Research Article

Committee Machines for Hourly Water Demand Forecasting in Water Supply Systems

Julia K. Ambrosio,[1] **Bruno M. Brentan** ⓘ**,**[2] **Manuel Herrera** ⓘ**,**[3] **Edevar Luvizotto Jr.,**[2] **Lubienska Ribeiro,**[1] **and Joaquín Izquierdo** ⓘ[4]

[1]*School of Technology, Universidade Estadual de Campinas, Campinas, Brazil*
[2]*LHC-School of Civil Engineering, Universidade Estadual de Campinas, Campinas, Brazil*
[3]*Institute for Manufacturing, Department of Engineering, University of Cambridge, UK*
[4]*Fluing-Institute for Multidisciplinary Mathematics, Universitat Politècnica de València, Valencia, Spain*

Correspondence should be addressed to Joaquín Izquierdo; jizquier@upv.es

Prediction models have become essential for the improvement of decision-making processes in public management and, particularly, for water supply utilities. Accurate estimation often needs to solve multimeasurement, mixed-mode, and space-time problems, typical of many engineering applications. As a result, accurate estimation of real world variables is still one of the major problems in mathematical approximation. Several individual techniques have shown very good estimation abilities. However, none of them are free from drawbacks. This paper faces the challenge of creating accurate water demand predictive models at urban scale by using so-called committee machines, which are ensemble frameworks of single machine learning models. The proposal is able to combine models of varied nature. Specifically, this paper analyzes combinations of such techniques as multilayer perceptrons, support vector machines, extreme learning machines, random forests, adaptive neural fuzzy inference systems, and the group method for data handling. Analyses are checked on two water demand datasets from Franca (Brazil). As an ensemble tool, the combined response of a committee machine outperforms any single constituent model.

1. Introduction

More than half of the world's current population live in cities, with a growth of 1,500 million people in the last 20 years, and the United Nations predicts that this trend will continue. This population size scenario, together with the continuously under-stress natural resources, makes it paramount to count on accurate and efficient methods for estimating urban water demand [1]. This is nowadays possible as there are a huge quantity of available data and suitable big data tools to deal with it. However, there are still several open challenges regarding the study and analysis of big data. Not only may new mathematical techniques and machine learning paradigms improve the quality of estimation techniques but also efforts to obtain a better understanding of what the methods do in detail and which effects are caused by changing their parameters can give an insight into how several models

could work together. Among them, approaches related to optimal and automatic tuning processes for the models' hyperparameters [2, 3] should be highlighted. For short-term water demand forecasting processes, machine learning techniques such as artificial neural networks (ANNs) and support vector machines (SVMs) are widely applied [4].

Machine learning methods are able to map highly non-linear spaces and to accurately estimate the ensuing output space. The results tend to be strongly linked to the preestablished definitions of their architectures (hyperparameters for each method), which are usually defined by the users. In this sense, models may require much time to be built with no guarantees of being optimal. Another interesting paradigm investigates the so-called ensemble learning [5] of several individual methodologies (machine learning methods or, plainly, machines) to generate a single combined model or committee machine. The latter, an approach not enough

explored yet, should reach better predictive performance than that obtained from any of the constituent algorithms by themselves [6].

Water demand forecasting has been explored through machine learning techniques. For the short and long term [8] proposed a model of water demand forecasting for summer peak consumption, using and comparing multiple linear regression, time series analysis, and ANNs. Prediction of water demand using a dynamic ANN model was proposed by Ghiassi et al. [9]. The authors modeled water demand data using the DAN2 method, reaching good results and showing that predictions do not depend on the explicit inclusion of weather variables. The study was applied in monthly, weekly, and daily models, obtaining a prediction accuracy of 99%, and for the hourly models obtained a precision above 97%. The model was also compared with autoregressive integrated moving average models and traditional ANN models.

This paper proposes the use of committee machines for the creation of predictive models for urban water demand. A committee machine mixes different-nature methodologies. The aim is to take advantage of each component's strengths, while avoiding its weaknesses when combined with other machine learning methods not necessarily based on the same algorithm [10]. For instance, a committee machine can reduce the influence of an accurate but not robust model by boosting the influence of a more robust algorithm for certain model scenarios. A combination of individual methods was proposed by Huang et al. [11]. The authors combined models, including wavelet transform and mean least squares partial-autoregressive moving average (KPLS-ARMA) to analyze the nonstationary behaviour of an annual urban series of water demand. The combined method proposed by the authors obtained improved accurate forecast of the city's urban water demand. Arandia et al. [12], also combined methods to predict short-term water demands; in this case, the authors combined seasonally integrated, self-correcting seasonal moving average (SARIMA) models with data assimilation. In their case study, forecasts were compared with actual volumes of water produced by the local utility.

The present paper compares several combinations of models that have already shown good performance by themselves on forecasting urban water demand. This is the case of multilayer perceptrons (MLPs) [13], support vector machines (SVMs) [14], extreme learning machines (ELMs) [15], random forests (RFs) [16], adaptive neural fuzzy inference systems (ANFIS) [17], and the group method for data handling (GMDH) [18]. For each combination, the committee machine integrates independent tuned machines together with some predefined ensemble rules to build the final model. After training, the performance of the models is analyzed through error indices, such as the mean square root error (RMSE) and the mean absolute error (MAE).

To justify the use of the previously mentioned machine learning techniques, let us mention that techniques such as ANFIS have already been used for online identification in control systems and for predicting future values in chaotic time series [17]. RFs have also been used, for example by Cheng et al. [19], in a demand prediction study focused on the energy sector; the authors use an ensemble method based

on the RF technique. Still within the scope of demand forecast in energy systems, Majumder et al. [20] proposed a study of solar energy forecast, through a decomposition in a hybrid empirical mode (EMD) and ELM. For the forecasting study of the demand, it is also possible to use the method for data handling (GMDH) [21]. The GMDH provided better results and performance than the obtained by applying SVMs.

There are a number of antecedents of proposals similar to committee machines on the creation of predictive models for various water engineering applications. This is the case of Barzegar et al. [22], which compares the accuracy of three neural computing techniques, such as MLPs, radial basis function neural networks (RBFNNs), and generalized regression neural networks (GRNNs) in the prediction of the groundwater salinity of the simple confined aquifer of Tabriz. The committee machine created by combining MLPs, RBFNNs, and GRNNs is showed to perform better than any of the individual techniques alone for predicting groundwater salinity. Lima et al. [23] generate a final committee machine by grouping a set of SVMs. Nadiri et al. [24] use supervised intelligent committee machines (SICMs) by combining several SVM models and neurodiffuse (NF) and gene expression programming (GEP) methods. Their aim was to evaluate groundwater vulnerability indexes of an aquifer. On urban water demand, Candelieri [25] uses SVMs in two stages, one for clustering and the other for forecasting short-term water demand. Brentan et al. [26] use a hybrid methodology combining SVMs with Fourier time series to approach near real-time urban water demand models. Combinations of self-organizing maps and RFs are also investigated in Brentan et al. [27].

In addition to the most used ANNs, other forms can also be used to predict water demand, as done by Guo et al. [28]. The authors proposed a comparison between a gated recurrent unit network (GRUN), a conventional ANN model, and a SARIMA model. The models aimed at predicting water demand for 24 hour horizon with a time interval of 15 min. In the case studied, the GRUN model obtained better performance compared to ANN and SARIMA. This type of approach shows the importance of studying not only models working individually but also together, as in the case of the ensemble methods.

Ensemble methods have been as well applied out to engineering related topics. To mention just a few, Johansson et al. [29] propose the use of a parallel model combination for predicting consumption of heating systems. Oliveira et al. [30] present a combination of ten MLP networks with different architectures and parameters. Polikar [31] shows that cluster-based systems may be more beneficial than their individual classifier counterparts. Ensemble methods are also used to predict nonstationary time series. This is the case of the work by Castro et al. [32], in which ANNs are trained with various parameter configurations and then grouped to provide a single solution.

The paper is organized as follows. After the Introduction, next section presents the methodological aspects. Then, a new section describes the case study followed by a section thoroughly describing the obtained results. Finally, the paper is closed by the section of conclusions and the references.

The current paper presents an important step ahead for the implementation of cutting-edge machine learning developments to improve water distribution systems control and management. On the other way round, the work is of interest on expanding the ways and topics on which novel machine learning and data-driven models are applied. The paper creates ensemble models for committee machines based on six machine learning techniques. The paper shows a simple but theoretically sound way of how estimation problems can be solved with an efficient multiresolution technique. Each model is applied to investigate the demand on two water supply areas of a medium-size city in the State of São Paulo in Brazil. Apart from the superior results produced by the committee machines when compared with single models, the paper also opens an in-depth discussion about the influences, limitations, and applicability of each technique and model combination rule.

2. Materials and Methods

Machine learning techniques are able to learn patterns and solve complex problems just by processing (very often) large-size databases. Probably, the most classical machine learning approach is constituted by the artificial neural network (ANN) paradigm, especially MLPs. Over the years, ANNs have evolved towards other approaches for example, SVMs. A SVM maps the model input onto a high-dimensional feature space to ease further computations. ANN structures have also been adapted to new configurations. This is the case of ELMs in which some parameters do not need to be tuned. ANFIS, in its turn, is a variant of ANNs that uses fuzzy inference, which is powerful in approaching hybrid methodologies for parameter tuning. GMDH methods split a problem into manageable pieces where to apply regression techniques to thus produce simpler problems than the original. In a natural way, some versions of GMDH can be considered as variations of ANNs.

This section briefly introduces MLPs, SVMs, ELMs, ANFIS, and GMDH. RFs are also introduced for further combination within committee machines.

2.1. Multilayer Perceptron (MLP). MLPs have been extensively studied in the literature and applied in several areas, in particular in studies on water supply systems and demand forecasting [33, 34]. These networks are based on interconnections of their calculation units, called perceptrons, organized in various layers. MLPs count on an input layer, and the input data is available to all its nodes. They also have an output layer providing the outcome of the process. Then, one or more hidden (or inner) layers facilitate the internal computations required for reaching optimal estimation for the problem to solve. The outputs of the neurons of one layer are distributed to only the inputs of the neurons of the next layer. Thus, the input signal propagates through the network in a progressive (feedforward) way.

The information processing in a MLP flows in two modes. One is related to its (forward) propagation through the network, layer by layer, until the corresponding output is produced. This stage is directly related to the network performance. The other mode regards the network adaptation. The MLP connection weights are modified using the backpropagation algorithm, which is based on the observed error at comparing the estimated output and the real value available. The error is propagated backwards from the output layer to the input layer, and the connection weights at the hidden layer(s) can be consequently adjusted so as to minimize the difference between the estimated and the real value [35].

Considering a simple MLP with one inner layer with M neurons, and N neurons in the input layer, the output, y_i, can be written in terms of the input vector, x_i, as

$$f(x_i) = f\left(\sum_{q=1}^{M} w_q f\left(\sum_{p=1}^{N} w_p w_i\right)\right) = y_i \quad (1)$$

In (1) w_p are the input weights, and w_q the weights of the hidden layer. Each linear combination is processed by an activation function, f, responsible for further nonlinear transformations. The sigmoid function is the most popular activation function for MLPs.

2.2. Support Vector Machine (SVM). Support vector machines (SVMs) were developed focusing on nonlinear separable data problems [36, 37]. To achieve this separation, a SVM finds the ideal hyperplane that maximizes the distance between two groups, thereby minimizing the margin error. To do this, the input data is projected onto a higher dimensional space where the SVM is able to linearly separate the nonlinearly separable data in the original space. The transformation of dimension and data separation is schematized in Figure 1.

Figure 1(a) represents a set of nonlinearly separable data. To achieve the separation of these data, the input space is projected onto a higher dimension space (Figure 1(b)). This transformation may be hard, mainly to find the correct dimension where the data are separable. For this purpose, it is used the so-called kernel trick, which applies a nonlinear transformation on the input space [38]. According to Rizk et al. [39], the kernel trick is used to reduce the computational complexity of the prediction demonstrated by the SVM.

Support vector regression (SVR) uses the same principles as SVM for classification. However, there are some differences, as the main objective is not to find the best hyperplane to separate the data, but the best data regression hyperplane.

2.3. Extreme Learning Machine (ELM). An ELM is based on a feedforward ANN with a single hidden layer [40]. For ELMs the input and hidden hyperparameters are randomly determined, and the proper learning process takes place on the output layer. ELM networks have some advantages over other networks such as MLPs. Among them, faster learning speed and less human interference on the network's architecture design must be highlighted.

Considering an N-dimensional input dataset processed by an ELM with M neurons in the hidden layer, the output of the network can be written as in (1). However, taking into account the random approach, this equation can be interpreted as a linear combination of (processed) input data
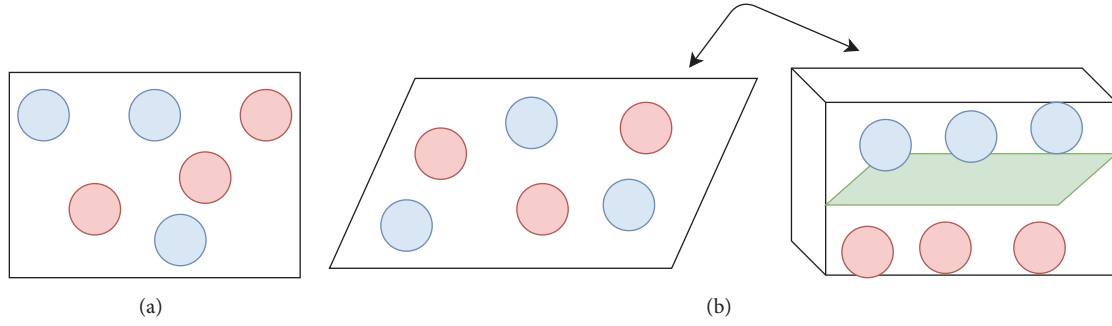
FIGURE 1: Representation of (a) a nonlinearly separable data set in a low dimension space and (b) the dimension transformation turning data separable.

(2), thus becoming close to a least-square problem where the parameters to be adjusted are the weights of the output layer.

$$f(x_i) = \sum_{q=1}^{M} w_q L\left(x_{i,p}\right) = y_i \qquad (2)$$

Here, $L(x_{i,p})$ is the nonlinear function, processed at the input and hidden layers. According to Huang et al. [34], the main advantage of ELMs is on creating a bridge between the ANN universal approximation theory and the SVM tuning process.

2.4. Random Forest (RF).

The random forest is a classification technique directly related to tree-like models [41]. To classify the data, the method combines the results of a number of decision trees through a voting mechanism. The voting for the classes is given by each tree, and the final classification corresponds to the class that receives the biggest number of votes among all the trees [16].

During an RF process, a new set is created from the initial training set, with which a tree, based on this new subset, which will be built with a random selection of attributes. At each node of the tree, a subset of m attributes is randomly selected and subsequently evaluated. The attribute that has the best performance is chosen to split the node. The value of m is set for all the nodes. An RF defines a margin function that measures the extent to which the average number of votes for the correct class exceeds the average of votes for any other class present in the dependent variable. The result of this measure aids to solve forecasting problems and also constitutes a way of associating a measure of confidence with these forecasts. RFs consider the average of the predictions of the trees to perform inference.

RFs are used for regression purposes and are made up by the growth of simple trees, each of which is capable of producing a numerical response value. In this case, the predictor set is randomly selected from the same distribution and for all the trees.

2.5. Adaptive Neural Fuzzy Inference System (ANFIS).

ANFIS is a hybrid technique of artificial intelligence that uses fuzzy logic and ANN learning processes. The ANFIS associated network has architecture with six layers interconnected by unit weights (Figure 2). Each layer is responsible for a specific operation resulting into a single output [17].

The first layer is responsible for reading the input data. The second layer processes these input values by a membership function aiming to identify the compatibility degree of each input with its respective fuzzy input sets. The interaction operation between the input membership functions occurs on the third layer. The fourth layer normalizes the previous layer outputs for their consequent transformation through activation functions, according to

$$f_i(w) = \frac{w_i}{\sum_{j=1}^{m} w_j} \qquad (3)$$

in which w_i is the output of neuron i from the previous layer. The fifth layer is responsible for the calculation of the values of the consequences of the rules. In this layer, the function responsible for the activation of the neurons is the Sugeno function, a p-th order combination of the input signals. The sixth layer sums through the obtained results, resulting into a single output for the network.

2.6. Group Method for Data Handling (GMDH).

The GMDH technique uses networks that perform nonlinear processing through a polynomial combination; in this combination, the adjustment of the polynomial coefficients occurs through training in batches [42]. Thus, the technique is applied when the separation method is applied to multiparametric models [18].

This polynomial support function is given by the Kolmogorov-Gabor polynomial. The function is represented by

$$y = a_0 + \sum_{i=1}^{m} a_i x_i + \sum_{i=1}^{m}\sum_{j=1}^{m} a_{ij} x_i x_j + \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} a_{ijk} x_i x_j x_k \qquad (4)$$

in which $x_1, x_2, \ldots, x_m$ are the input vectors of the variables and $a_1, a_2, \ldots, a_m$ are the method coefficients. From (4), it can be seen that the complexity of the relations between variables increases as a function of the number of terms.

The GMDH network is considered a constructive network; that is, it is a network which exhibits good organization, and this characteristic becomes an advantage over the most common neural networks.
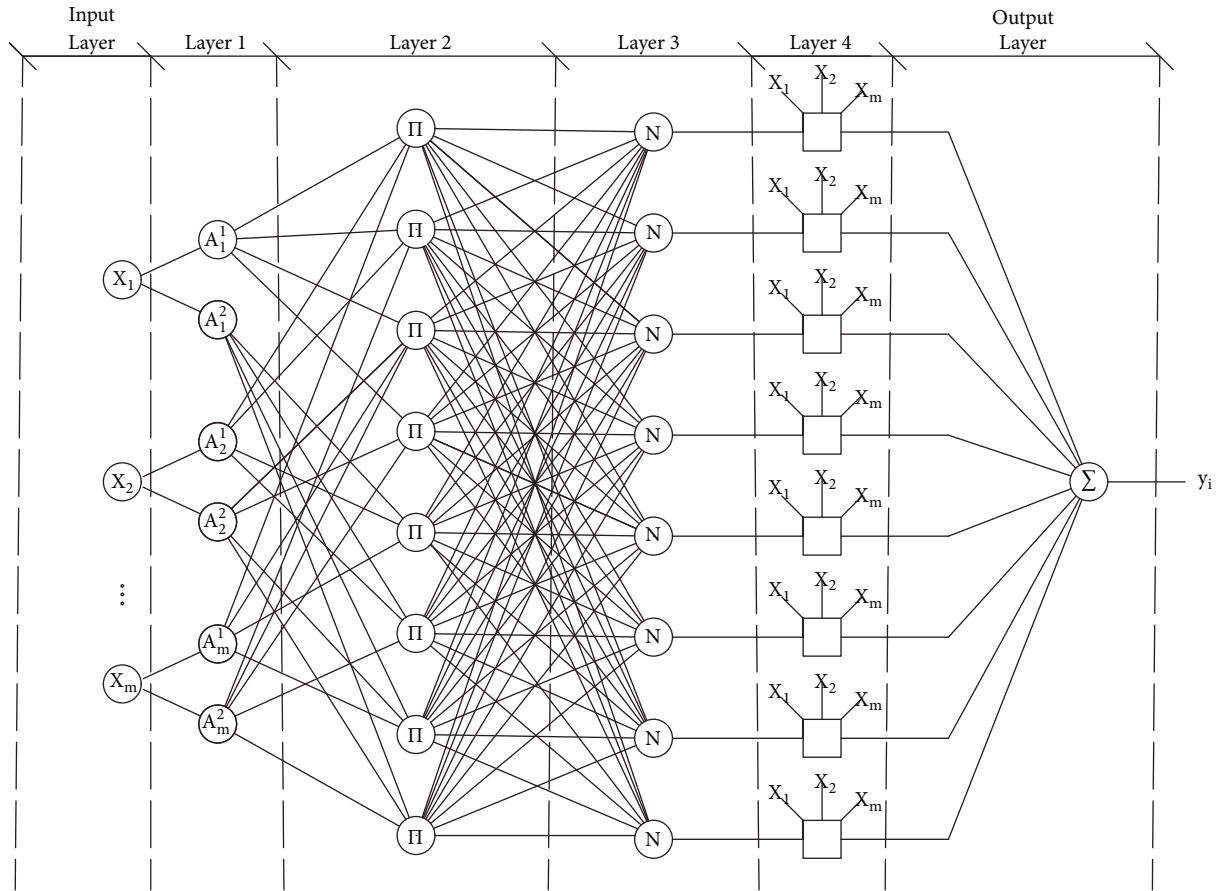
FIGURE 2: Architecture of an ANFIS network: Brentan [7].

2.7. Committee Machine-Ensemble Method. Committee machines attempt to minimize the errors of individual learning algorithms or machines by grouping them and making them to work synergistically. The ensemble is a more robust model than the model represented by any individual machine.

This paper proposes three merging rules to produce ensembles: arithmetic mean, geometric mean, and linear combination techniques (based on minimum least squares).

In addition, two other approaches for committee machines are introduced. Firstly, committee machine made of MLPs and ELMs with different architectures is proposed. This is not a mix of machines. Indeed, the MLPs and the ELMs are combined using the previous rules. This first strategy is used for two reasons: the first reason is to scrutinize the structure of the model (input variables) in the estimation process, which must be analyzed to know which variables influence most the system; and second, we intend to make comparisons with the second strategy, which eventually will provide the best results, namely, the one that combines a larger number of machines, specifically, MLP, ELM, SVM, GMDH, ANFIS, and RF. This combination process is represented by Figure 3.

Selection of Model Structures and Committees of Single Machines. Correlation analyses presented in the literature point out the strong link between weather and hourly water

TABLE 1: Model structure in terms of input variables.

Model	Input Variables
Model A	Time, weekday, holiday, temperature, humidity
Model B	Time, weekday, holiday, temperature
Model C	Time, weekday, holiday, humidity
Model D	Time, weekday, holidays

demand. This clearly advises the use of these variables as an essential part of the input for predictive models of water demand. In addition, machine learning approaches can correlate some time series demand anomalies with specific weather conditions, thus helping improve the reproduction of some scenarios in the future. Nevertheless, using weather inputs might be at risk of providing not reliable results if the nonpredictable nature of the weather is taken into account. In this sense, committee machines make it possible to obtain reliable models, while simultaneously using all the available information.

A variable selection process for ELM and MLP networks is first used to build four different models, in order to verify, which model will perform better. Details are presented in Table 1.
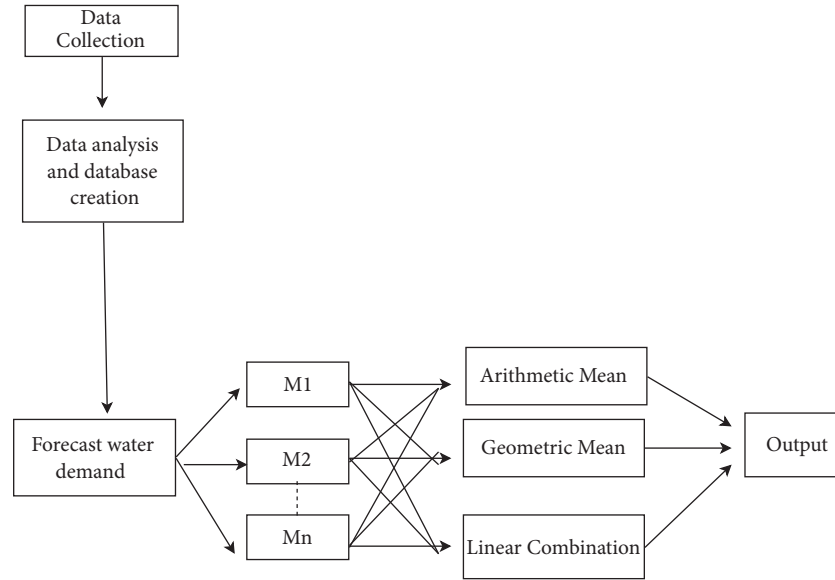
FIGURE 3: Representation of the ensemble methods for a committee machine.

For each proposed model, 50 architectures with two hidden layers were trained for the MLP and ELM networks. The final results were combined by using three rules, namely, arithmetic and geometric means and linear regression.

Database is divided into three parts: training dataset, ensemble dataset, and validation dataset. The training dataset is used for the learning process of each machine, while the ensemble dataset is used to obtain the linear regression coefficients. Finally, the validation data set is used to evaluate the performance of each model structure and the different combinations of single committee machines.

The performance of each method is evaluated by using the RMSE and the MAE, represented by (5) and (6), respectively,

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})} \tag{5}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y_i}}{y_i}\right| \tag{6}$$

where y_i is the measured demand and $\widehat{y_i}$ is the estimated demand at time step i.

3. Case Study

The experimental evaluation developed in this paper, to obtain accurate water demand predictive models for two district metered areas (DMAs) of the water distribution system of Franca (State of São Paulo, Brazil), shows that our method yields excellent results.

Let us first pinpoint that this municipality is of major economic importance for the state. Franca population is approximately 318,640 strong (IBGE, 2010) and is distributed in an area of 3,439 km (IBGE, 2010). The DMAs of the municipality of Franca that are studied in this work are
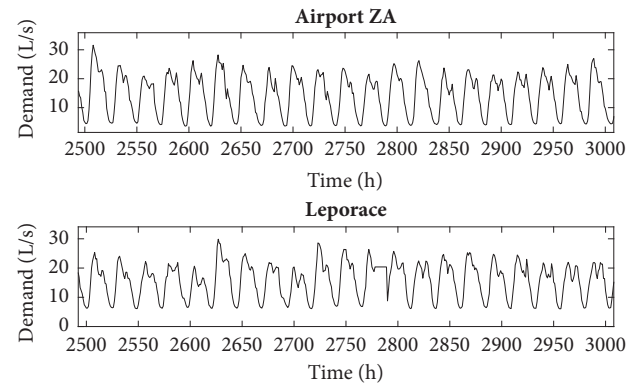

FIGURE 4: Typical demand patterns for each studied DMA.

TABLE 2: Statistical parameters of the time series.

	Leporace DMA	AirportZA DMA
Mean (l/s)	18.18	15.83
Standard deviation (l/s)	6.48	6.71

known as AirportZA and Leporace. Both are residential areas, each of them counting on 2,168 and 2,728 household connections, respectively. For the Leporace DMA a time series with 18166 data points is available, and the time series for AirportZA DMA includes 12576 data points. The data used for the study were provided by SABESP (Basic Sanitation Company of the State of São Paulo), the agency responsible for water supply in the municipality. Figure 4 presents a typical demand curve for each DMA.

Table 2 presents the mean and standard deviation for each DMA. As observed, the standard deviations correspond to 35% or more of the mean values.

TABLE 3: RMSE for validation of MLP applied to the AirportZA DMA (L/s).

Method	Model A	Model B	Model C	Model D
Arithmetic mean	4.57	3.22	3.23	3.66
Geometric mean	4.92	3.23	3.20	3.70
Linear regression	5.00	3.40	6.37	4.08

TABLE 4: MAE for validation of MLP applied to the AirportZA DMA (%).

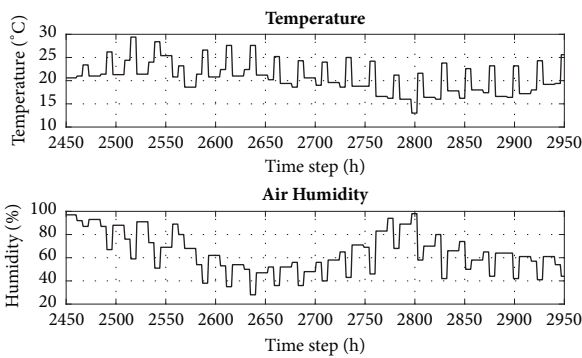| Method	Model A	Model B	Model C	Model D
Arithmetic mean	1.57	1.20	1.22	1.35
Geometric mean	1.64	1.21	1.22	1.36
Linear regression	1.58	1.21	1.42	1.33

FIGURE 5: Representation of weather variables (for both DMAs).

Social/calendar and weather inputs enhance the historical time series of hourly water demand. As shown by Tian et al. [43], who propose a study using an analogue approach with a developed reforecast of a numerical weather forecast (NWP), this may improve forecasting of urban water demand in the short term. The authors claim that short-term urban water demands are influenced by climatic conditions.

For this study hourly data, from April 2013 to December 2015, was used. The advantages of using these covariables have been widely explored in the literature, as done by Bakker et al. [44] and Praskievicz et al. [45], in which the authors analyze the influence of climatic and social variables in the prediction studies of water demand. Brentan et al. [26] recently studied the possible correlations between water demand and climatic and social variables. The social/calendar inputs taken for the models herein are day of the month, day of the week, time of the day, year, and holidays. The weather inputs are temperature and relative humidity. Weather variables were collected from the National Institute of Meteorology (INMET) database and arranged hourly (see Figure 5).

4. Results

In this section we perform a thorough presentation of results, with clear discussion on the model structure and the machine selection.

The database is split into three parts: training dataset with 75% of the data, ensemble dataset with 12.5% of the data, and validation dataset with 12.5% of the data [35]. First, each machine is trained by using the training dataset; then, regression coefficients are obtained by using the ensemble dataset; and, finally, the performance of each model structure and the different combinations of single committee machines is evaluated by using the validation data.

To show the performance of each proposed model and for the sake of brevity, we just present the AirportZA DMA study, since the Leporace DMA results provide much identical outcomes and interpretations.

The results for the MLP committee machine applied to the AirportZA DMA are presented in Tables 3 and 4.

Several MLP architectures varying the number of hidden layers from 5 up to 100 nodes have been used. As can be easily seen, the arithmetic mean, coupled with the use of model B, presents the smallest error.

As seen in Table 1, model B uses the variables time of day, day of week, holidays, and temperature. Model B clearly presents the best results for all the combination rules (look at the rows of Tables 3 and 4). However, in terms of committee machine rules (columns of each table), it is not possible to verify large differences in the results.

Now, the results for ELMs and the AirportZA DMA are in Tables 5 and 6.

The ELMs are built in a grid search process, by varying the number of hidden nodes from 50 to 2500. By analyzing the errors obtained in each combination performed from the ELMs, it can be seen that, again, the combination using arithmetic mean and model B presents the smallest error.

In both cases, it is important to highlight the difference between the RMSE magnitudes for the MLP and ELM models, pointing towards better accuracy of the ELM method.

Committee of Multiple Machines. Among other combinations, a more complex committee machine has also been created from the blend of MLP, SVM, ELM, RF, ANFIS, and GMDH. For the MLP and ELM cases, whose architectures have shown higher influence on the final results, due to their complexities, the configurations that provide the best performance have been selected. Both DMA data are used for the application of this committee machine, using the three given rules. Three rule-different committee machines, noted here as MA, MG, and MLR, have thus been considered. The results are again evaluated in terms of RMSE and MAE.

TABLE 5: RMSE for validation of ELM applied to the AirportZA DMA (L/s).

Method	Model A	Model B	Model C	Model D
Arithmetic mean	2,67	1,75	2,07	2,73
Geometric mean	2,73	1,75	1,89	2,78
Linear regression	2,53	1,87	1,93	2,52

TABLE 6: MAE for validation of ELM applied to the AirportZA DMA (%).

Method	Model A	Model B	Model C	Model D
Arithmetic mean	1.97	1.16	1.24	1.96
Geometric mean	2.01	1.17	1.22	1.99
Linear regression	1.85	1.24	1.21	1.83

TABLE 7: Linear regression coefficients for each DMA.

DMA/Method	MLP	SVM	ELM	RF	ANFIS	GMDH
AirportZA	-0.1715	0.9525	-0.0346	-0.0142	0.0961	0.2065
Leporace	0.3587	0.3148	0.0356	-0.0312	0.0851	0.2013

TABLE 8: Error parameter for the committee machines applied to AirportZA DMA-RMSE in L/s and MAE in %.

	MLP	SVM	ELM	RF	ANFIS	GMDH	MA	MG	MLR
RMSE	3.74	3.67	1.73	4.72	5.84	5.65	3.16	3.22	1.61
MAE	1.33	1.38	2.70	2.16	1.81	2.41	1.22	1.23	1.06

Each machine learning methodology is tuned and used to assess the demand individually. After this step, a linear regression process is applied to the data set in order to combine the values predicted by the individual machines, thus obtaining the real average value. The linear regression coefficients for each method are shown in Table 7.

For the AirportZA DMA, Table 8 shows the error parameters RMSE and MAE for each individual machine and for the three rule-based committee machines MA, MG, and MLR. These three committee machines are able to find a good trade-off between the most and the less accurate models. Above all, the committee machine using linear regression, which exhibits the lowest RMSE and MAE, stands out. Comparing the same data estimation, using ELMs or MLPs previously presented in Tables 3–6, it becomes clear that combining various machines improves the results in terms of error analysis.

Figure 6 shows the water demand for the validation dataset together with the estimated demand for each rule-based complex committee machine. It is worth highlighting the robustness of these committee machine models when an anomaly happens in the system. Typically, dynamic models, which are models using external input and sliding windows, such as recurrent networks (NARX), are able to predict water demand with high accuracy in normal conditions. However, they may be affected by anomalies, since the most common dynamic models work online using incoming new measurements. In contrast, the linear regression committee machine is able to estimate water demand with good accuracy, while not using dynamic components.
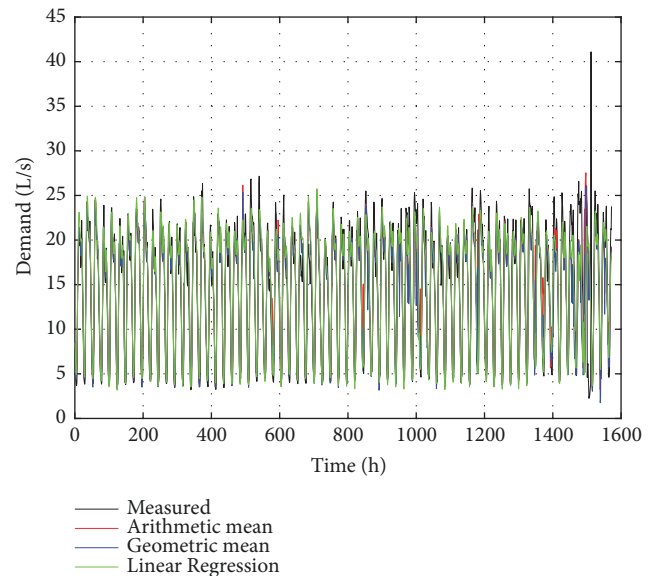

FIGURE 6: Demand estimation for the three complex committee machines for the validation data set of the AirportZA DMA.

To enhance the previous ideas, we also provide now the results for the Leporace DMA.

Table 9 shows the error parameter for each individual machine and for the three committee machines in this DMA. The relation of trade-off between good and not so good

TABLE 9: Error parameter for the committee machine applied to Leporace DMA-RMSE in L/s and MAE in %.

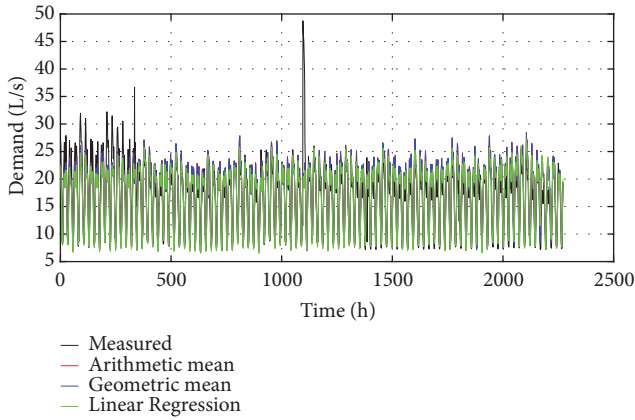|	MLP	SVM	ELM	RF	ANFIS	GMDH	MA	MG	MLR
RMSE	1.47	2.26	1.48	6.23	3.97	5.47	2.70	2.69	1.42
MAE	0.68	0.87	2.57	1.24	1.23	1.85	1.06	1.06	0.67

FIGURE 7: Demand estimation for the three complex committee machines for the validation data set of the Leporace DMA.

models is found again. The linear regression is also the most accurate committee machine for this DMA.

Figure 7 shows, now for the Leporace DMA, the validation dataset and the respective estimations using the three proposed rule-based committee machines. It is worth mentioning that the best performance of the linear regression is mainly at the consumption peaks, where averaged combinations usually do not have good enough performance.

5. Conclusions

Increased monitoring in water distribution systems allows the application of data mining and machine learning techniques to better understand the system and estimate future hydraulic states. Optimization algorithms aid to find good architectures for various forecasting techniques. This is the case of the number of hidden layers for ANNs, or the best set of hyperparameters for kernel-based functions (SVR). However, the robustness of the current machine learning methods could be exposed to extrapolation problems. For instance, models for water demand forecasting using weather variables, which can hardly predict the demand. In this case, machine learning algorithms lose their estimation abilities and the results are poor if not invalid. This paper presents the use of committee machines as a way to reduce failures in the estimation process for input data at the extrapolation boundaries and to improve the robustness of the models by combining several algorithms. Committee machines maximize the good performance of their individual machines, minimizing negative contributions, and, ultimately, reaching more accurate and stable estimations.

In this work, various committee machines are applied to predict the demand of a real water medium-size Brazilian

city. In a first stage, it has been determined the best model structure in terms of the input variables. MLP and ELM algorithms are used in this stage, in which various architectures for both algorithms were trained. Both algorithms allow a similar conclusion about the input model structure. For this dataset, the model using time of day, day of week, holidays, and temperature results in the lowest values of RMSE and MAE. The second step of the proposal applied six (individually trained) machine learning methods and combined their estimations to obtain the final demand using three rules. In this case, the linear regression combination results in better results for the two analyzed DMAs.

Machine learning methods can bring forward powerful tools for water distribution companies, aiding them in the operation and management of their systems. The evaluation of the best model structure in terms of the error can help reduce the number of inputs; it also reduces the algorithm execution computational time and last but not least the instabilities at the extrapolation bounds. The combination of various machine learning techniques can be useful to reduce the optimal architecture search process, which can be useful for further automation of various processes.

The aim is to keep on working on enhancing this research avenue on forecasting model development by ensemble methods. Other combination rules, such as the cascade combination process, where the output of a model is the input for another algorithm, should also be studied in future works.

Data Availability

Climate and demand data used to support the findings of this study are available from the first author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] D. A. Moreira, *Administração da Produção e Operações*, Cengage Learning, São Paulo, Brazil, 2nd edition, 2011.

[2] O. Golovidov, P. Koch, B. Wujek, and S. Gardner, "Automated hyperparameter tuning for effective machine learning," in *Proceedings of the SAS Global Forum 2017 Conference*, 2017.

[3] I. Montalvo, J. Izquierdo, R. Pérez-García, and M. Herrera, "Improved performance of PSO with self-adaptive parameters for computing the optimal design of Water Supply Systems," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 5, pp. 727–735, 2010.

[4] E. A. Donkor, T. A. Mazzuchi, R. Soyer, and A. Roberson, "Urban water demand forecasting: review of methods and

models," *Journal of Water Resources Planning and Management*, vol. 140, no. 2, pp. 146–159, 2014.

[5] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2012.

[6] C. Soares, P. Brazdil, C. G. Carrier, and R. Vilalta, *Metalearning: Applications to data mining*, Springer Science & Business Media, 2008.

[7] B. M. Brentan, *Sistemas de abastecimento de água inteligentes: Previsão de demanda e controle ótimo em tempo real*, Tese de Doutorado, FEC-UNICAMP, Campinas, 2017.

[8] J. F. Adamowski, "Peak daily water demand forecast modeling using artificial neural networks," *Journal of Water Resources Planning and Management*, vol. 134, no. 2, pp. 119–128, 2008.

[9] G. Ghiassi, D. K. Zimbra, and H. Saidane, "Urban water demand forecasting with a dynamic artificial neural network model," *Journal of Water Resources Planning and Management*, vol. 134, no. 2, pp. 138–146, 2008.

[10] R. T. Clemen, "Combining forecasts: a review and annotated bibliography," *International Journal of Forecasting*, vol. 5, no. 4, pp. 559–583, 1989.

[11] Y. Peng, L. Huang, C. Zhang, and H. Zhou, "Application of a combination model based on wavelet transform and KPLS-ARMA for urban annual water demand forecasting," *Journal of Water Resources Planning and Management*, vol. 140, no. 8, 2014.

[12] B. Eck, E. Arandia, A. Ba, and S. McKenna, "Tailoring seasonal time series models to forecast short-term water demand," *Journal of Water Resources Planning and Management*, vol. 142, no. 3, 2016.

[13] C. Bishop and C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, New York, NY, USA, 1995.

[14] M. Herrera, J. C. García-Díaz, J. Izquierdo, and R. Pérez-García, "Municipal water demand forecasting: tools for intervention time series," *Stochastic Analysis and Applications*, vol. 29, no. 6, pp. 998–1007, 2011.

[15] G. Huang, S. Song, and K. You, "Trends in extreme learning machines: a review," *Neural Networks*, vol. 61, pp. 32–48, 2015.

[16] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[17] J. S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 665–685, 1993.

[18] A. G. Ivakhnenko and G. A. Ivakhnenko, "The review of problems solvable by algorithms of the group method of data handling (GMDH)," *Pattern Recognition and Image Analysis C/C of Raspoznavaniye Obrazov I Analiz Izobrazhenii*, vol. 5, pp. 527–535, 1995.

[19] Y.-Y. Cheng, P. P. K. Chan, and Z.-W. Qiu, "Random forest based ensemble system for short term load forecasting," in *Proceedings of the International Conference on Machine Learning and Cybernetics, ICMLC '12*, pp. 52–56, China, July 2012.

[20] M. K. Behera, I. Majumder, and N. Nayak, "Solar power forecasting using a hybrid EMD-ELM method," in *Proceedings of the Proceedings of the International Conference on Circuit, Power and Computing Technologies, (ICCPCT '14)*, 2014.

[21] M. Najafzadeh, G.-A. Barani, and H. M. Azamathulla, "Prediction of pipeline scour depth in clear-water and live-bed conditions using group method of data handling," *Neural Computing and Applications*, vol. 24, no. 3-4, pp. 629–635, 2014.

[22] R. Barzegar and A. Asghari Moghaddam, "Combining the advantages of neural networks using the concept of committee machine in the groundwater salinity prediction," *Modeling Earth Systems and Environment*, vol. 2, no. 1, 2016.

[23] A. L. V. Coelho, C. A. M. Lima, and F. J. Von Zuben, "Ensembles of support vector machines for regression problems," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '02)*, vol. 3, pp. 2381–2386, Honolulu, Hawii, USA, 2002.

[24] R. Khatibi, S. Sadeghfam, A. A. Nadiri, M. Gharekhani, and A. A. Moghaddam, "Groundwater vulnerability indices conditioned by Supervised Intelligence Committee Machine (SICM)," *Science of the Total Environment*, vol. 574, pp. 691–706, 2017.

[25] A. Candelieri, "Clustering and support vector regression for water demand forecasting and anomaly detection," *Water (Switzerland)*, vol. 9, no. 224, 2017.

[26] M. Herrera, E. Luvizotto Jr., B. M. Brentan, G. Meirelles, and J. Izquierdo, "Correlation Analysis of Water Demand and Predictive Variables for Short-Term Forecasting Models," *Mathematical Problems in Engineering*, vol. 2017, Article ID 6343625, 10 pages, 2017.

[27] M. Herrera, J. Izquierdo, B. M. Brentan, E. Luvizotto Jr., and R. Pérez-García, "Hybrid regression model for near real-time urban water demand forecasting," *Journal of Computational and Applied Mathematics*, vol. 309, pp. 532–541, 2017.

[28] J. Li, R. Zhou, G. Guo, Y. Wu, and X. Zhu, "Short-term water demand forecast based on deep learning method," *Journal of Water Resources Planning and Management*, vol. 144, 2018.

[29] D. Geysen, O. Somer, N. Lavesson, C. Johansson, M. Bergkvist, and D. Vanhoudt, "Operational demand forecasting in district heating systems using ensembles of online machine learning algorithms," *Energy Procedia*, vol. 116, pp. 208–216, 2017.

[30] A. J. Silva Neto, R. C. Oliveira, N. I. A. Acevedo, and L. Biondi Neto, "Application of a committee of artificial neural networks for the solution of inverse problems in radiative transfer," *A Publication of the Brazilian Society of Applied and Computational Mathematics*, vol. 11, no. 2, pp. 171–182, 2010.

[31] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits and Systems Magazine*, vol. 6, no. 3, pp. 21–45, 2006.

[32] A. P. de Braga, C. L. de Castro, and A. V. Andrade, "Aplicação de um modelo ensemble de redes neurais artificiais para previsão de séries temporais não estacionárias," in *In XXV Congresso da Sociedade Brasileira de Computação*, Brasil, São Leopoldo, 2005.

[33] K. Kamiński, W. Kamiński, and T. Mizerski, "Application of artificial neural networks to the technical condition assessment of water supply systems," *Ecological Chemistry and Engineering S*, vol. 24, no. 1, pp. 31–40, 2017.

[34] R. P. Ferreira, A. Martiniano, A. Ferreira, A. Ferreira, and R. J. Sassi, "Study on Daily Demand Forecasting Orders using Artificial Neural Network," *IEEE Latin America Transactions*, vol. 14, no. 3, pp. 1519–1525, 2016.

[35] S. Haykin, *Redes Neurais: Princípios e Prática*, Bookman, Porto Alegre, Brazil, 2nd edition, 2001.

[36] S. Haykin, *Neural Networks and Learning Machines*, Pearson Prentice Hall, São Paulo, Brazil, 3rd edition, 2008.

[37] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[38] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and beyond*, The MIT Press, Cambridge, Mass, USA, 2001.

[39] N. Mitri, Y. Rizk, and M. Awad, "An ordinal kernel trick for a computationally efficient support vector machine," in *Proceedings of the International Joint Conference on Neural Networks, IJCNN '14*, pp. 3930–3937, China, July 2014.

[40] D. H. Wang, G. Huang, and Y. Lan, "Extreme learning machines: a survey," *International Journal of Machine Learning and Cybernetics*, vol. 2, no. 2, pp. 107–122, 2011.

[41] P. Tikmani, V. Ghosal, and P. Gupta, "Face classification using gabor wavelets and random forest," in *Proceedings of the Canadian Conference on Computer and Robot Vision, IEEE Computer Society, (CRV '09)*, pp. 68–73, Canada, May 2009.

[42] A. G. Ivakhnenko, "Heuristic self-organization in problems of engineering cybernetics," *Automatica*, vol. 6, no. 2, pp. 207–219, 1970.

[43] D. Tian, C. J. Martinez, and T. Asefa, "Improving short-term urban water demand forecastswith reforecast analog ensembles," *Journal of Water Resources Planning and Management*, vol. 142, no. 6, 2016.

[44] K. Van Schagen, J. Vreeburg, M. Bakker, H. Van Duist, and L. Rietveld, "Improving the performance of water demand forecasting models by using weather input," *Procedia Engineering*, pp. 93–102, 2014.

[45] S. Praskievicz and H. Chang, "Identifying the relationships between urban water consumption and weather variables in Seoul, Korea," *Physical Geography*, vol. 30, no. 4, pp. 324–337, 2009.