

Jornadas de Automática

Desarrollo de una herramienta de trazabilidad y visualización causal en EDARs

García-Cruz, María^a, Sánchez, Luis^a, Revollar, Silvana^b, Francisco, Mario^b, Vega, Pastora^a

^a Dpto. de Informática y Automática, Facultad de Ciencias, 37008 Salamanca (Salamanca).

^b Dpto. de Informática y Automática, ETSII de Béjar, 37700 Béjar (Salamanca).

To cite this article: García-Cruz, María, Sánchez, Luis, Revollar, Silvana, Francisco, Mario, Vega, Pastora. 2025. Development of a traceability and causal visualization tool in WWTPs. *Jornadas de Automática*, 46. <https://doi.org/10.17979/ja-cea.2025.46.12256>

Resumen

Las estaciones depuradoras de aguas residuales (EDAR) generan grandes volúmenes de datos operativos que habitualmente se utilizan para supervisión y control en tiempo real. Este trabajo presenta una herramienta basada en datos históricos para la trazabilidad y visualización causal en EDAR, aplicada a una planta de Castilla y León. La metodología propuesta permite detectar incumplimientos normativos en el efluente, caracterizar el estado del sistema mediante el análisis de variables operativas, identificar perturbaciones externas utilizando percentiles estacionales y evaluar las acciones correctoras implementadas a través de análisis estadísticos y residuos dinámicos. Los resultados se integran en diagramas de Sankey que representan las relaciones causa-efecto-respuesta, facilitando la interpretación de los eventos y la toma de decisiones operativas. La herramienta, desarrollada en Python, constituye un apoyo en el diagnóstico, la formación de operadores y la mejora continua de la operación de EDAR bajo condiciones reales.

Palabras clave: Detección y diagnóstico de fallos, Monitorización del rendimiento de procesos/control estadístico de procesos, Procesos de tratamiento de aguas residuales, Sistemas expertos en la industria de procesos.

Development of a traceability and causal visualization tool in WWTPs

Abstract

Wastewater treatment plants (WWTPs) generate large volumes of operational data typically used for real-time supervision and control. This work presents a historical data-based tool for traceability and causal visualization in WWTPs, applied to a facility in Castilla y León, Spain. The proposed methodology enables the detection of regulatory non-compliance events in the effluent, the characterization of the system state through operational variable analysis, the identification of external disturbances using seasonal percentiles, and the evaluation of corrective actions by statistical tests and dynamic residual analysis. The results are synthesized into Sankey diagrams that represent cause-effect-response relationships, facilitating the interpretation of events and operational decision-making. The tool, implemented in Python, supports diagnostic tasks, operator training, and the continuous improvement of WWTP operations under real conditions.

Keywords: Fault detection and diagnosis, Process performance monitoring/statistical process control, Wastewater treatment processes, Expert systems in process industry.

1. Introducción

Las estaciones depuradoras de aguas residuales (EDAR) son sistemas complejos, con comportamientos dinámicos no lineales, afectados por perturbaciones externas como la lluvia,

la temperatura o vertidos puntuales. Su operación eficiente requiere estrategias de control adaptativas capaces de responder a la variabilidad diaria y estacional del influente en caudal y carga contaminante (Gernaey et al., 2010; Gernaey et al., 2021).

Actualmente, las EDAR disponen de sistemas de monitorización continua y plataformas SCADA que registran la evolución temporal de las principales variables operativas. En la práctica, esta información se utiliza principalmente para supervisar el proceso y aplicar estrategias de control en tiempo real, desaprovechando su potencial para profundizar en la dinámica del proceso y mejorar las estrategias de operación a largo plazo.

El desarrollo de estrategias de supervisión y control más avanzadas requiere un conocimiento profundo del comportamiento dinámico del sistema y de su respuesta ante perturbaciones. Este conocimiento no puede inferirse únicamente a partir de observaciones puntuales o supervisión tradicional, sino que exige el análisis sistemático de datos de entrada y salida de planta, bajo condiciones reales de operación.

A pesar de que se han realizado numerosas investigaciones en este campo, muchas de ellas se sustentan en modelos teóricos o simulaciones con datos sintéticos en modelos teóricos o datos sintéticos generados por simulación, lo que introduce idealizaciones que limitan la representación de las interacciones entre las variables de planta y comprometen la aplicabilidad práctica de los resultados. Una revisión detallada de las estrategias de control desarrolladas hasta la fecha, centrada en modelado y simulación de datos, se recoge en Vilanova et al. (2017).

La priorización del análisis de datos históricos reales ha permitido impulsar técnicas de control avanzadas basadas en redes neuronales, lógica difusa y control tolerante a fallos. Un trabajo precursor en esta línea es el de Fuente y Vega (1999) en el que desarrollaron un sistema de detección de fallos en EDAR basado en modelos no lineales y redes neuronales, que permitía distinguir fallos funcionales bajo condiciones de ruido e incertidumbre, utilizando datos reales para discriminar entre perturbaciones y anomalías significativas. En trabajos posteriores, el análisis de datos ha sido determinante para el estudio del comportamiento de la planta, la detección de fallos y el diseño de sistemas de control (Santín et al., 2015; Santín et al., 2016; Sánchez et al., 2024, Revollar et al., 2025).

Pese a estos avances, sigue siendo necesario un mayor esfuerzo en el desarrollo de herramientas orientadas no solo a la detección de incumplimientos normativos, sino también a la identificación de sus causas y a la evaluación sistemática de las respuestas adoptadas por los sistemas de control frente a perturbaciones.

En este contexto, el presente trabajo propone una metodología de análisis retrospectivo de eventos de no conformidad en EDAR, basada en el tratamiento estadístico de datos históricos y la identificación de sus causas. La herramienta permite detectar patrones operativos recurrentes, evaluar las respuestas adoptadas y generar visualizaciones intuitivas útiles para el diseño de estrategias de control. Además de su utilidad técnica, la herramienta presenta un claro potencial didáctico. Al integrar perturbaciones, condiciones operativas y decisiones de control en una única

estructura causal, permite al alumnado comprender el funcionamiento de una EDAR y analizar críticamente las decisiones operativas. Esto la convierte en un recurso útil para facilitar la trazabilidad operativa, apoyar la formación de operadores y promover la mejora continua de la operación.

2. Tratamiento del conjunto de datos

Para llevar a cabo un análisis riguroso del funcionamiento de una estación depuradora de aguas residuales (EDAR), es necesario definir un conjunto reducido y representativo de variables operativas. La selección se realizó atendiendo a tres criterios fundamentales: su relevancia en el desempeño del proceso biológico, la fiabilidad de su medición y su disponibilidad en registros de monitorización continua.

Bajo estos criterios, se agruparon las variables en tres categorías (Figura 1) según su función en el análisis causal: (i) variables para la identificación de incumplimientos normativos (no conformidades); (ii) variables causales, que explican el origen de las desviaciones (clasificadas a su vez en perturbaciones y variables operativas); y (iii) variables controladas, que permiten identificar una intervención operativa como respuesta al incumplimiento.

2.1. Variables identificadoras de incumplimientos

Las variables indicadoras de fallos operativos son aquellas cuyo incumplimiento implica una violación directa de los requisitos establecidos por la normativa vigente para el vertido de aguas residuales urbanas (Directiva 91/271/CEE y su transposición en la legislación nacional). Destacan especialmente los sólidos en suspensión totales (SST), la concentración de nutrientes como nitrógeno y fósforo total (N_{total} y P_{total} , respectivamente), así como la carga de materia orgánica medida mediante la demanda química de oxígeno (DQO) y la demanda biológica de oxígeno (DBO). Estas variables garantizan que el efluente tratado cumple con los estándares de calidad exigidos para la protección del medio receptor.

2.2. Perturbaciones

Se definen como perturbaciones aquellas variables externas a la planta cuya variabilidad responde a factores climáticos, estacionales o de origen antrópico. Entre ellas se incluyen el caudal de entrada y la temperatura del agua en los reactores biológicos (Sánchez et al., 2025) y las concentraciones de DBO, DQO, SST, nitrógeno total y fósforo total en el influente.

2.3. Variables operativas afectadas

Se consideran variables operativas aquellas que forman parte del funcionamiento interno de la planta y que, aunque son influenciadas por perturbaciones externas, pueden ser modificadas mediante intervención operativa. En ausencia de una acción anticipatoria, estas variables evolucionan en función de la dinámica natural del sistema.

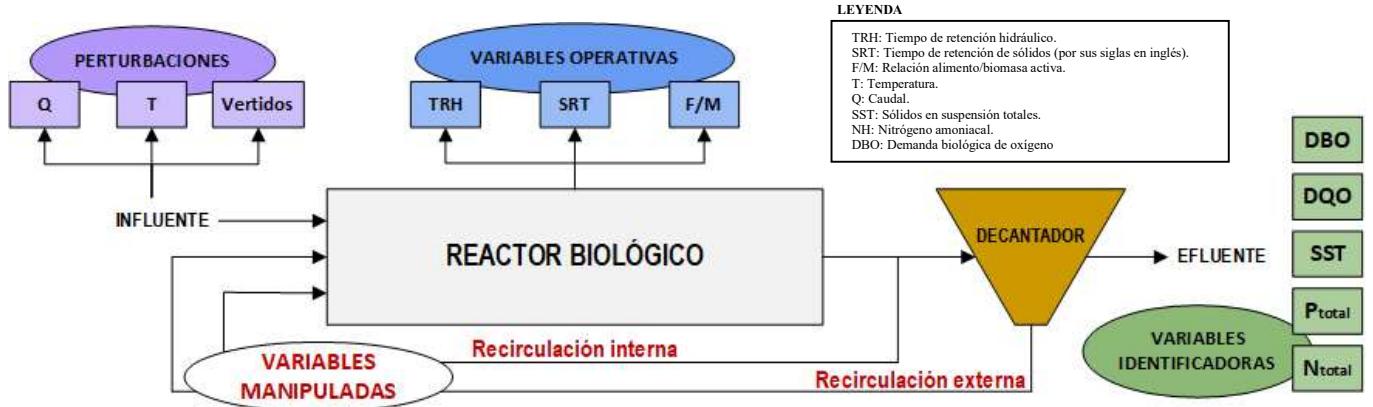


Figura 1: Esquema del proceso en el que se indican las variables consideradas.

Las variables seleccionadas en este grupo son: el tiempo de retención hidráulico (TRH), el tiempo de retención de sólidos —también conocido como tiempo de retención celular— por sus siglas en inglés (SRT), y la relación F/M, que representa la proporción entre alimento disponible y biomasa activa.

2.4. Variables controladas

Se consideraron como variables de control aquellas directamente manipulables desde la operación de la planta. En concreto, se analizaron la recirculación interna y externa, al tratarse de actuaciones habituales utilizadas para modificar las condiciones del proceso en respuesta a desviaciones.

La recirculación interna afecta principalmente a la redistribución de nitrógeno, optimizando la desnitrificación al controlar la disponibilidad de nitratos en las distintas etapas del tratamiento. Por su parte, la recirculación externa actúa sobre la gestión del caudal returned desde los decantadores secundarios, modificando el balance hidráulico y la carga orgánica en el reactor biológico (Sánchez et al. 2024).

3. Metodología

La metodología seguida se resume en el diagrama de flujo mostrado en la Figura 2. El proceso comprende: (i) la detección de incumplimientos en variables de calidad del efluente mediante umbrales reglamentarios; (ii) el análisis del estado del sistema en el momento del fallo, evaluando estadísticamente desviaciones significativas; (iii) la identificación de posibles perturbaciones externas utilizando percentiles estacionales; y (iv) la identificación de las acciones correctoras implementadas.

3.1. Detección de anomalías en el efluente

Se define un evento de incumplimiento como aquel en el que, en un instante de tiempo t , alguna de las variables de calidad del efluente $x_i(t) \in \{DBO, DQO, SST, P_{total}, N_{total}\}$ supera su correspondiente umbral reglamentario x_i^{lim} , es decir, $\exists i \in \{1, \dots, n\} / x_i(t) > x_i^{lim}$.

El cumplimiento de esta condición activa el procedimiento de análisis retrospectivo.

3.2. Caracterización del estado de la planta

Una vez identificado el momento en el que se produce un incumplimiento en el efluente, se procede a la caracterización del estado operativo de la planta. Este análisis se basa en el comportamiento de las variables operativas internas: tiempo de retención hidráulico (TRH), tiempo de retención celular (SRT) y relación F/M, evaluadas mediante criterios estadísticos dinámicos.

El objetivo es determinar si, en el momento del incumplimiento, alguna de estas variables presenta un cambio significativo respecto a su comportamiento habitual. Para ello, se comparan los valores registrados el día del evento con una muestra de días previos. Esto permite evitar sesgos derivados de fluctuaciones, asegurando que las comparaciones reflejen cambios relevantes no atribuibles a variabilidad estacional. Para ello se aplican dos pruebas complementarias entre sí:

- (i) El test t de Student para muestras independientes compara las medias de dos grupos bajo la hipótesis de normalidad y homogeneidad de varianzas; en caso contrario, se aplica la corrección de Welch para ajustar las varianzas desiguales.
- (ii) El test Mann–Whitney U, un test no paramétrico que compara las distribuciones de dos muestras independientes, adecuado cuando no se puede asumir normalidad.

Para cada variable operativa, se consideran dos subconjuntos de datos: uno formado por los valores registrados en un intervalo de k días previos al evento y otro correspondiente al valor observado el día del incumplimiento.

Un p-valor inferior a 0.05 se interpreta como evidencia estadísticamente significativa de un cambio entre ambos grupos, lo que permite atribuir la diferencia a una alteración real en el estado operativo de la planta y no al azar. En función de este análisis, cada variable es categorizada como alta o baja respecto a su media previa.

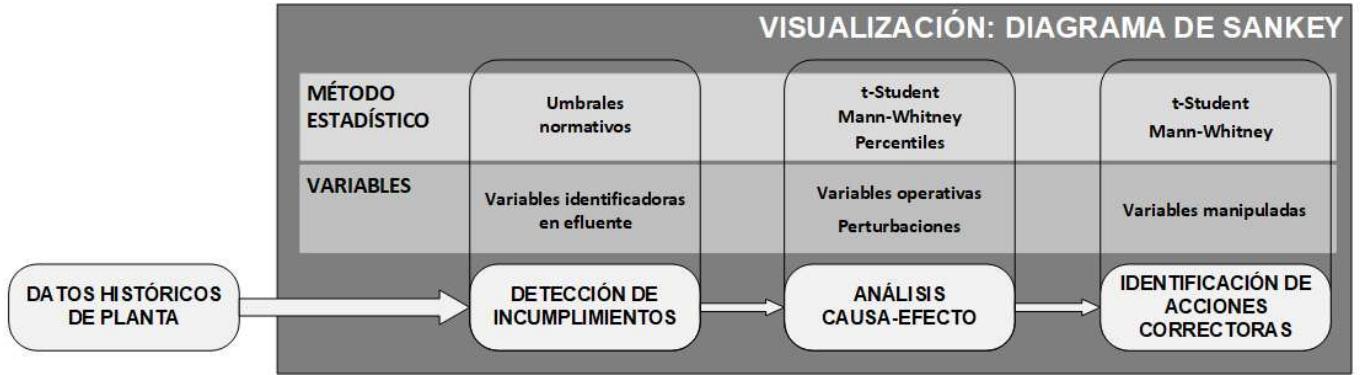


Figura 2: Diagrama de flujo sobre el proceso de detección, análisis estadístico y visualización implementado.

Para la aplicación de estos test, se calcula la media y desviación típica del grupo de k elementos que antecede a la no conformidad, siguiendo las expresiones indicadas en las ecuaciones (1) y (2). Sea x_t el valor de la variable en el día t , y x_{t-k}, \dots, x_{t-1} los valores en los k días anteriores al fallo. Se define:

$$\bar{x}_{pre} = \frac{1}{k} \sum_{i=t-k}^{t-1} x_i \quad (1)$$

$$\sigma_{pre} = \sqrt{\frac{1}{k-1} \sum_{i=t-k}^{t-1} (x_i - \bar{x}_{pre})^2} \quad (2)$$

Para detectar cambios abruptos en variables operativas como respuesta a un fallo, se emplea el cálculo del residuo dinámico como se indica en (3).

$$r_t = x_t - \bar{x}_{pre} \quad (3)$$

Se considera que en el instante t ha ocurrido un cambio local significativo si el residuo r_t supera un umbral proporcional a la desviación típica local, tal y como se expresa en (4).

$$|r_t| \geq \alpha \cdot \sigma_{pre} \quad (4)$$

El parámetro α determina el nivel de sensibilidad del detector. En este estudio, se adopta inicialmente $\alpha = 2$, correspondiente a un umbral estadístico de aproximadamente el 95 % bajo la hipótesis de normalidad.

3.3. Identificación de perturbaciones

Para el análisis de las posibles perturbaciones se consideran las variables externas definidas previamente: caudal de entrada, temperatura del agua residual, así como las concentraciones en el influente de DBO, DQO, SST, nitrógeno total y fósforo total.

El análisis se realiza aplicando un criterio estadístico basado en percentiles. Se establece que una variable $x(t)$ se clasifica como perturbación cuando su valor en un instante de tiempo t supera el percentil 90 o queda por debajo del percentil 10 de su distribución histórica estacional, es decir: $x(t) > P_{90}(x)$ o $x(t) < P_{10}(x)$.

donde $P_{90}(x)$ y $P_{10}(x)$ representan los percentiles 90 y 10 de la variable x , calculados estacionalmente para considerar adecuadamente la variabilidad asociada a los cambios climáticos y a las condiciones de operación a lo largo del año. Utilizar estos umbrales permite identificar eventos poco frecuentes, tanto altos como bajos, que pueden afectar el comportamiento de la planta sin restringirse únicamente a situaciones extremas.

3.4. Análisis de acciones operativas correctoras

Se analiza si, tras la detección de un incumplimiento, se implementó alguna acción correctora en la planta. Para ello, se evalúan cambios en las variables controladas, recirculación externa (R_{ext}) e interna (R_{int}), mediante los mismos contrastes estadísticos aplicados previamente.

4. Representación gráfica

Para representar de forma estructurada las relaciones entre fallos, causas y respuestas operativas identificadas a partir del análisis retrospectivo, se ha optado por el uso de diagramas de Sankey. Este tipo de representación resulta especialmente adecuado para visualizar flujos de información y relaciones jerárquicas entre variables, permitiendo ilustrar con claridad la secuencia completa de un evento: desde la detección de una no conformidad en el efluente hasta la intervención operativa registrada.

A diferencia de las visualizaciones convencionales ofrecidas por plataformas tipo SCADA —que representan las variables de forma aislada o en series temporales—, los diagramas de Sankey permiten sintetizar en un único gráfico la conexión entre distintos niveles del sistema. Para su construcción se ha utilizado la librería plotly de Python, que permite personalizar la visualización de flujos y jerarquías de

variables, facilitando su integración con los resultados del análisis estadístico desarrollado.

El ancho de las conexiones entre nodos refleja la frecuencia o intensidad con la que se han identificado dichas relaciones en los eventos analizados, lo que facilita la detección de patrones operativos recurrentes.

La Figura 3 muestra un ejemplo representativo del análisis causal aplicado a diversos valores típicos de variables operativas de planta, con el objetivo de ilustrar la funcionalidad de la herramienta y su capacidad para representar de forma estructurada las relaciones entre condiciones de operación, causas y respuestas implementadas.

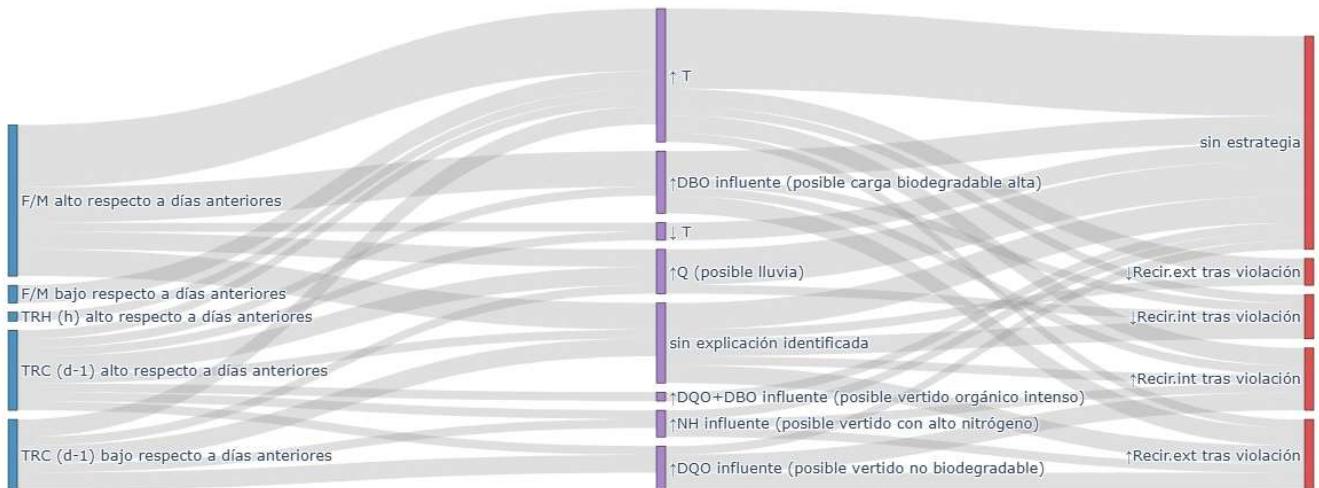


Figura 3: Diagrama Sankey representando la relación causa-efecto-acción correctora (análisis del incumplimiento de DBO anual).

5. Aplicación docente y profesional

El potencial formativo de la herramienta se refleja especialmente en la representación causal de eventos operativos complejos. Esta estructura permite al docente proponer dinámicas realistas para su análisis, permitiendo al alumnado enfrentarse a situaciones en las que deben identificar causas, evaluar estrategias de respuesta y justificar sus decisiones.

Desde una perspectiva profesional, esta misma funcionalidad permite al operador recurrir a la herramienta como apoyo en la toma de decisiones, en lugar de basarse únicamente en prácticas operativas habituales que, si bien están asentadas por la experiencia, no siempre resultan ser la estrategia más adecuada. Así, ante un fallo recurrente —como un aumento puntual de nitrógeno amoniacal en los meses de verano—, el operador puede consultar registros de eventos similares, identificar patrones causales frecuentes y evaluar la efectividad de las medidas correctoras aplicadas previamente. Esta consulta informada no solo mejora la calidad de la respuesta operativa, sino que refuerza un enfoque sistemático y basado en datos para la gestión de incidencias.

6. Conclusiones

Actualmente, la herramienta identifica las acciones implementadas tras un incumplimiento, pero no verifica si estas acciones lograron corregir la violación. Como desarrollo futuro, se plantea analizar si, tras la intervención, las variables que causaron el incumplimiento vuelven a situarse por debajo de los límites reglamentarios en los días posteriores,

fortaleciendo el aprendizaje operativo y apoyando la mejora continua de la gestión en planta.

Este análisis examinará la evolución de dichas variables tras la acción correctora y controlará la influencia de perturbaciones externas, evitando atribuciones erróneas a la operación. El objetivo es determinar si la corrección es atribuible a la estrategia aplicada o a la desaparición espontánea de la causa externa.

El enfoque propuesto establece un marco analítico transferible a otros procesos industriales caracterizados por un alto grado de interacción y complejidad operativa similar al presente en las EDARs. Esta metodología posibilita, en etapas posteriores, la creación de modelos inteligentes capaces de aprender dinámicamente de los datos recogidos, posibilitando la implementación de sistemas de recomendación automática de acciones correctivas al personal encargado de la planta.

A largo plazo, estos modelos podrían alcanzar niveles crecientes de autonomía, llegando incluso a tomar decisiones de manera independiente.

Agradecimientos

Los autores agradecen a: Grupo FCC Aqualia S.A por compartir su experiencia en la operación de EDARs. Al gobierno de España y al Ministerio de Ciencia, Innovación y Universidades por los proyectos: TED2021-129201B-I00, MICINNU/AEI/10.13039/501100011033 y FS/14-2024.

Referencias

- Fuente, M. J., Vega, P., 1999. Neural networks applied to fault detection of a biotechnological process. *Engineering Applications of Artificial Intelligence* 12, 569–584.
DOI: 10.1016/S0952-1976(99)00028-7
- Gernaey, K., Jeppsson, U., Vanrolleghem, P., Copp, J., Steyer, J. Benchmarking of control strategies for wastewater treatment plants., IWA Publishing, Colchester, UK, 2010.
- Gernaey, K., Rosen, C., Benedetti, L., & Jeppsson, U. (2021). BSM2: A model for dynamic influent data generation. *Wastewater Modelling*.
- Revollar, S.; Vilanova, R.; Vega, P.; Francisco, M.; Meneses, M. (2020). Wastewater Treatment Plant Operation: Simple Control Schemes with a Holistic Perspective. *Sustainability*, 12, 768.
DOI: 10.3390/su12030768
- Santín, I., Pedret, C., Vilanova, R., 2015. Fuzzy control and Model Predictive Control Configurations for Effluent Violations Removal in Wastewater Treatment Plants. *Industrial and Engineering Chemistry Research* 54 (10), 2763–2775.
DOI: 10.1021/ie504079q
- Santín, I., Pedret, C., Vilanova, R., Meneses, M., 2016. Advanced decision control system for effluent violations removal in wastewater treatment plants. *Control Engineering Practice* 279, 207–219.
DOI: 10.1016/j.conengprac.2016.01.005
- Vilanova, R; Santín, I., Pedret, C. (2017) Control en Estaciones Depuradoras de Aguas Residuales: Modelado y Simulación. *Revista Iberoamericana de Automática e Informática Industrial (RIAI)*. 14 (3): 217-233
DOI: 10.1016/j.riai.2017.09.001