



UNIVERSITAS
Miguel Hernández

UNIVERSIDAD MIGUEL HERNÁNDEZ
FACULTAD DE CIENCIAS SOCIALES Y JURÍDICAS DE ELCHE
TRABAJO DE FIN DE GRADO EN ESTADÍSTICA EMPRESARIAL

Biblioteca

**HERRAMIENTAS PARA EL ANÁLISIS
ESTADÍSTICO DE DATOS DE UNA RED
DE SENSORES: PASOS PREVIOS**

AUTOR:

JAVIER PÉREZ GUARDIOLA

TUTORES:

CARLOS PÉREZ VIDAL

JOSÉ VICENTE SEGURA HERAS

CURSO 2021-2022

ÍNDICE

RESUMEN	2
1. INTRODUCCIÓN	3
1.1 Sensores y redes de sensores	3
1.2 Uso de sensores en la industria	4
1.3 Uso de sensores en procesos industriales de procesos químicos	5
1.4 Sensores en procesos de producción energética y de biomasa	5
1.5 Técnicas de análisis de los datos generados por sensores	6
2. OBJETIVOS	8
3. ESTRUCTURA DE LA PLANTA DE FABRICACIÓN	8
3.1 Descripción y funcionamiento de la planta	8
3.2 Descripción de los dispositivos	9
3.3 Interfaz de comunicación	15
3.4 Estructura de la base de datos	16
3.5 Metodología estadística	18
4. ANÁLISIS ESTADÍSTICO	19
4.1 Introducción	19
4.2 Análisis on-line	22
5. CONCLUSIONES	39
6. BIBLIOGRAFÍA	42
7. ANEXOS	46
7.1 Anexo 1: Explicación de las tablas de la base de datos	46

RESUMEN

En este Trabajo de Fin de Grado se aborda el análisis de la base de datos generada por una planta de revalorización de residuos, con el objetivo de entender mejor su estructura, hacer un análisis descriptivo de la misma, encontrar posibles mejoras, tanto de la base de datos, como de los análisis que se hacen después, y sentar las bases para un posterior estudio de mayor envergadura que incluya técnicas estadísticas de predicción.

Se analiza la información recogida por sensores de temperatura, caudal o presión, entre otros, cada 30 segundos. Se establecen protocolos de acceso al sistema de recogida y transformación de la información, Programmable Logic Controller, y se realizan consultas mediante un protocolo Server Management Studio (SQL) a través de una VPN proporcionada por la empresa. Los análisis se han realizado con R y/o Power Bi.

Este análisis ha permitido introducir técnicas de preprocesamiento de datos para, además de subsanar posibles errores en los mismos, identificar mejoras, como la necesidad de resamplear las exportaciones de datos, actualizar y completar las tablas maestras que componen la base de datos, o automatizar la carga de datos a R, al igual que se ha hecho en Power Bi, y que se deben aplicar para facilitar la implementación de las técnicas estadísticas de inteligencia artificial. Es necesario completar esta primera toma de contacto con el problema de una manera más exhaustiva para poder plantear los modelos predictivos adecuados.

1. INTRODUCCIÓN

1.1 Sensores y redes de sensores

Un sensor es un dispositivo electrónico capaz de medir magnitudes físicas o químicas, como temperatura, presión o humedad, para transformarlas en variables eléctricas o en datos. Se suele integrar en una red de sensores o SN (por sus siglas en inglés Sensor Networks).

Una SN es un tipo de red integrada por una determinada cantidad de sensores que se caracterizan por tener poca capacidad de procesamiento. Los sensores que componen la red son pequeños sistemas informáticos embebidos, dedicados, como hemos comentado, a recoger datos del entorno donde se encuentran.

Los datos, tras ser recolectados por los sensores, son transmitidos hacia un nodo central conocido como *puerta de enlace* (Gateway). Este nodo se encarga de recopilar la información procedente de todos los nodos de la red y transmitirla a un dispositivo capaz de almacenar la gran cantidad de datos generados. Este dispositivo actúa como enlace entre la SN y una red TCP/IP que monitoriza los datos recolectados [1].

Sin embargo, para que estas redes de sensores resulten útiles, requieren de un diseño adecuado para el propósito con el que se crearon. Este problema se puede abordar desde distintas perspectivas, como la creación de una red fiable y de bajo consumo energético [2], el uso de algoritmos de Sensor Network Design (SND) para determinar los lugares óptimos para la instalación de sensores, maximizando la eficiencia de la planta con un determinado nivel de precisión [3], la creación de un conjunto subóptimo de redes de sensores mediante una búsqueda combinatoria de los mismos, cada red con un tamaño que a medida que aumenta, aumenta la precisión [4], la eliminación de sensores redundantes que puedan causar ruidos de medición [5], o la simulación para instalar los sensores en los lugares donde más averías o incidencias señale la simulación [6] [7].

Estas redes de sensores generan una gran cantidad de datos que requieren sistemas de almacenamiento y gestión ágiles, además, de plataformas que

permitan aprovechar esta información para el monitoreo de los procesos. Desde hace más de una década se están desarrollando estudios relacionados con la materia, mostrando buenos resultados. Podemos ver en [8] un ejemplo donde se muestra el rendimiento de una red de sensores inalámbricos usando el estándar ZigBee para sistemas de control industrial.

1.2 Uso de sensores en la industria

En los últimos años el uso de sensores y la implementación de SN en entornos industriales se ha multiplicado, permitiendo a muchas empresas de todos los sectores mejorar el control que tienen sobre sus procesos. Esto ha sido debido a factores como la reducción de tamaño y precio de los sensores, la evolución en los sistemas de comunicación o la mejora en la capacidad de procesamiento de grandes cantidades de datos.

Podemos encontrar aplicaciones de las SN en muchas áreas, como por ejemplo en la industria automovilística donde encontramos cientos de sensores en cada vehículo que hacen que la conducción sea más limpia, más segura, más económica y más cómoda [9] [10]; en la supervisión de grandes áreas de terreno, como ciudades, donde se obtienen datos más precisos mediante sensores ambientales que facilitan el estudio de la contaminación del medio ambiente o la evolución meteorológica [11] [30]; en el monitoreo de del tráfico en ciudades [29], así como la vigilancia, seguridad y control de entornos industriales [2].

Otro ejemplo de la utilización de sensores para recabar datos ambientales lo encontramos en la ciudad de Chizhou, donde se usa una red de sensores, extendida por toda la ciudad, para controlar la calidad del aire y detectar de forma precisa los focos de contaminación. Esto permite agilizar la toma de decisiones y tomar medidas concretas como el paro de las industrias cercanas a esos puntos de contaminación, en lugar de implementar soluciones que afecten a toda la industria de la ciudad [12].

Otra de estas aplicaciones es la industria nuclear, dónde se requiere de un elevado control del proceso para mantener la seguridad. Las centrales nucleares cuentan con extensas redes de sensores muy precisos que

monitorean muchos parámetros relacionados con la temperatura, las presiones, los niveles de radiación o las concentraciones de gas, entre otros. [\[13\]](#).

1.3 Uso de sensores en procesos industriales de procesos químicos

La industria química es uno de los entornos de producción en los que existe una mayor implementación de redes de sensores para evitar posibles accidentes mediante un adecuado monitoreo de todos los procesos.

Una de las áreas de aplicación es la supervisión de gasoductos. Estas canalizaciones requieren de un minucioso control para evitar que se produzcan fugas que puedan causar explosiones, intoxicaciones o una contaminación del entorno cercano a las mismas. Todo esto, obviamente, tiene una fuerte implicación económica para la empresa responsable. Podemos encontrar varios trabajos enfocados a detectar los puntos de fuga de la tubería, mediante una red de sensores y un adecuado sistema de monitorización en tiempo real [\[7\]](#), [\[14\]](#), [\[15\]](#).

Otro de los campos en los que se suele aplicar es en el control de la composición de distintas sustancias químicas que intervienen en los procesos productivos. Este monitoreo puede ser de interés para corregir errores que puedan conducir a accidentes o defectos en la producción. Es posible el control, mediante sensores, del contenido de azufre que hay en el gas natural que se va a utilizar en una planta petroquímica, lo cual permite ponerla en marcha de forma segura y evitar accidentes [\[16\]](#).

Encontramos, también, el uso de sensores para el monitoreo del estado de las bebidas y los alimentos, para medir la concentración de CO₂ en determinados espacios, para el control de procesos biológicos [\[17\]](#), así como para evaluar la concentración de vapores explosivos en procesos químicos [\[18\]](#).

1.4 Sensores en procesos de producción energética y de biomasa

Como se ha expuesto en apartados anteriores, tener un sistema de monitoreo de las plantas que sea fiable y eficaz es clave para asegurar que los procesos se desarrollen de forma adecuada. En este sentido, hay múltiples trabajos que

desarrollan sistemas de monitorización de redes de sensores con el objetivo de conocer mejor cómo se desarrolla el proceso.

Muchas empresas de producción de electricidad a partir de biomasa requieren de este tipo de sistemas. Una extensa red de sensores y un sistema de monitoreo y control adecuado puede aumentar significativamente la eficiencia de la planta, al mismo tiempo que se gana flexibilidad en la producción y se reducen las emisiones de gases contaminantes [\[19\]](#).

Muchas plantas de producción de biogás no cuentan con instrumentos para medir el volumen de gas producido por el reactor, por lo que se desarrollan sistemas de monitoreo mediante sensores para tener un mejor control del proceso. Algunas investigaciones recogen la temperatura, las presiones y el Ph, otras se centran en la composición de metano o en el caudal de gas que pasa por el reactor [\[20\]](#).

Al mismo tiempo, conocer la composición química de los fluidos que intervienen en el proceso es también importante. Se han desarrollado estudios que proponen medir la composición química de gas que se produce en tiempo real mediante una combinación de sensores basados en óxidos metálicos y un esquema de detección fotoacústica [\[21\]](#), o que se centran en la medición de las concentraciones de metanol, acetona o isopreno [\[22\]](#). También hay estudios que abarcan otros parámetros como concentraciones de polvo, temperaturas o presiones [\[23\]](#).

Muchos de estos sensores monitorean variables que son clave para la calidad y seguridad del proceso. Esto lleva a la instalación de lo que se conoce como sensores redundantes, esto es, dos o más sensores que miden una misma magnitud, en el mismo lugar, para asegurarse de que no se tomen mediciones erróneas de alguna de las variables clave del proceso [\[24\]](#).

1.5 Técnicas de análisis de los datos generados por sensores

Cuando se han recogido los datos, éstos deben ser procesados y aprovechados por la empresa. Una de las técnicas que se puede aplicar para su tratamiento es el análisis de series temporales. Una serie temporal no es

más que una sucesión de datos autocorrelados observados en tiempos consecutivos. Pueden distinguirse dentro de la misma componentes de tendencia o estacionalidad. Una herramienta para analizar este tipo de datos son los modelos de Suavizado Exponencial [1]. Con estos modelos podemos aprovechar todo el histórico de datos generado por los sensores para predecir lo que puede ocurrir en el futuro. Esta modelización da un mayor peso a las observaciones más recientes.

Otra forma de analizar este tipo de datos es recurrir a medias móviles ponderadas exponencialmente (EWMA). Este tipo de modelos realiza un preprocesamiento de los datos para eliminar el ruido de la muestra, y para la calibración de una red neuronal [25]. También se puede aplicar un modelo de regresión de componentes principales para hacer predicciones sobre la concentración de metano y dióxido de carbono [26].

También se puede hacer análisis más complejos y completos. Por ejemplo en [27], se expone un análisis en dos fases. Un primer análisis on-line, basado en el monitoreo continuo del funcionamiento de la planta, y otro análisis off-line, en el que se extraen los datos y se crean modelos predictivos, aplicando técnicas de clustering y redes neuronales para identificar el momento o lugar de fallo y tratar de prevenirlo, reduciendo así los costes de fallo o mantenimiento.

En [28], para la primera fase se realiza un clustering y aprendizaje automático mediante redes neuronales, y a continuación en una segunda fase, se ajustan modelos ARIMA de predicción de redes neuronales.

En este trabajo presentamos en la Sección 2, después de haber realizado la introducción en esta primera sección, los objetivos de este trabajo. En la sección 3 la descripción de la entrada y salida de información de la planta de fabricación objeto de estudio, y de las bases de datos que construimos a partir de la información recabada por los sensores. En la Sección 4 planteamos el análisis estadístico de esos datos, dejando para la Sección 5 las conclusiones del estudio.

2. OBJETIVOS

Analizar la estructura y contenido de un conjunto de sensores de una planta de revalorización de residuos para conocer mejor su funcionamiento, y establecer las modificaciones que se le pueden aplicar para facilitar su análisis, y adquirir los conocimientos que permitan aplicar técnicas más complejas.

3. ESTRUCTURA DE LA PLANTA DE FABRICACIÓN

3.1 Descripción y funcionamiento de la planta

La planta de la empresa hace uso de distintos residuos, como lodos de depuradoras, residuos sólidos urbanos, industriales o biomasa. Estos insumos, entran en primer lugar en un reactor rotatorio de pirólisis, dónde se extraen los componentes volátiles por acción del calor y en ausencia de oxígeno. El residuo carbonoso resultante, denominado Char, se introduce de nuevo en el mercado como un subproducto reciclado. Los componentes volátiles se envían a la cámara de condensación, donde en diferentes etapas se obtienen fracciones o subproductos de diferente valor. Con el producto que queda al final de la etapa de condensación se abastece térmicamente la planta de forma que sea autosuficiente, y mediante otros procesos, este producto se puede transformar en energía eléctrica o térmica. En la [Figura 1](#), se muestra el flujo del proceso a través de los equipos que componen la planta, según se ha explicado anteriormente.

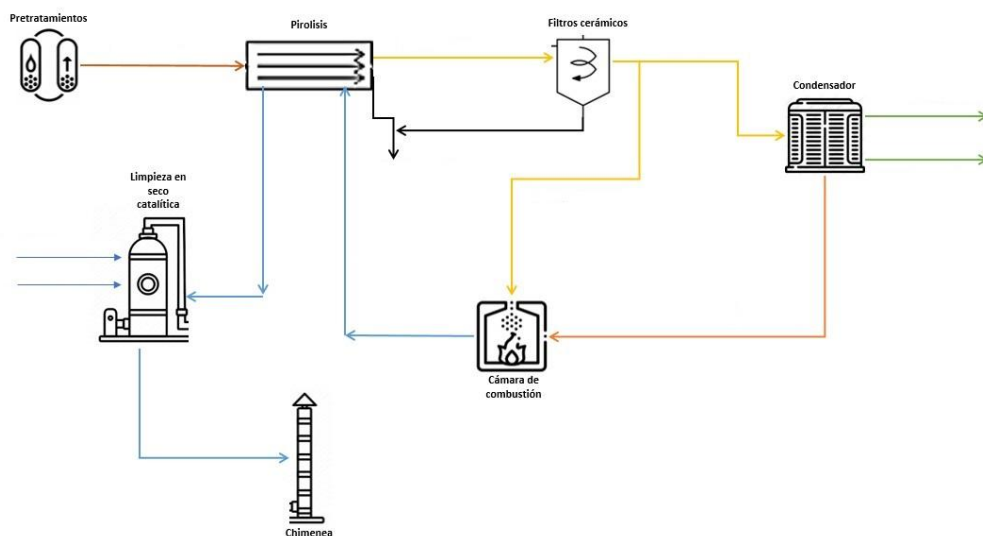


Figura 1: Esquema de los componentes de la planta de fabricación. Fuente: Greene

3.2 Descripción de los dispositivos

La planta de fabricación cuenta con varios tipos de sensores que se dedican a medir distintas magnitudes de tipo continuo, y que también registran el estatus de algunos equipos de la planta, en cuyo caso los sensores devuelven valores booleanos. También dispone de actuadores, que son dispositivos que modifican el comportamiento de la planta, por ejemplo, mediante la apertura o cierre de válvulas o el ajuste de ciertos parámetros, también devuelven valores continuos, como el porcentaje de apertura de una válvula, y booleanos, para indicar si se está actuando o no sobre algún elemento del proceso.

Se puede distinguir los siguientes tipos de dispositivos:

- Sensores de temperatura (TT). Por lo general, estos sensores son de tipo analógico, dado que recogen valores continuos de la temperatura en distintos lugares de la planta. Hay dos tipos de modelos, los primeros tienen un rango de medición entre 0°C y 1250°C, y los segundos entre 0°C y 1450°C. Sin embargo, como estos sensores analógicos generan señales eléctricas para dar sus mediciones, en muchos casos se limita la temperatura a rangos inferiores para conseguir mediciones más precisas. Algunos pueden tener asignada una cota de alarma para avisar al operador del problema asociado.
- Interruptores de temperatura (TS). Son un conjunto de sensores digitales que devuelven valores discretos, como el estado del dispositivo al que pertenecen, por ejemplo en qué posición está el interruptor.
- Sensores de presión (PT). Miden datos de tipo continuo. Hay mucha variedad de este tipo de sensores que dependen de las necesidades de la planta en cada punto del proceso, por eso cada sensor está configurado de una forma distinta para adaptarse a las necesidades en cada punto del mismo. Algunos pueden tener asignada una cota de alarma para avisar al operador del problema asociado.
- Interruptores de presión (PS). Se trata de un conjunto de sensores encargados de registrar el estado del dispositivo al que pertenecen, por ejemplo en qué posición se encuentra el interruptor.

- Sensores de caudal (QT y FT). Miden datos de tipo continuo. Las magnitudes utilizadas pueden ser milibares o kg/h. Estos sensores generan dos tipos de tags, el primer tipo, cuya codificación termina en .Val, almacena las mediciones directas de la sonda y la magnitud medida está en milibares, a continuación, a éstos se les aplica una serie de transformaciones y se ponen en valor absoluto para generar un segundo tipo de tags, que tienen el sufijo _Flow, y que se miden en Kg/h.
- Sensores de cromatógrafo (YT). Estos sensores toman datos de tipo continuo. Estas medidas representan el porcentaje de concentración en el aire de distintos compuestos como CO₂, O₂, N₂, H₂ o CO.
- Interruptores de posición (ZS). Se recoge su estado booleano, es decir, en qué posiciones se encuentran los sensores.
- Interruptores de nivel (LS). Se recoge el estado booleano de los interruptores, es decir, si están accionados o no.
- Actuadores. Incluyen electroválvulas, válvulas motorizadas, motores, motores con variadores de velocidad, quemadores, contactores o relés.
 - Electroválvulas (EV), se registran datos como el estatus del dispositivo (si están abiertas, cerradas, o en una determinada posición), las órdenes que se dan por programa (por ejemplo, ajustes de ciertos parámetros del proceso por parte del propio sistema como apertura o cierre de válvulas) o por parte del operador (ajustes de ciertos parámetros del proceso por parte del operador que controla la planta en cada momento como apertura o cierre de válvulas).
 - Válvulas motorizadas (VM), se registran datos como los ajustes de parámetros que se hacen por parte del programa o del operador, por ejemplo cambiar el % de apertura de una determinada válvula, otros como el porcentaje de apertura de las mismas en cada momento, y valores discretos como el estatus del dispositivo (válvula abierta o cerrada o si se está alimentando ciertas partes del proceso)
 - Motores (MK) se toman datos referentes a las órdenes que se dan por parte del operador o por parte del programa para el ajuste de ciertos parámetros como el arranque o parada de motores, así

como el estatus del dispositivo, esto es, si el motor está en funcionamiento o no.

- Motores con variadores (MV) se registran el estatus del dispositivo, esto es, si está en funcionamiento o no, valores continuos referentes a la velocidad del motor, medida en porcentaje, así como los comandos que dan el programa de control de la planta o el operador, para ajustar variables de control o parámetros del proceso.
- Quemadores (YKS), de éstos, se registran los datos de las órdenes que se dan por programa o por parte del operador, como el arranque o parada de los quemadores, así como su estatus, es decir, si están en funcionamiento o no.
- Contactores o relés (KM), de éstos, se registran las órdenes de encendido y apagado que dan los operadores o el programa de control de la planta, así como su estado (encendido/apagado) en cada momento.

En la [Tabla 1](#) se pueden observar algunos ejemplos de límites de medición que han recogido una serie de sensores considerando los valores mínimos, máximos y la escala de medida. Cada uno de estos sensores, a su vez, genera uno o varios tags, que son las distintas variables con las que se trabaja, como por ejemplo el valor que devuelve o el estado booleano del equipo.

Tabla 1: Límites de medición programados en varios tipos de sensores. PVEUMin es el límite inferior de medición, PVEUMax es el límite superior de medición. Fuente: Elaboración propia. Esta tabla, junto con el resto de tablas y figuras son de elaboración propia salvo que se indique lo contrario.

Tipo	TagName	PVEUMin	PVEUMax	Unidades
Temperatura	TT4120	0	1250	°C
	TT3870	0	300	°C
	TT3860	0	1450	°C
Presión	PT3770	-0.5	5	mbar
	PT3760	-0.5	25	mbar
	PT3700	-0.5	10	mbar
	PT3690	-0.5	50	mbar
	PT2960	-100	0	mbar
	PT2610	0	2000	mbar
	PT2560	0	200000	mbar
	PT1620	-50	0.5	mbar
	PT1540	-50	50	mbar
Caudal	QT3250	0	30	mbar
	QT2640	-50	50	mbar
	QT1630	0	30	mbar
	QT1520	0	300	mbar
	QT1041	0	50	mbar
	FT3252	-5	30	mbar
	FT1050	0	2000	Kg
	FT60	0	1400	Kg
Cromatógrafo	YT3470	0	20	%
	YT3420	0	3	%
	YT3410	0	20	%
	YT3380	0	5	%
	YT3370	0	30	%
	YT3360	0	90	%

Los sensores también se pueden clasificar por tipo, en concreto, en la planta han llegado a utilizarse 365 sensores a lo largo del tiempo que han generado un total de 1413 tags, pero en estos momentos no todos estos sensores están activos. En la [Tabla 2](#) se muestra, por tipo, el número de sensores que ha habido en la planta desde que empezó a funcionar.

Tabla 2: Cantidad de sensores de cada tipo en la base de datos.

Tipo de dispositivo	Cantidad
Sensores de temperatura (TT)	112
Interruptores de temperatura (TS)	7
Sensores de presión (PT)	32
Interruptores de presión (PS)	12
Sensores de caudal (QT, FT)	14
Sensores de cromatógrafo (YT)	10
Interruptores de posición (ZS)	2
Interruptores de nivel (LS)	3
Electroválvulas (EV)	18
Válvulas motorizadas (VM)	23
Motores (MK)	36
Motores con variadores (MV)	28
Quemadores (YKS)	8
Contactores o relés (KM)	13
Otros (KH, M, YF, PID, RAMP, Q_, Analiz_, D26, PI, HT, FS, V)	47

Por lo que respecta a su distribución, los sensores se encuentran repartidos a lo largo de la planta en sus diferentes equipos. En la [Tabla 3](#) se muestran los tags presentes en la base de datos, clasificados por equipo al que pertenecen. Algunos de los sensores que generan estos tags ya no están instalados en la planta porque el equipo al que pertenecen se puede haber desinstalado de la misma, pero los datos sí que están almacenados en la base de datos para análisis de puestas en marcha del pasado.

Tabla 3: Cantidad de tags de cada equipo presentes en la base de datos.

Equipo	Sensores	Equipo	Sensores
DEFAULT_SYSTEM	21	SF03_Limpieza_Syngas	14
Exhaust Gas	1	SF04_Extraccion_Syngas	54
SB1 Alimentacion MP	56	SF05_Filtrado_Final	1
SC02_Pirolizador	200	SF06_Enfriamiento_Final	2
SC03_Gasificador	167	SF08_Extraccion_Gases_Escape	43
SD01.1_Camara_Combustion_Deguisa	90	SG01_Extraccion_Cenizas	26
SD01.2_Camara_Craqueo	135	SH01_Instalacion_Neumatica	21
SE01.2_Intercambiador_Syngas-Aire_Alta	13	SH03_Instalacion_de_Gas	36
SE02_Intercambiador_Syngas-Humos	4	SI05_Antorcha	28
SE03_Intercambiador_Syngas-AireGasif	35	SJ02.1_Cuadro_1	4
SE04_Intercambiador_Syngas-Vapor	11	SJ02.2_Cuadro_2	2
SE05_Intercambiador_Syngas-AireCraqueo	54	SJ02.3_Cuadro_3	2
SE06_Caldera_Vapor	8	SK07_Analizador_Syngas	13
SE07.1_Intercambiador_Syngas-Agua_G02	57	No asignados	292
SE09_Intercambiador_GasEscape_Aire	29		

La planta objeto de estudio es una planta experimental, que cambia constantemente y que no está sujeta a la legislación que si aplica a las plantas comerciales, por lo tanto, no se está haciendo una clasificación de sensores críticos (aunque los ingenieros sí que los tienen identificados). Por la misma razón, no se aplican sensores redundantes en la planta, aunque sí se prevé hacerlo en las plantas comerciales.

3.3 Interfaz de comunicación

Los datos recabados por los sensores y actuadores se recogen en un Programmable Logic Controller (PLC) como señal eléctrica, donde se convierten en una magnitud física medible y se exportan a la base de datos operacional, almacenada en un servidor en la planta. Se realizan exportaciones cada 30 segundos de la base de datos operacional a la base de datos normalizada, almacenada en los servidores de Greene, y de ésta, tras realizar una serie de modificaciones, se exportan a la base de datos dimensional, que es la que se va a usar para este trabajo, el flujo de los datos puede verse en la [Figura 2](#). Para acceder a la base de datos es necesario hacer consultas mediante un protocolo Server Management Studio (SQL) a través de una VPN proporcionada por Greene. Los datos se cargan de forma distinta dependiendo del software en el que se utilicen.

Para Power Bi se hacen consultas a la base de datos mediante protocolo SQL, y se importan las tablas necesarias. Además, también se cargan datos de Excel, de un fichero proporcionado por Greene que recoge los límites de medición de los sensores y los rangos programados para las alarmas.

Los datos que se van a analizar mediante R, se obtienen cargando las tablas que contienen los datos a ficheros Excel tras aplicarles unos cambios en su estructura mediante Power Bi, para facilitar el análisis en R. Luego, se cargan estos ficheros a R para su posterior análisis.

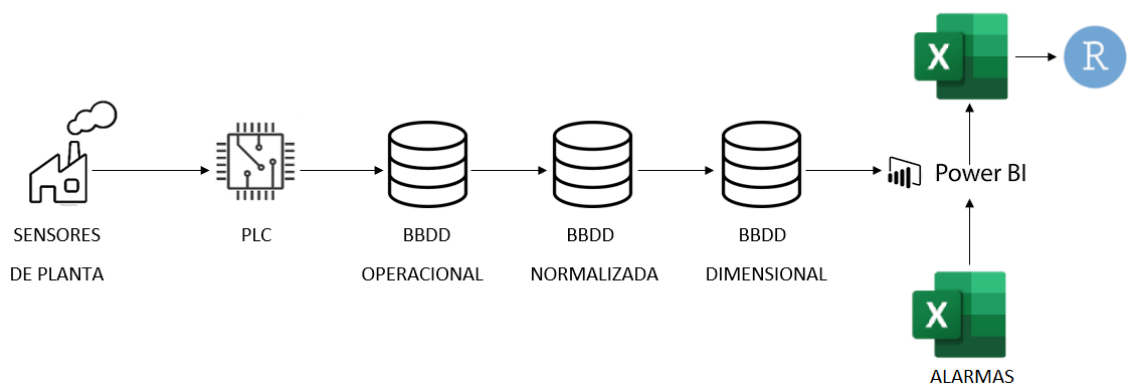


Figura 2: Diagrama de flujo de los datos.

3.4 Estructura de la base de datos

En este trabajo nos centraremos en el análisis de algunas tablas de la base de datos dimensional, en particular en aquellas que registran la telemetría del proceso, siguiendo un esquema dimensional a base de tablas maestras (D) y transaccionales (H). En el [Anexo 1](#) se puede consultar el detalle de todas las tablas que componen esta base dimensional.

- D_TAGS: Tabla maestra de tags. Esta tabla recoge todos los datos relativos a los tags o variables que hay presentes en la planta, como por ejemplo su identificador, nombre, descripción, sistema y dispositivo al que pertenece o magnitud que miden. La tabla está formada por 12 columnas:
 - Device_Code: Código del dispositivo, esto es, del sensor o actuador.
 - Device_Desc: Descripción del dispositivo.
 - Device_Type: Tipo de dispositivo, es parte del Device Code e indica qué clase de dispositivo es.
 - EU: Magnitud o dato que mide el dispositivo.
 - System_Desc: Descripción del sistema al que pertenece el dispositivo.
 - Tag_Code: Código del tag. Cada dispositivo puede tener varios tags o variables, por ejemplo, una que mida temperatura y otra el estatus del dispositivo.
 - Tag_Desc: Descripción del tag.
 - Tag_description: Descripción del tag, se trata de un campo duplicado que está pendiente de ser eliminado de la BBDD.
 - Tag_Id: Código identificativo del tag.
 - Tag_Type: Tipo de tag, los hay que recogen valores, estatus, ordenes por programa o operador a la planta, etc...
 - Device_EntryDate: Fecha de instalación del dispositivo.
 - Device_LeavingDate: Fecha de retiro del dispositivo.

- H_TAG_VAL: Es el registro histórico de las señales analógicas o tags, es decir, de la telemetría de la planta. Es en esta tabla transaccional donde se guardan las mediciones de los distintos sensores, así como las modificaciones en la apertura de las válvulas, y está formada por 5 columnas.
 - Id: Identificador de la entrada o registro.
 - Timestamp: Instante de tiempo en el que se toma el dato.
 - Tag_Id: Identificador del tag al que pertenece el registro.
 - Val: Valor o registro medido, puede ser un valor discreto o continuo.
 - Tag_Code: Código del tag al que pertenece el registro.
- D_NAPAS: Tabla maestra de estados de las NAPAs (Notificaciones, alarmas, ...) formada por 8 columnas. Contiene datos como el código, descripción o prioridad de las distintas NAPAs.
 - NAPA_Id: Identificador del NAPA.
 - NAPA_Code: Código del NAPA.
 - NAPA_Desc: Descripción del NAPA.
 - NAPA_Priority: Prioridad del NAPA.
 - NAPA_Severity: Severidad del NAPA.
 - NAPA_Type: Tipo de NAPA (Notificaciones, Alertas, Petición o Alarma).
 - NAPA_Limit: Límite del NAPA.
 - NAPA_Area: Área de actuación.
- H_NAPA: Es una tabla transaccional y registra el histórico de los cambios en los estados de las NAPAs, está formada por 14 columnas que recogen información.
 - Id: Identificador de la entrada o registro.
 - Timestamp: Momento en el que se toma el registro.
 - NAPA_Id: Código identificador del NAPA.
 - NAPA_Code: Código del NAPA.
 - NAPA_Value: Valor que ha tomado el NAPA (de la magnitud que lo activa).
 - NAPA_Status: Código que identifica el estado del NAPA (activo, reconocido, no activo).

- NAPA_Status_Desc: Descripción del estado del NAPA.
- NAPA_UserComment: Comentarios sobre el NAPA por parte del operador.
- NAPA_Active: Indica si el NAPA está activo.
- NAPA_Acked: Indica si el NAPA ha sido reconocido por el operador.
- NAPA_Disabled: Si el NAPA se ha desactivado por un fallo del propio dispositivo.
- NAPA_Supress: Indica si el NAPA ha sido suprimido por parte del operador de forma permanente.
- NAPA_Shelved: Indica si el NAPA ha sido suprimido por parte del operador de forma temporal.
- NAPA_EventChanged: Campo en desuso.

3.5 Metodología estadística

Se comienza realizando un análisis descriptivo de la base de datos objeto del estudio con el software Power Bi. Esta herramienta es un servicio de análisis de datos que proporciona Microsoft® y está orientado al diseño de gráficos interactivos a partir de los sistemas de información de las empresas. La elección de este tipo de software se debe a que es una herramienta con la que ya viene trabajando Greene.

Inicialmente se realiza un análisis online que permite visualizar las series temporales asociadas a cada variable, pudiendo así observar la evolución durante el proceso de fabricación de las magnitudes de los dispositivos utilizados, ya sea mediante datos discretos o líneas de eventos de actuadores o alarmas. Además, se han analizado las principales características de cada serie mediante parámetros como el mínimo, máximo, principales cuartiles, la media y desv. típica, y se ha realizado también un análisis de posibles outliers.

Para complementar el análisis, también se ha hecho uso de R. Este programa, dedicado a la computación estadística y ampliamente utilizado, permite realizar todos aquellos cálculos que no se pueden realizar con Power Bi. Se ha calculado el coeficiente de correlación de Pearson y en aquellos casos en los que la correlación entre dos sensores es elevada, se ha calculado también la

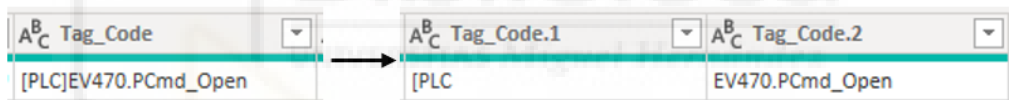
correlación cruzada. Este parámetro nos permite ver el grado de asociación entre dos sensores considerando varios retardos.

4. ANÁLISIS ESTADÍSTICO

4.1 Introducción

Para el análisis estadístico se utilizan las tablas H_TAG_VAL y D_TAGS de la base de datos dimensional, donde, como ya se ha comentado, se recogen las mediciones de los dispositivos. El fichero de Excel Rangos_Alarmas.xlsx contiene los rangos de alarma programados para cada tag, así como los intervalos de medición que admite cada uno. El primer paso consistirá en un preprocesamiento, algo habitual en la minería de datos, sobre los datos brutos para transformarlos en datos que tengan formatos que sean más fáciles de utilizar. A continuación se enumeran los cambios realizados:

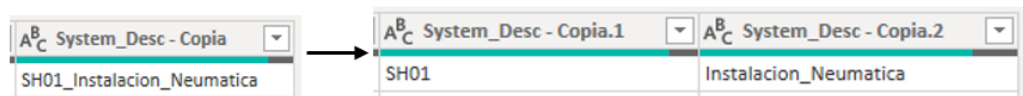
- Tabla D_TAGS
 - Tag_Code: Se elimina un prefijo genérico para poder aplicar posteriormente diferentes filtros y facilitar la lectura. (Figura 3).



A ^B _C Tag_Code	A ^B _C Tag_Code.1	A ^B _C Tag_Code.2
[PLC]EV470.PCmd_Open	[PLC]	EV470.PCmd_Open

Figura 3: Transformación aplicada a la columna Tag_Code.

- System_Desc: Se hace una copia, y en la copia se desdobra en dos campos la información de forma que aparezca el nombre de cada sistema en una columna distinta. (Figura 4).



A ^B _C System_Desc - Copia	A ^B _C System_Desc - Copia.1	A ^B _C System_Desc - Copia.2
SH01_Instalacion_Neumatica	SH01	Instalacion_Neumatica

Figura 4: Transformación aplicada a la columna System_Desc.

- H_TAG_VAL
 - De H_TAG_VAL se hacen 3 consultas distintas a SQL, una para extraer los datos de los tags de temperatura, otra para los de presión y otra para los de caudal, cada uno se almacena en una tabla distinta en Power Bi, llamadas H_TAG_VAL temperatura, H_TAG_VAL presión y H_TAG_VAL caudal respectivamente.

Luego, de cada una de ellas, se extraen 3 consultas más, una en la que se calcularán los datos de la descriptiva, otra en la que se calcularán los outliers y otra para las correlaciones, en la que se cambiará la estructura de la tabla. (Figura 5).

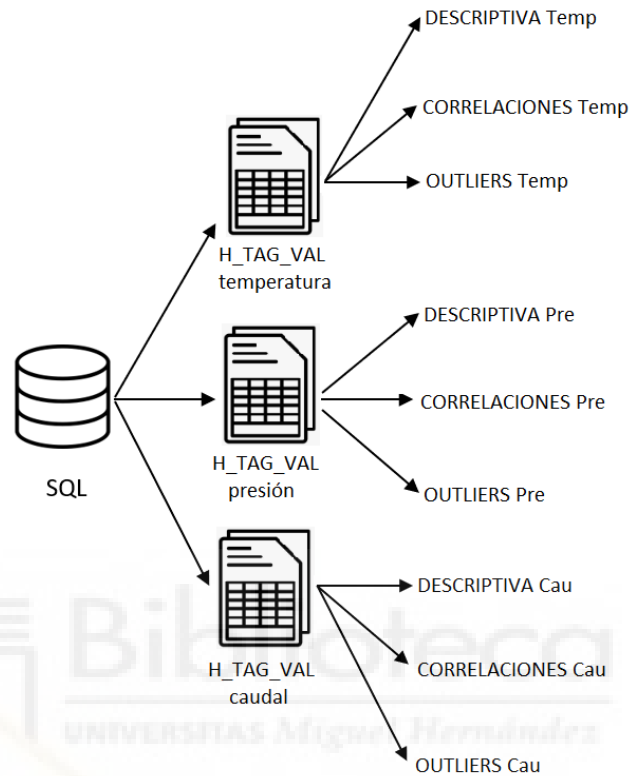


Figura 5: Flujo de datos de SQL a Power Bi.

- A las 3 consultas y a las 3 subconsultas que se hace a cada una, se les va a aplicar el mismo preprocesamiento.
- H_TAG_VAL temperatura, H_TAG_VAL presión, H_TAG_VAL caudal
 - Tag_Code: Se divide la columna Tag_Code para eliminar el prefijo genérico “[PLC]” y aplicar posteriormente diferentes filtros, así como facilitar la lectura. (Figura 6).

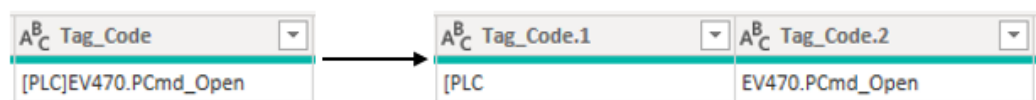


Figura 6: Transformación aplicada a la columna Tag_Code.

- CORRELACIONES Temp, CORRELACIONES Pre, CORRELACIONES Cau

- Timestamp: Se genera una variable llamada FechaHora, a partir de la variable Timestamp que agrupa los registros al intervalo de 30 segundos al que pertenecen.

Timestamp	FechaHora
04/04/2022 12:01:58	04/04/2022 12:01:30
04/04/2022 12:01:59	04/04/2022 12:01:30

Figura 7: Transformación aplicada a la columna Timestamp.

Esta columna se genera porque durante el proceso de toma de datos, se hace una toma de todos los sensores cada 30 segundos. Sin embargo, esta toma de datos no se registra correctamente, ya que como se observa en la [Figura 7](#), los datos se registran en los segundos 28 y 58 en algunos sensores, y 29 y 59 en otros, en lugar de en el mismo instante. Esto provoca que posteriormente, cuando se transpone la tabla por Timestamp para agrupar todos los registros que se han registrado en la misma toma en 1 fila, no se agrupen correctamente porque no aparecen tomados en el mismo instante, y se acaban generando 2 filas.

- Cambios en R: Tras importar los datos de Excel, se somete a éstos a un preprocesamiento para poder calcular las correlaciones cruzadas. Los datos se importan por separado, una tabla para los datos de temperatura, otra para los de presión y otra para los de caudal. El preprocesamiento es el mismo para las 3 tablas de datos, por lo que se va a explicar el que se aplica a los datos de caudal a modo de ejemplo.
 - Valores perdidos: En primer lugar se comprueba que en las variables que se quiere representar no hay valores perdidos, para no tener problemas cuando se quiera obtener el coeficiente de correlación cruzada. Para ello, se aplican las funciones `is.na()` y `any()`, que devuelven TRUE si hay algún valor perdido en cada una de las columnas.

`any(is.na(datoslimpiosCAUDAL$QT1630_Flow))`

`any(is.na(datoslimpiosCAUDAL$FT3252_Flow))`

Cuando se ha comprobado que hay valores perdidos en las columnas que se va a utilizar, se realiza un análisis más completo

en busca de valores perdidos, para ello, se usa la función `apply()` en combinación con `is.na()` para buscar los valores perdidos en las matrices de datos completas, en este caso denominada `datosCAUDAL`. Esta devuelve una lista con los valores perdidos que hay en cada columna.

```
apply(is.na(datosCAUDAL),2,sum)
```

Cuando aparece la lista con los valores perdidos, se aplica la función `na.omit()`, que elimina las filas en las que hay un valor perdido en alguna de las columnas. Se aplica este tipo de depurado porque se ha comprobado que se elimina una cantidad de filas muy pequeña en comparación con el tamaño de la base de datos.

```
datoslimpiosCAUDAL<-na.omit(datosCAUDAL)
```

Posteriormente, se vuelve a aplicar la funciones `apply()` y `is.na()` al nuevo conjunto de datos para verificar que no quedan valores perdidos.

4.2 Análisis on-line

4.2.1 Parámetros descriptivos

A continuación, se va a realizar la descriptiva de los datos o análisis on-line, para ello, se presenta un conjunto de figuras con datos referentes a las principales magnitudes recogidas en la base de datos, como temperatura, presión y caudal.

Se recogen en las tablas los valores mínimo y máximo observados, el primer y tercer cuartil y la mediana, la media y la desviación típica, el número de observaciones (n), así como el coeficiente intercuartílico (CIQ), que es el rango que hay entre el primer y tercer cuartil. Estas tablas incluyen también una referencia a los valores atípicos, utilizando como cota de los mismos:

$$LimInf = Q_1 - 1,5 * (Q_3 - Q_1) ; LimSup = Q_3 + 1,5 * (Q_3 - Q_1)$$

Se ha distinguido entre valores atípicos inferiores y superiores porque los primeros suelen estar vinculados a la puesta en marcha del proceso, donde los sensores comienzan y acaban midiendo en 0.

También incluye los rangos de alarmas que tienen codificados algunos tags. Las columnas LoLim y HiLim son los valores inferiores y superiores de alarma respectivamente, mientras que LoLoLim y HiHiLim son un segundo rango de alarmas, siendo LoLoLim el rango inferior y HiHiLim el superior. Este segundo rango de alarmas presenta valores más extremos, por lo que sobrepasarlos provoca una alarma de mayor gravedad.

Tabla 4: Análisis descriptivo y rango de las alarmas de los tags de caudal.

Tag_Code	Minimo	Cuartil 1	Mediana	Cuartil 3	Maximo	Media	Desv. Típica	CIQ	LimInf	LimSup	n	%outliersup	LoLoLim	LoLim	HiLim	HiHiLim
FT3252_Val	-0.01	0.63	0.75	0.77	0.96	0.62	0.29	0.15	0.41	0.99	11097	0				
FT3252_Flow	6.02	131.12	140.01	141.88	155.69	119.08	44.72	10.77	114.97	158.03	11101	0				
FT60_Val	-12.72	52.6	127.22	246.75	340.08	140.15	87.9	194.15	-238.63	537.98	11077	0		75		
QT1040_Val	-0.02	0.01	0.02	0.6	12.1	0.34	0.74	0.59	-0.86	1.48	11078	4.78				
QT1040_Flow	0	9.53	11.01	63.01	283.07	32.28	35.05	53.47	-70.67	143.22	11079	1.2				
QT1041_Val	-0.05	18.35	18.54	18.81	39.2	16.39	7.73	0.46	17.66	19.5	11097	20.41				
QT1041_Flow	0	616.23	619.47	623.89	902.25	540.39	217.72	7.67	604.72	635.39	11098	20.41				
QT1140_Val	-0.01	0	0.44	1.42	12.75	0.86	1.22	1.43	-2.14	3.56	11075	1.78				
QT1140_Flow	0	4.61	40.41	71.51	226.82	40.71	40.45	66.9	-95.74	171.86	11095	0.24				
QT1520_Val	-74.74	-74.74	-74.74	-74.74	-74.74	-74.74	0	0	-74.74	-74.74	11036	0				
QT1520_Flow	9.05	167.96	210.28	257.96	668.36	207.04	82.29	90	32.97	392.95	11100	1.45				
QT1630_Val	-0.04	10.09	11.77	12.49	13.5	9.73	4.58	2.4	6.49	16.09	11094	0				
QT1630_Flow	0	647.15	663.76	674.25	915.09	569.85	221.22	27.1	606.5	714.91	11100	2.72				
QT2640_Val	-0.86	-0.22	0.43	0.83	3.68	0.58	0.97	1.05	-1.8	2.41	11097	8.23				
QT2640_Flow	0.81	2.88	3.59	4.5	9.13	4.04	1.72	1.62	0.45	6.92	11100	13.77				
QT3250_Val	-0.05	0.82	1.3	1.92	12.92	1.44	1.06	1.1	-0.83	3.57	11101	2.39				
QT3250_Flow	5.7	105.72	132.35	162.36	420.68	130.31	51.8	56.64	20.75	247.33	11100	1.45				
QT380_Val	-0.06	-0.05	-0.04	-0.03	-0.02	-0.04	0.01	0.02	-0.08	0.01	11052	0				
QT380_Flow	20.09	24.6	30.13	32.54	36.9	28.95	4.56	7.94	12.69	44.46	11056	0				
QT540_Val	-0.04	0	0.02	1.23	30.3	0.98	1.84	1.23	-1.85	3.07	11091	11.31				
QT540_Flow	0	2.32	3.09	26.51	125.94	15.92	18.94	24.19	-33.97	62.8	11098	1.28				
QT680_Val	-0.37	0.11	0.55	1.28	10.23	0.83	0.89	1.17	-1.64	3.03	11086	1.83				
QT680_Flow	0	11.01	19.48	28.99	84.09	20.54	11.69	17.97	-15.94	55.94	11101	0.42				
QT680_Flow_GP	0	7.98	14.84	21.01	84.06	15.11	8.59	13.02	-11.55	40.55	11102	0.47				
QT680_Flow_GP_TT515	0	8.01	14.84	21.01	84.06	15.12	8.59	13	-11.49	40.51	11101	0.46				
QT890_Val	-0.01	-0.01	-0.01	0.43	8.08	0.2	0.54	0.44	-0.67	1.09	11062	2.93				
QT890_Flow	0.25	0.51	0.57	3.5	15.15	1.62	1.87	2.99	-3.97	7.98	11066	1.07				

La [Tabla 4](#) muestra los datos de todos los tags que miden caudales. Se observa en la misma que el tag QT1520.Val es constante en un valor de -74.74, por lo que será eliminado para los siguientes análisis. Se encuentran, además del anterior, 13 tags cuyo valor mínimo es negativo, lo que significa que, bien se trata de errores de precisión, puesto que la mayoría de ellos tienen valores inferiores a 0.1, como el QT890.Val, QT1630.Val o el QT1640.Val, o bien no están midiendo de forma correcta, ya sea por defectos de calibración, errores en la instalación u otras causas. Llama la atención el tag FT60.Val, que tiene valores negativos más grandes, debido a que su mínimo es de -12.72, y su cota inferior de alarma de 75, esto indica que puede haber habido un incidente durante el proceso.

Por otro lado, hay dos tags que presentan una gran variabilidad, QT1630_Flow y QT1041_Flow debido a que tienen numerosos picos en los que alcanzan valores muy elevados de forma temporal, como se observa en la [Figura 17](#),

esto puede deberse a fallos de calibración, comunicación entre el sensor y el PLC, errores en las transformaciones para crear estas variables a partir de las .Val o anomalías que se haya producido durante el funcionamiento de la planta.

Se observa también que muchos tags presentan un porcentaje elevado de valores atípicos por exceso, con 4 de ellos superando el 10% de los mismos, como por ejemplo el tag QT540.Val (11.31%). Este tag no tiene asociada una cota de alarma, por lo que podemos concluir que, pese a tener un número importante de valores atípicos, estos no supondrán ningún problema para el funcionamiento de la planta, lo mismo ocurre para los tags QT1041.Val (20.41%), QT1041_Flow (20.41%) y QT2640_Flow (13.77%).

Tabla 5: Análisis descriptivo y rango de las alarmas de los tags de presión.

Tag_Code	Mínimo	Cuartil 1	Mediana	Cuartil 3	Máximo	Media	Desv. Típica	CIQ	LimInf	LimSup	n	%outliersinf	%outliersup	LoLoLim	LoLim	HiLim	HiHiLim
PT100.Val	-44.76	-2.37	-0.86	-0.22	15.84	-1.82	2.72	2.15	-5.6	3.01	11096	11.02	0.13				0
PT111.Val	-0.04	4.07	4.21	4.32	4.44	3.5	1.61	0.25	3.7	4.69	11098	17.53	0		2.5		
PT1115.Val	-45.1	-6.9	-4.99	-4.29	4.36	-5.34	3	2.62	-10.82	-0.36	11100	1.4	8.14				0.5
PT118.Val	-0.06	3.87	4.06	4.17	4.28	3.35	1.57	0.3	3.42	4.62	11093	17.54	0				
PT1210.Val	0	0	0	0	0	0	0	0	0	0	11037	0	0				
PT124.Val	-44.68	-2.51	-0.92	-0.29	0.55	-1.87	2.7	2.22	-5.84	3.04	11101	11.06	0				0.5
PT1290.Val	-44.8	-8.26	-6.58	-4.85	3.16	-6.39	3.29	3.42	-13.39	0.28	11098	1.34	0.26				
PT1390.Val	-42.15	-5.17	-4.34	-2.43	3.71	-3.94	2.87	2.74	-9.28	1.68	11099	1.44	0.14				
PT1450.Val	-25.82	-6.99	-5.8	-4.46	2.54	-5.42	2.81	2.52	-10.77	-0.68	11101	1.4	8.23				0.5
PT1540.Val	-2.7	-0.11	-0.05	0.04	13.9	0.03	0.39	0.14	-0.32	0.25	11098	3.99	12.75				
PT1590.Val	-10.67	-9.21	-8.86	-7.56	3.91	-7.24	3.39	1.65	-11.68	-5.09	11100	0	21.44				
PT1620.Val	-31.27	-28.73	-27.63	-22.25	2.81	-22.47	10.68	6.48	-38.45	-12.53	11100	0	19.69				
PT215.Val	-1.14	0.03	0.21	1.14	12.03	0.99	1.84	1.11	-1.64	2.82	11093	0	10.75				
PT229.Val	-45	-6.72	-4.81	-2.87	3.98	-4.98	3.15	3.85	-12.5	2.9	11098	1.35	0.01				0.5
PT232.Val	-41.26	-3.49	-0.77	0.69	14.37	-1.24	3.25	4.18	-9.77	6.96	11096	0.56	1.35				2
PT235.Val	-44.72	-5.88	-3.72	-1.97	58.68	-3.64	5.67	3.91	-11.75	3.9	11097	1	1.09				0.5
PT2560.Val	-137.31	-122.05	-106.8	-91.54	-45.77	-99.38	25.68	30.51	-167.82	-45.77	11050	0	0				
PT2610.Val	4.42	11.13	13.27	15.1	19.22	12.87	3.08	3.97	5.19	21.05	11096	1.02	0				
PT2750.Val	-49.16	-14.99	-11.56	-6.49	1.07	-11.42	6.44	8.51	-27.75	6.27	11098	1.4	0				
PT2960.Val	-50.21	-15.59	-12.22	-6.52	1.32	-11.9	6.66	9.07	-29.19	7.08	11100	1.13	0				
PT3690.Val	-0.19	1.5	2.4	3.09	22.39	2.46	1.75	1.59	-0.88	5.47	11098	0	2.87				20
PT3700.Val	-0.16	0.4	0.49	0.57	1.19	0.47	0.16	0.16	0.16	0.81	11100	7.67	0.91				
PT3760.Val	-0.06	-0.03	-0.02	-0.01	0.07	-0.02	0.02	0.03	-0.07	0.03	11094	0	4.24				
PT3770.Val	-0.14	0	0	0.01	0.44	0.01	0.03	0.02	-0.03	0.04	11099	0.81	9.23				
PT480.Val	-40.52	-2.46	-0.95	-0.34	10.61	-1.9	2.7	2.12	-5.65	2.85	11098	11.15	0.13	-25		0.5	0
PT520.Val	-44.83	-6.36	-3.83	-2	4.23	-4.38	3.32	4.36	-12.91	4.54	11099	1.16	0				0.5
PT525.Val	-0.05	3.59	4.46	6.07	10.31	5	1.96	2.48	-0.14	9.79	11093	0	2.9				
PT861.Val	0.03	0.07	0.07	0.08	0.22	0.07	0.01	0.01	0.05	0.1	11061	4.34	0.55				

La [Tabla 5](#) muestra los datos de los tags que miden presiones. Se puede observar que el tag PT1210.Val es constante en el valor 0, por lo que se elimina para posteriores análisis.

Por otro lado, la mayor parte de los tags tienen su valor mínimo negativo, como el tag PT2560.Val, cuyo valor máximo también es negativo, mientras que el resto tienen valores máximos positivos. La anomalía que se produce con este tag puede deberse a fallos en la calibración o la colocación del sensor.

Por lo que respecta a las desviaciones típicas, tan solo dos de ellas son destacables, en comparación al resto, siendo éstas las de los tags PT2560.Val (25.68) y la del PT1620.Val (10.68).

En la columna de outliers superiores vemos que hay 4 tags con más de un 10% de outliers superiores entre sus registros, estos son el PT1590.Val (21.44%), PT1620.Val (19.69%), PT1540.Val (12.75%), PT215.Val (10.75%), sin embargo, si se atiende a las medias, se puede ver que los tags PT1590.Val y PT1620.Val tienen una media negativa, lo que indica que estos tags tienen una serie que, partiendo de valores próximos a 0, disminuye y se desarrolla en valores negativos. Esto se ejemplifica en la [Figura 18](#) con el tag PT1590.Val. Por su parte, ni PT1540.Val, ni PT215.Val tienen asociado un coeficiente de alarma, por lo que probablemente estos outliers no supongan un riesgo para el funcionamiento de la planta.

De entre los tags que tienen cota de alarma, el tag PT235.Val, presenta un máximo de 58.68 cuando su cota máxima de alarma era de 0.5, esto puede deberse a algún evento extraordinario que se haya producido en la planta, y algo similar ocurre con el tag PT480.Val, debido a que su mínimo es de -40.52, muy por debajo de la cota de alarma mínima de -25, mientras que su máximo es de 10.61, bastante por encima de la cota máxima de alarma de 0. En este caso, además, se detecta una contradicción en las cotas de alarma superiores, dado que la primera cota para el tag PT480.Val es de 0.5, mientras que la segunda es de 0, cuando ésta debería ser mayor a la primera.

Tabla 6: Análisis descriptivo y rango de las alarmas de los tags de temperatura.

Tag_Code	Minimo	Cuartil 1	Mediana	Cuartil 3	Maximo	Media	Desv. Típica	CIQ	LimInf	LimSup	n	% outliers	LoToLim	LoLim	HiLim	HHiLim
TT080.Val	13	227.62	302.61	332	397.74	265.72	111.57	104.38	71.05	488.57	11087	0				
TT095.Val	11.73	16.21	17.45	21.65	81.36	18.48	3.95	5.44	8.06	29.81	11095	0.01			100	150
TT096.Val	12.49	19.08	22.6	25.47	30.33	21.85	4.72	6.39	9.49	35.05	11099	0			100	150
TT1090.Val	0	591.03	693.48	708.77	723.79	566.28	246.71	117.75	414.41	885.4	11089	0			1000	
TT1100.Val	0	593.96	695.5	710.88	726.17	568.1	247.42	116.92	418.57	886.27	11095	0			1070	1400
TT1130.Val	0	640.18	725.25	739.94	792.06	590.99	264.3	99.75	490.56	889.56	11095	0	575			
TT1135.Val	14.83	832.54	1025.49	1127.49	1270.99	908.28	324.65	294.96	390.1	1569.93	11095	0	750		1300	
TT115.Val	12.57	53.06	185.56	196.18	202.84	145.01	75.25	143.13	-161.63	410.87	11098	0				
TT1150.Val	13.92	477.33	705.93	761.2	912.27	583.89	258.82	283.86	51.54	1186.99	11096	0			1050	
TT1201.Val	14.4	212.74	335.34	410.5	670.59	316.24	150.59	197.76	-83.91	707.14	11096	0				
TT125.Val	12.91	314.97	407.9	416.88	501.02	336.51	136.82	101.91	162.11	569.74	11090	0				
TT1250.Val	14.86	366.19	573.08	630.83	725.25	476.49	208.02	264.63	-30.76	1027.78	11093	0				1040
TT126.Val	13.18	260.4	349.03	357.91	405.98	280.39	122.29	97.51	114.13	504.18	11087	0				
TT1260.Val	14.52	275.04	484.37	528.09	584.33	397.99	176.2	253.05	-104.54	907.67	11094	0				1040
TT1265.Val	16.4	307.68	425.52	444.74	513.6	353.73	139.09	137.06	102.1	650.32	11096	0			700	
TT127.Val	12.54	56.77	78.47	155.06	290.34	109.73	77.6	98.29	-90.67	302.5	11087	0				
TT1270.Val	14.4	40.06	54.56	69.91	76.87	52.3	20.06	29.85	-4.72	114.69	11096	0			700	
TT1275.Val	14.52	278.48	459.94	513.28	561.5	384.38	167.39	234.8	-73.71	865.48	11099	0				1040
TT1280.Val	14.59	50.26	70.01	78.02	92.23	61.63	21.03	27.75	8.63	119.65	11094	0				700
TT1300.Val	14.59	212.83	367.01	402.01	476.98	306.78	131.98	189.18	-70.94	685.78	11094	0			350	
TT136.Val	14.01	257.47	352.24	359.56	423.1	287.7	116.41	102.09	104.33	512.7	11099	0			650	
TT1400.Val	12.69	28.71	100.81	165.29	236.82	99.31	63.47	136.58	-176.16	370.15	11100	0				
TT141.Val	14.53	31.5	51.09	116.44	419.04	96.1	102.11	84.94	-95.91	243.84	11095	10.22				
TT1600.Val	14.02	608.86	691.17	706.36	716.69	563.77	249.93	97.51	462.6	852.62	11096	0		450	840	
TT1610.Val	14.94	360.88	419.53	431.24	463.89	346.32	148.98	70.36	255.35	536.77	11098	0				
TT200.Val	12.96	21.72	27.1	29.99	38.48	25.74	6.33	8.26	9.33	42.38	11100	0			200	
TT223.Val	13.51	22.51	28.46	31.44	36.01	26.3	6.76	8.93	9.12	44.83	11095	0				
TT224.Val	13.64	197.68	261.04	285.3	334.47	221.95	94.74	87.62	66.24	416.74	11097	0			650	
TT226.Val	12.64	579.9	661.16	676.36	687.26	537.96	241.76	96.46	435.21	821.05	11099	0				
TT227.Val	13.58	242.62	325.5	351.77	626.05	275.72	113.27	109.15	78.9	515.49	11092	0.06			600	
TT234.Val	13.89	108.04	143.41	230.96	377.61	164.13	84.95	122.92	-76.34	415.34	11098	0			600	
TT243.Val	12.59	18.46	19.91	22.81	26.55	20.05	3.9	4.35	11.94	29.34	11100	0				
TT244.Val	12.28	18.54	20.07	23.27	26.93	20.3	4.29	4.73	11.45	30.37	11099	0				
TT246.Val	14.1	220.02	274.77	282.74	346.38	226.53	99.05	62.72	125.94	376.82	11097	0				
TT252.Val	13.8	112.13	113.35	114.15	119.21	94.5	38.44	2.01	109.11	117.17	11082	0.12				
TT253.Val	11.48	18.56	20.81	23.68	28.81	20.68	4.18	5.13	10.87	31.37	11088	0				
TT254.Val	13.06	19.66	22.34	26.55	33.02	22.7	4.85	6.9	9.31	36.9	11074	0				
TT2550.Val	13.44	18.95	20.65	24.95	29.19	21.36	4.33	6	9.96	33.95	11099	0				
TT2600.Val	13.25	21.38	23.53	28.52	33.6	24.04	5.34	7.14	10.67	39.23	11099	0				
TT2620.Val	12.59	23.18	25.47	31.86	37.39	26.09	6.78	8.68	10.16	44.87	11097	0				
TT2751.Val	13.76	51.45	86.62	103.57	137.9	78.3	31.23	52.12	-26.73	181.75	11097	0		130		
TT280.Val	15.19	298.9	364.22	374.25	392.63	299.18	127.01	75.35	185.88	487.27	11098	0				
TT2801.Val	12.06	24.38	27.83	31.18	35.71	26.67	6	6.8	14.18	41.37	11100	0				94
TT290.Val	15.02	548.11	637.96	644.51	671.45	517.69	226.44	96.39	403.52	789.1	11098	0				
TT2901.Val	13.46	26.36	30.12	33.1	37.95	28.56	6.3	6.74	16.26	43.2	11100	0			55	70
TT2971.Val	13.96	24.74	29.02	31.34	35.21	27.39	5.51	6.59	14.86	41.23	11099	0				
TT317.Val	14.94	376.77	467.49	475.24	517.23	377.44	167.66	98.47	229.06	622.95	11099	0				
TT3240.Val	14.35	24.47	28.84	31.73	37.38	27.46	5.7	7.26	13.59	42.61	11099	0				
TT335.Val	15.7	384.49	472.7	483.19	525.66	383.98	170.33	98.7	236.43	631.25	11097	0				
TT3715.Val	13.23	62.6	76.45	78.97	89.5	65.56	24.01	16.37	38.06	103.52	11099	0				
TT3800.Val	0	40.2	49.08	56.77	360.84	49.87	31.79	16.57	15.34	81.63	11097	13.9				
TT3820.Val	15.38	944.09	1149.4	1257.38	1423.45	1034.01	337.1	313.29	474.15	1727.32	11094	0		1150	1400	1450
TT3850.Val	16.04	860.31	1080.7	1182.48	1318.56	956.63	339.81	322.17	377.05	1665.74	11097	0			1400	1450
TT3860.Val	14.6	843.6	1059.56	1138.67	1250.97	923.12	327.45	295.07	401	1581.27	11095	0	950			
TT3870.Val	12.31	165.27	183.76	189.83	200.98	155.08	63.02	24.56	128.43	226.67	11099	0		130	210	
TT4120.Val	13.26	34.03	89.75	132	157.66	83.23	49.82	97.98	-112.94	278.96	11096	0				
TT490.Val	14.88	211.74	290.71	301.08	360.22	234.16	101.45	89.35	77.71	435.11	11098	0		400	650	
TT500.Val	0	215.41	272.97	295.74	779.19	235.47	112.03	80.33	94.92	416.24	11099	0.59				
TT501.Val	0	9.57	11.27	14.99	91.39	13.16	8.81	5.42	1.43	23.13	11090	7.94				
TT502.Val	0	94.03	136.49	196.25	774.91	134.32	69.42	102.21	-59.29	349.57	11099	0.01				
TT505.Val	0	50.72	88.39	137.81	223.47	93.9	56.67	87.09	-79.91	268.45	11098	0				
TT510.Val	0	22.2	26.78	31.13	78.38	27.73	11.88	8.93	8.81	44.52	11094	5.3				
TT515.Val	12.36	21.84	24.58	27.97	53.93	24.94	7.28	6.13	12.64	37.17	11094	4.65				
TT519.Val	13.86	315.38	408.67	416.73	511.64	336.19	143.55	101.34	163.37	568.74	11100	0				
TT530.Val	15.93	257.3	358.42	367.68	423.18	286.8	128.33	110.37	91.74	533.24	11096	0			650	
TT560.Val	-0.6	9.58	11.36	15.1	106.09	13.98	12.51	5.52	1.31	23.37	11091	8.2				
TT565.Val	0	7.05	9.67	13.73	106.04	11.53	6.29	6.68	-2.96	23.74	11087	7.22				
TT590.Val	12.31	20.56	23.9	26.69	32.05	22.85	5.36	6.13	11.35	35.89	11098	0				

En la [Tabla 6](#) se presenta el análisis descriptivo de los datos de los tags de temperatura. En este caso hay hasta 27 tags cuyo valor es constante a 0, bien porque ya no se usan en las puestas en marcha o porque no se han registrado correctamente, por este motivo, y para reducir el tamaño de la [Tabla 6](#) y facilitar su lectura, se han eliminado en este y posteriores análisis. Los tags eliminados del estudio son: TT1800.Val, TT1870.Val, TT1880.Val, TT1880.Val, TT1910.Val,

TT1920.Val, TT1930.Val, TT1950.Val, TT1960.Val, TT1970.Val, TT1980.Val, TT2000.Val, TT2010.Val, TT2030.Val, TT2040.Val, TT2300.Val, TT2410.Val, TT251.Val, TT295.Val, TT300.Val, TT305.Val, TT310.Val, TT315.Val, TT320.Val, TT325.Val, TT330.Val y TT506.Val.

De los tags restantes, todo salvo el TT560.Val tienen un valor mínimo igual o mayor a 0, siendo el del TT560.Val de -0.6, mientras que en los máximos hay mucha variabilidad, teniendo tags con máximos de alrededor de 30 grados, como por ejemplo los tags TT243.Val, TT244.Val o TT096.Val, y otros que superan los 1200, como el TT3820.Val o el TT1135.Val, esto puede deberse a que los distintos equipos que componen la planta están sometidos a condiciones de temperatura muy diferentes.

Las desviaciones típicas también son relevantes habiendo 12 tags que tienen una desviación típica mayor a 200 grados, con 4 de ellos superando los 300, en concreto el TT3850.Val (339.81), TT3820.Val (337.1), TT3860.Val (327.45) y TT1135.Val (324.65). Atendiendo a los máximos y mínimos de estos tags, es muy probable que se encuentren en las partes del proceso que alcanzan mayores temperaturas, puesto que partiendo de temperatura ambiente en sus mínimos, todos sus máximos superan los 1200 grados.

En la columna %outliersup, sólo hay 2 tags que superen el 10% de valores atípicos, es el caso de los tags TT3800.Val (13.9) y TT141 (10.22), aunque son tags que no tienen asociada cota de alarma.

4.2.2 Análisis de correlaciones

A continuación, se muestran un conjunto de figuras que representan los coeficientes de correlación de Pearson entre los tags de la planta. El color azul indica correlación positiva y el color rojo correlación negativa, además, a mayor tamaño del círculo e intensidad del color, mayor será el coeficiente de correlación entre los tags.

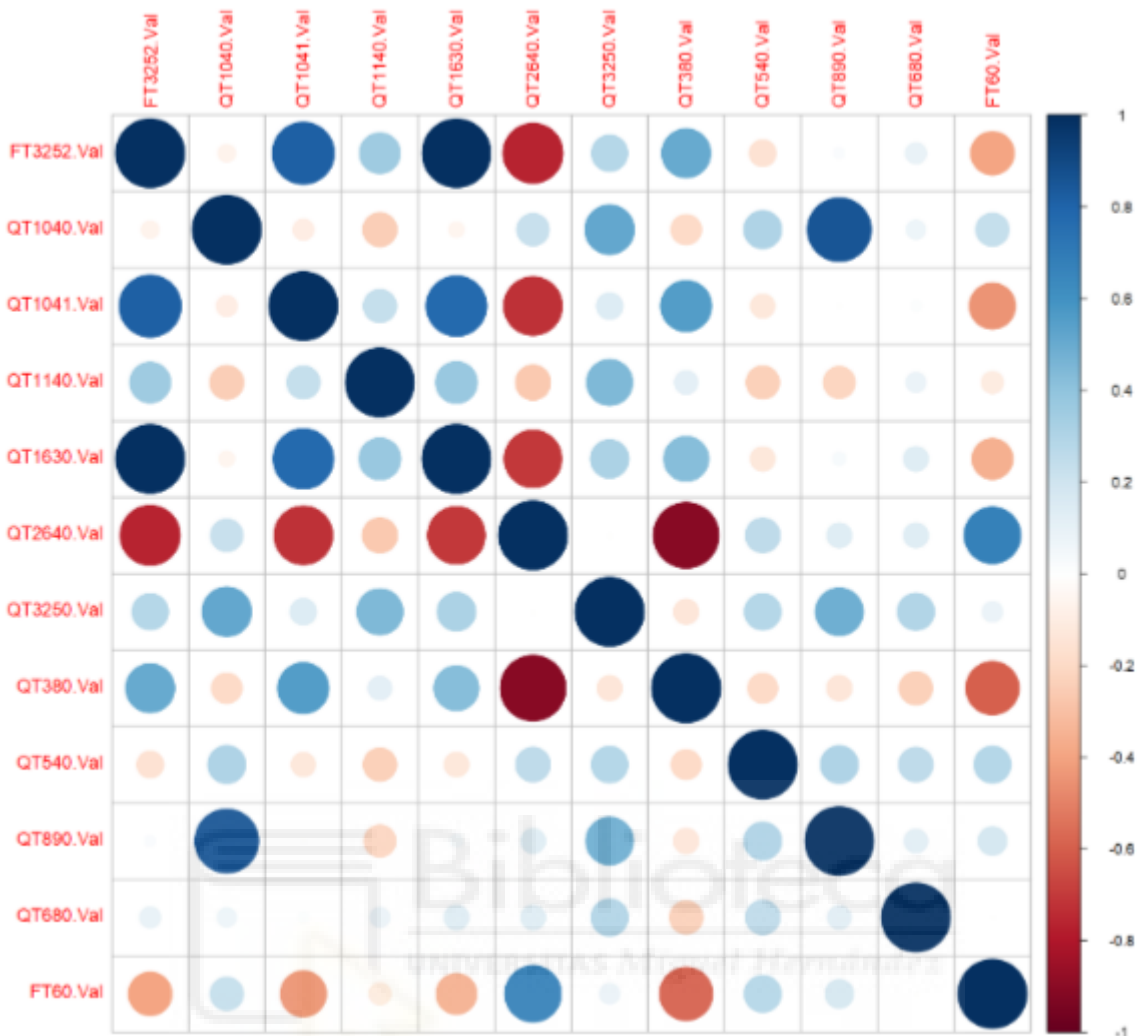


Figura 8: Gráfico de correlaciones de los tags de caudal medidos por los sensores.

En la [Figura 8](#) se muestran las correlaciones existentes entre los tags de caudal presentes en la base de datos. Se aprecia la existencia de correlaciones negativas entre distintos dispositivos, por ejemplo el tag QT2640.Val está correlado negativamente con los tags FT3252.Val, QT1041.Val, QT1630.Val, o el tag QT380.Val, mientras que está correlado positivamente con el FT60.Val. Por otro lado, también hay fuertes correlaciones positivas entre varios tags, como el QT890.Val con el QT1040.Val, o el QT1041.Val con el QT1630.Val, o el FT3252.Val con el QT1041.Val y el QT1630.Val. La correlación entre estos 3 últimos dispositivos también está relacionada con el hecho de que los 3 tengan una correlación negativa con el QT2640.Val como se ha comentado anteriormente, puesto que si están correlados entre ellos, sus correlaciones con el resto de dispositivos será similar.

Por último, se va a representar el gráfico de correlación cruzada entre dos variables altamente correladas. Se han elegido, a nivel ilustrativo, los tags QT380.Val y QT2640.Val que presentan un coeficiente de correlación de Pearson aproximadamente de -0.9, con un coeficiente R^2 de aproximadamente 0.81, esto quiere decir que el comportamiento de QT380.Val viene explicado en un 81% por el comportamiento de QT2640.Val y viceversa. En la [Figura 9](#), que representa la correlación cruzada entre ambos sensores, se observa que la correlación cruzada más alta entre ambos se encuentra en el lag 0, lo que indica que las dos series presentan una alta asociación (-0.9) en el mismo instante. El rango de cada una de estas series no permite una visualización clara de la asociación detectada.

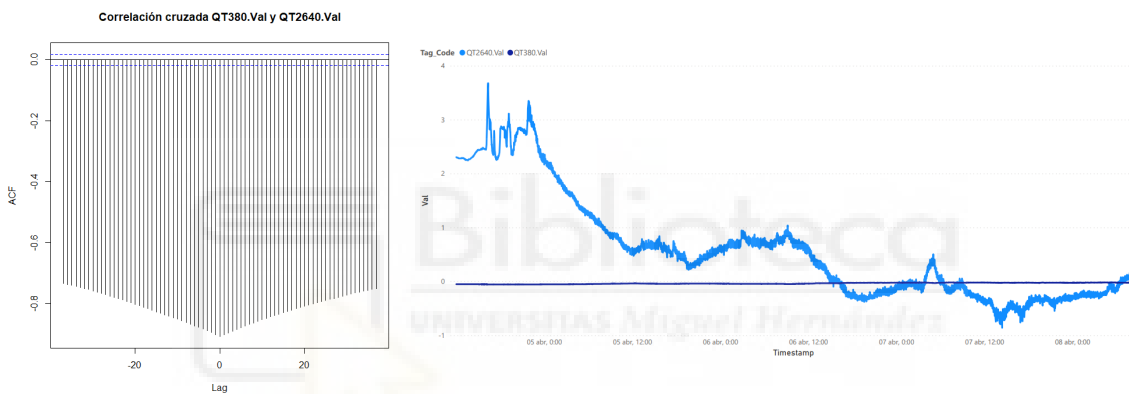


Figura 9: Gráfico de la correlación cruzada entre los tags QT380.Val y QT2640.Val.

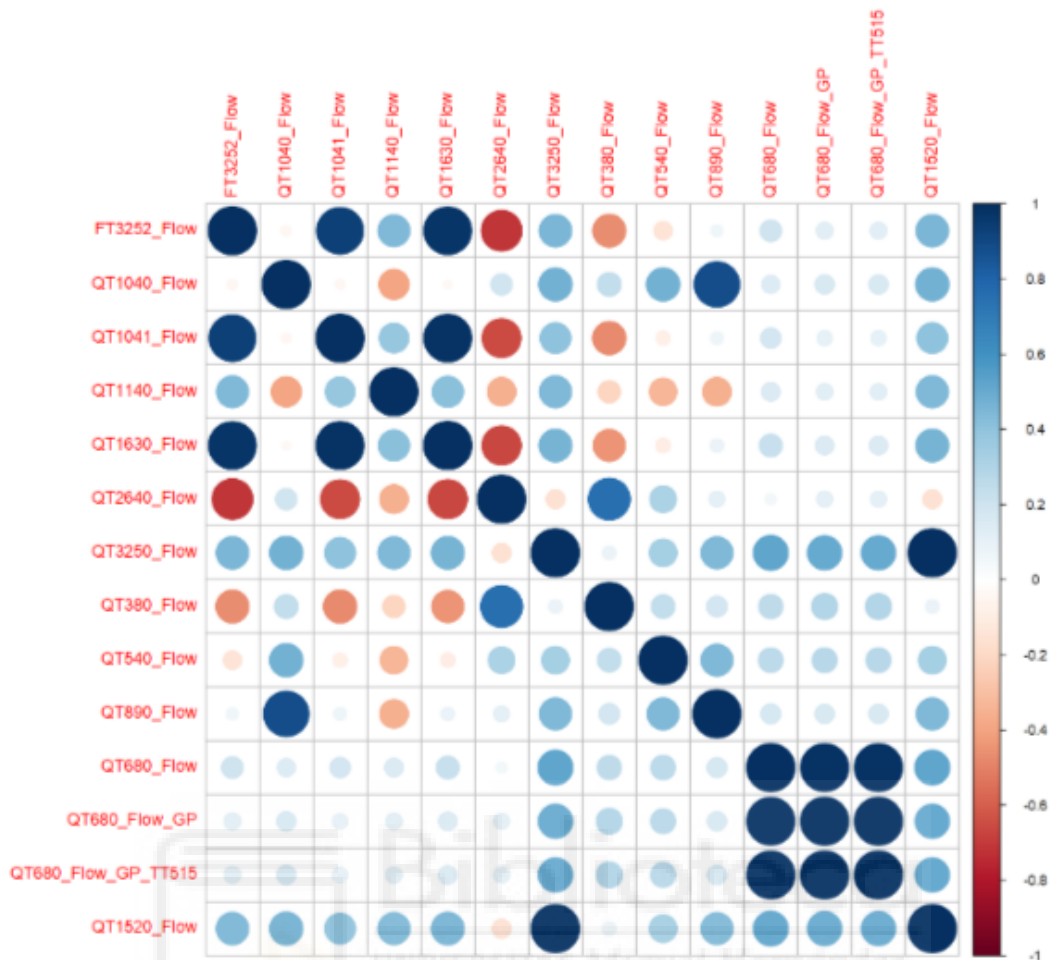


Figura 10: Gráfico de correlaciones de los tags de caudal acabados en _Flow.

En este caso, en la [Figura 10](#) se representan los tags de caudal acabados en _Flow. Al compararlo con los tags .Val, destaca el caso del tag QT380_Flow, que está correlado negativamente con los tags FT3252_Flow, QT1041_Flow, QT1140_Flow y QT1630_Flow, justo al contrario que el QT380.Val con los tags acabados en .Val de esos mismos dispositivos. Esto se debe a que el tag QT380.Val tenía registros negativos, y como el QT380_Flow es el resultado de una serie de transformaciones en valor absoluto del primero, QT380_Flow tendrá valores positivos, por lo que están correlados de forma importante, pero de forma inversa. Esta misma transformación se aplica a todos los tags de caudal, puesto que los tags acabados en .Val recogen las mediciones originales de los sensores, en milibares, para después aplicar una serie de transformaciones a partir de las cuales se crean los _Flow, medidos en kg/h y en valor absoluto.

También destaca el dispositivo QT680, que tiene 3 tags distintos acabados en `_Flow`, `_Flow_GP` y `_Flow_GP_TT515`, y que como es de esperar, están altamente correlados entre sí, puesto que proceden del tag QT680.Val.

En este caso también se mantienen las correlaciones negativas, por ejemplo, el tag QT2640_Flow está correlado negativamente con los tags FT3252_Flow, QT1041_Flow, QT1630_Flow, mientras que ahora, la correlación con el tag QT380.Val es positiva, dado que como se ha comentado anteriormente, al crearse el QT380_Flow éste está en valor absoluto, cambiando el signo de la correlación en este caso. El resto de tags que están correlados positivamente en la [Figura 8](#), también están correlados positivamente en la [Figura 10](#) en sus respectivos tags `_Flow`.

Ahora, se va a analizar la correlación cruzada entre 2 de los tags que tienen una alta correlación de Pearson. En este caso, se ha calculado la correlación cruzada para los tags FT3252_Flow y QT1630_Flow. Como se aprecia en la [Figura 11](#), el máximo de correlación cruzada se alcanza en el retardo 0, coincidiendo con el coeficiente de correlación lineal de Pearson, el representado en la [Figura 10](#) puesto que las series ocurren simultáneamente, no hay retardo o desfase entre los datos.

El coeficiente de correlación de Pearson, que es de aproximadamente 0.94, implica un coeficiente de R^2 de aproximadamente 0.88, esto quiere decir que el comportamiento de FT3252_Flow viene explicado en un 88% por el comportamiento de QT1630_Flow y viceversa.

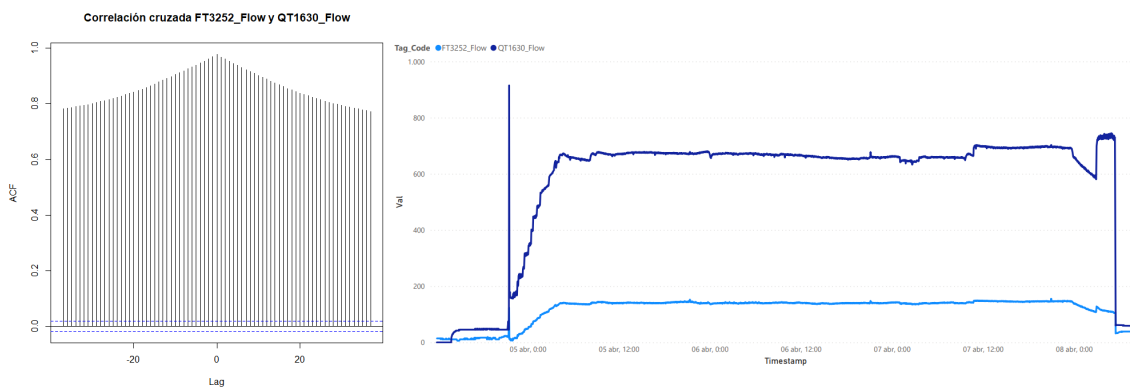


Figura 11: Gráfico de la correlación cruzada entre los tags FT3252_Flow y QT1630_Flow.

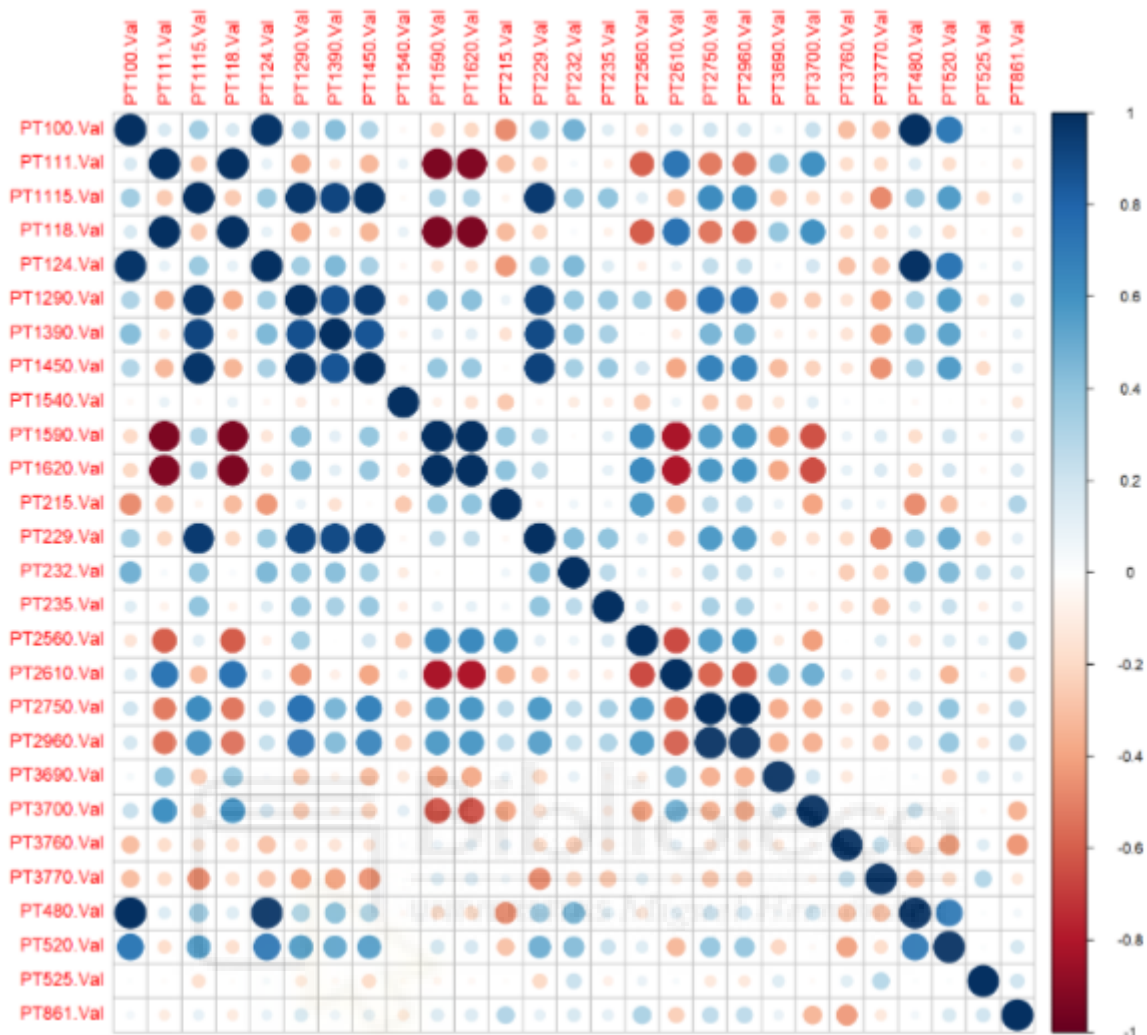


Figura 12: Gráfico de correlaciones de los tags de presión.

En la [Figura 12](#), se muestra la matriz de correlaciones para cada uno de los tags de presión presentes en la planta, en este caso, hay uno por sensor. Se aprecian importantes correlaciones positivas entre los tags PT1115.Val, PT1290.Val, PT1390.Val, PT1450.Val y PT229.Val, así como entre los tags PT1590.Val y PT1620.Val, entre PT2750.Val y PT2960.Val o entre el PT480.Val con los tags PT100.Val y PT124.Val, entre otros.

Por otro lado, también existen correlaciones negativas, como las que hay entre PT1590.Val y PT1620.Val con los tags PT111.Val, PT118.Val y PT2610.Val, que a su vez están correlados positivamente entre sí. Esto puede deberse a que estos grupos de sensores se encuentran en la misma parte del proceso o en dos partes muy conectadas entre sí.

A continuación, se analiza la correlación cruzada entre dos tags de presión altamente correlados en la [Figura 13](#). Se han seleccionado los tags PT100.Val y PT124.Val que, de acuerdo con el gráfico de correlaciones de Pearson representado en la [Figura 12](#), tienen una correlación por encima del 90%. En este caso, la correlación cruzada más alta se da en el retardo 0 del gráfico, lo que significa que la asociación máxima entre ambas series temporales se produce en el mismo instante.

En este caso, el coeficiente de correlación de Pearson, que es de aproximadamente 0.97, implica un coeficiente R^2 de aproximadamente 0.94, esto quiere decir que el comportamiento de PT124.Val viene explicado en un 94% por el comportamiento de PT100.Val y viceversa.

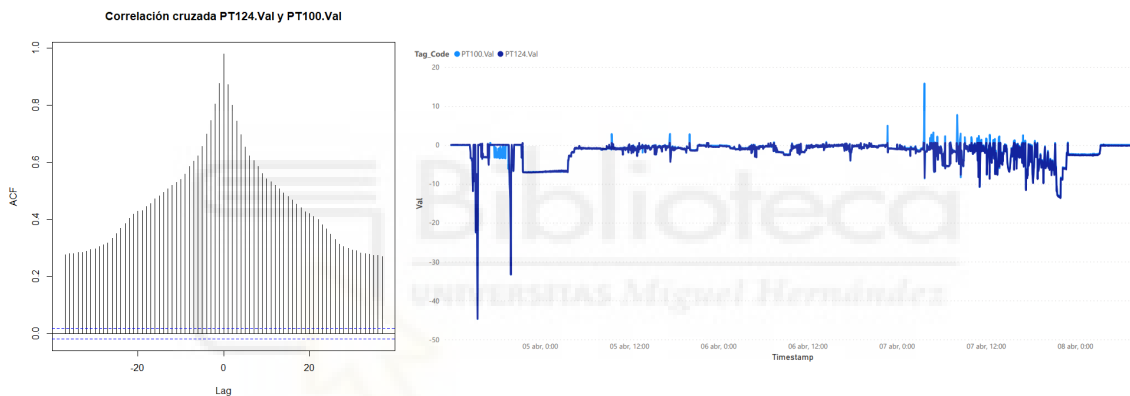


Figura 13: Gráfico de la correlación cruzada entre los tags PT124.Val y PT100.Val.

Por último, se va a analizar la correlación entre los tags de temperatura.

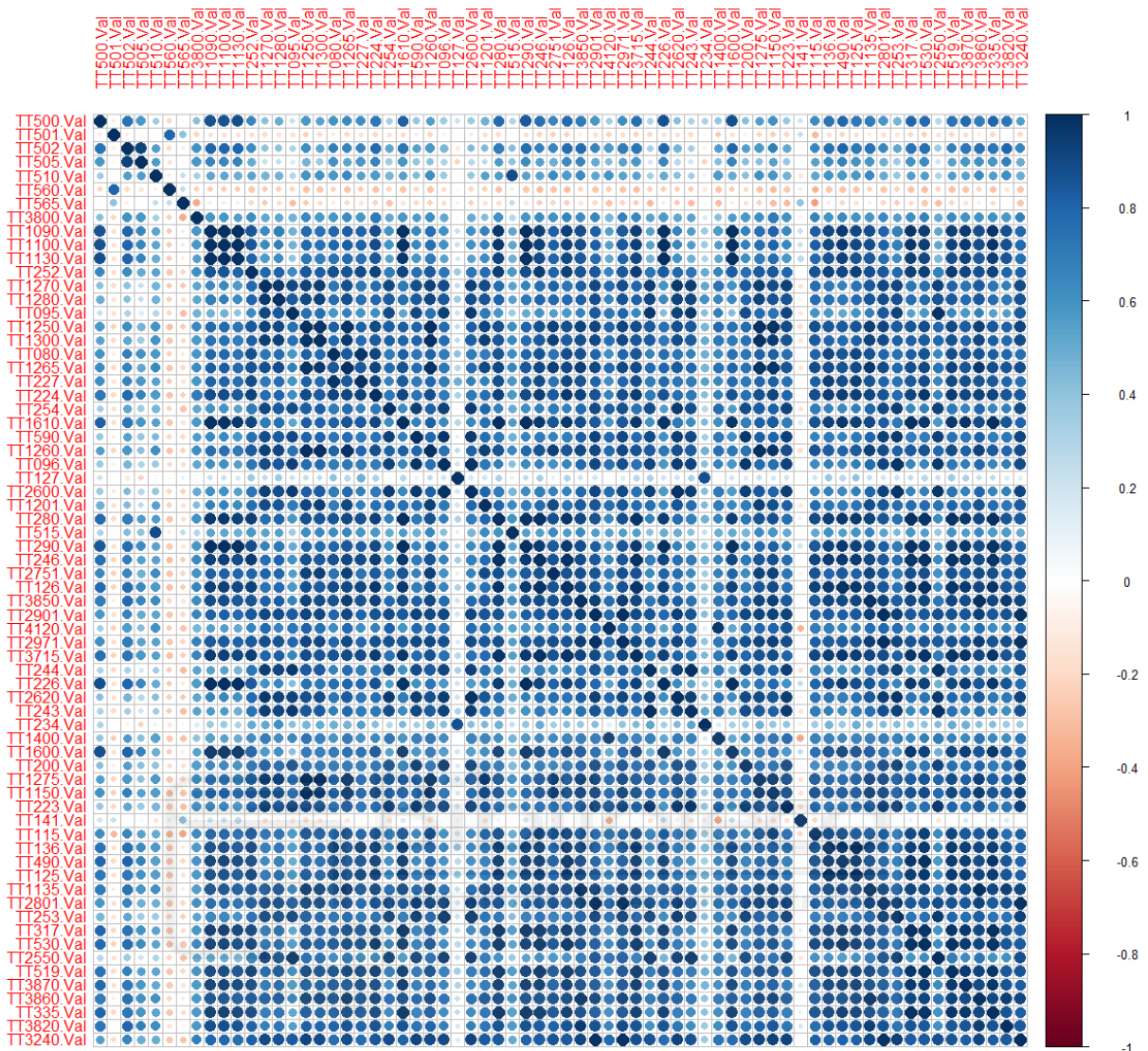


Figura 14: Gráfico de correlaciones de los tags de temperatura.

La [Figura 14](#) muestra las correlaciones entre los distintos tags de temperatura. En este caso se muestra que hay correlaciones positivas entre la mayor parte de los tags presentes en la planta, sin embargo, los tags TT127.Val, TT141.Val están correlados de una forma más débil con todos los demás, mientras que los tags TT501.Val, TT560.Val y TT565.Val están correlados de forma negativa, aunque también débil, con el resto. Esto puede significar que estos tags se encuentran en alguna parte del proceso que funciona de forma ligeramente diferente a los demás.

Por último, se va a analizar la correlación cruzada entre dos tags de temperatura altamente correlados, para ello, se ha seleccionado los tags TT096.Val y TT590.Val. En el gráfico de correlaciones de Pearson, [Figura 14](#),

se observa que ambos tags están altamente correlados de forma positiva, superando el 90% de correlación. Ahora, con la correlación cruzada calculada y representada en la [Figura 15](#), se comprueba que la correlación cruzada es muy superior al 90%. Además, el valor más alto se da aproximadamente en el retardo 0. Esto también se puede comprobar con la serie temporal representada en la [Figura 15](#).

En este caso el coeficiente de correlación de Pearson es de aproximadamente 0.96, implica un coeficiente R^2 de aproximadamente 0.92, esto quiere decir que el comportamiento de TT096.Val viene explicado en un 92% por el comportamiento de TT590.Val y viceversa.

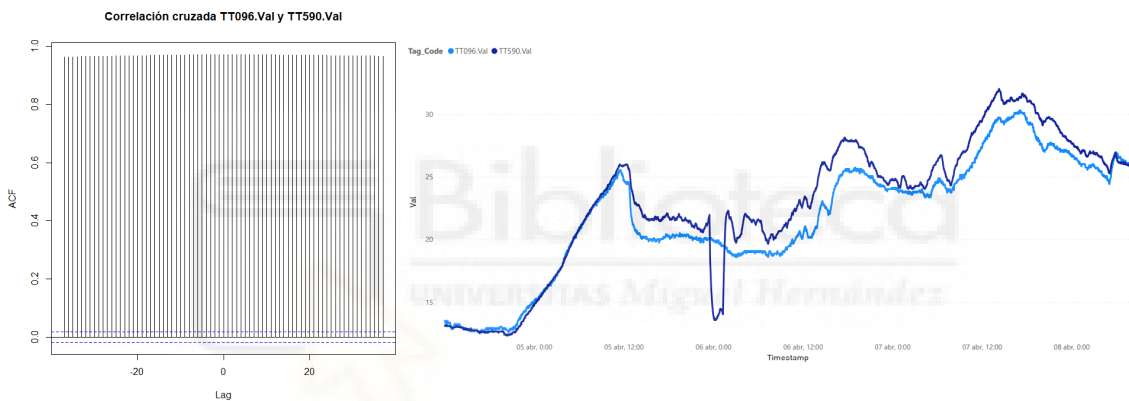


Figura 15: Gráfico de la correlación cruzada entre los tags TT096.Val y TT590.Val.

4.2.3 Visualización en Power Bi

Como se ha comentado en el apartado 4.2.2, los sensores de caudal generaban una primera serie de tags cuyo sufijo era .Val, medidos en mbar, a los que posteriormente se les aplicaba una serie de transformaciones y se generaba otro conjunto de sensores acabados en _Flow, medidos en Kg/h. Por este motivo, para los dispositivos de caudal se va a representar dos gráficos, uno para cada tipo de tag.

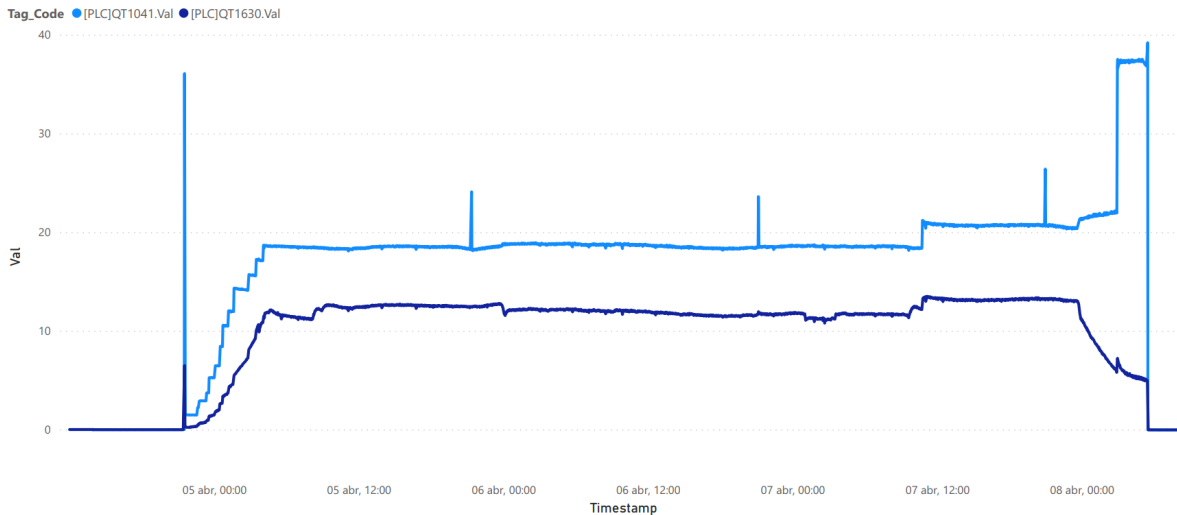


Figura 16: Gráfico de la serie temporal de los tags de caudal QT1041.Val y QT1630.Val. En la [Figura 16](#), se representa la evolución temporal de los tags QT1041.Val y QT1630.Val. Como se ha comentado en la [Figura 8](#), estos dos tags están altamente correlados, por lo que ahora, al representarlos, se puede comprobar que esto era cierto, puesto que siguen, con valores distintos, un patrón similar en su evolución a lo largo del tiempo. También se puede observar que la serie temporal comienza en 0, crece durante un periodo de tiempo, y posteriormente vuelve a bajar hasta volver a 0, esto se debe al funcionamiento de la propia planta, puesto que al ser una planta experimental, se realizan distintos ensayos que implican arranques y paradas. Para entender a que se deben los picos que se producen en la serie QT1041.Val, sería necesario hacer un análisis en profundidad, analizando los eventos que han ocurrido en ese instante y los anteriores.

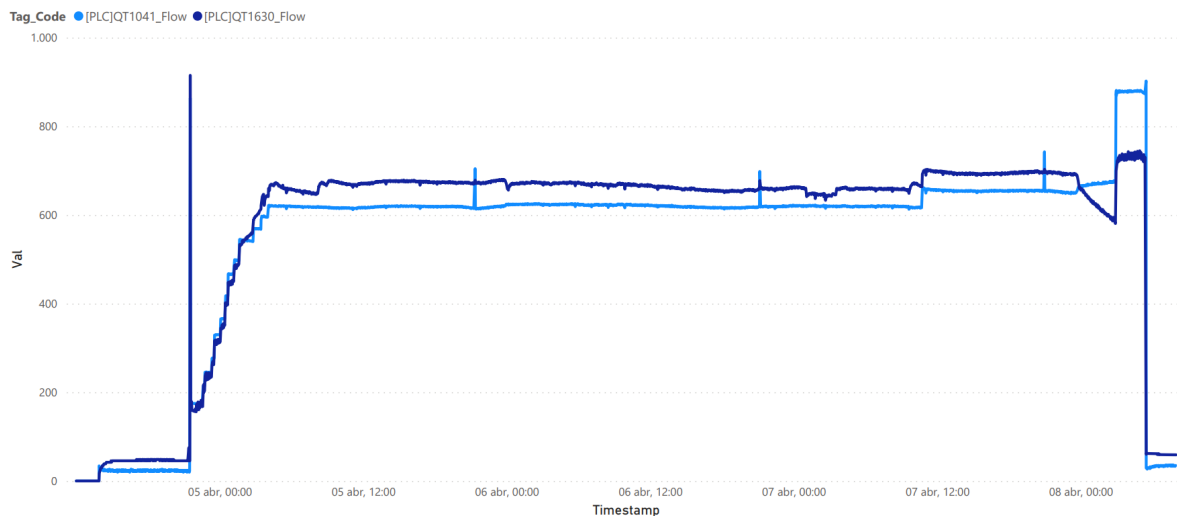


Figura 17: Gráfico de la serie temporal de los tags de caudal QT1041_Flow y QT1630_Flow.

En la [Figura 17](#) aparece representada la serie temporal de los tags de caudal QT1041_Flow y QT1630_Flow, estos pertenecen a los mismos sensores que los que se representa en la [Figura 16](#), pero tras aplicarle esas transformaciones, las series han sufrido cambios. En primer lugar, si antes el máximo de las dos series era inferior a 40 unidades, ahora sobrepasa las 800, además, también han incrementado las similitudes entre sus patrones de comportamiento.

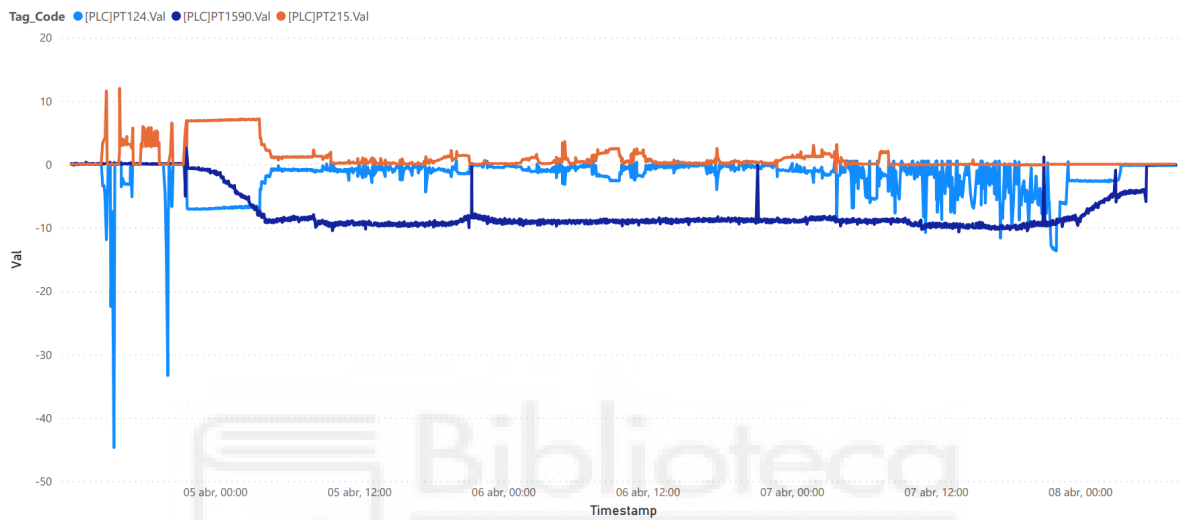


Figura 18: Gráfico de la serie temporal de los tags de presión PT124.Val, PT1590.Val y PT215.Val.

La [Figura 18](#), representa la evolución temporal de un conjunto de tags de presión, asociados cada uno a un sensor diferente. En este caso, la [Figura 18](#) ayuda a comprobar lo que se había calculado en la [Figura 12](#), en la que se obtenía que los tags PT124.Val y PT215.Val estaban altamente correlados de forma inversa. Por otro lado, el tag PT1590.Val, muestra una evolución más regular a lo largo del tiempo, y como ya se comentó en la descriptiva en la [Tabla 5](#), hay ciertos tags de presión que tienen una serie temporal que se desarrolla en valores negativos.

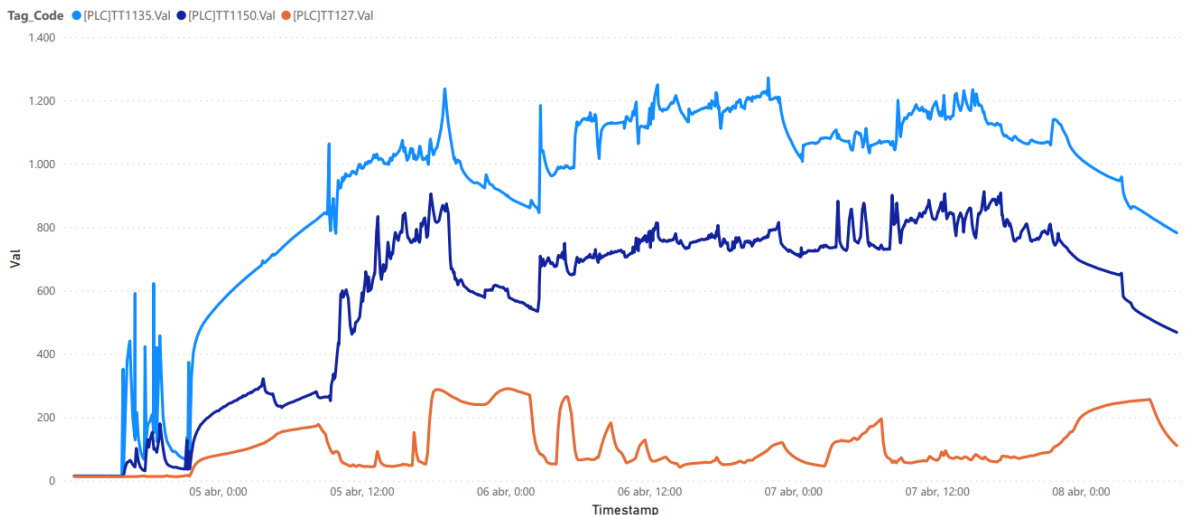


Figura 19: Gráfico de la serie temporal de los tags de temperatura TT127.Val, TT1135.Val y TT1150.Val.

La [Figura 19](#) representa la evolución temporal de 3 tags de temperatura, correspondientes a 3 sensores distintos. Al igual que en la [Figura 16](#) y la [Figura 18](#), se ha representado dos tags que según la [Figura 14](#) están correlados y otro que no lo está. Efectivamente, los tags TT1135.Val y TT1150.Val están correlados entre sí, y además, su evolución no se parece tanto a la del tag TT127.Val, tal y como se había comprobado en la [Figura 14](#) de correlaciones de tags de temperatura, donde se obtenía que el tag TT127.Val no estaba altamente correlado con el resto de tags.

En el apartado en el que se analizaba las correlaciones de los tags de temperatura, se ha comentado que había un grupo de 3 tags (TT501.Val, TT560.Val y TT565.Val), que estaban correlados de forma negativa con el resto. En la [Figura 20](#) se representa la serie temporal de los tags TT560.Val y TT1150.Val, que estaban correlados negativamente entre sí. Efectivamente, se puede comprobar que, aunque de forma débil, el tag TT560.Val desciende cuando el TT1150.Val aumenta, y viceversa.

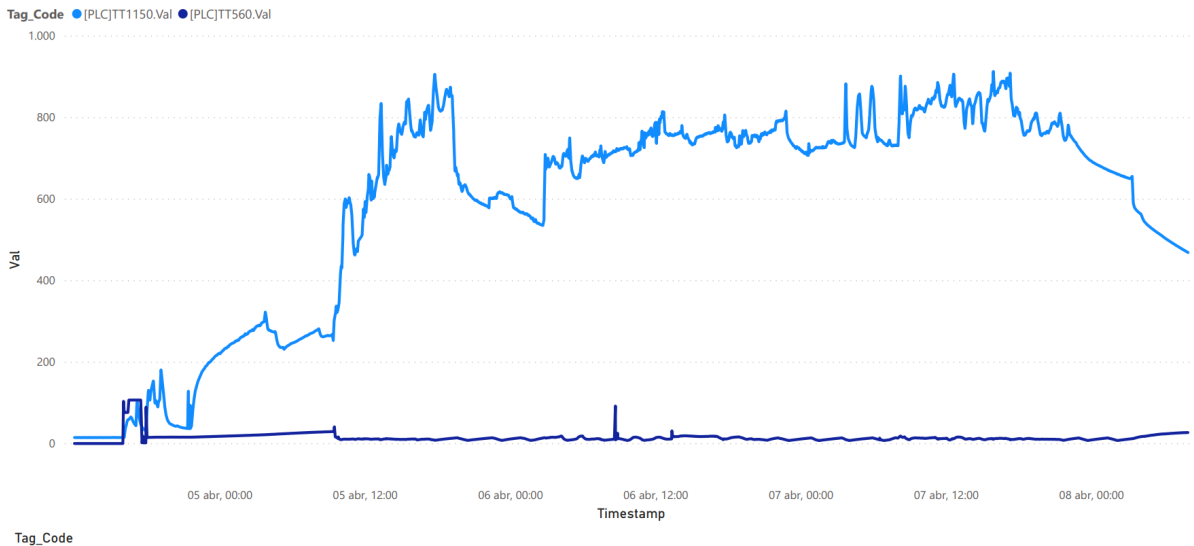


Figura 20: Gráfico de la serie temporal de los tags de temperatura TT560.Val y TT1150.Val.

5. CONCLUSIONES

A modo de cierre de este trabajo, se puede señalar que tras haber hecho una investigación exhaustiva de la base de datos y sus principales características, y haber concluido el análisis descriptivo de la misma, se ha adquirido un mejor conocimiento del funcionamiento de ésta, lo que facilitará la continuación de futuros análisis, y ayudará a seleccionar los mejores métodos para éstos.

Al mismo tiempo, el desarrollo del trabajo también ha permitido detectar una serie de mejoras que se deberían aplicar para facilitar éste y otros análisis, y que se exponen a continuación.

En primer lugar, una de las principales dificultades que se ha encontrado ha sido el manejo de la estructura de la base de datos. Las tablas transaccionales, que guardan el histórico de los datos, toman por norma general los registros de magnitudes y estados cada 30 segundos, y los cambios en los parámetros de válvulas y actuadores se registran en el momento en que se producen. Sin embargo, al no registrarse todos los datos en el mismo instante, los cambios que se aplican a la estructura de la tabla, como transponerla para generar una columna por cada tag y una fila cada 30 segundos se dificultan de forma significativa. Es por esto por lo que sería conveniente aplicar un remuestreo a la columna Timestamp de la tabla H_TAG_VAL y otras tablas que contengan este tipo de columna, para facilitar los cambios que se necesitan hacer en la

estructura de los datos cuando se importan a Power Bi o R. El remuestreo consistiría en que los registros que se hayan tomado se redondean al intervalo de 30 segundos al que pertenecen.

En segundo lugar, en este trabajo se han analizado las principales magnitudes de la tabla transaccional H_TAG_VAL, es decir, los registros históricos de temperaturas, presiones y caudales. Sin embargo, cuando se encuentran valores atípicos, o se sobrepasan los rangos de alarmas en alguno de los sensores, no se pueden analizar conjuntamente con los eventos de alarma que se han producido en la planta. Por este motivo, sería recomendable incluir en los siguientes análisis las tablas H_NAPA y D_NAPAS, que son las tablas transaccionales y maestras de las alarmas. Esto ayudaría a detectar alarmas y ver sus implicaciones en la evolución de las magnitudes.

Además de lo anterior, para realizar un análisis completo, también se debe incluir los eventos que se registran en la tabla H_TAG_VAL, como apertura o cierre de válvulas o ajustes de parámetros del proceso.

En el trabajo también se han detectado faltas de actualización de algunas columnas de la base de datos, como por ejemplo las columnas Device_EntryDate y Device_LeavingDate, que ha impedido determinar con exactitud que dispositivos estaban instalados en la planta y cuáles no, debido a que numerosos tags tenían vacíos los dos campos, por lo que sería recomendable actualizar estos campos.

Otra problemática detectada durante el trabajo ha sido la dificultad para automatizar la carga de datos en R, puesto que no ha sido posible crear una conexión directa entre R y los servidores de SQL, y tampoco de aplicar el preprocesamiento requerido para darle a los datos la forma que se necesitaba, por lo que se ha recurrido a copiar las tablas que se generaban en Power Bi tras aplicarles el preprocesamiento.

Como conclusión, son necesarias más investigaciones, aplicando los cambios señalados en este apartado para facilitar los análisis futuros, y profundizando en el análisis de las relaciones entre tags, ampliándolo a relaciones entre distintos tipos de tags. Esto ayudará a tener una visión más amplia del

funcionamiento de la planta, y de cómo unos tipos de tags influyen en otros, puesto que hasta ahora solo se ha analizado las correlaciones entre tags del mismo tipo. Este tipo de análisis nos permitiría identificar, por ejemplo, las válvulas que influyen en la presión que se detecta en un tag para de este modo, saber dónde actuar cuando se desee aumentar o reducir la presión en esa parte del proceso.

Por otro lado, estas investigaciones pueden ayudar a decidir que tipo de técnicas se puede aplicar en el futuro dependiendo de los objetivos de Greene, y que como se ha visto en la literatura, pueden ir desde el análisis de componentes principales para reducir el número de variables, en este caso de sensores, hasta la aplicación de modelos de predicción basados en redes neuronales o técnicas de análisis de series temporales como modelos ARIMA o EWMA.



6. BIBLIOGRAFÍA

1. Cantero Diaz A, Goire Castilla MM, Quintana Cassulo Y. Sistema para la gestión y análisis de datos de una red de sensores inalámbricos basado en un almacén de datos. *Revista Cubana de Ciencias Informáticas*, vol 13, núm 3. 2019 Jul 9;76–90.
2. Aponte Luis J, Galan JA, Gómez Bravo F, Sánchez Raya M. An Efficient Wireless Sensor Network for Industrial Monitoring and Control. *Sensors*. 2018 Jan 1;18(1).
3. Paul P, Bhattacharyya D, Turton R, Zitney SE. Sensor Network Design for Maximizing Process Efficiency: An Algorithm and Its Application. *AIChE J*. 2015 Feb;61(2):464–76.
4. Rameh, H., Tran, C.-T., Zoughaib, A., Evans, M.-A., & Gourelia, J.-P. (2017). Optimal Sensor Network Design to Monitor the Energy Performances of a Process Plant. In A. Espuna, M. Graells, & L. Puigjaner (Eds.), *27th European Symposium on Computer-Aided Process Engineering (ESCAPE)* (pp. 1507–1512). *Computer Aided Chemical Engineering*.
5. Nabil, M., & Narasimhan, S. (2012). Integrated Sensor Network Design. *11th International Symposium on Process Systems Engineering (PSE)*, 1522–1526.
6. Loong, M. L. H., Hanafi, M. Z. M., & Ismail, F. S. (2015). Thermal Pipeline Sensor Placement Model For Fault Detection. In A. Soh, H. Selamat, R. Rahman, A. Ishak, S. Ahmad, H. Ramli, & A. Faudzi (Eds.), *10th Asian Control Conference (ASCC)*. *Asian Control Conference ASCC*.
7. Jiang, Y., Zhou, Q., Xiao, S., Qian, M., Jiang, Z., Liu, J., Yin, W., & Jia, L. (2020). Optimized deployment method and performance evaluation of gas sensor network based on field experiment. *JOURNAL OF AMBIENT INTELLIGENCE AND HUMANIZED COMPUTING*, 12(1), 729–744.
8. Khakpour, K., & Shenassa, M. H. (2008). Industrial Control using Wireless Sensor Networks. *3rd International Conference on Information and Communication Technologies*, 1927–1931.

9. D'Orazio, L., Visintainer, F., & Darin, M. (2011, March 16). Sensor Networks on the Car: State of the Art and Future Challenges. Design, Automation and Test in Europe Conference (DATE).
10. Marek, J., & Finkbeiner, S. (2004). Micromechanical sensors - Driver for new systems in automotive applications. Conference on Focus on Vehicle Electronics, 201–207.
11. Lan, S., Miao, Q., & Du, J. (2008). Architecture of wireless sensor networks for environmental monitoring. In L. Qi (Ed.), International Workshop on Education Technology and Training/International Workshop on Geoscience and Remote Sensing (ETT and GRS 2008).
12. Li, X., Sun, M., Ma, Y., Zhang, L., Zhang, Y., Yang, R., & Liu, Q. (2021). Using Sensor Network for Tracing and Locating Air Pollution Sources. IEEE SENSORS JOURNAL, 21(10), 12162–12170.
13. Yadav, M. C. (2019, September 4). Role of Sensors (Nano) in Nuclear Technology. 1st Symposium of Emerging Nuclear Technology and Engineering Novelty (SENTEN) - Discovering Science and Engineering Novelty for Improving Human Life Prosperity.
14. Zhou, K., Li, F., Cai, H., Jing, Y., Zhuang, J., Li, M., & Xing, Z. (2022). Estimation of the natural gas leakage source with different monitoring sensor networks in an underground utility Tunnel: From the perspectives of energy security. Energy and Buildings, 254.
15. He, S., Wang, J., & Sun, Z. (2018). An improved Wireless Sensor Network for Natural Gas Monitoring. Chinese Automation Congress (CAC), 2542–2547.
16. Zhu, L., Zou, B., Zhang, H., Wang, Z., & Jiang, M. (2015). Design of Multi-sensor Wireless Monitoring System and its Application in Natural Gas Purification Plant. IEEE International Conference on Mechatronics & Automation, 436–441.
17. Jung, C. Y., & Park, K. K. (2019). Two-channel CMUT Chemical Sensor for Carbon Dioxide Detection. JOURNAL OF THE KOREAN SOCIETY FOR NONDESTRUCTIVE TESTING, 39(2), 89–94.
<https://doi.org/10.7779/jksnt.2019.39.2.89>
18. Wojtas, J., Bielecki, Z., Cudzilo, S., Nowakowski, M., Rutecka, B., Niedbala, R., & Wesolowski, M. (2010). Project of explosive material

- vapours concentrator and pyrolyzer unit. PRZEGLAD ELEKTROTECHNICZNY, 86(11A), 225–228.
19. Blondeau, J., Museur, T., Demaude, O., Allard, P., Turoni, F., & Mertens, J. (2020). Cost-effective flexibilisation of an 80 MWe retrofitted biomass power plants: Improved combustion control dynamics using virtual air flow sensors. *CASE STUDIES IN THERMAL ENGINEERING*, 21.
 20. Abdurrahman, A. H., Kirom, M. R., & Suhendi, A. (2020). Biogas Production Volume Measurement and Internet of Things based Monitoring System. 9th IEEE International Conference on Communication, Networks and Satellite (IEEE Comnetsat), 213–217.
 21. Knobelspies, S., Bierer, B., Woellenstein, J., Kneer, J., & Palzer, S. (2015). Towards in-situ biogas sensing. In G. Urban, J. Wollenstein, & J. Kieninger (Eds.), *Conference on EUROSENSORS* (pp. 269–274). Elsevier Ltd.
 22. Franchi, G., Zollo, S., Pennazza, G., Santonico, M., Capocelli, M., Di Paola, L., & Piemonte, V. (2019). Characterization of innovative sensors for volatile organic compounds trace compounds in biogas. *ASIA-PACIFIC JOURNAL OF CHEMICAL ENGINEERING*, 14(4).
 23. Zhou, X., & Peng, T. (2020). Application of multi-sensor fuzzy information fusion algorithm in industrial safety monitoring system. *SAFETY SCIENCE*, 122.
 24. Wang, Z., & Chiang, L. (2019). Monitoring Chemical Processes Using Judicious Fusion of Multi-Rate Sensor Data. *Sensors*, 19(10).
 25. Wang, J., Li, B., Lei, B., Ma, P., Lian, S., Wang, N., Li, X., & Lei, S. (2021). Design and application of mixed natural gas monitoring system using artificial neural networks. *Sensors (Switzerland)*, 21(2), 1–13. <https://doi.org/10.3390/s21020351>
 26. Eccleston, R., & Bongards, M. (2018). Measuring Biogas Composition with a Compact Microelectromechanical System-Based Spectrometer. *CHEMICAL ENGINEERING & TECHNOLOGY*, 41(4), 680–686.
 27. Surya Kameswari, U., & Ramesh Babu, I. (2015). Sensor data analysis and anomaly detection using predictive analytics for process industries. 2015 IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions (WCI), 1–8.

28. Kanawaday, A., & Sane, A. (2017). Machine learning for predictive maintenance of industrial machines using IoT sensor data. 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), 1–4.
29. Nellore, K., & P. Hancke, G. (2016). A Survey on Urban Traffic Management System Using Wireless Sensor Networks. Sensors, 16(2), 157. <https://www.mdpi.com/1424-8220/16/2/157>
30. Sharma, P. (2014). Wireless Sensor Networks for Environmental Monitoring. Computer Science. <https://www.semanticscholar.org/paper/Wireless-Sensor-Networks-for-Environmental-Sharma/f8501a47830512c7056470de88bcad92a1ae9507?p2df>



7. ANEXOS

7.1 Anexo 1: Explicación de las tablas de la base de datos

- **D_OPERATORS:** Es la tabla maestra de operadores, es decir, de los operarios que controlan la planta y está formada por 6 columnas o campos. Esta tabla recoge datos como el nombre o identificador de los operadores.
- **H_OPERATOR:** Esta tabla transaccional se corresponde con el histórico de registro de operadores, es decir, que operador ha habido en cada momento, y está compuesta por 4 campos.
- **D_TAGS:** Tabla maestra de tags. Esta tabla recoge todos los datos relativos a los tags o variables que hay presentes en la planta, como por ejemplo su identificador, nombre, descripción, sistema y dispositivo al que pertenece o la magnitud que miden. La tabla está formada por 12 columnas.
- **H_TAG_VAL:** Es el registro histórico de las señales analógicas o tags, es decir, de la telemetría de la planta. Es en esta tabla transaccional donde se guardan las mediciones de los distintos sensores, así como las modificaciones en la apertura de las válvulas, y está formada por 5 columnas.
- **H_STATE:** Consiste en 5 campos y es el registro histórico de los cambios de estado de los distintos sistemas que componen la planta y de la propia planta.
- **D_NAPA:** Tabla maestra de estados de las NAPAs (Notificaciones, alarmas, peticiones y alertas) formada por 8 columnas. Contiene datos como el código, descripción o prioridad de las distintas NAPAs.
- **H_NAPA:** Es una tabla transaccional y registra el histórico de los cambios en los estados de las NAPAs, está formada por 14 columnas que recogen información.
- **D_BATCHES:** Es la tabla maestra de los lotes de programación que intervienen en el funcionamiento de la planta y contiene registros de distintas características de los mismos, como el tipo, el código o la descripción. La tabla se compone de 7 columnas.

- H_BATCH: Tabla compuesta por 10 campos, recoge información sobre el histórico de cambios de estado de los distintos lotes de programación de la planta.
- D_ESSAYS: Es la tabla maestra de ensayos y se compone de 4 columnas, recoge información sobre las fechas de los ensayos y la identificación de éstos.
- TIME_MASTER: Tabla compuesta por 12 campos que funciona como tabla maestra de las fechas del resto de tablas de la base de datos. Recoge la información de fechas y horas en distintos formatos.

De todas estas tablas, para nuestro análisis vamos a utilizar las tablas H_TAG_VAL y D_TAGS, puesto que son las que registran la telemetría del proceso y las que nos permitirán crear series temporales que nos ayuden a entender mejor cómo funciona la planta.

- D_TAGS: Tabla maestra de tags. Esta tabla recoge todos los datos relativos a los tags o variables que hay presentes en la planta, como por ejemplo su identificador, nombre, descripción, sistema y dispositivo al que pertenece o magnitud que miden. La tabla está formada por 12 columnas:
 - Device_Code: Código del dispositivo, esto es, del sensor o actuador.
 - Device_Desc: Descripción del dispositivo.
 - Device_Type: Tipo de dispositivo, es parte del Device Code e indica qué clase de dispositivo es.
 - EU: Magnitud o dato que mide el dispositivo.
 - System_Desc: Descripción del sistema al que pertenece el dispositivo.
 - Tag_Code: Código del tag. Cada dispositivo puede tener varios tags o variables, por ejemplo, una que mida temperatura y otra el estatus del dispositivo.
 - Tag_Desc: Descripción del tag.
 - Tag_description: Descripción del tag, se trata de un campo duplicado que está pendiente de ser eliminado de la BBDD.
 - Tag_Id: Código identificativo del tag.

- Tag_Type: Tipo de tag, los hay que recogen valores, estatus, ordenes por programa o operador a la planta, etc...
- Device_EntryDate: Fecha de instalación del dispositivo.
- Device_LeavingDate: Fecha de retiro del dispositivo.
- H_TAG_VAL: Es el registro histórico de las señales analógicas o tags, es decir, de la telemetría de la planta. Es en esta tabla transaccional donde se guardan las mediciones de los distintos sensores, así como las modificaciones en la apertura de las válvulas, y está formada por 5 columnas.
 - Id: Identificador de la entrada o registro.
 - Timestamp: Instante de tiempo en el que se toma el dato.
 - Tag_Id: Identificador del tag al que pertenece el registro.
 - Val: Valor o registro medido, puede ser un valor discreto o continuo.
 - Tag_Code: Código del tag al que pertenece el registro.
- D_NAPAS: Tabla maestra de estados de las NAPAs (Notificaciones, alarmas, ...) formada por 8 columnas. Contiene datos como el código, descripción o prioridad de las distintas NAPAs
 - NAPA_Id: Identificador del NAPA.
 - NAPA_Code: Código del NAPA.
 - NAPA_Desc: Descripción del NAPA.
 - NAPA_Priority: Prioridad del NAPA.
 - NAPA_Severity: Severidad del NAPA.
 - NAPA_Type: Tipo de NAPA (Notificaciones, Alertas, Petición o Alarma).
 - NAPA_Limit: Límite del NAPA.
 - NAPA_Area: Área de actuación.
- H_NAPA: Es una tabla transaccional y registra el histórico de los cambios en los estados de las NAPAs, está formada por 14 columnas que recogen información.
 - Id: Identificador de la entrada o registro.
 - Timestamp: Momento en el que se toma el registro.
 - NAPA_Id: Código identificador del NAPA.
 - NAPA_Code: Código del NAPA.

- NAPA_Value: Valor que ha tomado el NAPA (de la magnitud que lo activa).
- NAPA_Status: Código que identifica el estado del NAPA (activo, reconocido, no activo).
- NAPA_Status_Desc: Descripción del estado del NAPA.
- NAPA_UserComment: Comentarios sobre el NAPA por parte del operador.
- NAPA_Active: Indica si el NAPA está activo.
- NAPA_Acked: Indica si el NAPA ha sido reconocido por el operador.
- NAPA_Disabled: Si el NAPA se ha desactivado por un fallo del propio dispositivo.
- NAPA_Supress: Indica si el NAPA ha sido suprimido por parte del operador de forma permanente.
- NAPA_Shelved: Indica si el NAPA ha sido suprimido por parte del operador de forma temporal.
- NAPA_EventChanged: Campo en desuso.

