Applying XAI based unsupervised knowledge discovery for operation modes in a WWTP. A real case: AQUAVALL WWTP☆

Alicia Beneyto-Rodríguez [a] , Gregorio I. Sainz-Palmero [a],*, Marta Galende-Hernández [a] , María J. Fuente [a] , José M. Cuenca [b]

[a] Department of System Engineering and Automatic Control, School of Industrial Engineering, Universidad de Valladolid, Paseo Prado de la Magdalena 3-5, 47011, Valladolid, Spain
[b] Department of R&D, AQUAVALL, Camino Viejo de Simancas 229, 47008, Valladolid, Spain

ARTICLE INFO

ABSTRACT

Water reuse is a key point when fresh water is a commodity in ever greater demand, yet also becoming more accessible. Furthermore, the return of clean water to its natural environment is also mandatory. Therefore, wastewater treatment plants (WWTPs) are essential in any policy focused on these serious challenges. WWTPs are complex facilities which need to operate at their best to achieve their goals. Nowadays, they are largely monitored, generating large databases of historical data concerning their functioning over time. All this implies a large amount of embedded information which is not usually easy for plant managers to assimilate, correlate and understand; in other words, for them to know the global operation of the plant at any given time. At this point, the intelligent and Machine Learning (ML) approaches can give support for that need, managing all the data and translating them into manageable, interpretable and explainable knowledge about how the WWTP plant is operating at a glance. Here, an eXplainable Artificial Intelligence (XAI) based methodology is proposed and tested for a real WWTP, in order to extract explainable service knowledge concerning the operation modes of the WWTP managed by AQUAVALL, which is the public service in charge of the integral water cycle in the City Council of Valladolid (Castilla y León, Spain). By applying well-known approaches of XAI and ML focused on the challenge of WWTP, it has been possible to summarize a large number of historical databases through a few explained operation modes of the plant in a low-dimensional data space, showing the variables and facility units involved in each case.

1. Introduction

Wastewater treatment plants (WWTPs) are dynamic and complex systems operating under varying operations, due to fluctuations in influent characteristics, environmental factors and operation adjustments. Understanding the different operation modes within a WWTP is essential for efficiently optimizing the treatment, reducing energy consumption and ensuring compliance with environmental regulations. However, due to the complexity of the nature of WWTP processes, traditional analytical approaches can often fall short in interpreting these operation modes. On the other hand, these facilities undergo large scale monitoring, supplying large databases about plant operation. This implies a need for a data summary to simply explain the WWTP operation mode that can be available at all times for the managers (see Fig. 1).

Data-driven techniques have enabled a more systematic analysis of WWTP operations by extensively using real-world data collected from monitoring systems. Some publications address this, taking into account different aims, such as Machine Learning (ML) for predicting wastewater quality in three WWTPs in China [1], an evolutionary data-intelligence model, integrating the Extreme Learning Machine (ELM) with Kernel Principal Component Analysis (KPCA) to predict the performance of a WWTP in Nigeria [2], or the development of a prediction using the Wastewater Quality Index (WWQI) for a regional WWTP in Malaysia, and Principal Components Analysis (PCA) to handle high-dimensional process data [3].

Clustering approaches are incorporated in [4], combining a $K-means$ based technique with a process simulation model to assess the

Fig. 1. Target overview.

impact of changing climatic operations on the influent characteristics of a WWTP in Italy. On the other hand, [5] develops an autoencoder based deep clustering method (SMEL) to detect abnormal data with no labels in a WWTP in China. All these approaches in real cases have a lack of explainability; moreover, they do not address the global operation of the plant and neither do they take into account the data from the monitoring systems, which is an essential and mandatory goal for any manager in charge of such facilities.

Here, the target is to know the operation modes of the WWTP facility using its monitoring data collected over time, and then locate at a glance every new monitored data sample regarding those already known or discovered WWTP operation modes as a point in a reduced dimension space of variables, i.e., summarizing the large database of the WWTP monitoring systems as points in that reduced dimension space of variables. All this being explained by the plant variables involved in every mode and case. A density based algorithm, DBSCAN (Density-Based Spatial Clustering of Applications with Noise), is used to discover the operation modes. Before all that, a mandatory dimensionality reduction is carried out to manage the most relevant information in a reduced input space, so as to deal with the *curse of dimensionality* for the ML techniques and to improve the explainability and interpretability of the outcome. Therefore, expert knowledge, principal component analysis (PCA), and Shapley Additive Explanations (SHAP) are used.

Some works have shown the effectiveness of DBSCAN in clustering for environmental and industrial process data, such as [6] to localize illegal industrial discharges of polluting wastewater in sewer networks. However, as DBSCAN effectively groups similar operation states, it does not inherently provide any explainability about the generated clusters in the domain terms. To bridge this gap, SHAP is applied to explain the contribution of individual features to the identified operation modes. This technique is the most popular within the scope of the eXplainable Artificial Intelligence (XAI) methods [7–10]. The concept of explainability is becoming ever more relevant, even mandatory when AI methodologies are applied to different fields, such as [11–13], but water systems are also considered a critical domain [14,15].

According to the goals for this work described above, the main contributions of this work are as follows:

1. An explainable machine-learning based methodology is created to summarize large and high dimensional WWTP monitoring databases as a point in a reduced space of operation modes.
2. This approach allows the current WWTP operation mode regarding the known modes to be located at all times.
3. Every WWTP operation mode is explained by the variables of the WWTP plant.
4. The validation of clustering results through expert knowledge, ensuring that the detected operation modes are meaningful and useful for WWTP operators in practice.

The rest of the paper is organized as follows. In Section 1, the introduction provides the context and motivation for the study, highlighting the challenges in monitoring WWTPs and the importance of developing explainable, data-driven methods for operation mode detection, as well

as a brief state of the art in this area. Section 2 provides the theoretical foundations necessary for this work, offering a brief overview of PCA for dimensionality reduction, DBSCAN for clustering, and SHAP for model interpretability. Section 3 describes the proposed methodology, while in Section 4, a real case study is presented, where the methodology is applied and the results are analyzed, validating the clustering results through expert knowledge. Finally, Section 5 summarizes the main findings, discusses the implications of the study, and outlines potential directions for future research.

2. Some foundations

The following subsections briefly introduce the theoretical items supporting this work.

2.1. Principal Component Analysis (PCA)

AQUAVALL data records contain a large number of variables with a high probability of correlation. This large number of variables (data dimensionality) is a challenge both for their computation and to know the specific relevance of each one. In this work, Principal Component Analysis (PCA) [16,17] is used to discover the relevance of each one, making a dimensionality reduction under some conditions, thus transforming the original variables into a new set of uncorrelated extracted variables which can be ranked by the data variance of each one. Then a selection can be carried out to create a subset of a small number of variables (principal components), while preserving most of the information.

Mathematically, PCA [18] begins by standardizing the dataset, X, to ensure that each variable contributes fairly to the analysis. Next, the covariance matrix C is computed. This matrix captures the relationships between variables. The principal components are then determined by solving the eigenvalue problem $Cv = \lambda v$, where λ represents the eigenvalues and v the corresponding eigenvectors. These eigenvectors (principal components) are ranked based on descending eigenvalues so as to be able to select the most significant components.

So, the raw data are transformed into a k-dimension subspace to be selected by choosing the k top eigenvectors of the covariance matrix, and therefore the k top principal components. These eigenvectors form the new basis for the data. However, this approach has a serious disadvantage: a lack of interpretability concerning their principal components and their meaning with regard to the problem domain.

2.2. DBSCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [19,20] defines clusters as dense regions of points separated by areas of lower density. It does this by classifying points as core points, border points, or noise, based on two main parameters: ε (*epsilon*), which determines the neighborhood radius; and $MinPoints$, the minimum number of points required to form a dense region.

This density-based approach makes DBSCAN highly effective for applications ranging from anomaly detection to spatial data analysis. Moreover, its flexibility, robustness to noise, and ability to discover non-spherical clusters, make it a widely used and valuable clustering algorithm in many real-world applications.

The ε-neighborhood of a point p is defined as:

$$N_\varepsilon(p) = \{q \in D \mid d(p, q) \leq \varepsilon\} \tag{1}$$

where D is the datset, $d(p, q)$ is the distance function (typically Euclidean distance), and ε is the neighborhood radius.

A point p is considered a core point (points that have at least $MinPoints$ other points within a distance of ε) if it satisfies the density operation:

$$|N_\varepsilon(p)| \geq MinPoints \tag{2}$$

where $|N_\varepsilon(p)|$ represents the number of points in the ε-neighborhood of p.

On the other hand, a point p is classified as noise if it is not density-reachable from any core point, meaning it does not belong to any cluster:

$$|N_\varepsilon(p)| < MinPoints \tag{3}$$

Therefore, DBSCAN groups all density-connected points into clusters, while points that do not meet the density criteria are labeled as noise points. This term, noise, is not considered for this work, as we are searching for events which could have happened once and cannot be considered as noise.

2.2.1. DBSCAN tuning

Clustering algorithms, unlike supervised learning, needs some methodology for tuning its own parameters in order to achieve success. Unlike traditional clustering methods, that assume clusters have a specific shape or require a predefined number of clusters, DBSCAN detects groups of varying structures and effectively handles noise within a dataset. Its effectiveness depends on the two parameters commented in the previous section: $MinPoints$ and ε (*epsilon*). There is no definitive formula to determine the ideal $MinPoints$ or ε values; however, the selection should be guided by the nature of the dataset and expert knowledge.

On the other hand, the search for a fair ε value also has to be addressed. One effective and well-known way to estimate ε is by analyzing the distances between each point and its k-nearest neighbors [21], where k corresponds to the chosen $MinPoints$. By checking out these distances in ascending order on a k-distance graph, the optimal ε can be identified at the point of highest curvature, often referred to as the "*elbow point*". In this way, it is possible to distinguish cluster boundaries by selecting a threshold that separates dense regions from sparse areas.

2.3. Shapley Additive Explanations (SHAP)

Nowadays, eXplainable Artificial Intelligence (XAI) is a goal, if not a mandatory goal, for AI applications. One of the options for the said target is the Shapley Additive Explanations (SHAP) method [22]. This is a powerful XAI method for explaining the output of machine learning models. It provides a way to attribute the prediction of a machine learning model to its individual input features in a fair and consistent manner, providing an alternative for addressing the classic and well-known black box challenge. SHAP values are an adaptation from the Shapley values, used in cooperative game theory, for explaining machine learning models.

Shapley values [23] were originally introduced in the context of cooperative game theory, where the goal was to fairly distribute the 'payout' (or value) generated by a team of players. In the case of machine learning, the 'game' involves predicting an outcome, and the 'players' are the individual features, or variables, of the model. Shapley values provide a way of assigning a contribution to each variable based

on its impact on the prediction. The Shapley value for each variable in a machine learning model is calculated by evaluating its contribution to all possible combinations of variables. Given a predictive model f and a set of m variables, the Shapley value for a variable i is calculated as:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(m - |S| - 1)!}{m!} [f(S \cup \{i\}) - f(S)] \tag{4}$$

where S is a subset of variables excluding the variable i, $f(S)$ is the model's prediction when using only the features in S, and $f(S \cup i)$ is the model's prediction when feature i is added.

However, SHAP simplifies this process by providing an efficient and scalable way to calculate Shapley values for machine learning models. Several versions of SHAP were introduced by [22]: DeepSHAP, KernelSHAP, LinearSHAP and TreeSHAP, developed for different types of models. In this work, KernelSHAP has been used to explain PCA and clustering outputs, as it is a model-agnostic method for estimating SHAP values efficiently, approximating Shapley values using a weighted linear regression approach.

3. Methodology

The main goal of this proposal is to understand at a glance the operation mode of the WWTP plant regarding its known standard operation modes at any time, and all this is provided in a reduced space of variables in an explainable way for the WWTP managers and operators. Consequently, this implies summarizing the large amount of high-dimensional data provided by the WWTP monitoring systems, explaining the main roots (variables) involved for each standard and non-standard operation modes. All this is based on a data-driven solution, so a sufficient quantity and quality of data, is mandatory for that goal. The proposed methodology (see Fig. 2) is based on a dimensionality reduction through expert knowledge and PCA, clustering by DBSCAN and explanations by SHAP, involving two main, usual stages for its setup: offline and online. All this is detailed in Section 4.1

3.1. Offline stage

A first and critical methodological step is focused on the preprocessing of the historical monitoring data to be used: data cleaning, resampling and standardization.

The data captured by the monitoring and control systems contain variables of different nature, some of them *useless* for this goal and with potential errors, so a cleaning procedure is mandatory. This challenge is solved by removing the *useless* ones, while leaving the variables monitoring the WWTP process, always based on the expert-knowledge of the plant. Then, another critical point is the sampling time, which has consequences for the outcome quality and its reliability/usability, but also for the computational efforts, so a resampling step must be addressed. Finally, the last part of the preprocessing involves the standardization of each of the acquired variables. Here, the *z-score* formulation is used for this goal.

Notwithstanding the above, the most relevant challenge to address is the data dimensionality, which is usually done through variable selection or extraction to manage the *curse of dimensionality* [24]. Then an extraction of variables/features is carried out to obtain new variables, concentrating the information about the WWTP. Here, the well-known Principal Components Analysis (PCA) [25] is involved. Nevertheless, another challenge appears as a consequence of this technique: the interpretability and explainability regarding the original variables is lost. This issue must be addressed, so SHAP is used for this goal. The dimensionality reduction from the principal components is based on two main criteria: the variability preserved in the remaining new extracted variables and the interpretability of the outcomes to be managed by the technicians in charge of the WWTP plant.
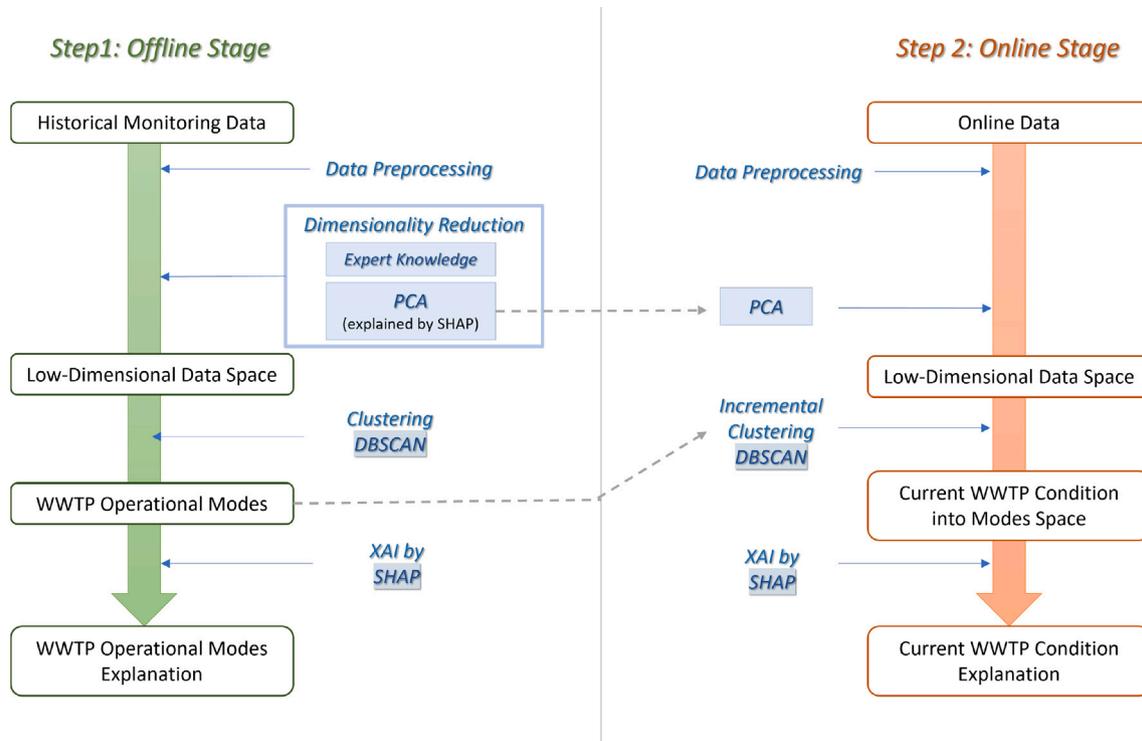
Fig. 2. Methodology scheme.

When the new input space is defined, then the stage devoted to finding operation modes is launched. Here, a well-known clustering approach is used, based mainly on density: DBSCAN. This algorithm is tuned to obtain the best data partitioning possible; while the clusters, or in this case the WWTP operation modes are explained by SHAP values in terms of WWTP variables. This XAI technique shows the relevance of each variable for each cluster, and thus for each operation mode of the WWTP plant.

3.2. Online stage

This stage involves putting in place the outcome from the previous stage. In modern plants, to know the current operation of the plant at any particular moment is a mandatory task for managers. Taking the information available from the monitoring system is the most usual option, showing the values on a screen, as with a standard SCADA application; in this work, however, these data are turned into a point in the low-dimensional data space of the WWTP operation modes, showing the current operation of the WWTP plant regarding the operation modes known so far, as well as having knowledge of the relevant variables of that mode.

The first step is based on the preprocessing of the new monitoring data, in the same way as in the offline stage: data cleaning, resampling and standardization. Then, after the dimensionality reduction based on the principal components obtained from the previous stage, an incremental clustering (Incremental DBSCAN) must be carried out, incorporating the new data for updating the operation modes of the plant and to compare the evolution of the plant regarding the previous operation modes. Finally, these modes are explained by SHAP, using the data samples included in each mode and the principal components obtained on the Off-line step.

If the original historical data were considered outdated due to the plant's evolution or after a long period of time, the offline stage would be restarted for a better performance.

4. Case study: AQUAVALL WWTP

The real WWTP used for this work as case study is run by AQUA-VALL,[1] the public consortium in charge of the integrated water system in the city of Valladolid (Castilla *y* León, Spain) (see Fig. 3). This consortium is in charge of the integrated water service for more than 344,600 households, industries, and other water consumers under optimal quality operations. The main facilities for these goals are two drinking water treatment plants (DWTP), a wastewater treatment plant (WWTP), and a laboratory located within one of the DWTP plants, where both the quality of drinking water and wastewater discharges into the network are monitored. Water is delivered through a network of pipelines spanning 639 km. Meanwhile, the wastewater generated in the city area reaches the WWTP via a 786 km sewer system. This WWTP processes around 122,000 m^3 of wastewater daily, being able to manage a maximum flow rate up to $\frac{m^3}{seg}$, equivalent to the organic load generated by 570,000 people. The AQUAVALL WWTP is organized into three processing lines: water, sludge, and gas, which are arranged to keep interconnected processes in close proximity. The water treatment line consists of several stages: an intake well, a storm-water tank, preliminary treatment, primary sedimentation, and a biological treatment process. A brief overview of all this is available at [26].

4.1. Experimental methodology

The proposal described in Section 3 has been implemented using real raw data obtained from the monitoring system of AQUAVALL WWTP, embracing a time period of 15 months with a sampling time of 1 min (658,084 data samples). Each sample records 3231 variables from the systems involved in the different sides of the WWTP facilities. Twelve months of data were used for the offline stage, and 3 months of data for the online stage.

---

[1] https://aquavall.es

Fig. 3. Water cycle at AQUAVALL - Valladolid (www.aquavall.es).

4.1.1. Offline stage

• Data preprocessing

This huge number of variables also implies a compulsory procedure of cleaning and removing of variables which do not concern the wastewater treatment, such as security systems and others. There are also variables with value 0 over the entire time, so they were removed, as well as variables with a level of missing values higher than 1%. In the opposite case related to missing values, the average between samples, the previous value or the next value, have been applied to deal with the issue. Other variables concerned the cumulative operation hours of devices deployed throughout the whole facility, such as electrical engines, sensors, valves, etc. All these variables were also removed because of their irrelevance for the WWTP operation. Finally, the number of remaining variables to be managed for the proposed goal were 453. This is a first step in the data dimensionality reduction.

On the other hand, the raw data were sampled every minute, which implies some issues for dealing with the goal of this work, in both a very high computational load and sampling time, one minute being very low for this WWTP process and its management. So, a resampling of data is needed to address both issues. In this proposal, two resampling times have been considered based on domain expertise:

- *12* h: considering the 12 h for resampling as 08:00–20:00 and 20:00–08:00.
- *24* h: considering the 24 h for resampling as 00:00–24:00.

A standardization is then carried out using the well-known standard score, or $z-score$, for the variables at each new resampling time.

• Data dimensionality reduction

As for the processing of the raw data, some of the available variables were removed due to different issues, leaving 453 variables. However, this reduction in the number of variables, or data dimension, is still not enough for machine learning approaches and others, which have to deal with the *curse of dimensionality*, so a further reduction is compulsory. This is addressed here by the well-known PCA [16,17]. In Section 2.1, a brief introduction to PCA has been done, showing that this extraction of variables is useful for dealing with the issue of dimensionality reduction. However, two new issues must be addressed:

1. The dimensionality of the new input space, or number of principal components to be kept: here, this number has been managed taking into account the variance accumulated by these components and the complexity of the model.
2. The explainability: the techniques for dimensionality reduction often have a serious issue, which is the loss of interpretation and explainability based on the original variables. Here, this challenge has been addressed by another well-known approach: SHAP [27,28].

All this concerns both resampling times.

• WWTP Operation modes: DBSCAN based clustering

In this work, we use $MinPoints = 1$, meaning that every point is considered a potential cluster core, and clusters can form with a single point. While this setting allows greater flexibility in detecting small or sparse clusters, it may also lead to increased sensitivity to noise.

Once $MinPoints$ have been set, an appropriate ε value is selected for each sample time and each stage of the methodology, using the k-nearest neighbors proposal mentioned in Section 3.

• Explanation of the operation modes: SHAP based XAI

Every cluster, or operation mode, discovered in the previous stage is explained by the most contributive WWTP variables through the SHAP values, showing these contributive variables and their corresponding WWTP units.

• WWTP operation modes into a reduced input space

The location of the WWTP operation modes in the input reduced space, according to the most relevant variables, shows the plant's known modes, and their closeness or similarity, to the managers and technicians in charge of its operation, as well as its explanation through its more contributive variables of the WWTP and the plant units involved.

4.1.2. Online stage

• WWTP operation mode at t_i

At every sampling time, new data are supplied by the monitoring system and, after its processing according to the previous steps, its prediction regarding the previously known modes is made and visualized, explaining the more contributive variables for this current operation mode. Here, the plant manager sees the current operation of the WWTP
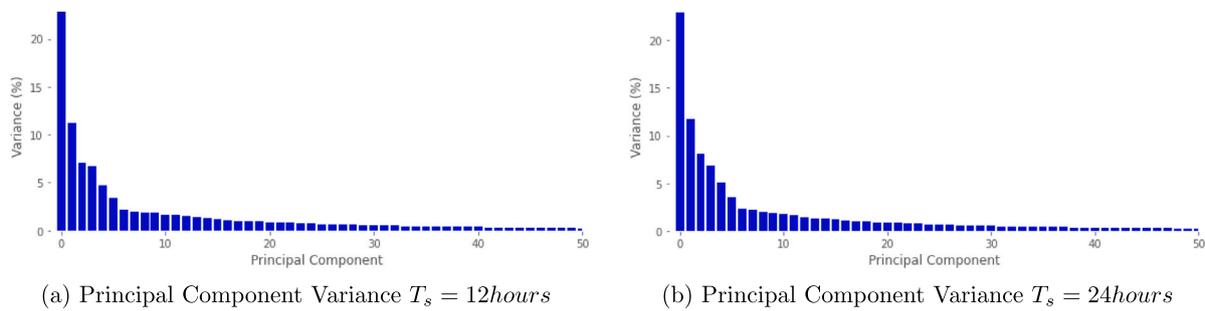
(a) Principal Component Variance $T_s = 12 hours$

(b) Principal Component Variance $T_s = 24 hours$

Fig. 4. Relevance of the principal components by PCA.

plant and its closeness to the known operation modes of the plant happening over time.

• **Model updating/rebuilding**

The model generated must be updated, or rebuilt, by incremental clustering based on DBSCAN to see the evolution of the plant over time. Here, the last 3 months of data are used for this goal, following the previous stages when necessary. On the other hand, when the WWTP mode model is considered outdated, all of the methodology can be carried out for a new mode model.

4.2. Results and analysis

The experimentation carried out in this work is based on 15 months of WWTP data: 12 months (from 2022/08/01 to 2023/07/31) are used for obtaining the mode model by tuning parameters, and 3 months (from 2023/08/01 to 2023/10/31) for simulating the online stage, considering in total 658,084 data samples from 3231 monitored variables, with 525,600 and 132,484 samples for building and testing, respectively.

These data were processed in order to remove useless variables and other issues, as explained previously, finally obtaining 453 variables of interest for this work. The following step was data resampling. Two data resamplings were carried out according to the available expert knowledge: 12 and 24 h, as described in Sub Section 4.1.1. The results of these resamplings are shown in Table 1 (Extra Supplementary Material).

This reduction in the number of variables, or dimensional input space, is highly significant, but is not enough to be applied successfully to machine learning techniques. So a process of dimensionality reduction is mandatory. This reduction in dimensionality was carried out by PCA. Fig. 4(a) and (b) show the variance supplied by each principal component for $T_s = 12$ h and $T_s = 24$ h, respectively. Here, the selection of the k principal components was based on the 'elbow' of the graphics, taking into account the variances preserved by this component selection and its number. In both cases, the number of principal components was 7, in the case of $T_s = 12$ h, preserving 58.04% of variance, and for $T_s = 24$ h, 60.51%. Therefore, this data dimensionality is reasonable for dealing with matching learning approaches.

However, when using these new extracted variables, or principal components, there is a lack of explainability; so we use SHAP to know the most relevant WWTP variables (i.e., the first ten) for the selection of the main principal components, as displayed in Table 4(a), and the relevant WWTP facilities/units in Table 4(b), both in Extra Supplementary Material.

The most influential variables obtained for each resampling time are ranked in a slightly different way, so the sample time influences the results of both DBSCAN operation modes and SHAP explanations. After the explanation of the 7 principal components, the following stage involves the detection of WWTP operation modes for each resampling time (T_s).

4.2.1. WWTP operation modes

The detection of the WWTP operation modes is based on the DB-SCAN algorithm, tuned according to the experimental methodology described in Section 3. In order to avoid a very long section, only the most relevant DBSCAN results are shown. Tables 2 and 3 (in Extra Supplementary Material) show a Silhouette Index (SI) [29] and Density-Based Clustering Validation Index (DBCVI) [30] based comparison of DBSCAN regarding to HDBSCAN [31] and K-Means [32], two well-known clustering algorithms.

• **Case:** $T_s = 12$ h, Offline

Here, the best tuning for the DBSCAN parameters was: $\varepsilon = 1.3$ and the $MinPoints = 1$. This last parameter was a hard restriction in order to detect operation modes and occasional events that seldom appear, even only once, but which must necessarily be known. Otherwise, the DBSCAN would show these cases as noisy cases, which is not adequate for the goal of this work.

Figure 11 (in Extra Supplementary Material) shows the effect of $MinPoints = 1$ and $MinPoints = 2$ on the DBSCAN clustering. The key difference is the lower number of detected modes when $MinPoints = 2$ and the treatment of isolated points. When $MinPoints = 1$, these points are presented as separate clusters (in different colors), whereas with $MinPoints = 2$, the same points are considered as noise and shown in a single color. DBSCAN setting $MinPoints = 1$ implies that even a single data sample if – sufficiently distant from others – can be considered as a distinct operation mode. This distinction highlights how the choice of MinPoints impacts the identification of rare or transient operation modes.

In Fig. 5(a), the detected WWTP operation modes are shown in the space of the 3 main principal components (1, 2 and 3); while Fig. 5(b) shows the same operation modes from another perspective: on the space of the principal components 5, 6 and 7. Here, 15 operation modes were detected, while most of the samples are logically clustered in *Operation Mode 0*, the rest are exceptional days that are considered as *Operation Modes* in order to know about similar situations in the near future. Fig. 5(c) shows the same data partition, but the sample colors are in accordance with the month of the samples; while in the blow-up of the *Operation Mode 0*, it is possible to see some distribution patterns over months.

However, all this has to be explained in terms of the WWTP domain, and therefore according to the variables managed by the WWTP managers. Table 5 (in Extra Supplementary Material) show an extract containing the variables of the facility with the greatest impact for every operation mode detected by DBSCAN, and the units of the WWTP facility involved. This table shows that input facility of the WWTP along with the plant's ability to treat the activate sludge, both in the bioreactors and the digestors for biogas production, are the main factors determining the different operational modes of the plant.

In this way, all the information from the monitoring systems of the WWTP, thousands of variables, is summarized as operation modes of the facility. Each one can be located at a glance as a point in a low-dimensional variable space, manageable by technicians in charge of the plant. Their closeness regarding other known WWTP operation modes
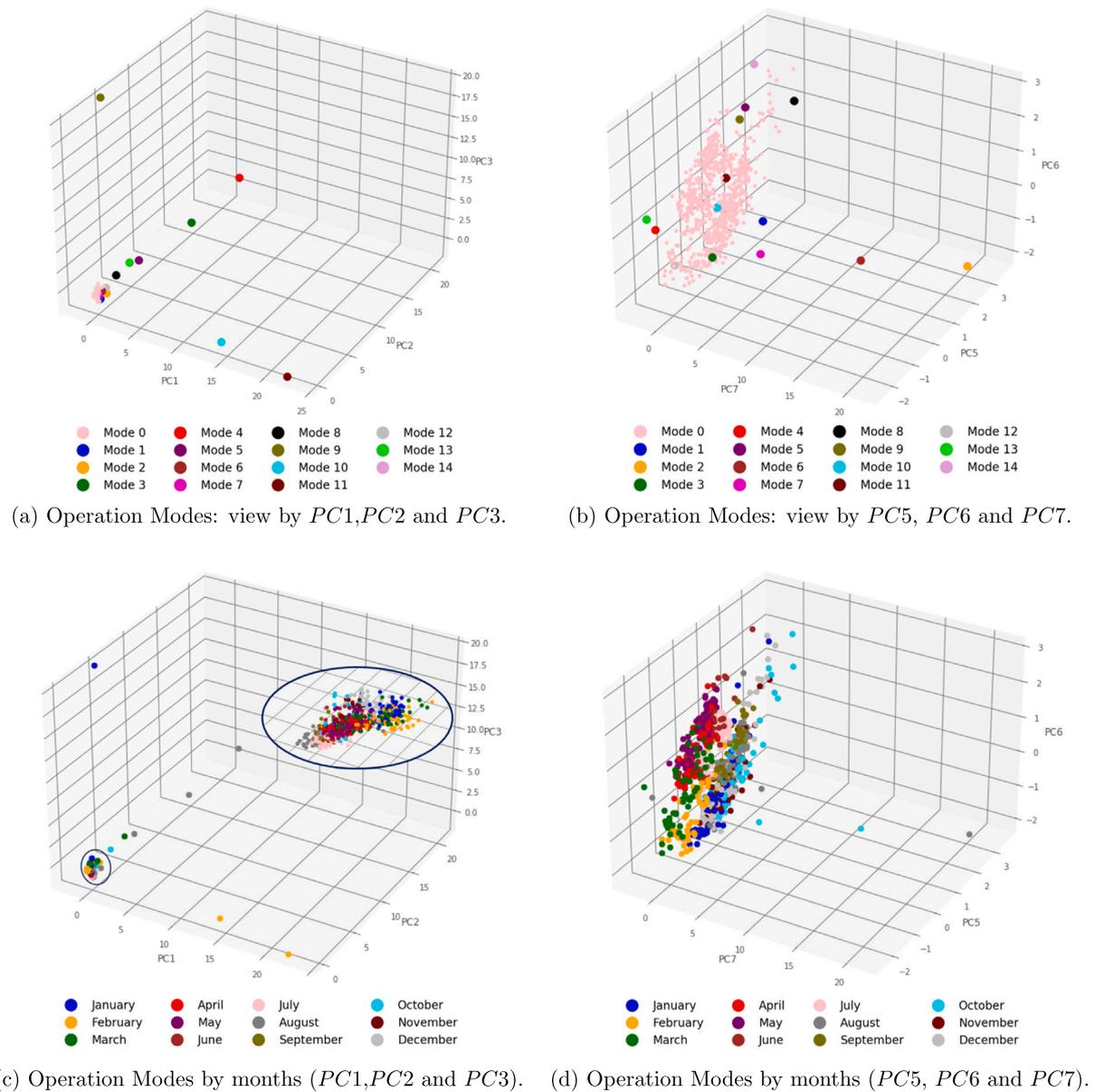
(a) Operation Modes: view by $PC1, PC2$ and $PC3$.

(b) Operation Modes: view by $PC5$, $PC6$ and $PC7$.

(c) Operation Modes by months ($PC1, PC2$ and $PC3$).

(d) Operation Modes by months ($PC5$, $PC6$ and $PC7$).

Fig. 5. Operation modes based on DBSCAN ($\varepsilon = 1.3$ and the $MinPoints = 1$).

and the impact of every monitored variable and corresponding WWTP unit can also be easily seen.

These results were also checked with the records of the WWTP facility in order to explain the unsupervised modes obtained by the available expert knowledge: *Operation Mode 0*, the most populated, concerns the regular operation of the plant over the year; while *Operation Modes 4 and 13* match with shutdown standbys of the plant due to maintenance and issues on power lines according to expert knowledge and records. Moreover, an arrival pumping variable is detected by SHAP for both as an explanation for these exceptional operation modes. *Operation Modes 5, 8, 9 and 14* concern episodes of heavy rain at the end of summer, in autumn, winter and summer, respectively. Therefore, variables related to the rain are highlighted by SHAP. Finally, *Operation Modes 1, 2, 6, 7 and 12* seem to be some type of sporadic disturbances at the edge of the regular operation. Here SHAP highlights the preprocessing and thickening facilities variables; while *Operation Modes 3, 10 and 11* are more distanced from the regular operation mode, and modes 3 and 11 share the influence of the pumping arrival facility variables according to SHAP.

- **Case:** $T_s = 12$ h, Online

Here, every new sample is managed in this mode model, checking the relationship regarding the known WWTP operation modes; while their nature regarding the WWTP variables is explained by SHAP. In short, each new sampling is turned into a point in the low-dimensional mode space and linked with operation modes known so far, being explained by the WWTP variables that have the greatest influence for this location.

Three months of data samples were used for this stage, being processed as in the previous step. Fig. 6(a) shows the online samples in red for the PC1, PC2 and PC3 input space. As can be observed in both Fig. 6(a) and (b), most of these new samples are in the space of the *Operation Mode 0*, the standard operation mode of the WWTP plant; while others, which are not linked to the known operation modes, such as the samples marked as 1, 2 and 3, are not known regarding the historical data. In Table 7 (Extra Supplementary Material), these new modes are explained by the WWTP variables and units involved. Samples number 1 and 2 match with a general shutdown for 2 h due to the change of power lines; so the most influential variables according to SHAP are the same, corresponding to the digestion and biological facilities. On the other hand, sample number 3 concerns a connection
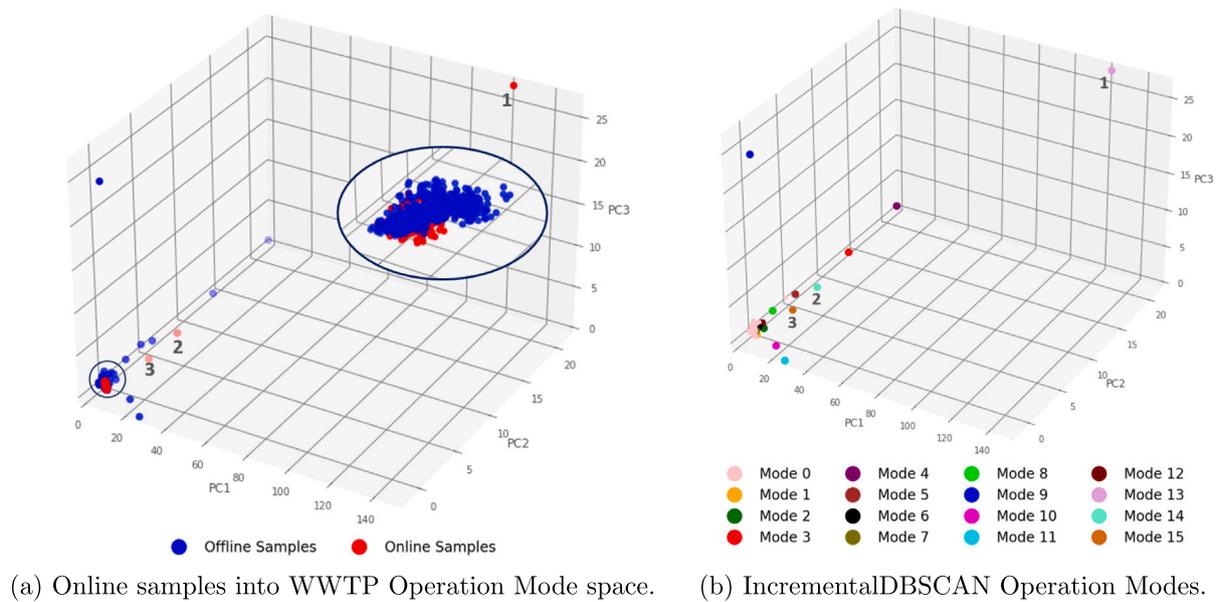
(a) Online samples into WWTP Operation Mode space.

(b) IncrementalDBSCAN Operation Modes.

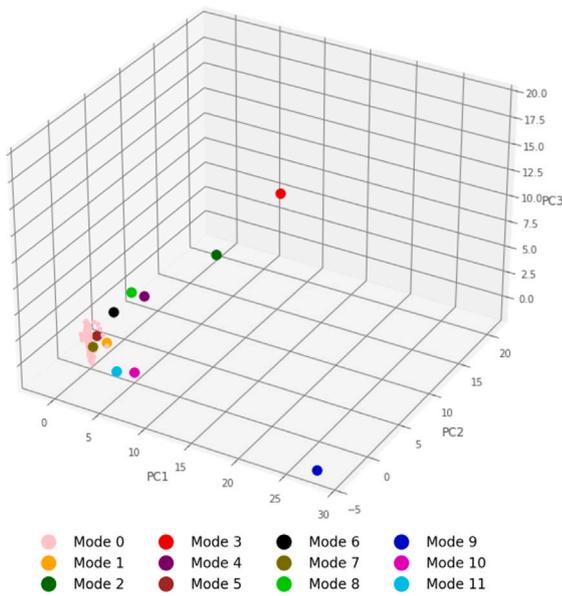**Fig. 6.** Online WWTP operation model ($T_s = 12$ h).

Fig. 7. Updated WWTP operation mode model.

failure, and therefore the appearance of a new variable related to the lighting. These new results have been checked with the experts on the WWTP in order to validate these new modes that appear in this online stage and the explanation given by SHAP.

When the WWTP operation mode model is considered outdated, then it can be rebuilt following the offline steps, generating a new WWTP operation model as shown in Fig. 7, taking into account 15 months of historical samples from the WWTP. Here, some previously exceptional mode samples have now transitioned into *Operation Mode 0* in this new scenario. Conversely, the majority of new samples fall into *Operation Mode 0*, while 3 new exceptional operation modes emerge. Once again, these results have been validated using available expert knowledge, supporting the obtained unsupervised operation modes.

• **Case:** $T_s = 24$ h, Offline

Here, a similar analysis has been made for the sample time $T_s = 24$ h. Once more, the parameter *MinPoints* was selected as only 1 point in

order to allow a better detection of minimal clusters, due to the WWTP operation. Then, ε is tuned at 1.1, following the tuning methodology described above.

In Fig. 8(a), the detected WWTP operation modes are presented in the space of the three main principal components (extracted variables); while Fig. 8(b) illustrates these same operation modes in the space of principal components 3, 5, and 6, to improve the visibility of the operation modes regarding *Operation Mode 0*.

Now, 13 operation modes have been detected: *Operation Mode 0*, into which almost all the samples are clustered, and the rest of the modes, which correspond to exceptional days according to the experts.

Under the WWTP operations, these points are not considered noise, as they show different situations in which the plant was working in comparison to its standard operation mode over the year. Fig. 8(c) shows the same as in Fig. 8(a), but here every sample is in a color according to its sample month, in order to visualize some patterns based on this time feature. Moreover, Fig. 8(d) shows the modes in the 3, 5 and 6 principal component spaces for a better discrimination of the *Mode 0*.

The explanation of these modes is newly obtained by SHAP, and shown in Table 6 (in Extra Supplementary Material). The text color is associated with the WWTP unit for each variable. All this allows us to know the operation modes as well as the facility variables and units involved, which allows us to locate a WWTP data sample in the operation mode model and its match regarding the known modes over time at any given moment.

These results are also checked regarding the available expert knowledge: *Operation Mode 0* corresponds to the regular operation of the plant over the year, while *Operation Modes 2 and 11* concern a general plant shutdown for maintenance. On the other hand, *Operation Modes 3, 4, 6, 7, 10 and 12* are heavy rainy days, with 4 and 6 being the heaviest; consequently, the intake flow variables are highly important for SHAP. Finally, *Operation Modes 1, 5, 8 and 9* seem to be an unknown event a short distance from the regular operation mode; while modes 1 and 5 seem to be a similar event because of the most important variables, related to the preprocessing and thickening facilities, as well as modes 8 and 9 which are related to the pumping arrival facility.
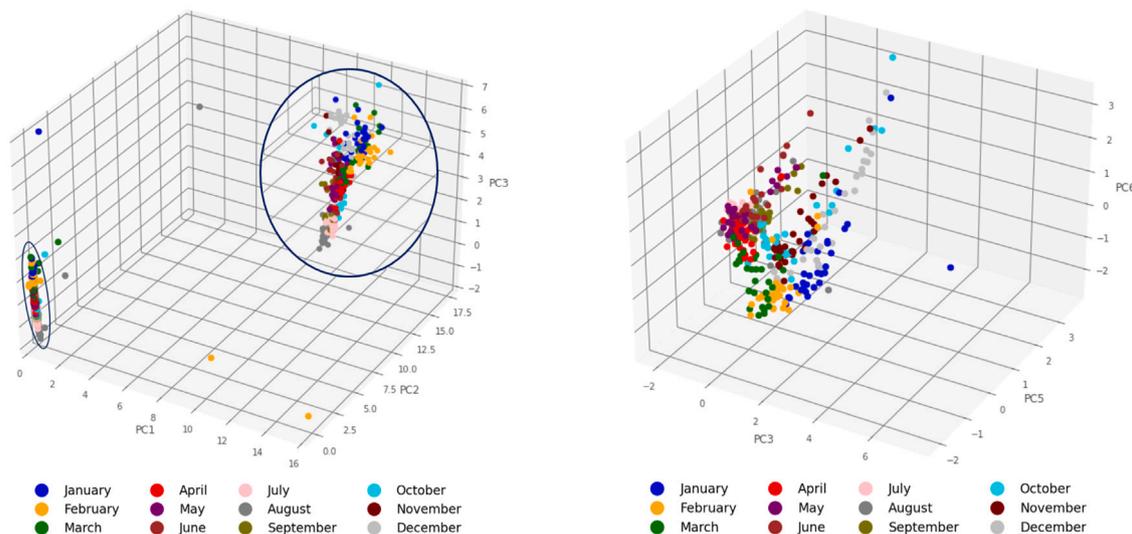
• **Case:** $T_s = 24$ h, Online

Now, each new data sample is shown in this modes model and input space to know its relationship or similarity to known operation modes.

Three months of samples are used for this stage, processed in the same manner as in the previous step. Fig. 9(a) shows the offline modes
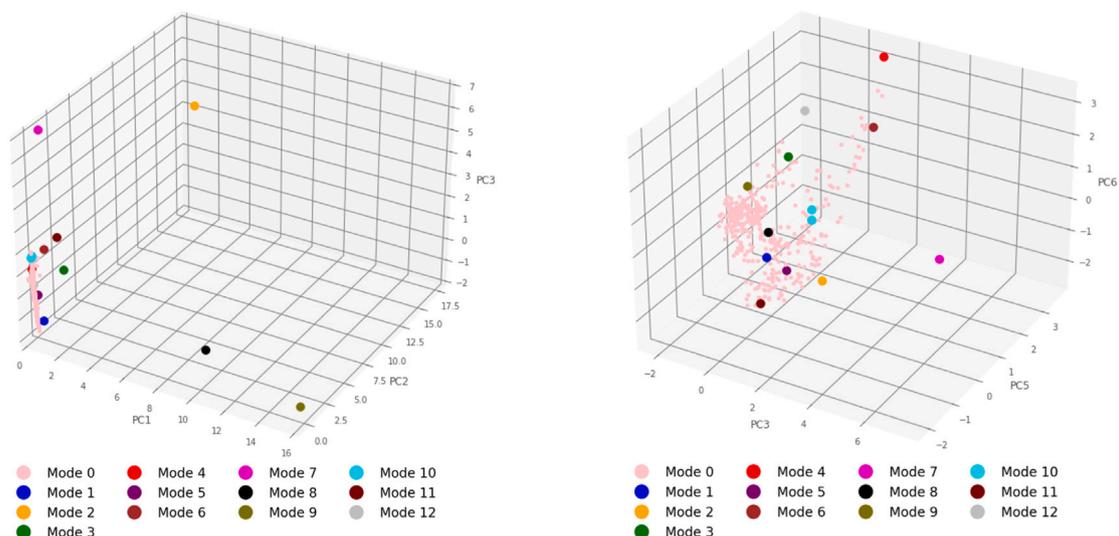
(a) Operation Modes: view by $PC1, PC2$ and $PC3$.

(b) Operation Modes: view by $PC3$, $PC5$ and $PC6$.

(c) Operation Modes by months ($PC1, PC2$ and $PC3$).

(d) Operation Modes by months ($PC3$, $PC5$ and $PC6$).

Fig. 8. Operation modes based on DBSCAN ($\varepsilon = 1.1$ and the $MinPoints = 1$).

in blue and online samples in red. The great majority of online samples go to *Operation Mode 0* (see Fig. 9(a) and (b)), while samples marked as 1 and 2 are far from the regular operation mode, being explained by SHAP in Table 8 (in Extra Supplementary Material). Sample number 1 concerns a general plant shutdown which lasted 2 h due to electrical matters, according to the WWTP experts. Meanwhile, its most influential variables according to SHAP are concerned with the biological units. On the other hand, sample number 2 matches with a connection failure in the plant, with the pumping arrival facility being affected, according to SHAP. These new results have also been validated by the experts on the WWTP.

If the WWTP operation modes model is in need of an update, this is carried out by the offline steps of this proposal. Fig. 10 shows a new WWTP operation model considering the 15 months of historical WWTP data samples. Regarding the DBSCAN operation modes obtained over 12 months, some exceptional mode samples of those months have become *Operation Mode 0* in this new scenario. On the other hand, most of the new samples belong to *Operation Mode 0* and five new exceptional operation modes have emerged. Once more, these results have been

checked with the available expert knowledge, justifying the obtained unsupervised modes.

5. Conclusions

This work introduces an engineering approach to know the WWTP operation modes applied to a real case, AQUAVALL WWTP in Valladolid (Spain). The goal is to translate the large and high-dimensional databases from WWTP monitoring systems, such as SCADA systems, into a point of a low-dimensional WWTP operation modes space, in which the WWTP operation modes are recorded over time as clusters, explained by their most influential WWTP variables.

All this would allow plant managers to know, at a glance, the standard operation modes, the new ones and their closeness and similarities, as well as the variables and units of the plant involved for each one. The large and high-dimensional database is summarized in a small number of operation modes or points in a reduced input space. All this has been carried out using the historical records of the monitoring system of the AQUAVALL WWTP over 15 months: 658084 records with 3231 monitored variables for each one.
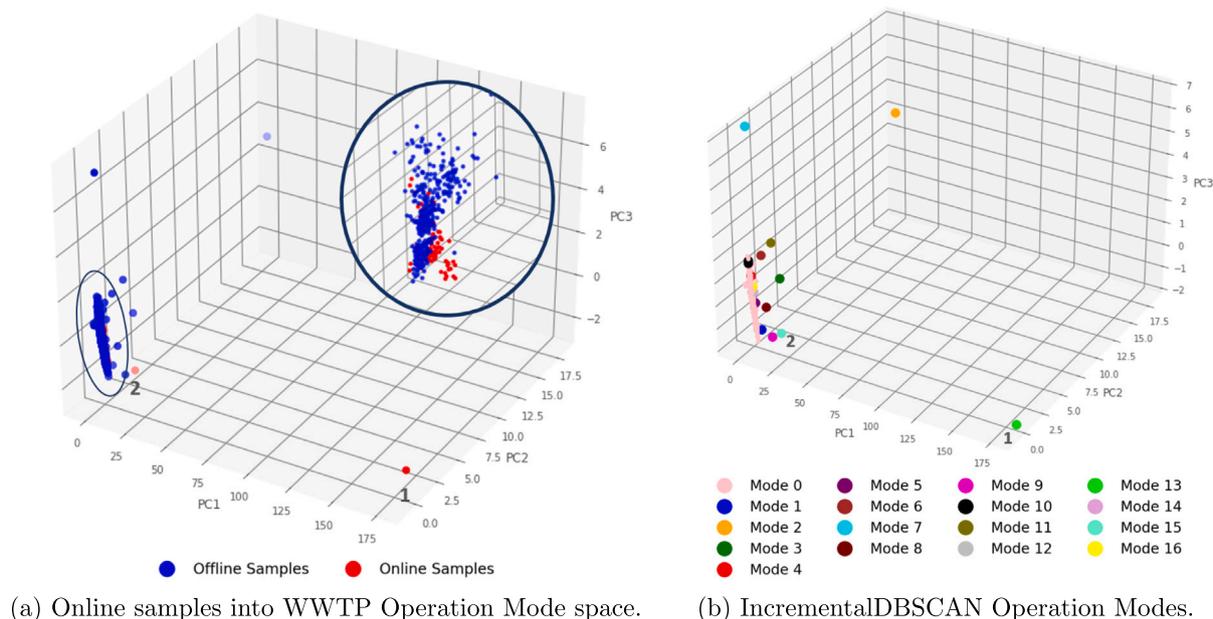
(a) Online samples into WWTP Operation Mode space.

(b) IncrementalDBSCAN Operation Modes.

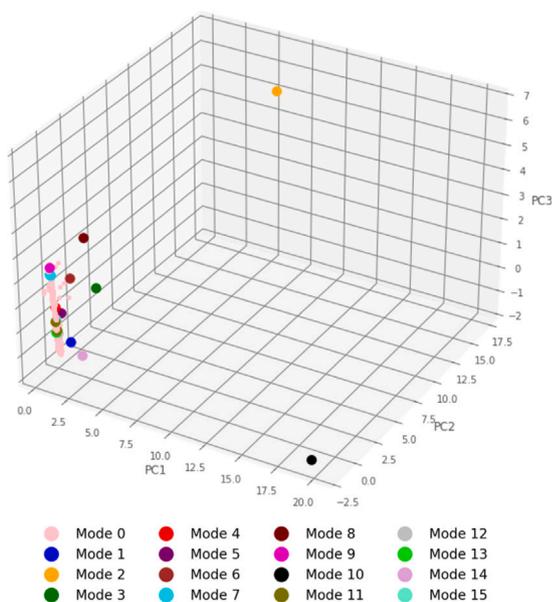**Fig. 9.** Online WWTP operation model (T_s = 24 h).

Fig. 10. Updated WWTP operation mode model.

Resampling, expert knowledge, PCA based dimensionality reduction, density based clustering DBSCAN, and SHAP values for explanation, are the compounds of the methodology applied to deal with the issues of the real WWTP case used in this work.

This methodology was tested with resampling times of 12 and 24 h. It was possible to detect operation modes by matching them with the expert knowledge and the events saved in the plant records, such as shutdown, electrical issues and heavy rain. Explaining all these operation modes by the WWTP variables and units involved, and by analyzing SHAP explanations, we confirmed that the most influential process variables were consistent with expert knowledge, reinforcing the fact that the detected clusters were not arbitrary and reflected real operation modes.

On the other hand, the results revealed some slight differences between the two resampling times: the 12 h sample time captured short-term variations in WWTP operations. The detailed patterns obtained are valuable for detecting transient states, sudden process fluctuations, and short-term anomalies. In contrast, while the 24 h sample time can help in identifying stable operation modes, it may have obscured rapid changes that could be crucial for early anomaly detection and process optimization.

Moreover, the offline stage has allowed a comprehensive cluster analysis; while the online stage has demonstrated the applicability of the method in real-time monitoring. Furthermore, the necessity for model updating has been demonstrated because of the dynamic operation of the WWTP, where treatment processes can fluctuate due to changes in the influent characteristics, seasonal variations, and operation adjustments, changing the operation modes obtained with DBSCAN.

This approach, as with other data-based approaches, relies on validation and interpretation provided by a WWTP expert to fully exploit the detected modes. All this provides new support and feedback for the usual decision making carried out by the plant managers, as well as new knowledge for updating plant management. On the other hand, if this WWTP expert was not available, the detected modes would provide a first and raw working view of the WWTP.

The work in progress is focused on the implementation of the methodology in the real WWTP to automate the operation mode detection and its interpretability, thus providing actionable insights for the operators. Considering this work, the effects of the sampling time for data monitoring systems are being analyzed considering the 'slow' dynamics of the WWTP plant, which would permit technical and economic resources to be saved, making a lighter data processing and computation possible. On the other hand, non-linear feature extraction is being analyzed for a better dimensional reduction of the input space to keep more and better information, but here the interpretability and explicability are serious challenges that need to be addressed. All these avoiding black and gray box approaches, so new XAI approaches are being tested to explain and interpret the predictions of complex machine learning models.

CRediT authorship contribution statement

Alicia Beneyto-Rodríguez: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Formal analysis, Data curation, Conceptualization. **Gregorio I. Sainz-Palmero:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Marta Galende-Hernández:** Writing – original draft, Validation, Supervision, Methodology, Investigation. **María J. Fuente:** Writing – original draft, Validation, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **José M. Cuenca:** Writing – original draft, Validation, Supervision, Data curation.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Gregorio I. Sainz Palmero reports financial support was provided by Spain Ministry of Science and Innovation. Gregorio I. Sainz-Palmero reports financial support was provided by Agencia Española de Investigación. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jwpe.2026.109915.

Data availability

The data that has been used is confidential.

References

[1] Q.V. Ly, V.H. Truong, B. Ji, X.C. Nguyen, K.H. Cho, H.H. Ngo, Z. Zhang, Exploring potential machine learning application based on big data for prediction of wastewater quality from different full-scale wastewater treatment plants, Sci. Total Environ. 832 (2022) 154930, http://dx.doi.org/10.1016/j.scitotenv.2022.154930.

[2] S. Abba, Q.B. Pham, A. Usman, N.T.T. Linh, D. Aliyu, Q. Nguyen, Q.-V. Bach, Emerging evolutionary algorithm integrated with kernel principal component analysis for modeling the performance of a water treatment plant, J. Water Process. Eng. 33 (2020) 101081, http://dx.doi.org/10.1016/j.jwpe.2019.101081.

[3] S. Rahmat, W.A.H. Altowayti, N. Othman, S.M. Asharuddin, F. Saeed, S. Basurra, T.A.E. Eisa, S. Shahir, Prediction of wastewater treatment plant performance using multivariate statistical analysis: A case study of a regional sewage treatment plant in Melaka, Malaysia, Water 14 (20) (2022) 3297, http://dx.doi.org/10.3390/w14203297.

[4] S. Borzooei, G.H. Miranda, R. Teegavarapu, G. Scibilia, L. Meucci, M.C. Zanetti, Assessment of weather-based influent scenarios for a WWTP: Application of a pattern recognition technique, J. Environ. Manag. 242 (2019) 450–456, http://dx.doi.org/10.1016/j.jenvman.2019.04.083.

[5] H. Han, M. Sun, F. Li, Z. Liu, C. Wang, Self-supervised deep clustering method for detecting abnormal data of wastewater treatment process, IEEE Trans. Ind. Informatics 20 (2) (2024) 1155–1166, http://dx.doi.org/10.1109/tii.2023.3268777.

[6] F. Solano, S. Krause, C. Wollgens, An internet-of-things enabled smart system for wastewater monitoring, IEEE Access 10 (2022) 4666–4685, http://dx.doi.org/10.1109/access.2022.3140391.

[7] J. Cohen, X. Huan, J. Ni, Shapley-based explainable AI for clustering applications in fault diagnosis and prognosis, J. Intell. Manuf. 35 (8) (2024) 4071–4086, http://dx.doi.org/10.1007/s10845-024-02468-2.

[8] T. Clement, H.T.T. Nguyen, N. Kemmerzell, M. Abdelaal, D. Stjelja, Beyond explaining: XAI-based adaptive learning with SHAP clustering for energy consumption prediction, 2024, http://dx.doi.org/10.48550/ARXIV.2402.04982.

[9] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, R. Ranjan, Explainable AI (XAI): Core ideas, techniques, and solutions, ACM Comput. Surv. 55 (9) (2023) 1–33, http://dx.doi.org/10.1145/3561048.

[10] M. Cilínio, M. Pereira, D. Duarte, L. Mata, P. Vieira, Explainable fault analysis in mobile networks: A SHAP-based supervised clustering approach, in: 2023 16th International Conference on Signal Processing and Communication System, ICSPCS, IEEE, 2023, pp. 1–9, http://dx.doi.org/10.1109/icspcs58109.2023.10261152.

[11] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Off. J. Eur. Union 2024 (1689) (2024) 1–144, URL https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689.

[12] C. Panigutti, R. Hamon, I. Hupont, D. Fernandez Llorca, D. Fano Yela, H. Junklewitz, S. Scalzo, G. Mazzini, I. Sanchez, J. Soler Garrido, E. Gomez, The role of explainable AI in the context of the AI Act, in: 2023 ACM Conference on Fairness, Accountability, and Transparency, FAccT '23, ACM, 2023, pp. 1139–1150, http://dx.doi.org/10.1145/3593013.3594069.

[13] The High-Level Expert Group on AI presented Ethics Guidelines for Trustworthy Artificial Intelligence, Ethics Guidelines for Trustworthy AI, European Commission, 2019, URL https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

[14] X.K. Polaine, R. Dawson, C.L. Walsh, J. Amezaga, M. Peña-Varón, C. Lee, S. Rao, Systems thinking for water security, Civ. Eng. Env. Syst. 39 (3) (2022) 205–223, http://dx.doi.org/10.1080/10286608.2022.2108806.

[15] K.R. Gamache, Critical infrastructure: Water and wastewater systems sector, in: L.R. Shapiro, M.-H. Maras (Eds.), Encyclopedia of Security and Emergency Management, Springer International Publishing, Cham, 2021, pp. 171–181, http://dx.doi.org/10.1007/978-3-319-70488-3_63.

[16] I. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, Philosophical Trans. R. Soc. A (2016) http://dx.doi.org/10.1098/rsta.2015.0202.

[17] I. Jollife, Principal component analysis, Springer-Verlag, 2002, http://dx.doi.org/10.1007/b98835,

[18] G.T. Reddy, M.P.K. Reddy, K. Lakshmanna, R. Kaluri, D.S. Rajput, G. Srivastava, T. Baker, Analysis of dimensionality reduction techniques on big data, IEEE Access 8 (2020) 54776–54788, http://dx.doi.org/10.1109/access.2020.2980942.

[19] D. Deng, DBSCAN clustering algorithm based on density, in: 2020 7th International Forum on Electrical Engineering and Automation, IFEEA, IEEE, 2020, pp. 949–953, http://dx.doi.org/10.1109/ifeea51475.2020.00199.

[20] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al., A density-based algorithm for discovering clusters in large spatial databases with noise, in: AAAI Press, 96, (34) 1996, pp. 226–231.

[21] J. Sander, M. Ester, H.-P. Kriegel, X. Xu, Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications, Data Min. Knowl. Discov. 2 (2) (1998) 169–194, http://dx.doi.org/10.1023/a:1009745219419.

[22] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems 30, Curran Associates, Inc., 2017, pp. 4765–4774, http://dx.doi.org/10.48550/arXiv.1705.07874.

[23] S. Hart, Shapley value, in: Game Theory, Palgrave Macmillan UK, 1989, pp. 210–216, http://dx.doi.org/10.1007/978-1-349-20181-5_25.

[24] F. Anowar, S. Sadaoui, B. Selim, Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE), Comput. Sci. Rev. 40 (2021) 100378, http://dx.doi.org/10.1016/j.cosrev.2021.100378.

[25] M. Greenacre, P.J.F. Groenen, T. Hastie, A.I. D'Enza, A. Markos, E. Tuzhilina, Principal component analysis, Nat. Rev. Methods Prim. 2 (1) (2022) 100, http://dx.doi.org/10.1038/s43586-022-00184-w.

[26] Aquavall, El ciclo urbano del agua, 2025, URL https://aquavall.es/captacion/.

[27] W.E. Marcílio, D.M. Eler, From explanations to feature selection: assessing SHAP values as feature selection mechanism, in: 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI, 2020, pp. 340–347, http://dx.doi.org/10.1109/SIBGRAPI51738.2020.00053.

[28] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intell. 2 (1) (2020) 2522–5839, http://dx.doi.org/10.1038/s42256-019-0138-9.

[29] K.R. Shahapure, C. Nicholas, Cluster quality analysis using silhouette score, in: 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA, 2020, pp. 747–748, http://dx.doi.org/10.1109/DSAA49011.2020.00096.

[30] D. Moulavi, P.A. Jaskowiak, R.J.G.B. Campello, A. Zimek, J. Sander, Density-based clustering validation, in: Proceedings of the 2014 SIAM International Conference on Data Mining, SDM, pp. 839–847, http://dx.doi.org/10.1137/1.9781611973440.96.

[31] R.J.G.B. Campello, D. Moulavi, J. Sander, Density-based clustering based on hierarchical density estimates, in: J. Pei, V.S. Tseng, L. Cao, H. Motoda, G. Xu (Eds.), Advances in Knowledge Discovery and Data Mining, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 160–172.

[32] K.P. Sinaga, M.-S. Yang, Unsupervised K-means clustering algorithm, IEEE Access 8 (2020) 80716–80727, http://dx.doi.org/10.1109/ACCESS.2020.2988796.