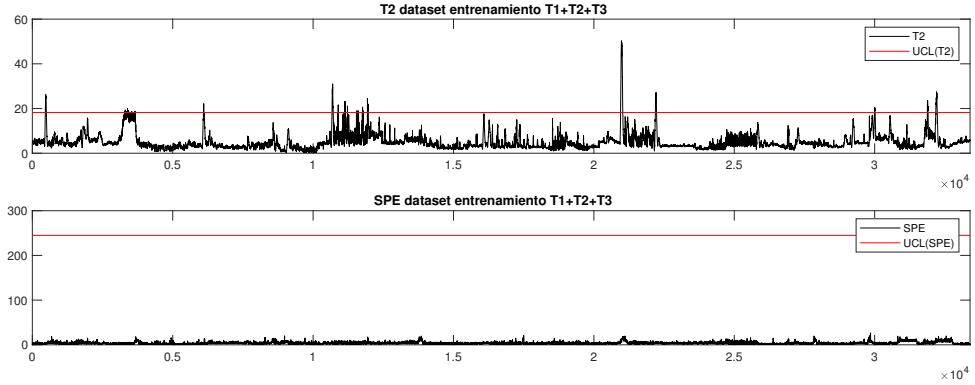

Figura 5.14: Gráficos de control T^2 y SPE del set de entrenamiento. Criterio 2.

En las Figuras 5.15, 5.16, 5.17 y 5.18 se muestran los gráficos de control T^2 y SPE trazados a partir de ambos criterios. Estos gráficos se han generado a partir del set 1 del FaultyCase6 de los archivos de Cranfield. Como se puede observar los resultados son muy similares siendo el segundo criterio más conservador que el primero. En este caso ambos serían totalmente válidos para proceder con su uso.

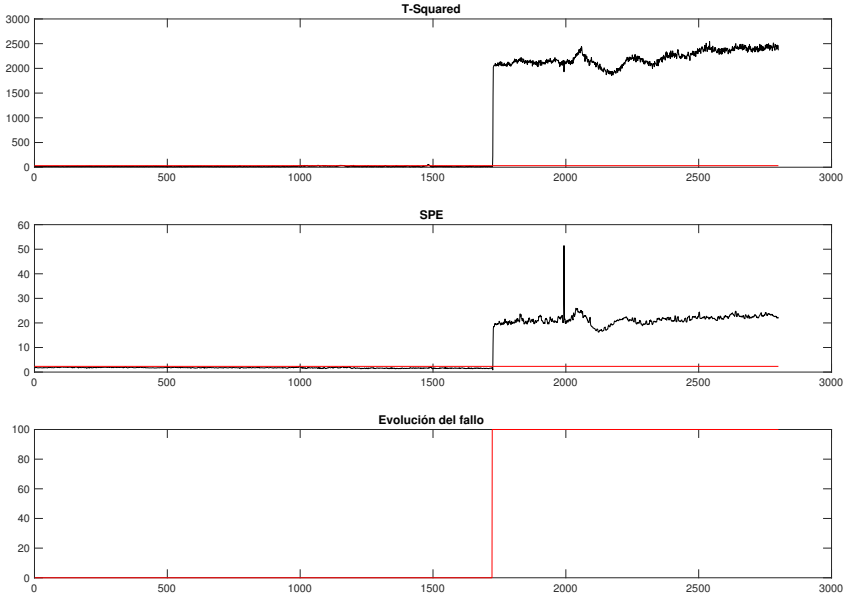

Figura 5.15: Gráficos de control T^2 y SPE del set 1 del FaultyCase6. Criterio 1.

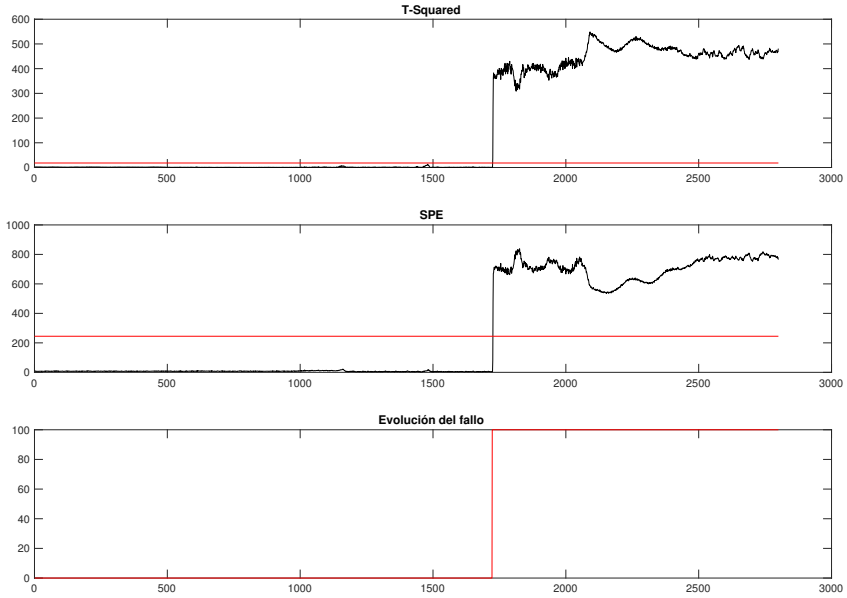

Figura 5.16: Gráficos de control T^2 y SPE del set 1 del FaultyCase6. Criterio 2.

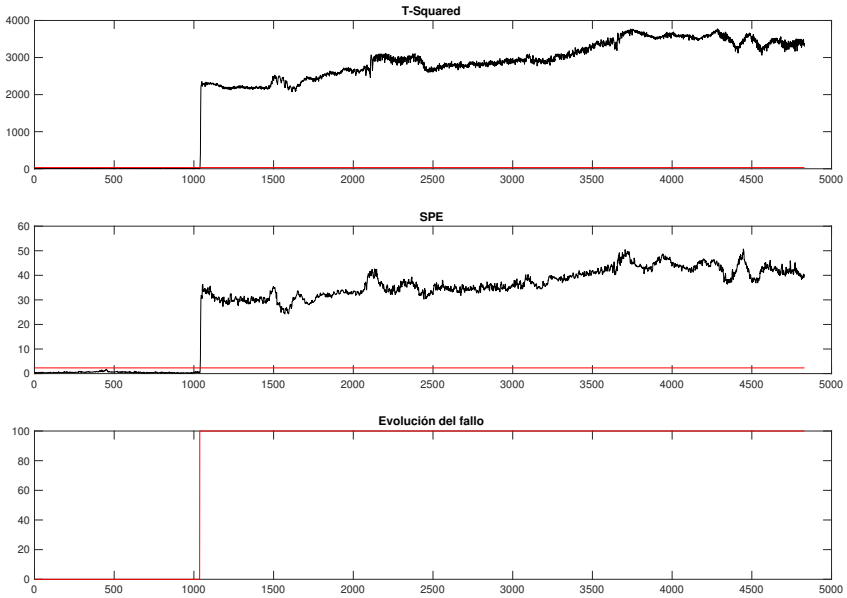

Figura 5.17: Gráficos de control T^2 y SPE del set 2 del FaultyCase6. Criterio 1.

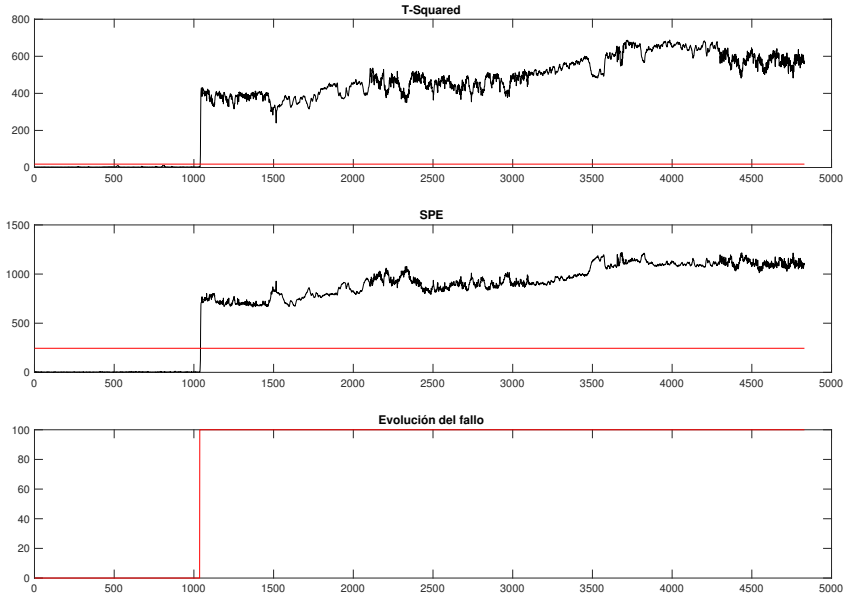

Figura 5.18: Gráficos de control T^2 y SPE del set 2 del FaultyCase6. Criterio 2.

5.2.2 Diagnóstico del fallo. Comparación de resultados

A continuación se adjuntan los resultados de los diagramas de contribuciones para los dos casos anteriores. Como se puede ver en ambos criterios se señala a la misma variable como la responsable del fallo, la variable 24. Teniendo en cuenta la causa del Fallo 6 tal y como se ha explicado al final del apartado 3.2.2. resulta lógico que dicha variable sea señalada por el diagrama de contribuciones.

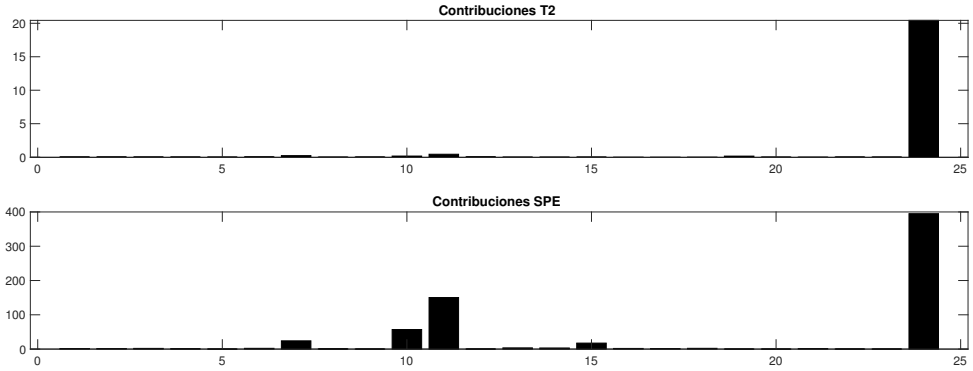

Figura 5.19: Diagrama de contribuciones del set 1 del FaultyCase6. Criterio 1.

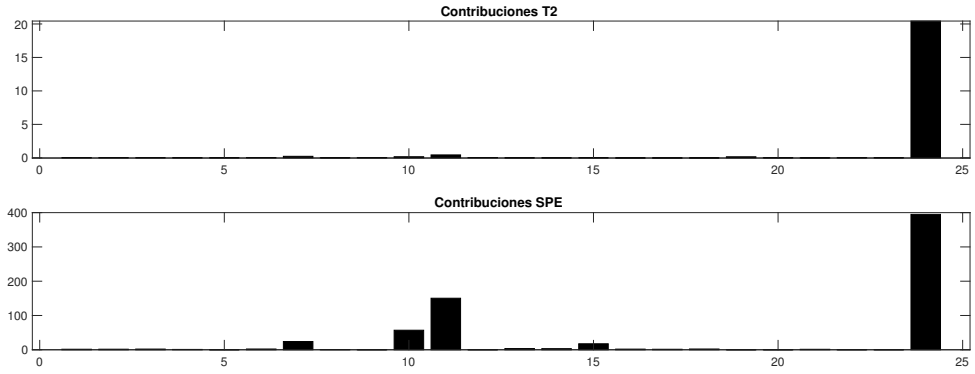

Figura 5.20: Diagrama de contribuciones del set 1 del FaultyCase6. Criterio 2.

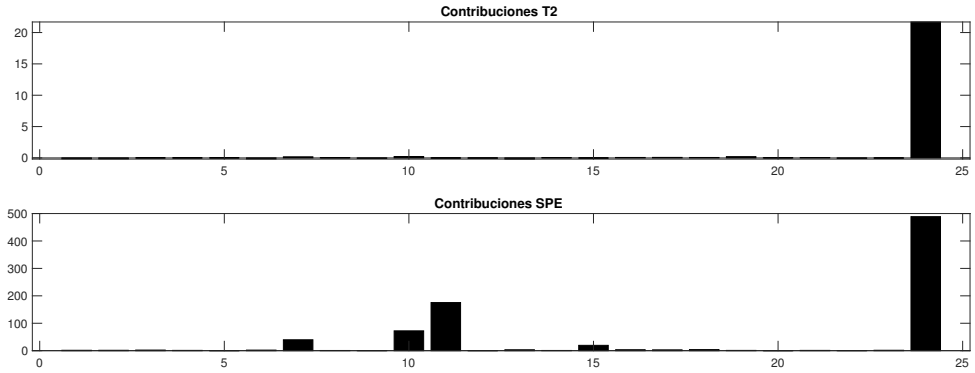

Figura 5.21: Diagrama de contribuciones del set 2 del FaultyCase6. Criterio 1.

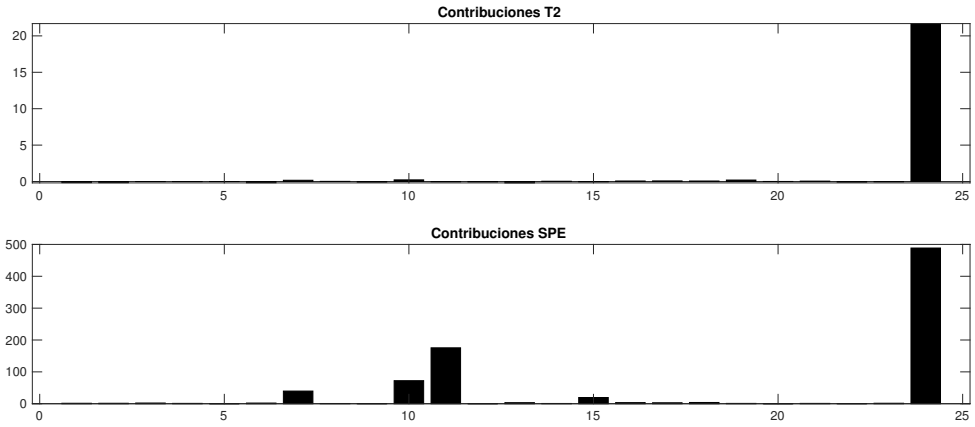

Figura 5.22: Diagrama de contribuciones del set 2 del FaultyCase6. Criterio 2.

Finalmente y teniendo en cuenta que se trata de técnicas estadísticas, como en el caso BATADAL se puede concluir que la detección de los fallos se realiza de manera aceptable. En cambio, los resultados de diagnóstico indican que el algoritmo no está ofreciendo la precisión que a priori se podría esperar de él. Por lo tanto, la parte de diagnóstico del fallo queda abierta a futuras modificaciones que permitan su mejora.

En línea con la metodología basada en el entrenamiento de modelos estadísticos que se ha seguido en este trabajo, existen diferentes técnicas que podrían realizar un diagnóstico de las incidencias con mayor precisión que los diagramas de contribuciones. En los últimos años se ha popularizado el uso y potenciado el desarrollo de técnicas relacionadas con la inteligencia artificial, concretamente con el aprendizaje automático o *machine learning*. En particular, considerar los clasificadores basados en el aprendizaje supervisado, como pueden ser las redes neuronales artificiales o las máquinas de vectores de soporte, es una buena estrategia para conseguir la deseada mejora del diagnóstico de fallos. Aunque estas opciones sean relativamente nuevas, su utilización es totalmente viable debido a la precisión y escalabilidad que presentan. Además, también resultan compatibles con las metodologías del campo del Control Estadístico de Procesos Multivariante. De hecho, en el aprendizaje automático también se suele utilizar el Análisis de Componentes Principales para reducir el tiempo de cálculo durante el entrenamiento de los modelos estadísticos. Así, mediante la combinación de múltiples modelos matemáticos entrenados de manera independiente se podría crear un mejor algoritmo. Lo que en la práctica se conoce como la técnica de *stacking*.

Capítulo 6

Conclusiones

Tras todo el trabajo que ha supuesto este proyecto, en general se puede concluir que se han cumplido los objetivos planteados desde el inicio. Al finalizar las pruebas con los diferentes conjuntos de datos se puede concluir que se han alcanzado los siguientes objetivos:

Análisis previo

- Se ha realizado un estudio de los dos casos considerados. Buscando información y documentándose acerca del contexto en el que se encuentran.
- Se ha realizado un análisis de los conjuntos de datos que se han proporcionado para ambos casos.
- Se han visualizado las diferentes variables y eliminado los datos anómalos si procedía.
- Se han visualizado los histogramas de cada una de las variables para saber si seguían la distribución normal, distribución que es requisito para la posterior utilización de las técnicas estadísticas multivariantes.
- Se han detectado patrones repetitivos temporales en algunas variables.
- La mayoría de las variables están autocorrelacionadas, como no podía ser de otra forma al tratarse de sistemas cerrados interdependientes.

Algoritmo y aplicación

- Se ha implementado la estandarización de las variables mediante el cálculo del valor Z-Score, centrándolas y escalándolas, consiguiendo media nula y desviación típica unitaria.
- Se ha descartado la realización del Análisis Dinámico de Componentes Principales debido a su dificultad de implementación.
- Se ha implementado el Análisis de Componentes Principales mediante la utilización de la matriz de covarianza y el cálculo de los valores y vectores propios.

- Se han implementado diferentes criterios para la selección del número de componentes principales. Concretamente el criterio de variabilidad mínima explicada del 99 %, y el de valores propios mayores que la unidad.
- Se ha descartado la implementación de los criterios basados en el Scree Plot y en la validación cruzada. Debido a que el primero no aportaría mejoras sustanciales a los ya implementados, y a que el segundo presenta una elevada dificultad de implementación. Con la validación cruzada posiblemente mejorarían los resultados.
- Se ha implementado el algoritmo que permite la generación del modelo estadístico basado en PCA, el cálculo de los gráficos de control Hotelling's T-Squared (T^2) y Squared Prediction Error (SPE, Q), de sus respectivos límites de control y de las contribuciones de las variables originales al valor de ambos estadísticos.
- Se ha realizado la elección justificada de las tecnologías para la implementación final de la aplicación. LabVIEW ha sido la herramienta elegida para realizar todo el desarrollo.
- Se han probado diferentes arquitecturas de software durante la implementación, siendo la arquitectura derivada de la conocida Maestro-Esclavo, Queued State Machine with event-driven Producer-Consumer la que mejor resultados ha dado y mayor escalabilidad ha tenido.
- Se ha diseñado e implementado una interfaz de usuario coherente y cohesionada con el problema en cuestión. Basada en tres pestañas, cada una con un propósito diferente. Con los suficientes controles que permitan el control total del usuario.
- La aplicación tiene un impacto mínimo en el sistema donde se esté ejecutando ya que solo consume tiempo de CPU cuando se realizan los cálculos necesarios.
- La pestaña en la que se realiza la carga del modelo permite al usuario la configuración de los parámetros necesarios para generar correctamente el modelo matemático, así como la elección del archivo con los datos de entrenamiento.
- También la visualización de los gráficos de control y sus respectivos límites tras generar el modelo, permitiendo así al usuario realizar las correcciones oportunas en caso de haber realizado el procedimiento incorrectamente.
- La segunda pestaña permite al usuario ver el resto de los parámetros del modelo generado. Matriz de cambio de espacio, medias y desviaciones típicas de las variables durante la fase de entrenamiento, y los autovalores (latent) junto con su porcentaje de variabilidad explicada acumulada.
- La pestaña del Modo Online permite al usuario la elección del servidor OPC del que quiere recibir los datos para ejecutar la monitorización en línea. También la visualización en tiempo real de los gráficos de control y de sus límites, así como de las alarmas y del registro de incidencias.

Simulación de una fuente de datos OPC

- Se ha creado un servidor OPC mediante el software MatrikonOPC Server for Simulation.
- Se ha implementado una pequeña aplicación en LabVIEW que permita simular la escritura en el servidor OPC a partir de un archivo con el histórico de datos de un proceso.
- La frecuencia a la que se actualizan los datos es variable y depende tanto de la interfaz de LabVIEW que lee el fichero de datos como del servidor OPC.

Resultados

- Se han expuesto los resultados del caso BATADAL, comparándolo con los resultados previamente publicados desde la organización del concurso.
- Se han expuesto los resultados del caso Cranfield obtenidos a través de MATLAB.
- Se ha realizado una breve comparación de los resultados.

Finalmente y dados todos los puntos anteriores, se puede concluir que los objetivos planteados para este Trabajo Final de Grado han sido alcanzados con éxito.

Bibliografía

[1] *A Benchmark Case for Statistical Process Monitoring - Cranfield Multiphase Flow Facility*. URL: https://es.mathworks.com/matlabcentral/fileexchange/50938-a-benchmark-case-for-statistical-process-monitoring-cranfield-multiphase-flow-facility?s_cid=ME_prod_FX (visitado 10-02-2018) (vid. pág. 28).

[2] Saurabh Amin y col. «Cyber Security of Water SCADA Systems—Part I: Analysis and Experimentation of Stealthy Deception Attacks». En: 21 (sep. de 2013), págs. 1963-1970 (vid. pág. 19).

[3] *Analysis of the Cyber Attack on the Ukrainian Power Grid*. URL: https://ics.sans.org/media/E-ISAC_SANS_Ukraine_DUC_5.pdf (visitado 08-04-2018) (vid. pág. 10).

[4] *BATADAL - BATTLE OF THE ATTACK DETECTION ALGORITHMS*. URL: https://batadal.net/challenge.html (visitado 10-02-2018) (vid. pág. 20).

[5] A.J. Ferrer-Riquelme. «1.04 - Statistical Control of Measures and Processes». En: *Comprehensive Chemometrics*. Ed. por Steven D. Brown, Romá Tauler y Beata Walczak. Oxford: Elsevier, 2009, págs. 97-126. ISBN: 978-0-444-52701-1. DOI: https://doi.org/10.1016/B978-044452701-1.00096-X (vid. pág. 14).

[6] Zhiqiang Ge y Zhihuan Song. *Multivariate Statistical Process Control*. Springer-Verlag London, 2013 (vid. pág. 12).

[7] *GitHub survived the biggest DDoS attack ever recorded*. URL: https://www.wired.com/story/github-ddos-memcached/ (visitado 02-05-2018) (vid. pág. 9).

[8] Paul Kocher y col. «Spectre Attacks: Exploiting Speculative Execution». En: *ArXiv e-prints* (ene. de 2018). arXiv: 1801.01203 (vid. pág. 8).

[9] Moritz Lipp y col. «Meltdown». En: *ArXiv e-prints* (ene. de 2018). arXiv: 1801.01207 (vid. pág. 8).

[10] J.F. MacGregor y T. Kourti. «Statistical process control of multivariate processes». En: *Control Engineering Practice* 3.3 (1995), págs. 403-414. ISSN: 0967-0661. DOI: https://doi.org/10.1016/0967-0661(95)00014-L (vid. págs. 13, 14).

[11] Joaquim Meléndez y Joan Colomer. «Statistical Process Control (SPC) and Multivariate Statistical Process Control (MSPC) for fault detection and diagnosis». IIiA-UdG, Slide 74, Escuela de Diagnostico, Peñaranda de Duero. 2006 (vid. pág. 14).

[12] Amin Rasekh y col. «Smart Water Networks and Cyber Security». En: 142 (feb. de 2016), pág. 01816004 (vid. pág. 19).

[13] C. Ruiz-Cárcel y col. «Statistical process monitoring of a multiphase flow facility». En: *Control Engineering Practice* 42 (2015), págs. 74-88. ISSN: 0967-0661. DOI: https://doi.org/10.1016/j.conengprac.2015.04.012 (vid. pág. 28).

[14] Riccardo Taormina y col. «Assessing the Effect of Cyber-Physical Attacks on Water Distribution Systems». En: (mayo de 2016), págs. 436-442 (vid. pág. 19).

[15] Riccardo Taormina y col. «BATtle of the Attack Detection ALgorithms (BATADAL)». En: (mayo de 2017) (vid. págs. 21, 25).

[16] Riccardo Taormina y col. «Characterizing Cyber-Physical Attacks on Water Distribution Systems». En: 143 (feb. de 2017) (vid. págs. 19, 21, 22).

[17] *Tennessee Eastman Challenge Archive*. URL: https://depts.washington.edu/control/LARRY/TE/download.html (visitado 10-02-2018) (vid. pág. 28).

[18] *WannaCry ransomware used in widespread attacks all over the world*. URL: https://securelist.com/wannacry-ransomware-used-in-widespread-attacks-all-over-the-world/78351/ (visitado 02-05-2018) (vid. pág. 9).

[19] *Wikipedia. Defense in depth (computing)*. URL: https://en.wikipedia.org/wiki/Defense_in_depth_(computing) (visitado 08-02-2018) (vid. pág. 20).

[20] *Wikipedia. Hadamard product*. URL: https://en.wikipedia.org/wiki/Hadamard_product_(matrices) (visitado 10-02-2018) (vid. pág. 17).

[21] *Wikipedia. Principal Component Analysis*. URL: https://en.wikipedia.org/wiki/Principal_component_analysis (visitado 08-02-2016) (vid. pág. 13).

[22] *Wikipedia. Singular-Value Decomposition*. URL: https://en.wikipedia.org/wiki/Singular-value_decomposition (visitado 10-03-2018) (vid. pág. 13).

Parte II

Presupuesto

En este documento se ha procedido a realizar un desglose completo del presupuesto necesario para la realización del diseño e implementación de la aplicación descrita en la memoria del proyecto.

PRECIOS DE MANO DE OBRA

Trabajo realizado por Ingeniero Graduado en Tecnologías Industriales.

Código			Precio (€/h)
001	h	Ingeniero Graduado en Tecnologías Industriales	30,00

PRECIOS DE MATERIALES

Se incluye el precio de los componentes informáticos necesarios para la realización de este proyecto. El precio de las licencias que se incluyen es el de licencia perpetua. MatrikonOPC Server for Simulation es gratuito y por lo tanto no se ha incluido en el cuadro siguiente.

Código			Precio (€)
002	u	Ordenador	1140,00
003	u	Sistema Operativo - Microsoft Windows 7 64 Bits	60,00
004	u	Licencia anual MATLAB 2018a	800,00
005	u	Licencia anual National Instruments LabVIEW 2017	400,00

PRECIOS UNITARIOS

Nº Actividad		Descripción	Medición	Precio	Importe
01		**DISEÑO DEL ALGORITMO**			
01.01	u	Análisis de los datos	1		1530,00 €
01.02	u	Implementación inicial MATLAB	1		3927,00 €
01.03	u	Validación en MATLAB	1		459,00 €
				Precio Total Unidad de Obra 01	**5916,00 €**
02		**IMPLEMENTACIÓN DE LA APLICACIÓN EN LABVIEW**			
02.01	u	Implementación del algoritmo	1		2703,00 €
02.02	u	Diseño de la interfaz de usuario	1		459,00 €
02.03	u	Implementación del Modo en Línea	1		1683,00 €
				Precio Total Unidad de Obra 02	**4845,00 €**

PRECIOS DESCOMPUESTOS

Nº Actividad	Código		Descripción	Rend.	Precio	Importe
01			**DISEÑO DEL ALGORITMO**			
01.01	**U01.01**	**u**	**Análisis de los datos**			
	001	h	Ingeniero Graduado en Tecnologías Industriales	10	30,00	300,00
	002	u	Ordenador	1	1140,00	1140,00
	003	u	Sistema Operativo - Windows 7	1	60,00	60,00
			Coste total de mano de obra y material			1500,00
		%	Costes directos complementarios	0,02		30,00
			Coste total de Unidad de Obra			**1530,00 €**
01.02	**U01.02**	**u**	**Implementación inicial MATLAB**			
	001	h	Ingeniero Graduado en Tecnologías Industriales	75	30,00	3050,00
	004	u	Licencia anual MATLAB 2018a	1	800,00	800,00
			Coste total de mano de obra y material			3850,00
		%	Costes directos complementarios	0,02		77,00
			Coste total de Unidad de Obra			**3927,00 €**
01.03	U01.03	u	**Validación en MATLAB**			
	001	h	Ingeniero Graduado en Tecnologías Industriales	15	30,00	450,00
			Coste total de mano de obra y material			450,00
		%	Costes directos complementarios	0,02		9,00
			Coste total de Unidad de Obra			**459,00 €**

Nº Actividad	Código		Descripción	Rend.	Precio	Importe
02			**IMPLEMENTACIÓN DE LA APLICACIÓN EN LABVIEW**			
02.01	**U02.01**	**u**	**Implementación del algoritmo**			
	001	h	Ingeniero Graduado en Tecnologías Industriales	75	30,00	2250,00
	005	u	Licencia anual LabVIEW 2017	1	400,00	400,00
			Coste total de mano de obra y material			2650,00
		%	Costes directos complementarios	0,02		53,00
			Coste total de Unidad de Obra			**2703,00 €**
02.02	**U02.02**	**u**	**Diseño de la interfaz de usuario**			
	001	h	Ingeniero Graduado en Tecnologías Industriales	15	30,00	450,00
			Coste total de mano de obra y material			450,00
		%	Costes directos complementarios	0,02		9,00
			Coste total de Unidad de Obra			**459,00 €**
02.03	**U02.03**	**u**	**Implementación del Modo en Línea**			
	001	h	Ingeniero Graduado en Tecnologías Industriales	55	30,00	1650,00
			Coste total de mano de obra y material			1650,00
		%	Costes directos complementarios	0.02		33,00
			Coste total de Unidad de Obra			**1683,00 €**

PRESUPUESTO

Nº Actividad		Importe
01	DISEÑO DEL ALGORITMO	5916,00 €
02	IMPLEMENTACIÓN DE LA APLICACIÓN	4845,00 €
	PRESUPUESTO DE EJECUCIÓN MATERIAL	10761,00 €
	13% Gastos Generales	1398,93 €
	6% Beneficio Industrial	645,66 €
	PRESUPUESTO DE EJECUCIÓN POR CONTRATA	12805,59 €
	I.V.A. 21%	2689,18 €
	PRESUPUESTO BASE DE LICITACIÓN	15494,77 €

El presupuesto asciende a la cantidad de:

QUINCE MIL CUATROCIENTOS NOVENTA Y CUATRO EUROS CON SETENTA Y SIETE CÉNTIMOS