Expert Systems With Applications 185 (2021) 115623

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Water end-use consumption in low-income households: Evaluation of the impact of preprocessing on the construction of a classification model

Karla Oliveira-Esquerre [a], Mariza Mello [a], Gabriella Botelho [a,c,*], Zikang Deng [b], Farinaz Koushanfar [b], Asher Kiperstok [a]

[a] Department of Chemical Engineering, School of Engineering, Federal University of Bahia, Brazil
[b] Electrical and Computer Engineering Department, University of California San Diego, USA
[c] Center of Science and Technology, Federal University of Recôncavo of Bahia, Brazil

ARTICLE INFO

Keywords:
Low-income water end use
Demand management
Random forest model
Adaptive KNN model
ERP measure applied to KNN
Dataset preprocessing

ABSTRACT

The challenge of transforming massive water flow data into desegregated smart information according to water end uses is an issue that has motivated many researchers. This challenge is even more difficult in low-income regions owing to the high variability of data because predominant hydraulic devices offer many activation possibilities for users as they are controlled by globe valves. Devices with standardized flow rates such as washing machines or dishwashers are exceptions. A common practice is to apply commercial software that classifies events at the end-use level and then to develop a personalized classification model with enhanced alignment with the database. If the preprocessing step is not performed properly, it can affect perceived device behaviors, which may lead to incorrect conclusions. To evaluate how this variability can interfere with commercial software responses, we developed classification models using a dataset preprocessed by Trace Wizard® as training data and then applied the trained models to a test dataset consisting of events that were authenticated by individual flow sensors. Our goal was to identify the degree of difference between the two datasets. The results demonstrate that when Trace Wizard® is applied, the features of each device differ from the original water consumption flow, indicating that data variability interferes with the credibility of feedback. Additionally, preprocessing tended to increase the volume, duration, and flow rates, giving the impression that the consumption was higher than the real scenario. The constructed models were not able to overcome the distortions introduced by Trace Wizard® classification. For example, fixtures had poor matches for several houses, with statistical measures below 50%.

1. Introduction

Water end use control and rationalization are essential for the universalization of water access. This concept has been proven in the cases of agricultural and industrial water use and must be expanded to urban and domestic water use. Many countries, particularly developing countries, still experience unacceptable water losses in their urban distribution systems. In some cases, more than 50% of the water produced at treatment facilities does not reach households. Additionally, significant water waste occurs in both rich and poor households with various population demographics.

Given current consumption patterns and habits, there will likely be an increase in residential water demand as a result of global urban growth (Cominola, Giuliani, Piga, Castelletti, & Rizzoli, 2015). Additionally, water pollution, urban development, agricultural irrigation, climate change, and droughts also contribute to disparities between the availability of quality water sources and consumption demand (Jorgensen, Graymore, & O'Toole, 2009). Therefore, in the face of water scarcity, information regarding how and when water is used can aid the development of policies aimed at reducing water consumption (Vašak, Banjac, & Novak, 2015).

Water consumption in buildings, residential areas, commercial enterprises, or institutional facilities depends on multiple factors as discussed by Kiperstok and Kiperstok (2017). Consumption depends on

* Corresponding author at: Network of Clean Technologies, Department of Environmental Engineering, Polytechnic School of Federal University of Bahia, Salvador, Bahia, CEP 40210-630, Brazil.

E-mail addresses: karlaesquerre@ufba.br (K. Oliveira-Esquerre), mariza.smello@gmail.com (M. Mello), gbotelho@ufrb.edu.br (G. Botelho), zkdeng@eng.ucsd.edu (Z. Deng), fkoushanfar@eng.ucsd.edu (F. Koushanfar), asher@ufba.br (A. Kiperstok).

https://doi.org/10.1016/j.eswa.2021.115623
Received 25 February 2021; Received in revised form 14 July 2021; Accepted 14 July 2021
Available online 19 July 2021
0957-4174/

technological, managerial, and behavioral issues. It is widely accepted that awareness is key to the rational use of water and that there can be no awareness without proper control. However, control is not possible without accurate measurements. Water distribution systems are being designed to tackle this issue and the hydro metering of all consumer units is either already conducted or is being actively pursued by water authorities worldwide. Combining information from residential water meters with regional or water sector flow and pressure records can allow urban water losses to be curtailed. Moving water control inside buildings and households is an important challenge for water authorities and consumers. Understanding how water is consumed and whether it is properly meeting a demand or being wasted allows consumers to adopt necessary measures to reduce consumption while satisfying their desires. It also allows authorities, researchers, and suppliers to design strategies to favor more rational equipment and practices.

To identify how water is consumed, wasted, or lost through building hydraulic installations, pipes, reservoirs, faucets, tubs, washing machines, or showers, two main methods are typically used: installing a water meter for each equipment or developing a means to interpret the flow signals from a central water meter[1]. Previous works (Mello, Oliveira-Esquerre, & Botelho, 2018; Soares, Oliveira-Esquerre, de Aguiar, Botelho, & Kiperstok, 2018) have applied both types of methods.

Over the past three decades, research has promoted the development of intelligent water meters for fostering the characterization of water consumption patterns according to end uses (Bennett, Stewart, & Beal, 2013; Liu, Giurco, & Mukheibir, 2016; Nguyen, Stewart, & Zhang, 2014). Several issues make water end use recognition challenging, such as whether an observed time series represents individual or combined events and how combined events can be separated. Additionally, it is difficult to handle multiple behaviors associated with the same fixture or with new user patterns. Currently, a common practice is to use commercial software such as Trace Wizard® to address these issues. Another possibility is to create custom models using preprocessed data, such as with data that are already treated (in cases with simultaneous uses) and labeled according to end-use equipment. This requires reliable software responses because processed data may not represent the real water consumption behaviors of each device. This is a significant issue because models (including secondary steps, if any) mold themselves to the characteristics of data, which influences the choice of pattern recognition techniques, decisions regarding the factors that influence consumption, information about the quantity of water used by hydraulic equipment, and user behavior. Therefore, it is crucial to have a means of verifying the preprocessing step.

This study aimed to explore the importance of having a dataset that is truly rated by end-use equipment to highlight the impact on data behavior when using Trace Wizard® preprocessing, as well as the impact when a model is constructed based on preprocessed data and then applied to classify truly rated data. Two models are explored: a random forest (RF) based on extracting features from time series and a 1-nearest neighbor (1NN) model using edit distance with a real penalty measure (ERP), which calculates the similarity between an unknown time series and reference time series dataset.

The remainder of this paper is organized as follows. The limitations of Trace Wizard® and their implications are discussed in Section 2. In Section 3, a concise literature review of related works on end-use classification methods is presented. In Section 4, the considered classification models are presented. In Section 5, information regarding water flow data is presented and the water consumption characterization results is discussed. Our methodology is discussed in Section 6. We present (a) comparisons between a dataset classified by Trace Wizard® and by individual flow sensors and (b) demonstrate that the selected models are not fully able to conform to data acceptably, when there is major differences between preprocess method responses. Additionally, models were constructed using training data from Trace Wizard® and tested on data classified by sensors. The results are discussed in Section 7. Finally, in Section 8, we summarize the main conclusions drawn from our experimental results.

2. Limitations of the Trace Wizard® application

Trace Wizard® (DeOreo, Heaney, & Mayer, 1996) is a commercial software that can split simultaneous device uses and can classify a time series of flow data into end uses. It uses a decision tree to perform event classification by evaluating similarity based on manually predefined parameters for each type of equipment. The use of this software requires attention to some key points. For example, it is highly dependent on human inputs for the choice of statistics derived from water flow series, such as the duration, volume, maximum flow, and average flow of each end-use equipment. Additionally, it requires a manual validation for all outputs. This software works with templates containing fixtures and their features, where the order in which fixtures appear is key for the classification step. The template order is manually defined in the software. The device positioned at the top of the list is the first to be classified, the second device is the second to be evaluated, etc. The devices that are positioned at the top of the list should be those with more restricted flows, durations, and volume intervals (i.e., devices with a small chance of being confused with other devices). This increases the chances of accurate device classification. Therefore, to achieve accurate classification results, it is important to have device features with well-defined intervals and low variability.

The features of unknown time series, such as volume, duration, mode, and peak flow rate, are compared to those defined in the templates. If there is compatibility within a constrained limit, then the time series is labeled for the corresponding fixture without additional analysis of other devices. However, this creates a limitation in that the software does not analyze all possible matching fixtures, even though this scheme does speed up analysis. Additionally, there may be intersections between the patterns of more than one fixture, such as shower use that could be confused with the use of a kitchen sink. The greater the variability of the patterns, the greater the likelihood of the occurrence of such intersections.

As mentioned previously, many studies that have analyzed time series data from smart meters have created their own classification models. One common approach (generally in supervised models) is the use of a commercial software classification model to perform a data labeling process (see the paper by Nguyen, Stewart, and Zhang (2013)), particularly if the database is large, which makes it difficult to classify all data by other means. However, because commercial software has limitations that increase with the variability of data, as demonstrated in Section 7.1, it may be more suitable to use a custom method to obtain information regarding end-use schedules, such as resident notes for specific periods or data from flow sensors installed at each point of use, which can be used over longer periods. The choice of methodology depends on the number of fixtures considered in a study, resident consent for taking notes, and the size of the team available for processing data. Based on independent information, it is possible to evaluate the classification performance of a software, verify whether the software can be used to classify an entire historical series, and determine if there are excessive distortions.

3. Background of End Use classification methods

For water consumption time series classification, the most popular approach is to use software that applies pattern recognition tools. Some suitable commercially available software are Identiflow® (Kowalski & Marshallsay, 2005), HydroSense® (Larson et al., 2012),

[1] When a water supply network fills a reservoir in a house and water flows from the reservoir to different consumption points, as commonly occurs in developing countries, additional measurement strategies are required. Such strategies are not discussed in this paper.

BuntBrainEndUses® (Pastor-Jabaloyes, Arregui, & Cobacho, 2018), and Trace Wizard®.

Identiflow® applies a decision tree, similar to Trace Wizard®, to identify and classify events based on discriminating information regarding the use of each fixture in a residence. Additionally, it considers the type of housing and plumbing arrangements. An analyst must then review and refine the analysis procedure, which is disadvantageous in terms of human error and computational time. Alternatively, **HydroSense®** processes data from a pressure sensor (installed at a single point in a house) and is able to disaggregate and identify water use events automatically based on a probabilistic classification model. According to Morrison and Friedler (2014), this software is sensitive enough to determine the uses of individual fixtures, as well as the use of hot or cold water. Unlike other techniques, HydroSense® is automatic, meaning that it does not require intervention from an analyst, thereby reducing the potential for human error (Morrison & Friedler, 2014). However, the software's authors did not consider situations in which the pressure in a water distribution network is unstable, so it is not possible to determine the potential effects of such variations on the performance and accuracy of HydroSense®.

In addition to existing commercial software, researchers have also explored and developed novel approaches that utilize other experimental strategies. Nguyen, Zhang, and Stewart (2011), Nguyen et al. (2013), Nguyen, Zhang, and Stewart (2013), Nguyen et al. (2014), Nguyen, Stewart, Zhang, and Sahin (2018) developed the software Autoflow®. Their data were gathered from smart meters installed with a resolution of 0.0014 L/pulse and read at intervals of five seconds. End uses were separated manually by analysts using Trace Wizard®. Time series from the same fixtures were gathered into small groups using a similarity measure called dynamic time warping (DTW) and representative time series were identified. Reducing the number of reference time series reduces the computational time of classification algorithms. However, the authors reported that their procedure is not applicable to fixtures whose consumption patterns exhibit significant variability. They constructed an algorithm based on a hidden Markov model (HMM) combined with device features (volume, duration, fashion, peak), DTW, probability of a particular time of day, and information provided by residents. Their latest version uses a "self-organizing maps" algorithm for single/combined event separation. Separated events are then classified using HMMs and an artificial neural network (ANN). This combination of techniques was applied in an unsupervised manner and did not rely on any pre-training steps (Nguyen et al., 2018). Wonders, Ghassemlooy, and Alamgir Hossain (2016) conducted a comparative study on the classification efficiencies of three different machine learning techniques: ANN, support vector machines (SVM), and K-nearest neighbors (KNN). They evaluated the effects of increasing the training database size through the generation of synthetic data. They also varied the parameters of each technique, as well as the number of considered variables. Their study was conducted on data from one bathroom of a two-resident house. The authors justified their single-room approach based on classification error reduction.

Pastor-Jabaloyes et al. (2018) also developed a technique for water consumption disaggregation based on automatic and semiautomatic steps. Classification was performed based on a partition around medoids (PAM). Given a dataset, this algorithm searches for k objects that should represent the medoids (centers) of clusters. The remaining data are assigned to the closest medoids using an arbitrary dissimilarity measure. This algorithm is similar to KNN. The features used were total volume and average flow rate. Duration was ignored owing to the noise associated with long water uses. Fontdecaba et al. (2013) used a multivariate Gaussian distribution with variables chosen based on non-random heuristics. Their time series were analyzed in full form, as well as at the beginning and end, by considering variables such as duration and volume. Vašak et al. (2015) employed a robust linear classification technique for feature vectors. They removed the least-used devices and considered multiple variables, including the duration, volume, maximum flow, mode, and the number of occurrences of modes. They also considered $sin2\pi t/T$ and $cos2\pi t/T$ (t is the number of minutes elapsed starting at midnight and T is the number of minutes in 24 h), as well as the squares of these values, to mitigate nonlinearity. By considering that fixtures of the same type can exhibit different behaviors between houses, they tested their algorithm by implementing it on a per-residence basis and with multiple residences combined. Rahim, Nguyen, Stewart, Giurco, and Blumenstein (2020) provided an important perspective on water measurement techniques, including the evolution of the methods used to classify datasets of water-use events since 2010.

Regarding the techniques presented above, there are some limitations to their application. For example, (a) Identiflow® is not available for use outside the UK, (b) HydroSense® deals with data from pressure smart meters, and (c) Trace Wizard® is not recommended for equipment that exhibits significant behavior variability. This last observation is connected to certain difficulties that have been encountered with the decision tree, PAM, ANN and SVM algorithms (also stated by researchers working with these algorithms), whose inputs are event features. Such methods exhibit poor classification performance owing to (1) the difficulty of separating events correctly in space and (2) similar features of different devices, leading to misclassification or unspecified uses. Autoflow® is only compatible with a specific type of smart meter, but other attempts have been made to use HMMs. However, HMMs only operate on the probability of occurrence of a specified flow range. Therefore, if an event exhibits a new flow value, an error will occur.

Based on the challenges discussed above, we developed two new classification methods, which are presented in the following section.

4. Developed classification models

The most commonly used hydraulic devices are faucets (kitchen, bathroom, outdoor areas), showers, and toilets. These types of end uses depend on human handling, which in turn depends on user behavior, level of awareness regarding the proper use of water, and the condition of hydraulic installations (Kiperstok & Kiperstok, 2017). Additionally, based on the characteristics of these types of devices, it is natural to assume that they will be used for performing several functions, contributing to an increase in the variation of durations and volumes for a given type of equipment. Among the techniques we analyzed and tested, the adaptive 1NN model with an edit distance with real penalty (ERP) measure and the RF model stood out when considering the variability of the target data. These models were implemented using the R programming language. In both cases, household models were constructed separately.

4.1. Adaptive 1NN model with an edit distance with real penalty (ERP) measure

This model (1NNERP) performs comparisons between unknown time series and sets of reference time series (that represent the behavior of each device) to determine the correct class. This concept is advantageous based on its capability of assimilating different uses related to the same equipment.

We constructed a classification model based on the concept of the KNN algorithm. As this learning algorithm is non-parametric and lazy (meaning that it memorizes a dataset, rather than creating a discriminative function), there are no parameters to be trained. For KNN classification, the output is the assignment of an object to the most common class among its k closest neighbors, where k is a natural number greater than zero (Wonders et al., 2016). If $k = 1$, then the object is assigned to the class of its nearest neighbor. The distance between objects is the most important measure for this process and is calculated using a (dis) similarity measure.

A similarity measure computes the strength of the relationship between two objects. ERP (Chen and Ng, 2004) comes from the field of character series comparison and computes the number of insertions, exclusions, or substitutions that are required to make two series of potentially different lengths identical to each other. For a numerical series, a distortion matrix containing $d(x_i, y_j)$ is constructed and the extent to which two series are dissimilar is calculated according to the best alignment that exists between them.

ERP (Eqs. (1) and (2)) uses a constant g to quantify possible gaps in a time series and uses the L1 distance (also known as the Manhattan distance) as the distance between elements (measure of dissimilarity). This measure satisfies the triangular inequality and is sensitive to noise, but its computational time is greater than Euclidean distance. However, in general, the time series are relatively short, and the time cost is related to the size of the reference group and number of tested time series.

Our initial tests included other measures such as the dynamic time warping (DTW) applied by Mello et al. (2018). In our final tests, ERP stood out as the most suitable measure.

$$E(i,j) = \begin{cases} \sum_{k=1}^{j} |y_k - g|, i = 0 \\ \sum_{k=1}^{i} |x_k - g|, j = 0 \\ min \begin{cases} d_{erp}(x_i - y_j) + E(i-1, j-1) \\ |x_i - g| + E(i-1, j) \qquad , i \neq 0 and j \neq 0 \\ |y_j - g| + E(i, j-1) \end{cases} \end{cases} \tag{1}$$

$$d_{erp}(x_i, y_j) = \begin{cases} |x_i - y_j|, if x_i, y_j not gaps \\ |x_i - g|, if y_j is a gap \\ |y_j - g|, if x_i is a gap \end{cases} \tag{2}$$

Based on the information above, an outline of the model is presented in Fig. 1. An unknown time series is compared to the reference database by using ERP to compute distances and the class is selected according to the closest neighbor. Additionally, the RF model is applied cooperatively to classify events declared as unknown by the 1NNERP model. This step was not present in the flowchart because it was rarely required.

In terms of optimization, the parameter g has not been modified because it is a reference related to the Cartesian system. There is the possibility of imposing global restrictions, assuming that an alignment away from the distortion matrix's diagonal is probably incorrect. Such restrictions can be imposed by creating an adjustment window with a given length r (Sakoe & Chiba, 1978). However, the size of this window is related to the time series length and because there is a wide variety of lengths, this strategy is not feasible.

This model directly utilizes time series. In contrast, the RF operates on features extracted from time series and uses them to recognize triggered devices.

4.2. RF model

The RF model is essentially a set of trees constructed randomly to predict classification labels based on a majority vote (James, Witten, Hastie, & Tibshirani, 2013). To understand the process used by RF models, it is necessary to understand the construction of classification decision trees. The goal of such a tree is to partition data into smaller and more homogeneous groups. When travelling down the tree, data are split into two possible responses that symbolize the branches of a tree. These responses are called nodes. To perform each partitioning operation, a commonly used metric is "Gini," which is a variable that allows RF models to partition the nodes of each tree into more homogenous groups that contain a larger proportion of one class in each subsequent node (Kuhn & Johnson, 2013). The construction of a tree, which consists of partitioning each node and ordering of the predictor variables in terms of importance within the tree, is performed based on Gini impurity. There are other metrics similar to Gini, such as cross entropy and misclassification error. However, Gini is the most commonly used metric (James et al., 2013).

The equation of the Gini index is provided in Eq. (3), where $k$is the number of classes in the model and $\hat{p}_{mk}$ is the probability of the occurrence of class k at node m. In this study, three different classes were evaluated based on types of hydraulic devices: toilets, showers, internal faucets. The sum of all probabilities is equal to one, as shown in Eq. (4).

$$G = \sum_{k=1} \hat{p}_{mk}\left(1 - \hat{p}_{mk}\right) \tag{3}$$

$$p_1 + p_2 + p_3 = 1 \tag{4}$$

Therefore, by substituting Eq. (4) into Eq. (3), Eq. (5) can be

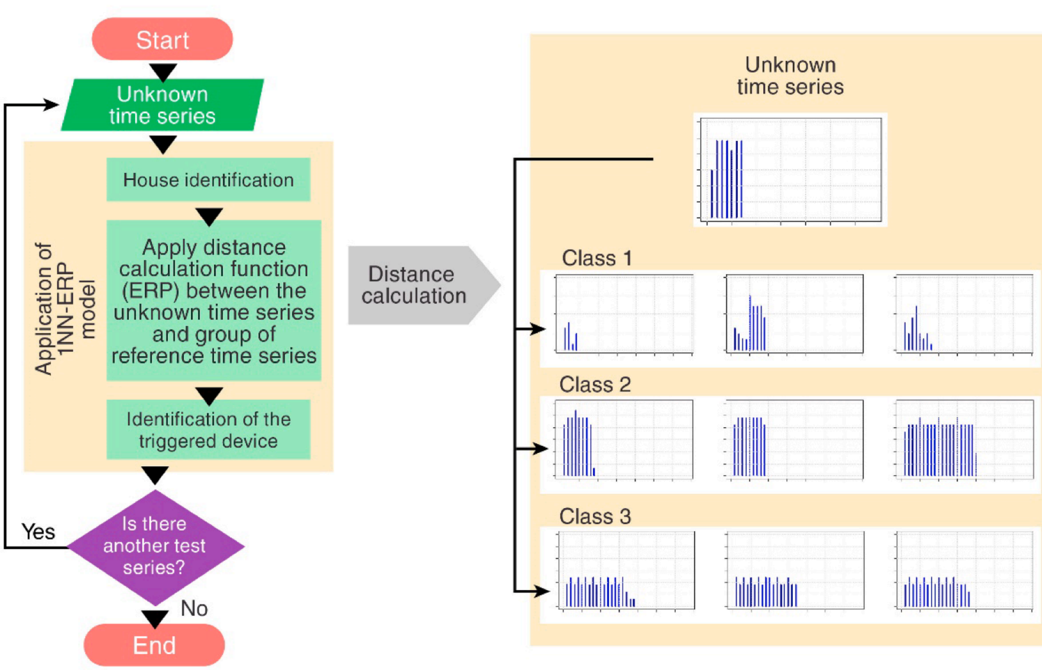

Fig. 1. 1NNERP model simplified scheme.

obtained.

$$G = 2(p_1p_2 + p_1p_3 + p_2p_3) \quad (5)$$

The main hyperparameters used for the construction of an RF model are the number of trees (*ntree*) and number of predictor variables or input data to be randomly sampled for each node in a tree (*m* or *mtry*). The literature recommendation is to begin with a number of trees equal to 1000 and to increase this number until model performance stabilizes. Regarding *mtry*, the literature recommendation for classification problems is to use the square root of the number of predictor variables. However, tests can be performed with different values to identify the model with the smallest error or greatest accuracy. These are called tuning tests for the evaluation of hyperparameters (Kuhn & Johnson, 2013).

The criteria for constructing trees using the RF model are based on a series of random selections of information from the original database with reinsertion. This means that the information drawn by one tree returns to the database to be part of the drawing process for additional trees. This process is called bootstrapping. Each tree is created with a bootstrapped sample from the original database and approximately one third of the original data are not considered (Breiman, 2001). The data that are not selected at the time of random selection are then used to test the classification performance of new sets of trees constructed based on out-of-bag observations. This process provides information regarding classification errors (James et al., 2013). After generating an RF with bootstrapped samples, the process of obtaining an output ensemble of trees follows four steps: (1) selecting *m* or *mtry* variables randomly from the *p* predictor variables, (2) choosing the best variable or cutoff point among the *m* or *mtry* variables sampled, (3) separating nodes into two child nodes, and (4) classifying the majority vote of the *b*th RF tree (Hastie & Tibshirani, 2009).

The selection of hyperparameters to maximize model performance can also be performed using optimization techniques. Grid search and manual search are the most commonly used approaches for hyperparameter optimization. However, the random search method is more efficient (Bergstra & Bengio, 2012). The grid search and random search methods are known as brute-force searches because they search for all possible known solutions while recording the solution that yields the best outcome. Other approaches known as smart searchers can also be used for hyperparameter optimization, such as Bayesian optimization (Wu, Chen, Zhang, Xiong, & Lei, 2019), which uses a probabilistic model to find the minimum of any function that returns a real value metric and uses it to select the most promising hyperparameters for evaluating the objective function. Other methods include *meta*-heuristics such as the arithmetic optimization algorithm (Abualigah, Diabat, Mirjalili, Abd Elaziz, & Gandomi, 2021), which is based on the distribution behavior of main arithmetic operators, grasshopper optimization algorithm (Abualigah & Diabat, 2020), and particle swarm optimization (Li & Zhang, 2020). These are bio-inspired methods. Another *meta*-heuristic method is the sine–cosine algorithm, which is a population-based optimization algorithm (Abualigah & Diabat, 2021).

Since the RF training is based in select sample data from database, it is important to evaluate if there is imbalanced data. If it occurs, it is likely that several trees work with only one type of class, which can make the model response biased. In this study, as will be seen in the posterior sections, there are fixtures visibly more triggered than others and, to overcome this issue, two RF models were constructed. One for the most used fixture and other to the remaining ones (Fig. 2). Here, the use of internal faucets is superior to the other fixtures in most residences; so the approach was to create a RF model focused in the internal faucet (RF1) and another one focused in the other fixtures (RF2). The first consisted in relabel the other fixtures as "X" and by doing that the balance between the classes was improved; the response is either internal faucet or X. If, for a given event, the first step response was "X" that means this event was related to the other fixtures. Then it would pass to the second model (RF2), which labeled the event based in the remaining fixtures.

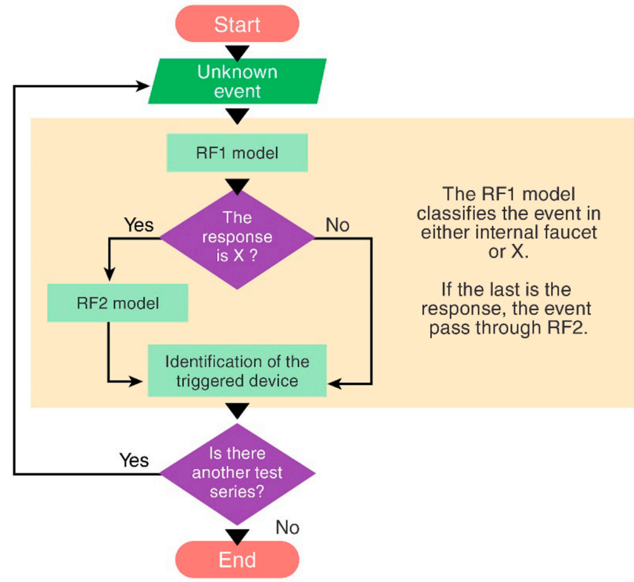

Fig. 2. RF model simplified flowchart.

5. Water flow data

5.1. Characteristics of the study area

Our study was performed in Plataforma, Salvador, Bahia. The study location is highlighted in Fig. 3.

Plataforma is one of the oldest districts in Salvador. This neighborhood can be characterized as residential based on the presence of only small- and medium-sized businesses. An important local characteristic is that the majority of the inhabitants have low purchasing power and little schooling, with most residents having only completed elementary school.

This information was confirmed in a survey conducted on 111 residences in Plataforma. Fig. 4 presents the statistics of this survey. One can see that 80% of the families have minimum wages of up to 538 USD, 60% have no schooling or incomplete elementary schooling, and more than 70% have more than two years of debt related to water and sewage services.

5.2. Data gathering

The dataset for our study was provided by the "Rational Use of Water

Fig. 3. Location of the study area.

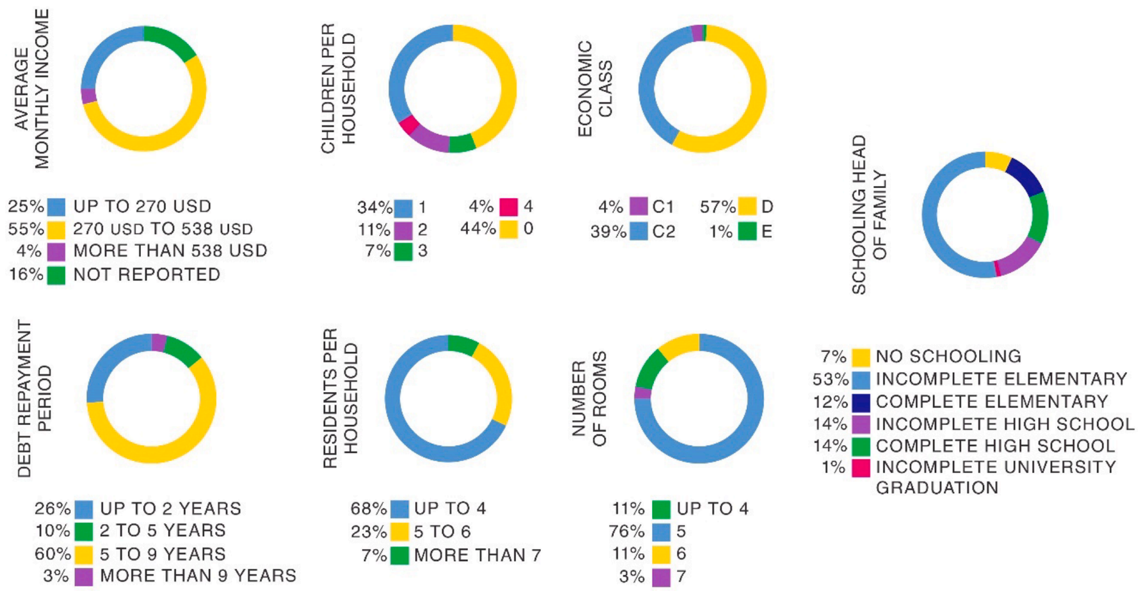

Fig. 4. Statistical information for the residents in the study location.

and Energy Efficiency in Housing of Social Interest" project as a response to the "Public Call: Environmental Sanitation and Housing 07/2009" by the Financier of Studies and Projects (FINEP) and Basic Sanitation Research Program (PROSAB). The data were collected by researchers from the Cleaner Technology Network of Bahia (Teclim-BA) of Bahia Federal University (UFBA).

Water consumption from seven residences was monitored using a system consisting of an LAO® device, DN 20 device, and metrological class-B multi-jet water meter installed at the reservoir outlet. The sensor emits one pulse for every 0.1 L of water used and records information every 10 s. All smart meters were calibrated and verified. Data were collected for approximately one year. The installation of the monitoring system in some households required adjustments to the original hydraulic installations, as shown in Fig. 5. The dashed lines indicate modifications made to the measurement systems.

Monitoring included an investigative phase of 5 to 14 days and a continuous monitoring phase of approximately one year. Table 1 presents information on the number of inhabitants, water meter situations, types of fixtures, and number of monitored days (total and investigative time period). In general, the common fixtures between residences were toilets, showers, and faucets. All residences had a water meter from the concessionary. Some residences had an irregular water meter situation, such as house C, where the water meter was broken. Irregularities were also observed for houses D and G, despite these houses having installed and functional water meters. This is probably because the residents found a way to bypass the device to avoid full payment of their water bills. In houses without a water meter (A, E, and F) or with a water meter that did not work (C), water consumption was estimated according to the average consumption in the region.

During the investigative time period, individual water flow meters for each fixture were also installed. YF-S201 sensors were installed and linked to a platform for collecting data. These sensors notified the data collection platform if a fixture was activated and reported the duration of each use and volume of water consumed. This information provided accurate event categories for the dataset.

The external faucet water consumption in houses B, C, and G was estimated based on hydric balance because the external faucets were located at points where the team identified that the measurement equipment could be stolen and based on a lack of power points for connecting the data concentrator modules.

5.3. Consumption routines and peculiarities of water use

Fig. 6 presents the monthly average consumptions of the seven residences for all monitored periods. One can see that five of the seven residences (A, B, E, F, and G) have monthly water consumption levels below 10 m^3/month. The value of 10 m^3/month was considered as the consumption threshold below which a unique minimum rate was charged. For the residents who benefited from social tariffs (those who were already registered to receive government benefits such as *Bolsa Família*), the water tariff represented 2.6% of the monthly minimum wage. However, residents not enrolled in the government programs paid an equivalent of 5.2% of the monthly minimum wage. The United Nations Development Program, in website publication water for life decade and human rights for water suggests that water costs should not exceed 3% of household income. However, in rate systems with minimum consumption thresholds, residents are not financially rewarded for their efforts to reduce water consumption, particularly for small families. Additionally, even if larger families consume the bare minimum for meeting basic needs per inhabitant of 50 L per person per day (Gleick, 1996), the total consumption is greater than the minimum rate. Therefore, a greater percentage of the family salary is spent on the water bill because there is an extra charge for each cubic meter consumed above the minimum value.

In 2017, there was a change in the tariff policy of the utility company in Bahia state that modified the minimum fixed rate charged for consumption to 6 m^3 of water per month. Although there was no increase in the price of the tariff, there were changes in the manner in which residents were charged. Fig. 6 reveals that only house E would pay the minimum water rate under this new policy. However, as discussed previously, fewer households would have the problem of not being rewarded with bill reductions when they manage to reduce consumption, while greater consumption would still be penalized.

Water usage was characterized by extracting features such as volume, mode, maximum flow, and average flow (Fontdecaba et al., 2013; Nguyen et al., 2013; Vašak et al., 2015). Table 2 lists some statistics regarding water use by each fixture in each house measured during the investigative period.

The use of toilets for most houses exhibits similar variations. However, two houses exhibit higher consumption as a result of leaks that were not promptly repaired (houses A and C). In these cases, the leaks were not isolated because it was not possible to separate them from the

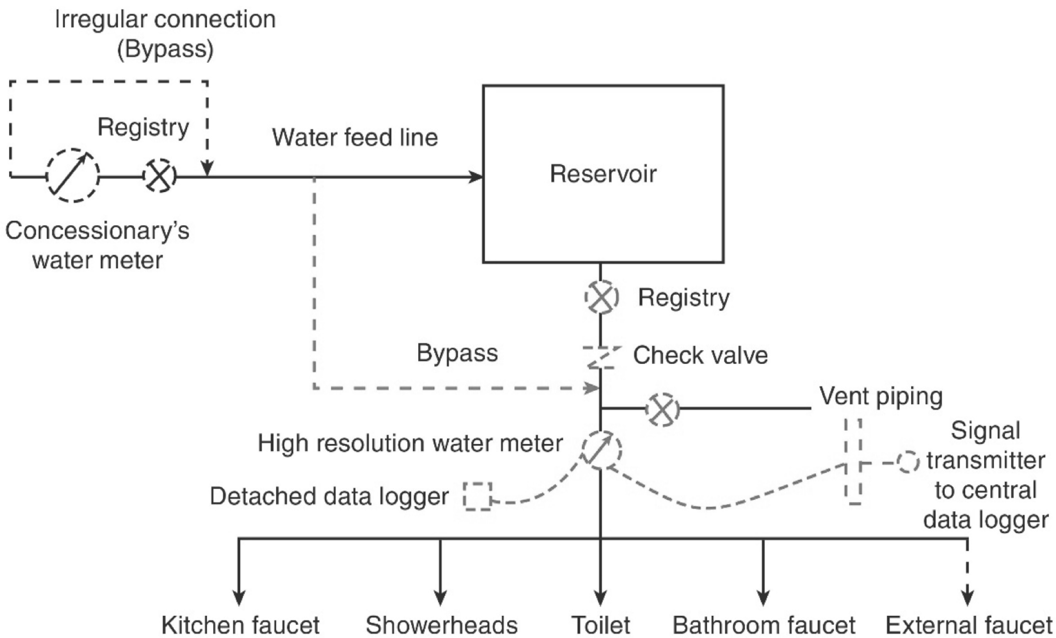

Fig. 5. Residential hydraulic installation system, before (black line) and after (red line) the installation of the smart meters.

Table 1
Characteristics of the monitored residences.

Residence		A	B	C	D	E	F	G
Number of inhabitants		2	2	6	2	2	3	2
Water meter situation		N	R	I	I	N	N	I
Total number of fixtures		4	5	9	7	5	4	6
Number of fixtures	Toilet	1	1	2	2	1	1	2
	Shower	1	1	2	2	1	1	1
	Internal faucet	2	2	4	2	2	1	2
	External faucet	0	1	1	1	1	1	1
Number of monitored days		320	396	236	222	307	341	246
Investigative time period	Number of monitoredDays	8	9	5	8	9	14	11

N: no water meter, R: regular meter situation, I: irregular meter situation.

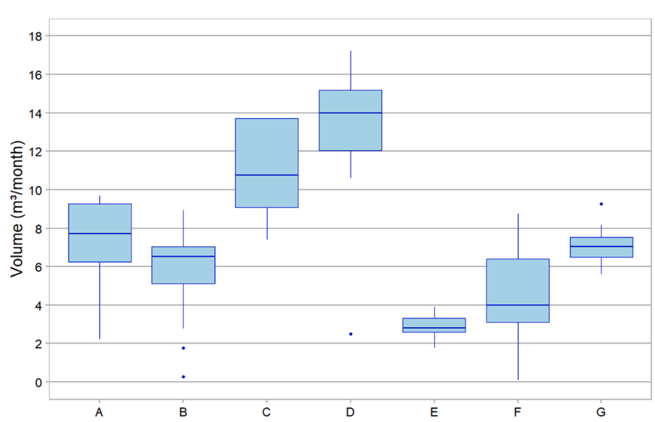

Fig. 6. Total average water consumption per month.

toilets correctly.

A comparative analysis of water consumption in toilets reveals that in three of seven residences (B, C, and E), the average volume per flush is greater than 6 L. This suggests a lack of calibration in either water flux or water reservoir storage.

Regarding the flow rates in showers, five of seven residences had a range between 3.0 and 8.5 L/min (C, D, E, F, and G). Additionally, the residents of house A also used their shower to wash clothes because they do not have an external faucet, which explains the large number of usage events identified. Another relevant finding is that none of the houses except D had hot water showers, which justifies the short shower durations observed in the data. Although substantial use of water is expected for houses with heated showers, the increasing cost of energy bills leads to a reduction in consumption in low-income areas.

Regarding the internal faucets, we noticed that families that were involved in economic activities such as making and selling food (A and G) exhibit a greater number of uses for internal faucets (average of 78 uses per day). For house G, in addition to a significant number of uses, a higher median volume and flow rate can be observed compared to the other houses, which could be attributed to extra cooking and cleaning processes. The flow rates of these events in all residences are between 1.2 and 2.4 L/min. Six of seven residences (A, B, D, E, F, and G) exhibit a median duration of less than 20 s. The family in house C, which confirmed that they never closed the faucet during dishwashing, is the only house to have a 30 s median duration of use. Not turning off the tap between uses during one activity increased the average median duration of consumption by 50% and resulted in one of the greatest mean volumes and flow rates. This house also contains the greatest number of inhabitants (six), which could increase the variability of the modes of use and number of uses (average of 46 uses per day). In the same interview, the family in house D confirmed that they did not close the faucet during dishwashing and that they tended to use a high flow rate, even for small tasks. Such habits could contribute to the relatively high

Table 2
Characteristics of water use in residences during the investigative period monitored by sensors.

Toilet	House	N (uses)	$\bar{V}$(L/use) ± σ_V($\tilde{V}$)	$\bar{D}$(s) ± σ_D ($\tilde{D}$)	$\bar{Q}$(L/min) ± σ_Q ($\tilde{Q}$)	Shower	N (uses)	$\bar{V}$(L/use) ± σ_V ($\tilde{V}$)	$\bar{D}$(s) ± σ_D ($\tilde{D}$)	$\bar{Q}$(L/min) ± σ_Q ($\tilde{Q}$)
	A	51	5,7 ± 5,1 (3,9)	82 ± 65 (70)	4,11 ± 1,68 (4,80)		157	2,9 ± 5,4 (0,2)	89 ± 220 (20)	1,60 ± 1,20 (0,60)
	B	45	8,7 ± 1,0 (8,5)	95 ± 13 (90)	7,01 ± 0,96 (7,20)		4	-	-	-
	C	13	8,0 ± 3,2 (8,0)	131 ± 38 (140)	3,65 ± 1,28 (3,60)		66	3,3 ± 3,3 (2,1)	53 ± 43 (40)	3,55 ± 1,97 (3,15)
	D	35	5,9 ± 7,1 (5,1)	103 ± 103 (90)	3,97 ± 0,73 (4,2)		51	11,2 ± 10,1 (9,2)	117 ± 86 (90)	5,71 ± 1,86 (6,00)
	E	41	6,1 ± 2,7 (6,8)	169 ± 72 (190)	2,18 ± 0,65 (2,40)		28	2,8 ± 1,8 (2,4)	57 ± 32 (55)	3,30 ± 1,29 (3,45)
	F	55	5,3 ± 1,5 (5,2)	238 ± 61 (240)	1,24 ± 0,20 (1,20)		46	9,6 ± 8,6 (6,5)	101 ± 80 (80)	5,65 ± 1,73 (6,00)
	G	44	3,9 ± 2,9 (4,4)	65 ± 36 (70)	3,12 ± 2,03 (2,85)		56	7,7 ± 6,7 (6,3)	60 ± 41 (50)	7,27 ± 3,65 (8,40)
Internal faucet	**A**	621	0,9 ± 1,4 (0,5)	29 ± 32 (20)	1,64 ± 1,00 (1,20)	**External faucet**	–	–	–	–
	B	460	0,9 ± 1,1 (0,6)	25 ± 27 (20)	2,12 ± 1,69 (1,80)		–	–	–	–
	C	228	2,1 ± 3,3 (1,0)	39 ± 38 (30)	2,62 ± 1,57 (2,40)		–	–	–	–
	D	274	2,3 ± 3,0 (0,8)	36 ± 41 (20)	2,92 ± 2,02 (2,40)		–	–	–	–
	E	285	0,6 ± 0,6 (0,4)	21 ± 12 (20)	1,59 ± 0,79 (1,50)		23	1,4 ± 1,2 (0,9)	37 ± 23 (30)	2,39 ± 0,87 (2,4)
	F	43	0,6 ± 0,3 (0,5)	21 ± 8 (20)	1,61 ± 0,65 (1,50)		20	5,9 ± 4,7 (5,6)	73 ± 47 (65)	4,74 ± 1,62 (5,4)
	G	883	1,1 ± 1,1 (0,8)	23 ± 18 (20)	2,65 ± 1,60 (2,40)		–	–	–	–

$\bar{V}$: mean volume, σ_V: volume standard deviation, $\tilde{V}$: median volume, $\bar{D}$: mean duration, σ_D: duration standard deviation, $\tilde{D}$: median duration, $\bar{Q}$: mean flow rate, σ_Q: flow rate standard deviation, $\tilde{Q}$: median flow rate.

values observed in the mean flow rate and volume per use of this house.

External faucets or laundry faucets in most parts of the houses were used to fill semiautomatic devices to wash clothes or for manual clothes washing. A laundry machine was only present in house D. During the study period, only the flow rates of the external faucets in houses E and F were measured and their median flow rates were 2.4 and 5.4 L/min, respectively. The median durations of use were 30 and 65 s for houses E and F, respectively, which are longer than most usage durations of internal faucets based on the types of uses of these devices (e.g., filling a bucket or washing clothes).

The variations mentioned above lead to a higher probability of misclassification based on the overlap of water consumption time series. Regarding the use of appliances, it is known that the habits of each family significantly influence the volume and flow rate per use. Personal habits such as turning off the tap when lathering or opening a faucet with a lower flow rate also impact consumption. When comparing the results between Fig. 6 and Table 2, one can observe a significant impact on the final consumption values caused by these factors. For example, house D, which only has two inhabitants, had a higher total consumption compared to a house with three inhabitants (house C) and houses with economic activities that use water (A and G) as a result of differences in personal habits.

6. Methodology

6.1. Comparison of databases using preprocessing methods

The investigative week was a period used to understand and label the water consumption values related to each fixture. The data collected during this period represents 2% to 4% of the training dataset for each residence. This dataset was classified using Trace Wizard® (TW-class) and individual flow sensors (FS-class). For TW-class, an experienced researcher inputted of the features and statistics required for classification. In FS-class, data on consumption were obtained from the YF-S201 sensors installed in each device. The events were then manually classified. The sensors provided the times at which the fixtures were activated. The data were then classified with real assignments based on cross-references between this time information and the flow meter data. The data were then compared to the total water consumption measured by a hydrometer. For both methods, there were difficulties in differentiating simultaneous fixture uses.

First, a comparison was made between the two methods to understand disparities and to discuss their impacts. Next, models were constructed with the goal of verifying if TW-class could overcome the disparities and provide reliable results.

6.2. Initial modeling tests

Initial tests were conducted to evaluate the performance of the adopted models based on data from FS-class. Based on the results obtained, it was possible to verify that the models were suitable for our case study. This step was performed by partitioning the data for training (80%) and testing (20%). This division accounted for the number of events per device (i.e., 80% of each set of events per device was used in the training phase). Unfortunately, this stage was negatively affected by the amount of data available for certain residences (Table 1). The fixtures evaluated were internal faucets, toilets (less in house C), and showers (less in house B) based on the available data.

For the RF model, the “caret” and “randomForest” packages were adopted. The “caret” package has a group of functions for training and plotting classification and regression models. Specifically, “createDataPartition” was used to separate the training and testing dataset at a 4:1 ratio. The “randomForest” package provides functions for classification and regression using the RF algorithm based on the study by Breiman (2001). The “randomForest” function was used to implement an RF algorithm. Regarding predictive variables, there were several choices for constructing the models, such as duration (time of each use per device), nmode (number of times the flow rate mode is repeated), volume (water consumed in each use), mode (more frequent flow rates), average, peak, median flow rate, resident, hours of day, and days of week. According to the result of sensitivity analysis, the optimal model component variables were selected. Next, tuning tests were performed to evaluate the hyperparameters to be used in the model to obtain the lowest computational cost and processing time while maintaining quality.

To tune the main model hyperparameters *mtry* and *ntree* the grid search technique was applied; the range for *mtry* was 1 to 17 and for *ntree* was 100 to 1000. Since two RF models were used to mitigate the imbalanced data, the tuning process was made in both. Following sensitivity analysis and hyperparameter tuning, our final model was implemented with the variables of resident, duration, volume, and median flow and the hyperparameters for RF1 was $ntree = 100$ and $mtry = 2$; and for RF2 was $ntree = 100$ and $mtry = 1$.

For the 1NNERP model, comparisons between unknown time series and the database were performed simultaneously to reduce the loop duration. The ERP function belongs to the package "*TSdist*," which contains a set of time series distance measures discussed in the literature that are useful for several types of applications, including classification algorithms. Specifically, the function *ERPDistance* was used in our model.

The RF accuracy was 87% in terms of events and 80% for volume consumed. The 1NNERP achieved accuracy of 88% for events and 85% for volume consumed. When the 1NNERP algorithm could not reach a conclusion regarding a given event (either not classifying or presenting more than one possible answer), the RF was used to resolve the issue (the RF algorithm was occasionally used).

In general, both models showed the ability to classify data accurately (Figs. 7 and 8), despite the similarities between some devices caused by certain types of uses. The performance of the models is better observed by analyzing the percentage of consumed volume correctly identified; this is the most relevant, since the future intention is to be able to derive information, which will be pertinent if they characterize most of the consumption. In terms of consumed volume, the classification ability of both models[2] was excellent (Fig. 8), which shows that the misclassified events were those with lower volumes. Internal faucets, which were the most used fixtures, yielded high statistical measure values for both models, in terms of events (Fig. 7). Also, the toilets were well classified in both models; however, the statistical measures for showers were average but, as said previously, the major consumed volume was correctly classified making the models reliable.

In addition to algorithm optimization, we also considered other improvements for enhancing classification, such as the pre-selection of fixtures based on the time of day, day of week, volume, and duration. None of these methods appeared to have an effect on classification accuracy. Wonders et al. (2016) suggested increasing the database size by synthesizing time series, but synthesis would be overly complex based on the variability in the types of use.

The 1NNERP model provides the ability to use a set of time series that represent the same fixture instead of becoming overly focused on a single time series. For example, a faucet can be used for washing hands, cleaning dishes, cooking, etc., so by having a set of potentially matching time series, the model more accurately reflects the data. Additionally, this model requires minimal human intervention (i.e., when results are inconclusive).

The RF has several advantages in that it can handle different types of data (categorical, numerical, binary) using a single model and numerical data do not need to be rescaled. Because each tree operates on a sample of the training data, the algorithm is computationally inexpensive. Another perspective on the algorithm is related to the multiple trees that receive randomly selected data subsets. This allows several ways to interpret an unknown event, without restricting it.

Accordingly, to our objectives, the models were developed with training data from TW-class. Tests were conducted on data from FS-class (Fig. 9). Additionally, testing was extended to the training data to analyze the highest statistical measure values that could be achieved.

6.3. Statistical measures

To evaluate model performance, we used the precision and sensitivity measures acquired from a confusion matrix. The confusion matrix is an $m \times m$ matrix, where m is the number of variables predicted by the model. The matrix rows correspond to the predicted results and the matrix columns correspond to the ground-truth results. The main diagonal of the matrix represents predicted information equal to the known results, which means that it represents accurate predictions.

Precision (Eq. (6)) is related to classifier quality and indicates the percentage of data that were assigned to the correct class. Sensitivity (Eq. (7)) indicates the percentage of original data of a given class that were classified correctly. Both metrics were evaluated based on the number of events and water volume. Ideally, a large portion of the volume consumed (high sensitivity) would be correctly categorized because our goal is to be able to derive new information in the future, which would be possible if most consumption was categorized properly. Additionally, it is important for the classifier to detect activated fixtures accurately (high precision) to instill confidence in a given answer.

$$Precision = \frac{True\ positive}{\sum\ Predicted\ condition\ positive} \tag{6}$$

$$Sensitivity = \frac{True\ positive}{\sum\ Condition\ positive} \tag{7}$$

7. Comparative analysis of results

7.1. Impact of preprocessing on the perception of equipment behavior

When comparing the classification results provided by the two methods, approximately 34.3% of the events were classified equally, corresponding to 33.8% of the water volume (with FS-class as the reference for classification). Tables 3 and 4 list the percentages of correspondence per house and fixture in terms of both events and consumed volumes. For the external faucet, house E was classified completely oppositely from Trace Wizard®, whereas 45% of the events in house F were classified correctly, corresponding to 57% of the consumed volume. The internal faucets exhibit correct correspondence between methods in general. However, for house C, there is only a 46% correspondence, accounting for only 23% of the volume consumed. Regarding the showers, for events in houses A and B, the methods exhibit no correspondence. Houses E and F exhibit values above 70% (for events and consumed volume) and houses C, D, and G exhibit low values per event with volume percentages above 50%. Regarding the toilets, in houses E, F, and G, the volume percentages are greater than 70%, while house A has a very low percentage.

TW-class contains some misclassifications, such as some shower uses being labeled as toilet uses based on the template hierarchy. There were also misclassifications between toilets and internal faucets, and between showers and internal faucets in the cases of short events. Additionally, there were simultaneous events that were recognized as single events.

From a behavioral perspective, these differences are reflected in the distributions of median flow rates (Fig. 10), volumes (Fig. 11), and durations (Fig. 12).

Regarding the external faucet, as mentioned previously, the flow from only the faucets in houses E and F were measured by the sensors. In other houses, parameters were obtained based on hydric balance. Therefore, only the two aforementioned houses have histograms based on FS-class. For house E, TW-class did not recognize the use of this

[2] Sensitivity represents the percentage of events correctly identified and precision represents model reliability

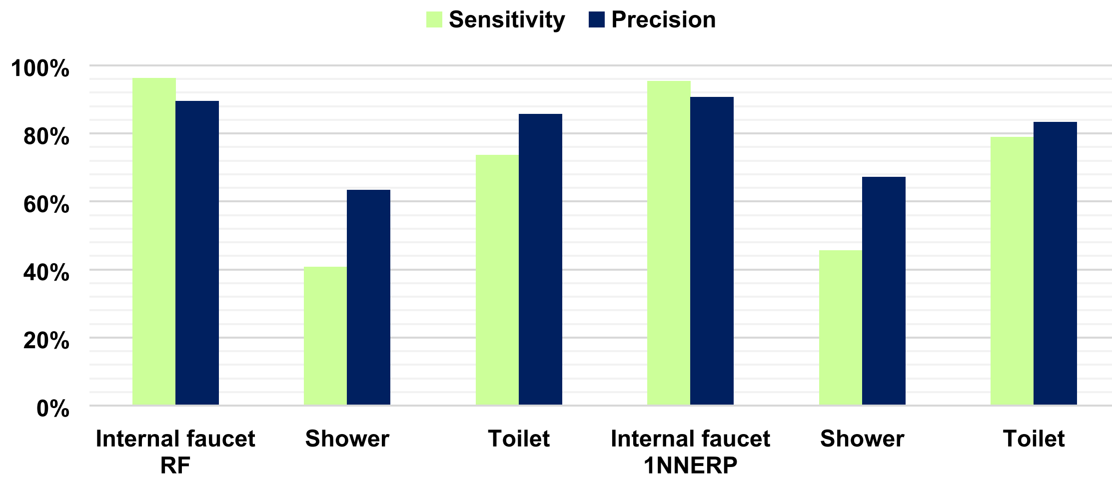

Fig. 7. General precision and sensitivity by fixture for RF and 1NNERP models.

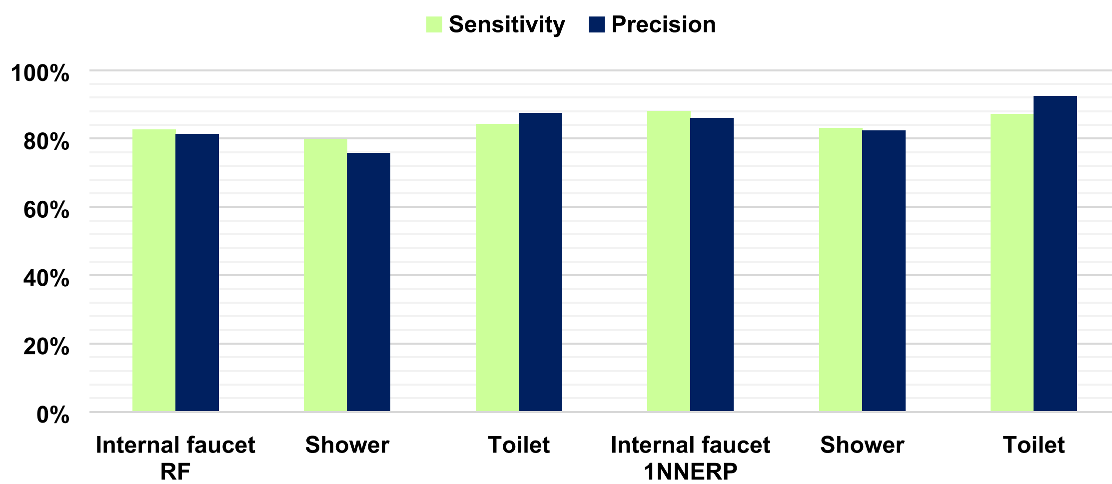

Fig. 8. General precision and sensitivity by fixture for RF and 1NNERP models. In terms of consumed volume.

fixture. When it was applied to house F, the most frequent median flow value was approximately 6 L/min and the duration was approximately 65 s. In contrast, FS-class exhibits a more distributed flow rate with a higher incidence of smaller values and durations distributed with a median of approximately 130 s.

Regarding the internal faucets, there was a significant correspondence between the methods for all houses, except for house C. In this house, there were two families, one in the main house and another in a secondary house. Therefore, there were two kitchen faucets and two washbasin faucets. Consequently, the variability of these fixtures was higher, which contributed to the classification difficulties of TW-class. We found that the internal faucets were often misclassified as showers, which can be attributed to the residents not closing faucets during dishwashing, as mentioned previously.

Regarding the showers, we initially observed that houses A and B exhibited no correspondence between methods. For house A, the first position in the template was occupied by the shower, which allowed more uses to be included in this event. Additionally, the shower had a duration restriction of 90 s, which reduced the uniqueness of shower features for this house. The analysis of house B was compromised by the small amount of data (four events). For houses C and D, the shortest uses were misclassified as internal faucets for TW-class. In the template of house C, the internal faucet appears before the shower and it had characteristically smaller volumes. Therefore, short shower uses were classified as internal faucet uses. In house D, despite the shower appearing before the internal faucet in the template, it only registered events with volumes above 2.5 L.

Finally, regarding the toilets, the correspondence between the methods was similar to that of the showers with some variations in results. For house A, where the correspondence between methods was virtually nonexistent, template analysis revealed that the shower (which appears before the toilet) restricted the mode and peak flows to intervals from 1.8 to 4.8 L/min and less than 5.4 L/min, respectively. However, this information is in contrast with the graphs for TW-class because both the shower and toilet have high rates of manual intervention in their classification, which could justify the differences observed in the graphs.

In general, the methods exhibited distinct differences in results ranging from minor to severe. Based on our analysis of the graphs, there are clear behavioral distortions. TW-class generally exhibits larger volumes, more intense flows, and longer durations, which would lead to misguided demand management measure suggestions. Internal faucets were an exception, potentially because they are the most-used fixtures, and exhibited the expected features.

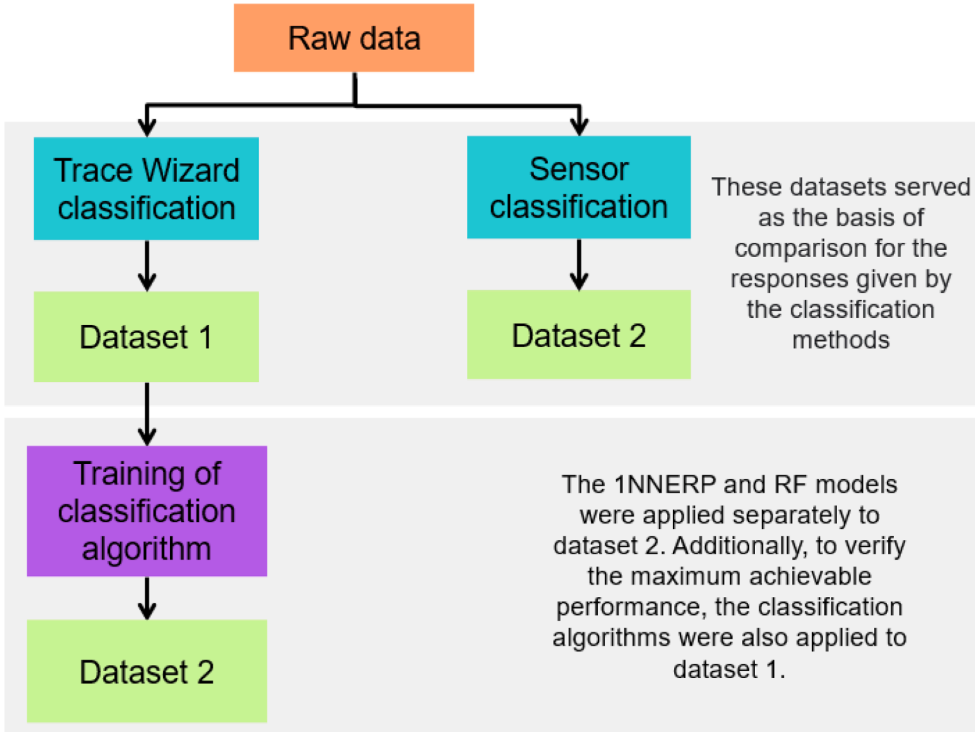

Fig. 9. Flowchart of the methodology used for pre-classifying raw data from the investigative week and testing the classification models. The purple box represents the RF and 1NNERP models, which were applied separately.

Table 3
Percentage correspondence in terms of single events.

	A	B	C	D	E	F	G
External faucet					0%	45%	
Internal faucet	80%	98%	46%	95%	94%	81%	92%
Shower	0%	0%	30%	35%	75%	70%	63%
Toilet	2%	67%	46%	83%	78%	75%	66%

Table 4
Percentage correspondence in terms of volume.

Fixture	A	B	C	D	E	F	G
External faucet					0%	57%	
Internal faucet	77%	92%	23%	81%	79%	86%	81%
Shower	0%	0%	56%	64%	88%	69%	84%
Toilet	2%	65%	46%	68%	92%	76%	83%

7.2. Model tests

All analyses were performed based on precision and sensitivity calculated from confusion matrices. Matrices were constructed for all residences, but given the number of matrices, they are not presented in this paper. These results were presented in tables.

- Internal faucets

In terms of sensitivity (Table 5), all houses except for house C had more than 80% of events and volume consumed correctly classified by both the 1NNERP and RF models. This behavior is expected considering that our analysis indicating that there was a significant correspondence between the pre-classification methods for this fixture type. House C exhibited a low percentage for volume classification because the internal faucet was misclassified as a shower.

In terms of precision (Table 5), both the RF and 1NNERP models were able to identify more than 70% of the events that were attributed to internal faucets correctly. For house F, the 78% precision in terms of events corresponded to 60% of the volume. This is because the models considered additional short consumption events that corresponded to other fixtures.

- Showers

In terms of sensitivity (Table 6), compared to the internal faucets, the percentage of accurately classified events was lower. The events percentages were very low to houses A, C and D; volume percentages were higher because longer uses were correctly classified. For house A, as expected based on our previous analysis, almost none of the events were correctly identified.

In terms of precision, both models were able to correctly identify most events for houses D, F, and G. For houses A, C, and E, the precision was relatively poor. These analyses are valid for both models because they exhibited similar responses.

- Toilets

In terms of sensitivity (Table 7), the toilets yielded significant performance for both volume and number of events compared to the internal faucets and showers. For house A, the results were unsatisfactory, as expected. For house G, the toilet was misclassified as an internal faucet (in terms of precision). For this house, the discharge of one of the toilets was improvised through the use of a bucket of water filled by a faucet. In general, the classification results for the other houses were adequate.

Again, these analyses are relevant to both models because they exhibited similar responses.

8. Conclusions

An investigative period is fundamental for the development and

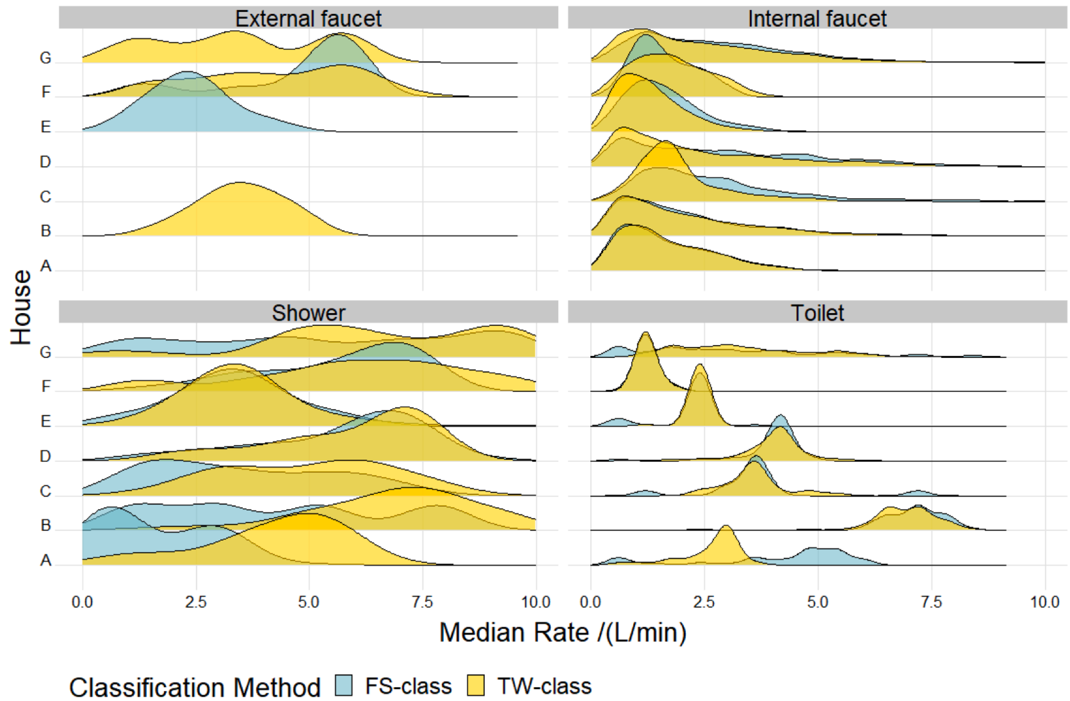

Fig. 10. Median rate histograms for each fixture in each house, differentiated by pre-classification methods.

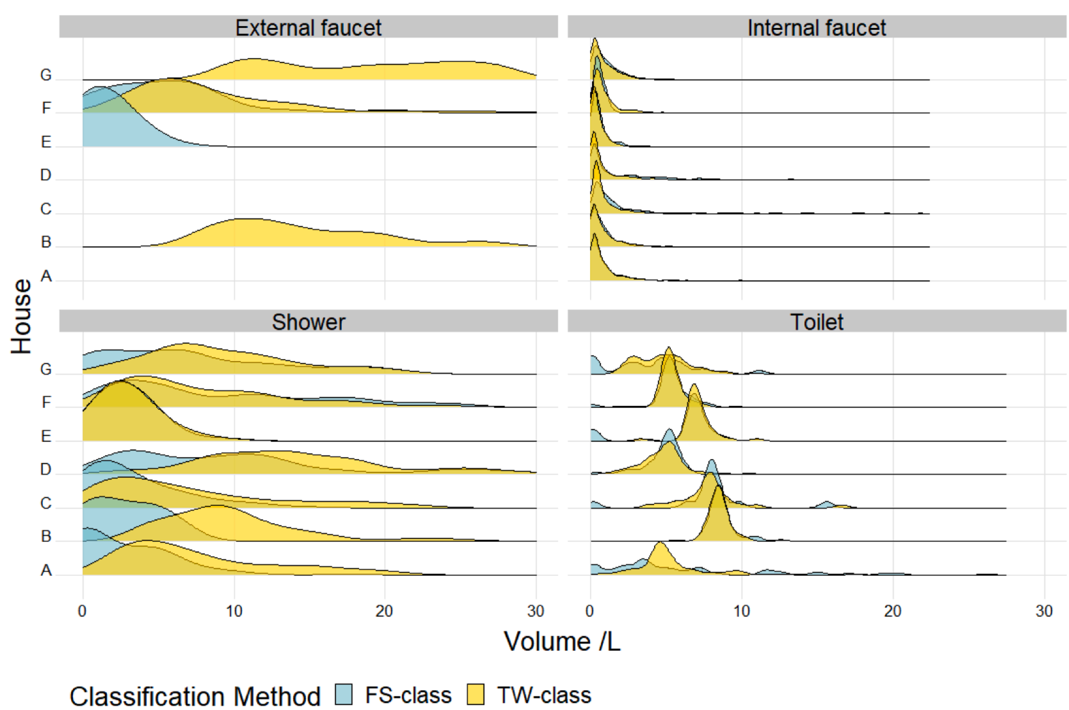

Fig. 11. Volume histograms for each fixture in each house, differentiated by pre-classification methods.

validation of supervised models, but also for the quantitative understanding of features per device and potential changes over time. However, even residences located in neighborhoods with similar architectural, socioeconomic, and climatic characteristics exhibit considerable water consumption variations per device, which makes generalization difficult. Therefore, a reliable prior labelling of time series is fundamental for constructing classification models for water end uses and ensuring their performance.

In this study, data from investigative periods were classified per hydraulic device using sensors that detected the activation of each device and by Trace Wizard®. Although the latter classification was conducted under the supervision of a researcher, who provided input information for templates, there were significant consumption behavior distortions per fixture for residences with large variations in consumption behavior. These differences could not be overcome by the RF and 1NNERP models.

This study does not propose the installation of individual water sensors for each device owing to the high costs involved and resident

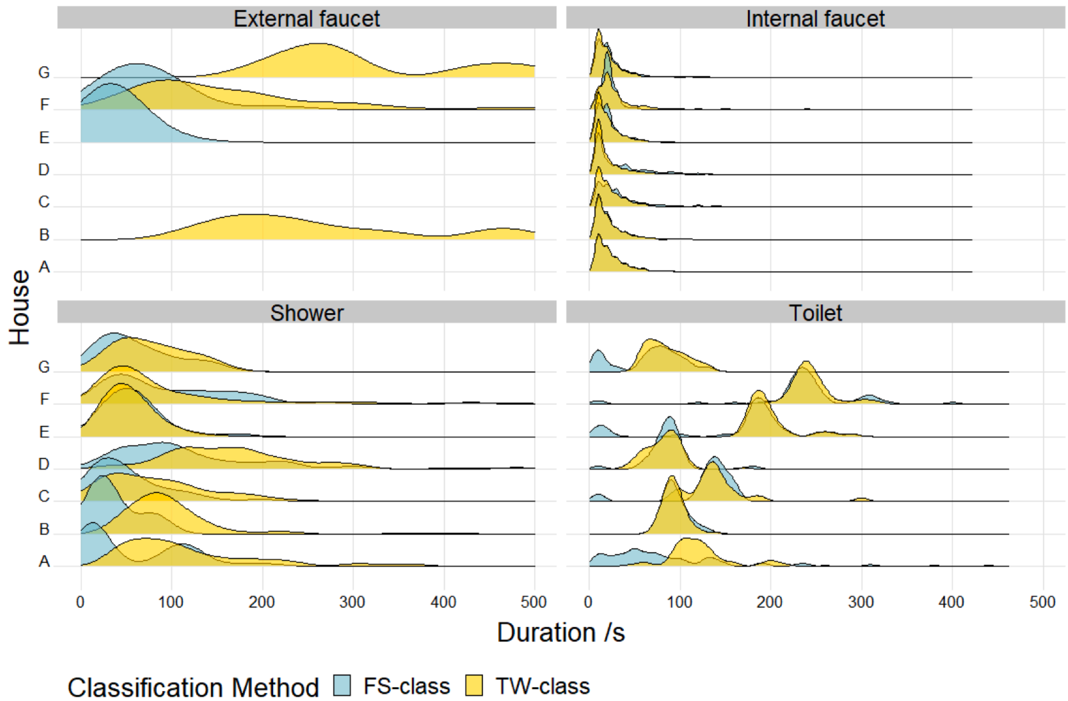

Fig. 12. Duration histograms for each fixture in each house, differentiated by pre-classification methods.

Table 5
Results of precision and sensitivity per house in terms of volume and number of events for internal faucets. The numbers in parentheses indicate the maximum potential of the models.

House	Model	Precision (events)	Sensitivity (events)	Precision (volume)	Sensitivity (volume)
A	RF	85% (99.6)	97.4% (99.8)	92.3% (99.3)	89.9% (99.5)
	1NN-ERP	85.1% (100)	96.9% (100)	93.3% (100)	88.8% (100)
B	RF	99.1% (100)	98.9% (100)	97% (100)	94% (100)
	1NN-ERP	99.1% (100)	98.9% (100)	97% (100)	94% (100)
C	RF	82.8% (98.9)	84.2% (99.8)	76.5% (99.1)	43.6% (99.9)
	1NN-ERP	82.4% (99.1)	84.2% (99.8)	75.1% (99.4)	43.2% (99.9)
D	RF	88.8% (99.8)	95.3% (100)	73.3% (99.4)	80.7% (100)
	1NN-ERP	89.1% (99.9)	95.3% (100)	73.5% (99.9)	80.7% (100)
E	RF	89.1% (99.2)	94.7% (99.9)	83% (98)	79.1% (99.4)
	1NN-ERP	89.7% (99.7)	95.1% (100)	84.7% (99.4)	80.2% (100)
F	RF	78.2% (100)	100% (100)	60.8% (100)	100% (100)
	1NN-ERP	79.6% (100)	100% (100)	64.9% (100)	100% (100)
G	RF	96.6% (99.8)	96.3% (100)	96.2% (99.6)	85.5% (100)
	1NN-ERP	96.7% (99.9)	96.3% (100)	96.5% (99.9)	85.5% (100)

Table 6
Results of precision and sensitivity per house in terms of volume and number of events for showers. The numbers in parentheses indicate the maximum potential of the models.

House	Model	Precision (events)	Sensitivity (events)	Precision (volume)	Sensitivity (volume)
A	RF	0% (100)	0% (94.4)	0% (100)	0% (96.1)
	1NN-ERP	0% (100)	0% (100)	0% (100)	0% (100)
C	RF	43.3% (98.7)	39.4% (93.8)	36.1% (99.9)	66.3% (99.3)
	1NN-ERP	43.1% (98.7)	37.9% (95.1)	35.7% (99.9)	64.4% (99.5)
D	RF	79.2% (100)	37.3% (96.7)	84.9% (100)	67.7% (99.2)
	1NN-ERP	79.2% (100)	37.3% (98.3)	84.9% (100)	67.7% (99.8)
E	RF	47.5% (98.5)	67.9% (91.5)	51.7% (99.1)	84.2% (96.7)
	1NN-ERP	51.2% (100)	75% (97.2)	53.5% (100)	87.8% (99.1)
F	RF	76.7% (93.9)	71.7% (98.6)	83.8% (96.5)	71% (96.6)
	1NN-ERP	79.1% (100)	73.9% (100)	84.7% (100)	71.6% (100)
G	RF	89.7% (98.1)	62.5% (87.9)	92.5% (97.9)	90.2% (94.6)
	1NN-ERP	92.7% (100)	67.9% (98.3)	95.1% (100)	93.4% (99.9)

reluctance. However, our results demonstrate that Trace Wizard® patterns generated for model training must be carefully analyzed because they may not represent realistic water consumption patterns.

Overall, the two fundamental goals of residential water consumption research are (1) providing information for public water management agencies to enable them to promote better initiatives for more rational water consumption by residents and (2) encouraging residents to become aware of the importance of rational consumption, which can be achieved based on information provided to them regarding their own water consumption. To achieve these two major goals, reliable consumption information is essential. This type of data processing is complex and cannot provide perfect solutions, but it is an important step to ensure coherent results and for designing accurate demand management measures.

Table 7
Results of precision and sensitivity per house in terms of volume and number of events for toilets. The numbers in parentheses indicate the maximum potential of the models.

House	Model	Precision (events)	Sensitivity (events)	Precision (volume)	Sensitivity (volume)
A	RF	1.4% (95.6)	2% (97.8)	1.8% (96.2)	2.4% (99.2)
	1NN-ERP	1.3% (100)	2% (100)	1.8% (100)	2.4% (100)
B	RF	100% (82.8)	82.2% (96.4)	100% (82.6)	79.9% (95.1)
	1NN-ERP	100% (100)	68.9% (100)	100% (100)	67.8% (100)
D	RF	73.2% (100)	88.2% (98.4)	70.7% (100)	93.8% (99.1)
	1NN-ERP	73.8% (100)	91.2% (100)	71% (100)	95.4% (100)
E	RF	97.1% (100)	80.5% (100)	98.6% (100)	94.7% (100)
	1NN-ERP	97.1% (100)	80.5% (100)	98.6% (100)	94.7% (100)
F	RF	100% (100)	89.3% (100)	100% (100)	92.8% (100)
	1NN-ERP	100% (100)	87.5% (100)	100% (100)	91.2% (100)
G	RF	46.9% (97)	68.2% (99.2)	48.7% (96.2)	82.4% (98.2)
	1NN-ERP	49.2% (100)	70.5% (100)	52.5% (100)	88.8% (100)

CRediT authorship contribution statement

Karla Oliveira-Esquerre: Conceptualization, Methodology, Project administration, Supervision, Writing - original draft, Writing - review & editing, Formal analysis, Funding acquisition. **Mariza Mello:** Software, Writing - original draft, Writing - review & editing, Visualization, Formal analysis, Validation, Data curation. **Gabriella Botelho:** Writing - original draft, Writing - review & editing, Data curation, Investigation, Formal analysis. **Zikang Deng:** Software, Writing - review & editing. **Farinaz Koushanfar:** Supervision, Writing - review & editing. **Asher Kiperstok:** Project administration, Resources, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge PROSAB and FINEP for providing financial support for our research, Teclim-BA for providing datasets, and the Coordination for the Improvement of Higher Education Personnel-CAPES (CAPES/PRINT - 41/2017, Proc. N. 88887.467907/2019–00) for their visiting scholarship at UCSD. Additionally, we wish to acknowledge PhD Kelly Fontoura for helping with data collection and the residents of the households who gave their time and allowed their water consumption to be monitored.

References

Abualigah, L., & Diabat, A. (2020). A comprehensive survey of the Grasshopper optimization algorithm: Results, variants, and applications. *Neural Computing and Applications, 32*(19), 15533–15556. https://doi.org/10.1007/s00521-020-04789-8

Abualigah, L., Diabat, A., Mirjalili, S., Abd Elaziz, M., & Gandomi, A. H. (2021). The Arithmetic Optimization Algorithm. *Computer Methods in Applied Mechanics and Engineering, 376*, 113609. https://doi.org/10.1016/j.cma.2020.113609

Abualigah, L., & Diabat, A. (2021). Advances in Sine Cosine Algorithm: A comprehensive survey. *Artificial Intelligence Review, 54*(4), 2567–2608. https://doi.org/10.1007/s10462-020-09909-3

Bennett, C., Stewart, R. A., & Beal, C. D. (2013). ANN-based residential water end-use demand forecasting model. *Expert Systems with Applications, 40*(4), 1014–1023. https://doi.org/10.1016/j.eswa.2012.08.012

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research, 13*, 281–305.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5–32. https://doi.org/10.1023/A:1010933404324

Chen, L., & Ng, R. (2004). On The Marriage of Lp-norms and Edit Distance. In *Proceedings of the 30th VLDB Conference* (pp. 792–803). https://doi.org/10.1016/b978-012088469-8/50070-x

Cominola, A., Giuliani, M., Piga, D., Castelletti, A., & Rizzoli, A. E. (2015). Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review. *Environmental Modelling and Software, 72*, 198–214. https://doi.org/10.1016/j.envsoft.2015.07.012

DeOreo, W. B., Heaney, J. P., & Mayer, P. W. (1996). Flow trace analysis to access water use. *Journal - American Water Works Association, 88*(1), 79–90. https://doi.org/10.1002/j.1551-8833.1996.tb06487.x

Fontdecaba, S., Sánchez-Espigares, J. A., Marco-Almagro, L., Tort-Martorell, X., Cabrespina, F., & Zubelzu, J. (2013). An Approach to disaggregating total household water consumption into major end-uses. *Water Resources Management, 27*(7), 2155–2177. https://doi.org/10.1007/s11269-013-0281-8

Gleick, P. H. (1996). Basic water requirements for human activities: Meeting basic needs. *Water International, 21*(2), 83–92. https://doi.org/10.1080/02508069608686494

T. Hastie R. Tibshirani J. Friedman The Elements of Statistical Learning: Data Mining 2009 In Springer Inference and Prediction 10.1007/b94608.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. *Springer*. https://doi.org/10.1007/978-1-4614-7138-7

Jorgensen, B. S., Graymore, M., & O'Toole, K. (2009). Household water use behavior: An integrated model. *Journal of Environmental Management, 91*(1), 227–236. https://doi.org/10.1016/j.jenvman.2009.08.009

Kiperstok, A., & Kiperstok, A. C. (2017). Technology Improvements or Influencing User Behaviour for Water Savings in Administrative and University Buildings: Which One Should Come First? In E. Ghisi (Ed.), Frontiers in Civil Engineering (Vol. 2, pp. 153–202). https://doi.org/10.2174/97816810848311170201.

Kowalski, M., & Marshallsay, D. (2005). Using measured microcomponent data to model the impact of water conservation strategies on the diurnal consumption profile. *Water Science and Technology: Water Supply, 5*(3–4), 145–150. https://doi.org/10.2166/ws.2005.0094

Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling [Hardcover]. *Springer*. https://doi.org/10.1007/978-1-4614-6849-3

Larson, E., Froehlich, J., Campbell, T., Haggerty, C., Atlas, L., Fogarty, J., & Patel, S. N. (2012). Disaggregated water sensing from a single, pressure-based sensor: An extended analysis of HydroSense using staged experiments. *Pervasive and Mobile Computing, 8*(1), 82–102. https://doi.org/10.1016/j.pmcj.2010.08.008

Li, Y., & Zhang, Y. (2020). *Hyper parameter estimation method with particle swarm optimization*. ArXiv: Retrieved from. https://arxiv.org/pdf/2011.11944.pdf.

Liu, A., Giurco, D., & Mukheibir, P. (2016). Urban water conservation through customised water and end-use information. *Journal of Cleaner Production, 112*, 3164–3175. https://doi.org/10.1016/j.jclepro.2015.10.002

Mello, M., Oliveira-Esquerre, K., & Botelho, G. (2018). Comparative study of similarity measures used to classify residential water flow pattern of low-income households in salvador - Brazil. *Computer Aided Chemical Engineering, 44*, 1405–1410. https://doi.org/10.1016/B978-0-444-64241-7.50229-9

Morrison, J., & Friedler, E. (2014). A critical review of methods used to obtain flow patterns and volumes of individual domestic water using appliances. *Urban Water Journal, 12*(4), 328–343. https://doi.org/10.1080/1573062X.2014.900090

Nguyen, K. A., Zhang, H., & Stewart, R. A. (2011). Application of dynamic time warping algorithm in prototype selection for the disaggregation of domestic water flow data into end use events. 34th IAHR Congress 2011 - Balance and Uncertainty: Water in a Changing World, Incorporating the 33rd Hydrology and Water Resources Symposium and the 10th Conference on Hydraulics in Water Engineering, (November 2017), 2137–2144.

Nguyen, K. A., Stewart, R. A., & Zhang, H. (2013). An intelligent pattern recognition model to automate the categorisation of residential water end-use events. *Environmental Modelling & Software, 47*, 108–127. https://doi.org/10.1016/j.envsoft.2013.05.002

Nguyen, K. A., Zhang, H., & Stewart, R. A. (2013). Development of an intelligent model to categorise residential water end use events. *Journal of Hydro-Environment Research, 7*(3), 182–201. https://doi.org/10.1016/j.jher.2013.02.004

Nguyen, K. A., Stewart, R. A., & Zhang, H. (2014). An autonomous and intelligent expert system for residential water end-use classification. *Expert Systems with Applications, 41*(2), 342–356. https://doi.org/10.1016/j.eswa.2013.07.049

Nguyen, K. A., Stewart, R. A., Zhang, H., & Sahin, O. (2018). An adaptive model for the autonomous monitoring and management of water end use. *Smart. Water, 3*(1). https://doi.org/10.1186/s40713-018-0012-7

Pastor-Jabaloyes, L., Arregui, F., & Cobacho, R. (2018). Water end use disaggregation based on soft computing techniques. *Water (Switzerland), 10*(1), 46. https://doi.org/10.3390/w10010046

Rahim, M. S., Nguyen, K. A., Stewart, R. A., Giurco, D., & Blumenstein, M. (2020). Machine learning and data analytic techniques in digitalwater metering: A review. *Water (Switzerland), 12*(1), 294. https://doi.org/10.3390/w12010294

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 26*(1), 43–49. https://doi.org/10.1109/TASSP.1978.1163055

Soares, E. S., Oliveira-Esquerre, K. P., de Aguiar, A. M., Botelho, G. L. P., & Kiperstok, A. (2018). Development of a model to identify combined use in residential water end use events. *Computer Aided Chemical Engineering, 44*, 1951–1956. https://doi.org/10.1016/B978-0-444-64241-7.50320-7

Vašak, M., Banjac, G., & Novak, H. (2015). Water use disaggregation based on classification of feature vectors extracted from smart meter data. *Procedia Engineering, 119*(1), 1381–1390. https://doi.org/10.1016/j.proeng.2015.08.992

Wonders, M., Ghassemlooy, Z., & Alamgir Hossain, M. (2016). Training with synthesised data for disaggregated event classification at the water meter. *Expert Systems with Applications, 43*, 15–22. https://doi.org/10.1016/j.eswa.2015.08.033

J. Wu X.Y. Chen H. Zhang L.D. Xiong H. Lei S.H. Deng Hyperparameter optimization for machine learning models based on Bayesian optimization Journal of Electronic Science and Technology 17 1 2019 26 40 https://doi.org/10.11989/JEST.1674-862X.80904120.