Short-Term Water Demand Forecasting: A review

Safa Ghannam[1] safaa.ghannam@uts.edu.au (S. Ghannam, ORCID: 0000-0001-9626-0883), Farookh Hussain[2] farookh.hussain@uts.edu.au (F. Hussain)
[1,2] University of Technology Sydney, Faculty of Engineering and Information Technology, Ultimo, NSW 2007, Australia

ARTICLE INFO

Keywords:
Water demand
residential
review
forecasting
water consumption
Short-term
urban

ABSTRACT

Water utilities rely on precise water demand predictions to ensure efficient water management and the dependable functioning of the water supply infrastructure. Accurate forecasting of water demand is crucial for effective water planning and maintaining the reliable operation of the water supply system. However, various water demand forecasting models have been proposed making it challenging to select the most appropriate forecasting model. This research presents a literature review on the relevant studies that address short-term water demand forecasting and the most frequently used methods, tools and techniques in the field of short-term water demand forecasting. Suggestions for future work in water demand forecasting are proposed.

1. Introduction

Precise prediction of water demand is crucial for water utilities to optimize water planning and schedule pumping operations effectively Zounemat-Kermani et al. (2020). A real-time system is necessary to ensure that water utilities have the confidence to meet the demand for water consumption Romano and Kapelan (2014). In this paper, we present a systematic literature review (SLR) to find and assess the relevant studies that address short-term water demand forecasting, propose forecasting models and discuss the related methods, tools and techniques. Water demand forecasting models differ based on the forecasting horizon, with short-term prediction ranging from hourly to monthly to one year. Short-term water demand prediction is particularly useful for management and scheduling pump operations and system operations Zhou et al. (2002). The remainder of this paper is organized as follows: Section 2 presents the review methodology. Section 3 explains how the selected articles contribute to short-term water demand prediction, highlighting the methods, techniques, and tools utilized in each article. Section 4 critically analyzes the selected studies, identifies the gaps and the findings, and finally, Section 5 draws the conclusion and suggests directions for future work.

2. Literature Review Methodology

We conducted an SLR to enhance re- search reliability and accuracy. The primary focus of this re- view is to find and assess the relevant studies that address short-term water demand forecasting, propose forecasting models and discuss the related methods, tools and techniques. The following steps were taken to conduct the SLR:

- Search protocol: the databases that were searched and the search terms are clearly defined.
- Define inclusion criteria and exclusion criteria: the inclusion criteria and exclusion criteria are used to define the boundaries of the search and to focus on retrieving the most relevant research.
- Study selection process: research papers that meet the inclusion criteria are accepted for this literature review.
- Data extraction: data is extracted after reviewing the accepted articles

2.1. Search Protocol

The aim of this SLR is to present a comprehensive review of the relevant studies in the literature that address the short-term water demand forecasting problem. The scientific databases that were searched for this SLR are Scopus, IEEE Xplore, SpringerLink and ACM Digital Library. These databases were chosen because of their well-known reputation in the engineering field. Furthermore, the four selected databases cover a wide range of journals and conferences. The Boolean operators "OR", "AND", asterisk symbol "*" and parentheses were used to ensure more relevant and productive results. The study statement includes two parts:

- The joining or the union of the same terms connected by the "OR" operator (e.g., water consumpt* OR water demand*) to broaden the search.
- The conjunction of the terms to narrow and filter the search to find studies relevant to this review. This is shown in Table 1.
- The asterisk (*) was used to broaden the search and to include all words that begin with the same letter and to find all words in different forms.

Search Category	Keywords
Short-term forecasting of water consumption.	(“short-term forecast*” OR “short-term predict*”) AND (“water consumpt*” OR “water demand*”)

Table 1: Search category and keywords

2.2. Inclusion Criteria and Exclusion Criteria

The following inclusion criteria and exclusion criteria were applied to identify the research papers that should be included or excluded:

Inclusion Criteria

- Is the date of publication of the research paper between 2010 to 2023?
- Is the full text of the research paper available?
- Is the article written in the English language?
- Is the article peer reviewed?
- Does the article address the short-term water demand forecasting problem?

Exclusion Criteria

- Articles not written in the English language.
- The full text of the article is not available.
- Duplicate articles.

2.3. Article Selection Process:

After we identified the keywords and the search terms for our research topic as detailed in section 2.1, articles were assessed for eligibility using the defined inclusion and exclusion criteria. As a result, 194 articles were retrieved from the databases and 8 articles were excluded because of duplication. Another filtration step based on the study title was applied on the remaining articles. The articles with irrelevant titles were excluded. The abstracts of 53 articles were reviewed to evaluate their eligibility. 40 studies remained after the abstract filtration step. The last filtration step was to review the introduction. Finally, 37 articles were included in the SLR as detailed in Figure 1.

2.4. Data Extraction

After a thorough examination of the literature, 37 relevant studies were selected for review. In the next section, we analyze and review the selected studies to investigate how they contribute to the short-term urban water demand prediction literature. Our ultimate goal is to highlight deficiencies, issues, or gaps in predicting urban water demand, as we perceive them.

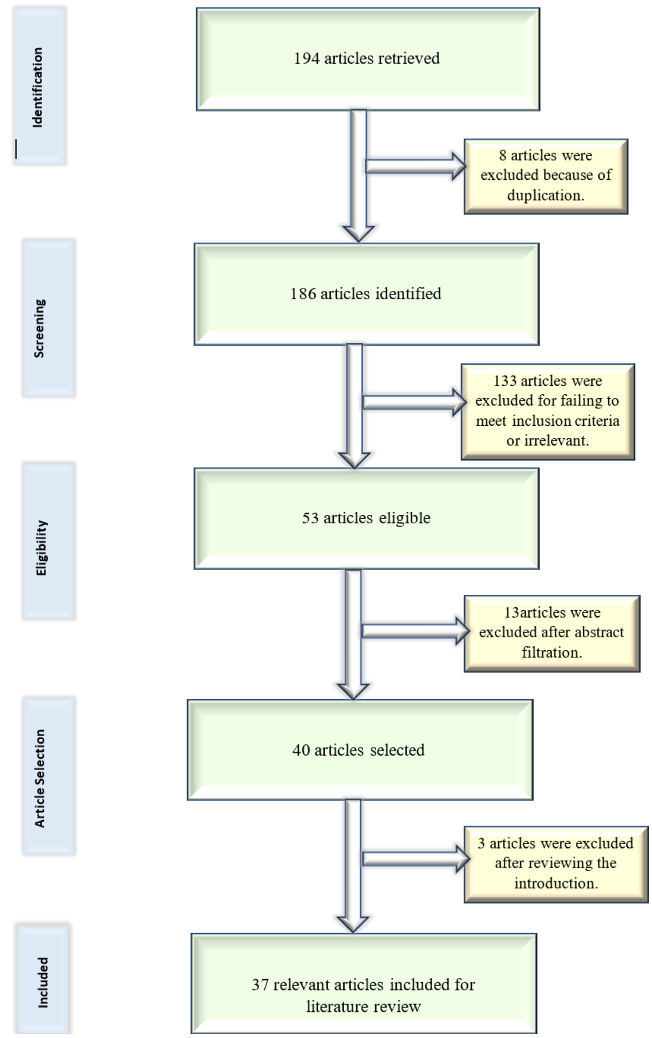

Figure 1: Article selection process

3. Literature Review for Short-Term Water Demand Forecasting

The purpose of this section is to analyze the selected studies and determine how they contribute to the investigation of short-term water demand forecasting. Additionally, we identify the methods, techniques, and tools utilized in each article. It is worth noting that many of the reviewed articles utilized artificial intelligence algorithms to predict short-term water demand. The following sections provide an overview of these approaches. Based on the classification presented in Rahim et al. (2020), we can categorize the selected articles into four groups according to their modeling method, as shown in Figure 2.

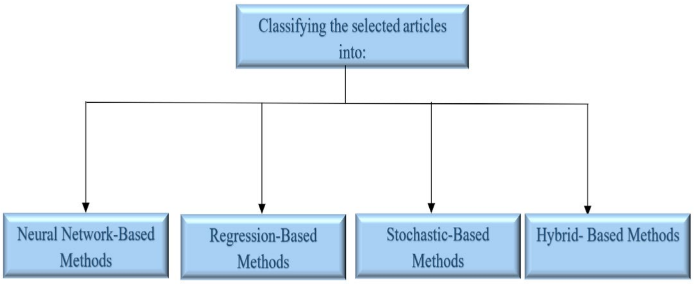

Figure 2: Article classification process

3.1. Neural Network-Based Methods

A study was conducted by Bennett et al. (2013) to develop water end-use demand forecasting models and to identify the key factors for a residential end-use forecasting model. The dataset that was utilized in this study comprised household demographic data, socio-economic data, and water appliance stock efficiency data. The models that were developed were two feedforward neural networks, two back propagation neural networks and one radial basis function network.

This study included various socio-economic variables, such as the income of the household, the number of adults, teenagers and children gender, age, and education and the efficiency of the water appliance stock (such as the toilet and clothes washer) as the key factors. The work showed that an artificial neural network (ANN) -based model was a feasible means of constructing water consumption forecasting models. Furthermore, it was shown that the most accurate feedforward neural network models with moderate prediction accuracies ranging between 0.33 to 0.6 for all end-use water demand (except for bath end-use water demand) were obtained by using a sigmoid activation function in the hidden layers and linear activation in the output layer. However, the study did not consider outdoor end-use (e.g., tap and garden irrigation) and household leakage when developing the models.

Pesantez et al. (2018) developed models to predict six-hour ahead water demand for individual accounts. The dataset that was used in this work included hourly water consumption reported by 24 smart meters for a period of 12 months and climate variables (e.g., temperature, dew point and humidity). Two machine learning algorithms and time clustering were deployed to propose short-term water demand forecasting models. Time series clustering using the k-means algorithm on the dataset improved forecasting accuracy by exploring similar consumption patterns.

Three feedforward neural networks layers, each layer with three neurons and a total of eight predictors, were used to forecast six-hour ahead water demand. A supervised learning algorithm regression tree with selected algorithmic parameters, number of trees and leaf size was applied in the study. The model's performance was assessed based on the mean absolute percentage error (MAPE) and standard deviation. The results in the study showed that a neural network approach outperformed the regression tree based on training errors. However, in terms of variability, the neural network method suffered from high values, so the method was unfit for further deployment.

Mouatadid and Adamowski (2017) evaluated various methods to identify the most reliable and accurate method for short-term urban water demand forecasting. They investigated various ANN methods, namely, a multi-layer perceptron feedforward network and a Leven- berg Marquardt backpropagation algorithm, the support vector regression method, the extreme learning machine method and the multiple linear regression method to predict urban water consumption for 1- and 3-days ahead. Their study utilized a dataset that was collected from 1999 to 2010, consisting of the average daily water consumption, max temp, total precipitation and occurrence of rain. This was used to train the proposed models, test them and validate them. In terms of prediction accuracy, the study demonstrated that the extreme learning machine model is an efficient learning method and outperformed the other forecasting methods.

All the proposed models achieved better performance in predicting 1-day lead time water demand than predicting 3-days ahead. However, the performance of all the proposed forecasting models deteriorated as the lead time increased. Furthermore, the study evaluated the overall performance of the models without considering whether some models performed better than other models on different calendar days (such as seasonality, holidays).

Liu et al. (2019) conducted separate research where they devised a forecasting model for short-term water consumption (forecasting up to five hours ahead) to examine the patterns in water demand. Their model, built upon the Bayesian regularized NAR neural network, employed min-max standardization to normalize the historical water consumption data. The effectiveness of the suggested model was evaluated by comparing it with a standalone backpropagation network and a standalone Bayesian regularized backpropagation network. The findings indicated that the developed model exhibited superior adaptability in predicting water demand compared to both the standard backpropagation neural network and the Bayesian regularized backpropagation neural network. However, it should be noted that this study relied on a relatively small database compared to other research studies, as it only included a single attribute, specifically water consumption values spanning a period of two weeks.

A comparison study to evaluate different models (i.e., exponential smoothing method, naïve methods, and ANN models) to forecast short-term water demand in Portugal was presented by Coelho and Andrade-Campos (2019).

The proposed models were evaluated on different datasets with different input variables to forecast hourly water demand for 1- and 24-hours ahead. In this work, the authors studied the impact of climate factors and anthropic factors in the water demand series using Pearson correlation coefficients analysis.

The results found that feedforward network-based models that had weather and anthropic data as input variables outperformed the exponential smoothing and naïve models. Furthermore, the authors stated that model accuracy was highly dependent on the right selection of inputs. However, the number of neurons in the neural network of the proposed model was determined by trial-and-error methods but the study didn't take into account the different water usage behaviors on different days of the week or the holidays.

A study was conducted by Vijai and Sivakumar (2018) to employ several methods to forecast water demand 1-hour, 12-hours and 24hours ahead. The researchers examined the performance of the developed models utilizing different factors such as water consumption values, weather factors and population.

The outcomes indicate that when considering forecast accuracy, ANN surpassed the other techniques to predict water demand in 1- hour, 12-hours and 24-hours ahead. In contrast, ANN had a higher computation time when compared to the other techniques except for deep neural networks.
Considering both accuracy and computational time, the least-squares support vector machine had better general performance compared to ANN. However, no validation process was performed on any of the proposed techniques. Also, the study didn't take into account the variability in water usage habits between the different days of the week.

Guo et al. (2018) investigated the utilization of a gated recurrent unit (GRU) model to predict water demand. To improve the accuracy of predicting 24-hour ahead water demand, a correction module was developed to decrease the forecasting errors. The correction module was based on a neural network method that consisted of one fully connected layer.
The developed framework comprised the following steps:

1. The water demand features were obtained and fed into each GRU model and ANN model to forecast 24-hour ahead water demand.

2. The outputs of the GRU and ANN model (i.e., forecasted values) were sent as input to the correction module to train the module to gain stable weight matrices, that is, to obtain a forecast value closer to the observed value. In the stage of testing the correction module, 96 outputs of the GRU model and ANN model were sent as inputs to the module and then these 96 values were multiplied by the dense layer weights to obtain 96 outputs.

More precisely, the correction module for 24-hour prediction was developed to predict water demand in daily basis with 96-time steps. The investigation revealed that concerning accuracy, the GRU model achieved superior performance compared to the other models for forecasting 15-min ahead water demand. Moreover, it was shown that the GRU- correction model had better performance to predict water demand 24-hours ahead. However, in terms of complexity, the structures of the ANN model and the GRU model were more complex and therefore the computational load was higher compared to the seasonal autoregressive integrated moving average model. Also, the predictability interval of the study was limited to a 15-min time step. In this study, no consideration was given to other variables such as calendar days or weather factors which was another drawback of the study.

The study conducted by Mu et al. (2020) proposed a model using the long short-term memory (LSTM) method to forecast hourly (i.e., 1-hour ahead) and daily (i.e., 24-hours ahead) urban water demand in Hefei, China. The authors compared the model's performance with various forecasting models such as autoregressive integrated moving average, support vector regression, and random forests. They evaluated the performance of the models using variant time resolution data (i.e., 15 mins, 1 hour and 24 hours). Also, the performance of the models was assessed on data with high uncertainty and abrupt changes. Furthermore, the study investigated the performance of the proposed model when external factors such as temperature and rainfall were excluded. Different metrics were applied to evaluate the performance of the developed forecasting models.
It was shown that in terms of forecast accuracy, the proposed model outperformed the others when predicting water demand with a 15-min and 1-hour time resolution. However, all models had a similar overall performance with a 24-hour time resolution. It was also shown that external factors had a limited effect on the models' performance for forecasting water demand with a 24-hour resolution. In addition, the authors compared the performance of the presented models to forecast water demand in multiple forecast horizons (i.e.,1 hour and 24 hours) using 15-min time resolution data. However, the performance of all the proposed forecasting models deteriorated as time increased; nevertheless, the performance of the LSTM- based model was better than the random forest, autoregressive integrated moving average and support vector regression models particularly when in- creasing the forecasting time. However, the study didn't take into account the different water usage behaviors between weekends and weekdays, which may affect the result for forecasting 24-hour water demand.

Xu et al. (2019) developed a novel continuous deep belief neural network forecasting model for daily urban water demand time series prediction in Zhuzhou, China. This study used one-step ahead forecasting. The developed model was based on the chaotic phase space reconstruction theory. The authors analysed the chaotic characteristic of the water demand time series using the power spectrum method and the largest Lyapunov exponent method. The C–C method was utilized to reconstruct the phase space, and the input of the constructed model was determined by calculating the best embedding dimension for the re- constructed phase. The forecasting model integrated two algorithms: the continuous deep belief network method was adopted to extract features at the bottom layer and the neural network method in the top layer was used for feature regression.

The research was conducted using only one input, specifically historical water demand values. The proposed model was compared to various models such as the support vector regression model, the generalized regression networks model and the feedforward network model.

The results show that the developed model achieved the best performance in predicting daily water demand compared to the others. However, the proposed model provided low prediction accuracy during peak daily water demand. In addition, the authors did not mention any information related to the computational time of the constructed model.

A novel deep model CDBESN was proposed by Xu Y et al. (2019) to forecast hourly urban water demand in Zhuzhou, China. The developed model combined a continuous deep belief network that was adopted for feature extraction in the bottom layer and an echo state network that was used for feature regression at the top layer. The proposed model used historical water demand on an hourly time scale as the only input data. Using three evaluation criteria, the forecasting accuracy of the developed model was contrasted with several forecasting models such as echo state network model, the continuous deep belief neural network model, and the support vector regression model. The results found that the constructed model outperformed the others in predicting hourly water demand by having the largest correlation coefficient value and the smallest MAPE and normalized root mean square error (RMSE) values. However, the study did not consider any external factors (such as weather factors) which may affect the results' accuracy. In addition, the authors did not mention any information regarding the computational time of the developed model.

Bata et al. (2020) proposed ANN models to predict 24-hour and one-week ahead water demand in Ontario, Canada. Five years of continuous hourly data were used from 2013 to 2017 in this study such as ambient temperature, absolute humidity solar radiation, dew point and station pressure. They investigated how input can impact model performance, using historical water demand as the only input compared to historical water demand and weather data as input for the other developed models. The results showed that forecasting 24-hours and one-week ahead using nonlinear models (i.e., non- linear autoregressive input and nonlinear autoregressive with exogenous input) have relatively better performance compared with the linear models. Moreover, the study showed that all the developed models with exogenous variables outperformed the models that were only fed with one input namely, historical water demand.

A hybrid model was proposed by Du et al. (2021) to predict daily urban water demand in Suzhou, China. The model was developed using the LSTM technique, the discrete wavelet transforms technique and principal component analysis to forecast water demand. They conducted the study using daily water consumption data from 2016 to 2020. To train the model, 998 data were utilized, and 662 data were applied to test the model's performance. The discrete wavelet transform technique was applied to eliminate noise from the water demand data, then principal components was applied to select the input. Moreover, two LSTM-based models were developed for daily water demand forecasting. The models' performance was assessed using three metrics namely, MAPEs of peaks, the MAPE, correlation coefficient and explained variance score. The results show that the proposed model achieved the best performance in predicting daily water demand compared to the other methods.

Kühnert et al. (2021) investigated the application of LSTM models in forecasting water demand for the next day. The researchers compared various approaches, including naïve predictor models, vector autoregressive models, and LSTM models, using RMSE as a performance metric. The findings indicated that LSTM models outperformed the other models, as they have the ability to retain relevant information such as the day type. The researchers also examined the transfer learning capability of LSTM models and confirmed that sensible results could be achieved by training the models with data from only a few days.

3.2. Regression-Based Methods

A comparative study was conducted by Herrera et al. (2010) to investigate various predictive machine learning models such as ANN, random forest, multivariate adaptive regression splines, support vector regression model and projection pursuit regression to predict hourly urban water demand using non-linear time series data. The analysis of the study was based on the hourly water consumption data of 5000 accounts from one hydraulic sector and daily meteorological information such as wind speed, temperature, dew point, rain and atmospheric pressure.

The input variables in the study were based on historical water consumption (i.e., the current hour and the previous hour water consumption values, and the target hour water consumption values in the previous week).

Based on the results, the study found that support vector regression models were the most accurate models compared to multivariate adaptive regression splines, projection pursuit regression and random forests respectively for predicting hourly urban water demand. In contrast, the experiments found that the ANN provided the worst prediction performance of all the other forecasting models. However, the water demand patterns were limited as the authors excluded weekends from this study, however unexpected water demand levels may occur on weekends which may affect the results.

Candelieri (2017) conducted a separate investigation focused on predicting hourly water demand by analyzing two distinct levels of data: aggregated data at the SCADA level (representing urban water demand) and individual customer consumption data at the AMR level (representing single customer data). The study employed time series clustering and support vector machine regression methods. Time series clustering through the application of the k-means algorithm was used on the data to investigate water demand patterns. The data at an individual level was collected for three months from 26 automatic meter readings that were installed in each building in a small area in Milan, Italy. The proposed model was based on the daily time series clusters to classify different water consumption patterns (i.e., similarities in time, shape and variations). At the SCADA level, six clusters were identified. After the clustering stage, each cluster was considered as input for the support vector machine regression model, each cluster was trained with a different support vector machine-based regression model based on the target hour. The findings of the study indicated that the support vector machine regression model was better suited for forecasting urban water demand compared to individual water consumption due to the variability in individual water usage patterns. However, it is important to note that the analysis in the study solely relied on water consumption values and did not account for other factors that could potentially influence the results, such as meteorological information. Additionally, the assessment of the support vector machine regression-based model's performance was based on a single metric, namely MAPE. In another investigation by Makki et al. (2015) to predict short-term water demand from a bottom-up end usage level, regression analysis was used to explore the determinant factors of indoor water consumption at the end-use level (i.e., bath, shower, toilet, dishwasher, tap, washing machine). The study investigated the influence of demographic, physical, and household makeup characteristics on water indoor end-use consumption. The authors proposed models to ensure the prediction accuracy for daily household indoor water consumption based on the identified determinant factors. It was shown that socio-demographic characteristics are the key determinant factors of all water end-use demand except for the toilet and the tap. The study reported that households with more working residents had higher indoor water consumption compared to those with retired residents. Also, the results showed that the education level of the household was a key determinant only for the end use of the dishwasher and shower.

Chen et al. (2019) carried out a unique framework that Integrated three stages: State assessment, forecasting, and error correction. The study focused on the application of the extended Kalman filter technique to forecast daily water demand in China. The developed forecasting model integrated the following three components:

- The prediction stage was based on multivariate local polynomial regression (MLPR) for data series. To increase the accuracy of the forecast prediction, the time series was decomposed into 4 subseries using discrete wavelet transformation.
- The relevance vector regression (RVR) technique was used for dynamic error correction prediction.
- The forecasted values of the MLPR and the error residual series which were obtained from the RVR model were utilized by the Kalman filter technique to calculate the parameters and to assess the state of a process to minimize the covariance error.

The study found that the developed model had better performance compared to MLPR, multi-scale relevance vector regression, autoregressive moving average, back propagation neural networks and multiple linear regression. The daily scale validation showed that the proposed model in terms of forecast accuracy had MAPE = 3.48, RMSE = 614.67, Nash-Sutcliffe coefficient = 0.93 and R-squared = 0.94. However, this study was conducted on a dataset that contained daily urban water consumption values only. Also, the high value of running time and the complexity of the model can be seen as another drawback of the proposed model.

Nguyen et al. (2021) investigated the possible factors that may influence residential water demand in Wallonia, Belgium using regression analysis. In this study, the data at the household level was collected using the Utility Survey and historical water consumption. They examined factors such as the number of adults, number of children, household size and income, living area and each resident's age. The study confirmed the importance of taking household characteristic into consideration when predicting water demand. Moreover, they emphasized that the household size factor has a greater effect on water use than household income.

Timotewos et al. (2022) conducted a study in Ethiopia to examine the influence of precipitation, relative humidity, and mean temperature on residential water demand. Multiple linear regression methods and principal component analysis were utilized in this study. The findings revealed a strong correlation between mean temperature and water demand in Arba Minch, while in Debre Birhan, residential water demand was significantly associated with humidity. However, for Ziway, the study found that none of the weather variables had an impact on water demand changes.

Stańczyk et al. (2022) conducted a study in Wroclaw, Poland, aimed at predicting water demand. The researchers proposed a methodology that combined the linear regression model with evolutionary algorithms to extract the weekly seasonality. The findings indicated that the proposed model achieved a significant reduction in MAPE, resulting in a 1.31% MAPE to predict the water demand.

3.3. Stochastic-Based Methods

Abadi et al. (2017) proposed a probabilistic method on the basis of non-homogeneous Markov models, with the aim of clustering the dynamics of daily water consumption behaviour into different groups to forecast daily consumption behaviours for each group. A log normalized approach was used to convert data to a normal distribution. The k-means approach was used for clustering the normalized data to classify the daily patterns of water consumption.

The performance of their proposed model was compared to the state independent model and homogeneous Markov model using various combinations of input factors such as water consumption values, temperature, and rainfall.

The proposed model achieved reliable performance with 80% accuracy for forecasting consumption behaviors.

In terms of prediction accuracy, the study showed that the proposed methods had better performance when only using consumption variables as input and not combining these with climate factors. However, the complexity (i.e., no. of parameters to be adjusted) of the proposed model was higher than the other models, namely, state independent model and homogeneous Markov model. The study did not include the consumers' socioeconomic data and demographic data which was another limitation of this work.

Blokker et al. (2010) developed a stochastic end-use model to forecast water demand on a small-time horizon scale and small residential scale based on statistical information about users and end use rather than using water demand measurements.

The stochastic end-use model was based on a Poisson rectangular pulse model. Furthermore, the authors assumed that there are several types of users such as children, teenagers, adults with or without a job away from home and seniors. The developed model used Monte Carlo simulations and the output was a water demand pattern.

The simulation results showed that the model achieved high accuracy in forecasting short-term water demand (i.e., R-squared of 0.93). However, the study did not consider weekend behavioral patterns and the end-use model was based on data that was retrieved from surveys.

Rathnayaka et al. (2017) conducted a study presenting a stochastic model to predict water demand over various spatial and temporal ranges. The findings revealed that the model demonstrated better prediction accuracy at the quarterly level compared to the hourly level.

When evaluating the model's performance at the hourly validation level, satisfactory predictions were achieved, with a Nash-Sutcliffe coefficient of 0.73 and an R-squared value of 0.75, resulting in an 85% accuracy. However, it should be noted that the assessment of the model's performance was based solely on the Nash-Sutcliffe coefficient and R-squared values.

Gagliardi et al. (2017) introduced two Markov chain models that employed statistical concepts to predict water demand 24 hours in advance with 1-hour intervals. In terms of deterministic forecast accuracy, the results demonstrated that the homogeneous Markov chain model outperformed the non-homogeneous model. It achieved comparable performance to the ANN model and surpassed the naïve models. Moreover, the proposed Markov chain model enabled probabilistic forecasts by estimating future water demand probabilities within a predetermined interval of up to 6 hours ahead. However, it is crucial to mention that the study relied on a solitary input variable, specifically water consumptions, and the evaluation of the model's effectiveness was conducted using just one metric, the Nash-Sutcliffe index.

3.4. Hybrid- Based Methods

To forecast potable water demand for the next 24-48 hours with 1-hour time steps, a parallel adaptive weighting strategy using univariate time series was proposed by Sardinha-Lourenço et al. (2018).The proposed model combined three forecasting methods, namely, heuristic model and two moving average and autoregressive models (one with classified data with regard to the type of day and the other without classification) and assigned weights to them according to their historical success to forecast water demand to reduce the MAPE error in the last 24 hours. The proposed strategy was validated using both a short-term forecast heuristic model and an autoregressive integrated moving average model.

The consumption of potable water and calendar days were used as the model's input in this study. The study showed that by using the proposed model, the prediction error and the average error reduced by 15.96% and 9.20% respectively. The research primarily focused on utilizing water demand values and evaluated the model's performance using two performance indices, namely MAPE and R-squared.

A model to predict daily water demand in Al-Khobar, Saudi Arabia was proposed by Al-Zahrani and Abo- Monasar (2015). The study was based on integrating the neural network model with a time series model. The authors con- structed various models to forecast daily water demand using standalone regression neural network models (GRNN), standalone best moving average and autoregressive techniques models and the proposed combined model based on three phases: Phase 1: time series forecasting using ARMA models only. Phase 2: to predict future water consumption using GRNN models only. Phase 3: to combine the models from phase 1 and phase 2. The proposed model (TS-GRNN) combined the best GRNN models with the best moving average and autoregressive technique models (ARMA). To ensure the model's accuracy, models with a high coefficient of correlation (i.e., greater than 0.55) were selected for integration. TS-GRNN was based on historical water demand values and climatological variables such as average temperature, precipitation and humidity. The authors concluded that the combined model outperformed the GRNN models and times series models with a coefficient of correlation close to 0.9. However, they assessed the performance of the combined forecasting technique using MAPE and the coefficient of correlation only. In addition, the authors did not provide any information related to the computation time or the complexity of the proposed model.

Bata et al. (2020) conducted a research investigation that assessed the efficacy of several models for short-term water demand forecasting. The researchers introduced a novel hybrid model by merging self-organizing maps with regression tree. The results demonstrated that the hybrid model exhibited superior predictive accuracy compared to both the regression tree and seasonal autoregressive integrated moving average models.

The hybrid model achieved a lower MAPE compared to the standalone regression tree model, with improvements ranging from 15% to 60%. However, it is important to note that the hybrid model, while demonstrating improved prediction accuracy, possessed greater complexity and necessitated longer execution time compared to both the regression tree and seasonal autoregressive integrated moving average models.

The impact of weather factors on water demand was investigated by Zubaidi et al. (2018). The authors proposed two hybrid approach models to forecast daily municipal water demand on yearly and seasonal data, (i.e., gravitational search technique-ANN and backtracking search technique-ANN). Several statistical techniques such as data screening, correlation matrix analysis, autocorrelation technique and variance inflation factor were adopted to select the appropriate input data of the NN model from the identified potential weather variables to decrease the uncertainty in the input factors. To determine the optimum parameter values of the ANN model (i.e., best learning rate and the number of neurons), the two algorithms' gravitational search technique and backtracking search technique were utilized in the ANN model. The results found that in terms of fitness function, the gravitational search technique-ANN outperformed the backtracking search technique-ANN for all yearly and seasonal data. Thus, the gravitational search technique-ANN was capable of forecasting daily water consumption with minimum error prediction for both yearly and seasonal water consumption data. However, the study did not consider the difference between weekend water usage behaviours and weekdays.

Wu et al. (2020) proposed a hybrid model to forecast water demand one day in advance, utilizing 15-minute time intervals. The proposed model utilized a least-squares support vector method to forecast daily water demand using 96 steps and an error correction module. The error correction module consisted of three stages:

1. Transform the error time series that was obtained from the initial least-squares support vector machine into chaotic time series. 2. Develop a forecasting model of the errors by adopting the least-squares support vector machine and chaotic time series. 3. Calculate the forecasted error at the next time step to correct the initial forecast value.

The presented model was validated in different metering areas in China. The proposed model was assessed against several other models, including the single prediction least-squares support vector machine model, as well as the moving average and autoregressive models.

The findings demonstrated that the adoption of the proposed hybrid model resulted in superior overall performance compared to the moving average and autoregressive models. However, it is important to note that the model's validation was limited to a forecast horizon of only 24 hours.

A study by Odan and Reis (2012) used ANN methods to forecast hourly water demand for Araraquara, in Brazil. They explored the best model using hourly water consumption values and weather data such as temperature and relative humidity. In this study, the estimation was based on 24-hour lead times with a 1-hour time step. Correlation analysis was used to select the input variables for the neural network models. The authors compared various algorithms, namely multilayer perceptron with backpropagation algorithm (MLP-BP), dynamic neural network (DAN2), and two hybrid neural network models that combined Fourier series with the ANNs. The hybrid models (i.e., ANN (MLP- BP)-H and DAN2-H) used the error formed by the Fourier series forecasting (i.e., difference between the Fourier series predictions and the observed water consumption) with weather factors as input variables to ANNs. The study showed that DAN2 performed better than MLP-BP. The proposed hybrid model that combined DAN2 with FS performed the best for predicting 1-hour and 24-hours ahead. Also, joining FS to MLP-BP did not improve the performance for forecasting next hour water demand, and some hybrid models of ANN (MLP-BP) had a worse performance compared to the non-hybrid models. The study also showed that different versions of the DAN2 forecasting models that were used to predict the first hour or 24-hours ahead did not need to use weather information. However, the model's accuracy was evaluated using only the mean absolute error (MAE) and Pearson correlation coefficient. A hybrid wavelet bootstrap-based neural network model was developed by Tiwari and Adamowski (2013) to forecast short-term urban water demand for daily, weekly, and monthly lead times for Montreal in Canada. This work utilized a dataset that consisted of average daily water consumption values, average monthly water consumption values, max temperature, and precipitation. The performance of the recently developed hybrid model was evaluated and compared against several other models in the study. These included the autoregressive integrated moving average model, the autoregressive integrated moving average model with exogenous input variables, regular neural networks (NNs), wavelet analysis-based NNs, bootstrap-based NNs (BNN), and a simple naïve persistence index model. It was demonstrated that in terms of prediction accuracy, the wavelet analysis-based NN and the hybrid wavelet bootstrap-based NN models provided significantly better prediction results than the other forecasting models. Also, it was shown that using precipitation and max temp improved the prediction accuracy of urban water demand using wavelet analysis. Moreover, the authors showed that the number of bootstrap samples affected the performance of the bootstrap-based NN models. The prediction accuracy of the bootstrap-based NN models can be improved by using a large number of bootstrap resamples. The authors also showed that using wavelets techniques improved the accuracy of the forecasting model and using bootstrap techniques reduced the uncertainty related to the forecasts, improved model reliability and ensured robustness. However, it can be noted that this pro- posed technique involved a large dataset to ensure accurate forecasting, which may not always be appropriate.

Tiwari and Adamowski (2017) proposed a hybrid wavelet, bootstrap-based neural network to forecast urban water demand for 1 day, 3 days and 5 days ahead. The dataset that was used in this study consisted of average daily water consumption for 3 years and meteorological variables (such as max temperature and precipitation) for Calgary in Canada. The authors compared the performance of the proposed hybrid model to traditional neural networks, wavelet-based NN, and bootstrap-based NN models. The study showed that the conjunction model hybrid wavelet, bootstrap-based neural network performed significantly better than the other proposed models. Furthermore, the hybrid model could efficiently predict water demand and assess uncertainty 1 day, 3 days and 5 days ahead. The model's performance was evaluated based on five indexes, namely, the coefficient of determination, RMSE, persistence index (PI), MAE, and peak percentage deviation (Pdv). However, it can be noted that this proposed technique involved a large dataset to ensure accurate forecasting, which may not always be appropriate.

Caiado, (2010) conducted a case study to investigate the prediction accuracy of individual and combined double seasonal univariate time series models to predict the daily water consumption in Granada, Spain with a multistep ahead forecasting. All possible combinations of double-seasonal Holt-Winters, moving average and autoregressive and generalized autoregressive conditional heteroskedasticity were considered using simple and optimal weights for the forecasts. The optimal weights of the forecasts were computed using the inverse of the MSE and squared errors of the three individual methods. Water pattern recognition was studied within double cycles: week cycle of seven days and a year cycle of 365 days. The performance of the models in predicting water demand with a multistep ahead was compared by computing the MSE index.

It was found that the optimal combination that was computed using weighted by inverse squared errors of the forecasting methods Holt-Winters, moving average and autoregressive, and generalized autoregressive conditional heteroskedasticity was more accurate than the several simple combinations, except for seven-step ahead forecasting. Also, the result showed that for forecasting 1-day lead time, the average MSE reduced by 8.33% using optimal combined forecasting models compared to individual methods (i.e., Holt-Winters, moving average and autoregressive and generalized autoregressive conditional heteroskedasticity) and for 2-days and 3-days ahead, the error reduced by 12.77% and 10.64%, respectively. However, the prediction performance of the proposed approach was limited to 7 days ahead. Also, this study utilized a dataset that was limited to one input (i.e., the daily urban water consumption series). Moreover, the model's performance was assessed using only one metric (i.e., MSE).

Nunes Carvalho et al. (2021) conducted a study in Fortaleza, Brazil which utilized three techniques, namely, integrated random forest method for variable selection and importance, a self-organizing map for clustering and to analyze water demand patterns, and ANN to predict water demand in two spatial-level aggregations with various information. The results showed that the central and eastern regions have a higher level of health environments and a better level of education with a high water consumption. However, the western and southern zones have a low average income and low water consumption.

A stack model with a bias correction technique was proposed by Xenochristou and Kapelan (2020) to improve daily water demand forecasting in the UK. The data was collected from October 2014 to September 2017 from 1793 properties. In this study, seasons, day type, historical water consumption, weather variables and area postcode were used as inputs for the developed model. The weather variables comprised air temperature, mean relative humidity, days with no rain and number of sunshine hours. The dataset was split into 70% for training the developed models and 30% for testing. The proposed methodology (i.e., the stack model) and the other developed models namely random forests, ANNs, deep neural networks, extreme gradient boosting, generalized linear models and gradient boosting machines were assessed using several metrics. The three metrics that were applied in this study to evaluate the model's performance are MAPE, coefficient of determination and mean squared error. The authors confirmed that the proposed methodology (i.e., the stacked model) which utilizes random forests, gradient boosting machines, deep neural networks, and generalized linear models had better performance compared to the other water demand forecasting methods.

Du et al. (2022) conducted research to predict water demand using an interval forecasting method that gives a range of potential values for the anticipated water consumption rather than one specific estimate. The research involved utilizing two distinct approaches to generate interval forecasts: a particle swarm optimization algorithm optimized kernel density estimation distribution, and an LSTM. The particle swarm optimization-optimized kernel density estimation distribution method involved utilizing historical water demand data to estimate a probability density function that can be applied to produce interval forecasts. The bandwidth parameter of the kernel density estimation distribution, which affects the width of the probability density function and the accuracy of the interval forecasts, is optimized using the particle swarm optimization algorithm. To generate interval forecasts, the LSTM utilized both historical water consumption data and other factors to train the model. In this work, the proposed model was compared with other classical models using data on urban water demand in Suzhou, China to evaluate its performance. The results showed that the proposed model was more efficient than the other prediction models and it had the potential to improve decision-making in man- aging urban water resources.

Zhou et al. (2022) proposed a novel hybrid frame- work to predict daily urban water demand in Suzhou, China with multiple variables. The frame- work, named the attention-based CNN-LSTM model, merges several deep learning techniques, including: The attention mechanism, CNN, LSTM, and an encoder decoder. The findings indicate that the hybrid model outperformed the other models in terms of prediction accuracy as indicated by its lower values for MAE, MAPE, RMSE and the coefficient of determination.

Tables 2, 3, 4, and 5 present a brief overview of the selected studies, including the methods, tools and techniques employed in each.

ID	Study Title	Author(s)/ Year	Journal/ Conference Name	Contributions	Techniques/methods used
N1	ANN-based residential water end-use demand forecasting model	Bennett et al., 2013	Expert Systems with Applications	Developed water end-use demand forecasting models and identified the key determinants for a residential end-use forecasting model.	Feedforward backpropagation net- works one radial basis function net- work, Sigmoid activation, linear activation, linear regression analysis.
N2	Modeling and Forecasting Short-Term Water Demand Reported by Smart Meters	Pesantez et al., 2018	WDSA-CCWI 2018 Joint Conference	Developed models to predict six-hour ahead water demand for in- dividual accounts.	Three-layer feedforward neural net- works, regression tree, time clustering, k-means algorithm.
N3	Using extreme learning machines for short-term urban water demand forecasting	Mouatadid and Adamowski, 2017	Urban Water Journal	Presented various machine learning methods to explore the most reliable and accurate method to predict daily urban water consumption 1- and 3-days ahead.	Multi-layer perceptron feedforward, Levenberg Marquardt, backpropagation algorithm, support vector regression, extreme learning machine, multiple linear regression models.
N4	Bayesian regularized NAR neural network based short-term prediction method of water consumption	Liu et al., 2019	E3S Web of Conferences	Implemented a predictive model for short-term water consumption time series, specifically forecasting five hours ahead.	Bayesian regularization algorithm and nonlinear autoregressive method, standard backpropagation NN, Bayesian regularized backpropagation, min-max standardization method.
N5	Short-term forecasting of hourly water demands – A Portuguese case study	Coelho and Andrade-Campos, 2019	International Journal of Water	A comparison study to evaluate different developed models to forecast hourly water demand 1 and 24 hours ahead in Portugal.	Pearson correlation coefficient analysis, exponential smoothing, naïve models, feedforward neural networks.
N6	Performance comparison of techniques for water demand forecasting	Vijai and Sivaku- mar, 2018	Procedia Computer Science	Different machine learning methods were utilized to assess the accuracy of predicting water demand at different interval.	Deep neural network, extreme learning machines, a simple feedforward neural network, random forest, Gaussian process regression, multiple regression, least-squares support vector ma- chine.
N7	Short-term water demand forecast based on deep learning method	Guo et al., 2018	Journal of Water Resources Planning and Management	The authors developed models to predict water demand at two different time intervals: 15 minutes and one day.	GRUN, feedforward propagation and backpropagation, dense layer weight matrix, seasonal autoregressive integrated moving average.
N8	Hourly and Daily Urban Water Demand Predictions Using a Long Short-Term Memory Based Model	Mu et al., 2020	Journal of Water Resources Planning and Management	LSTM technique was employed to develop a model to forecast 1 hour and 24 hour ahead urban water demand in Hefei, China. The effectiveness of the proposed model was evaluated by comparing its performance with several other forecasting models.	LSTM, autoregressive integrated moving average model, support vector regression model, random forest model, sensitivity analysis.

N9	Daily Urban Water Demand Forecasting Based on Chaotic Theory and Continuous Deep Belief Neural Network.	Xu et al. 2019	Water (Switzerland).	Constructed a novel continuous deep belief neural network forecasting model for daily urban water demand time series prediction in Zhuzhou, China.	Continuous deep belief neural network, RBM, backpropagation, chaotic theory, phase -space reconstruction method, C–C method, multilayer FFNN, power spectrum, largest Lyapunov exponent, support vector regression, GRNN.
N10	Hourly Urban Water Demand Forecasting Using the Continuous Deep Belief Echo State Network.	Xu Y et al. 2019	Water	A novel deep (CDBESN) was proposed to forecast hourly urban water demand in Zhuzhou, China.	A continuous deep belief network, echo state network, support vector regression models, sigmoid activation function
N11	Short-Term Water Demand Forecasting Using Nonlinear Autoregressive Artificial Neural Networks.	M. H. Bata et al., 2020	Journal of Water Resources Planning and Management.	The authors suggested the use of artificial neural network (ANN) models to forecast water demand in Ontario, Canada, for both 24-hour and one-week periods ahead.	Nonlinear autoregressive, seasonal autoregressive integrated moving average, nonlinear autoregressive with exogenous.
N12	Deep learning with long short-term memory neural networks combining wavelet transform and principal component analysis for daily urban water demand forecasting.	Du et al., 2021	Expert Systems with Applications.	A hybrid model was proposed to predict daily urban water demand in Suzhou, China.	Discrete wavelet transform, principal component analysis, LSTM.
N13	Application of LSTM Networks for Water Demand Prediction in Optimal Pump Control	Kühnert et al., 2021	Water	The study investigated the application of LSTM models for predicting water demand one day ahead. The authors conducted a comparative analysis of various approaches to evaluate the models' performance.	Naïve method, vector autoregressive and LSTM.

Table 2: Relevant articles on neural network-based methods and publication information

ID	Study Title	Author(s)/ Year	Journal/ Conference Name	Contributions	Techniques/methods used
R1	Predictive models for forecasting hourly urban water demand.	Herrera et al., 2010	Journal of Hydrology	Investigated various predictive machine learning models to predict hourly urban water demand using non-linear time series data.	ANN (one hidden layer feedforward neural network and the backpropagation algorithm). Random forest, multivariate adaptive regression splines, support vector regression model and projection pursuit regression.
R2	Clustering and Support Vector Regression for Water Demand Forecasting and Anomaly Detection.	Candelieri, 2017	Water (Switzerland)	A model was proposed to predict water demand on an hourly basis, considering two distinct data levels: the aggregated and the individual.	Time series clustering, k-means algorithm, support vector machine regression, mapping function.

R3	Novel bottom-up urban water demand forecasting model: Revealing the determinants, drivers and predictors of residential indoor end-use consumption.	Makki et al., 2015	Resources, Conservation and Recycling	To predict daily water demand and explore the key determinants of indoor water consumption at the end-use level.	Regression analysis, statistical techniques such as bootstrapping cluster analysis dummy coding, Pearson's chi- squared test.
R4	A Forecasting Frame-work Based on Kalman Filter Integrated Multivariate Local Polynomial Regression: Application to Urban Water Demand.	Chen et al., 2019	Neural Processing Letters	he study utilized the extended Kalman filter technique to predict daily water demand.	Kalman filter technique, MLPR, wavelet transformation, RVR, phase space reconstruction, c-c method.
R5	Factors influencing residential water consumption in Wallonia, Belgium.	Nguyen et al., 2021	Utilities Policy	Study investigated the possible variables that may impact urban water demand in Wallonia, Belgium using regression analysis.	Fixed effects regression with spatial predictors and mixed-effects regression with spatial random intercepts.
R6	The Assessment of Climate Variables and Geographical Distribution on Residential Drinking Water Demand in Ethiopia.	Timotewos et al., 2022	Water	In Ethiopia, a research investigation was conducted to examine the impact of weather variables on urban water demand.	Principal component analysis. Multiple linear regression.
R7	Improving short-term water demand forecasting using evolutionary algorithms.	Stańczyk et al., 2022	Scientific Reports	A technique that combines a linear regression model and evolutionary algorithms was proposed to forecast water demand in Wroclaw, Poland.	Evolutionary algorithms, support vector regression, linear regression, regression trees, multilayer perceptron.

Table 3: Relevant articles on regression-based methods and publication information

ID	Study Title	Author(s)/ Year	Journal/ Conference Name	Contributions	Techniques/methods used
S1	Predictive Classification of Water Consumption Time Series using Non-homogeneous Markov Models.	Abadi et al., 2017	IEEE	A probabilistic approach was proposed using non-homogeneous Markov models to cluster the patterns of daily water consumption behavior into distinct groups. The goal was to forecast daily consumption behaviors for each group.	k-means algorithm, non-homogeneous Markov models, state independent model homogeneous Markov log normalize.
S2	Simulating Residential Water Demand with a Stochastic End-Use Model.	Blokker et al., 2010	Journal of Water Resources Planning and Management.	A stochastic end-use model was developed to forecast water demand within a short time horizon and at a small residential scale.	Statistical information about users and end uses, Poisson rectangular pulse.

S3	Prediction of urban residential end-use water demands by integrating known and unknown water demand drivers at multiple scales II: Model application and validation.	Rathnayaka et al., 2017	Resources, Conservation and Recycling	A stochastic model was proposed to predict water demand for residential areas across various spatial and temporal scales.	A stochastic approach, negative binomial and Poisson probability distributions.
S4	A Probabilistic Short-Term Water Demand Forecasting Model Based on the Markov Chain Francesca.	Gagliardi et al., 2017	Water	A method was devised to estimate the probabilities of future water demand over a predefined interval up to six hours in advance.	Homogeneous and non-homogeneous Markov chain, multi-layer perceptron ANN, naïve method.

Table 4: Relevant articles on stochastic-based methods and publication information

ID	Study Title	Author(s)/ Year	Journal/ Conference Name	Contributions	Techniques/methods used
H1	Increased performance in the short-term water demand forecasting through the use of a parallel adaptive weighting strategy.	Sardinha-Lourenço et al., 2018	Journal of Hydrology	Proposed a methodology to fore-cast potable water demand for the next 48 hours with a 1h time step.	Parallel adaptive weighting strategy using univariate time series. Short- term forecast heuristic model, and autoregressive integrated moving average model.
H2	Urban Residential Water Demand Prediction Based on Artificial Neural Networks and Time Series Models.	Al-Zahrani and Abo-Monasar, 2015	Water Resources Management.	Developed a model to forecast daily water demand in Al-Khobar, Saudi Arabia.	GRNN and moving average and autoregressive techniques.
H3	Short-term water demand forecasting using hybrid supervised and unsupervised machine learning model.	M. Bata et al., 2020	Smart Water	explored the performance of short-term water demand forecasting models and proposed a hybrid model to forecast water demand 1 hour, 8 hours, 24 hours, and 1 week ahead.	Self-organizing map models, regression tree, seasonal autoregressive integrated moving average, Pearson correlation coefficient.
H4	Short-Term Urban Water Demand Prediction Considering Weather Factors.	Zubaidi et al., 2018	Water Resources Management	The purpose of this study was to investigate the impact of weather factors on water demand. The authors proposed two hybrid models to predict daily urban water demand.	Data pre-processing, correlation matrix, autocorrelation and variance inflation factor gravitational search algorithm, backtracking search algorithm, artificial backpropagation neural network, Levenberg-Marquardt.
H5	Hybrid Model for Short-Term Water Demand Forecasting Based on Error Correction Using Chaotic Time Series.	Wu et al., 2020	Water	A hybrid model was suggested to forecast water demand one day in advance with 15-minute time intervals. The model relied on water consumption data and calendar information.	Least-squares support vector machine, chaotic time series, Fourier series, moving average and autoregressive.

H6	Hybrid Water Demand Forecasting Model Associating Artificial Neural Network with Fourier Series.	Odan and Reis, 2012	Journal of Water Resources Planning and Management	The authors used ANN methods to forecast hourly water demand for Araraquara, Brazil.	Correlation analysis, ANN, MLP-BP, DAN2, Fourier series.
H7	Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models.	Tiwari and Adamowski, 2013	Water Resources Research	A hybrid wavelet, bootstrap-based neural network model was developed to forecast short-term urban water demand for daily, weekly, and monthly lead times.	Autocorrelation, cross-correlation, Levenberg-Marquardt, wavelet analysis, bootstrap technique, multilayer feedforward NN.
H8	An ensemble wavelet bootstrap machine learning approach to water demand forecasting: a case study in the city of Calgary, Canada.	Tiwari and Adamowski, 2017	Urban Water Journal	A hybrid wavelet, bootstrap-based neural network was developed to forecast urban water demand for 1 day, 3 day and 5 day lead times.	Autocorrelation, cross- correlation, Levenberg-Marquardt, wavelet analysis, bootstrap technique, multilayer feedforward NN.
H9	Performance of combined double seasonal univariate time series models for forecasting water demand.	Caiado, 2010	Journal of Hydrologic Engineering	To investigate the prediction accuracy of various individual and combined univariate time series models to predict daily water consumption in Spain with multi-step (i.e., 1 to 7 days) ahead forecasting.	Holt Winters, moving average, autoregressive generalized autoregressive conditional heteroskedasticity, recursive technique.
H10	Urban Water Demand Modeling Using Machine Learning Techniques: Case Study of Fortaleza, Brazil.	Nunes Carvalho et al., 2021	Journal of Water Resources Planning and Management	To analyze water demand patterns and to predict water demand in Fortaleza, Brazil.	Random forest, self-organizing map and ANN.
H11	An ensemble stacked model with bias correction for improved water demand forecasting.	Xenochristo and Kapelan, 2020	Urban Water Journal	Proposed a stack model with a bias correction technique to improve daily water demand forecasting in UK.	Random Forests, ANNs, deep neural networks, extreme gradient boosting, generalized linear models and gradient boosting machines.
H12	Interval forecasting for urban water demand using PSO optimized KDE distribution and LSTM neural networks.	Du et al., 2022	Applied Soft Computing	The work investigated the efficiency of two techniques - particle swarm optimization algorithm optimized kernel density estimation distribution, and a LSTM for interval forecasting o f urban water demand.	Particle swarm optimization, optimized kernel density estimation, LSTM
H13	IA Hybrid Framework for Multivariate Time Series Forecasting of Daily Urban Water Demand Using Attention-Based Convolutional Neural Network and Long Short-Term Memory Network.	Zhou et al., 2022	Sustainability	The research paper presented a hybrid framework for predicting daily urban water demand in Suzhou, China.	CNN, LSTM, encoder decoder, attention mechanism, Pearson correlation co- efficient.

Table 5: Relevant articles on hybrid-based methods and publication information

3.5. Relevant Studies Analysis

Table 6 summarizes the analysis of the relevant articles based on the defined criteria to identify research gaps in the literature. A score of 1 indicates that the study has met a particular criterion, while a score of 0 means that the study has not met that criterion. Each row corresponds to a specific study, and the columns describe different aspects of the study evaluation. The columns denote the following:

- ID: identifier for the study
- Model Validation: a binary value (0 or 1) indicating whether the proposed model has been validated using a validation dataset. Model validation refers to the process of evaluating the model's performance using data that was not used during the model's training.
- Multiple Input: a binary value indicating whether the model can handle multiple input variables.
- Real DB: a binary value indicating whether the model has been tested, using real-world data.
- Multi-Horizon: a binary value indicating whether the model can make forecasts for multiple time horizons.
- Multi-Periodicity: a binary value indicating whether the model can handle data with multiple periodicities (e.g., daily and weekly data)
- Benchmark: a binary value indicating whether the model has been compared to other models.
- Behavioural Changes: a binary value indicating whether the model can handle changes in water usage behavior (e.g., due to weekend, public holidays, droughts and water conservation campaigns)
- Evaluation Indexes >= 3: This criterion involves assessing model performance using a minimum of three specific criteria. In our review, the evaluation criteria include measures such as forecasting accuracy, MAPE, RMSE, R-squared, and other relevant indicators. These metrics are employed to assess the performance of the developed water demand forecasting model. Thus, 'Evaluation Indexes >= 3' signifies that the model's performance is evaluated based on multiple quantitative criteria, ensuring a comprehensive assessment of its accuracy and reliability.
- Forecast Urban Water Demand: The binary value indicating whether the model can forecast urban water demand was determined based on the specific objectives outlined in the reviewed papers. We assigned a value of 0 if the study's objectives did not explicitly aim to forecast urban water demand, indicating that the model developed in that study was not intended for such forecasts. Conversely, a value of 1 was assigned if the study explicitly aimed to forecast urban water demand or if the model's scope encompassed urban water demand forecasting. For instance, if a study focused on evaluating hydrological factors' impact on water consumption or identifying variables affecting residential water consumption without specifying urban forecasting or predicting water demand at the individual level, we assigned a value of 0.

By using this type of analysis, the differences and limitations of each study can be easily identified and compared. This can help in assessing each study as well as in identifying areas where further research is needed to improve the accuracy and reliability of urban water demand forecasting. Additionally, this type of analysis can provide insights into the strengths and weaknesses of different modeling approaches and help to guide the development of more effective and efficient models for urban water demand forecasting.

ID	Model Validation	Multiple Input	Real DB	Multi Horizon	Multi Periodicity	Bench-mark	Behavioural Changes	Evaluation Indexes >= 3	Forecast urban Water Demand
N1	1	1	1	0	0	1	0	1	0
N2	1	1	1	0	0	1	1	0	0
N3	1	1	1	1	0	1	0	0	1
N4	1	0	1	0	0	1	0	1	1
N5	1	1	1	1	0	1	0	1	1
N6	0	1	0	1	0	1	0	1	1
N7	1	0	1	1	0	1	0	1	1
N8	1	1	1	1	1	1	0	1	1

N9	1	0	1	0	0	1	0	1	1
N10	1	0	1	0	0	1	0	1	1
N11	1	1	1	1	0	1	0	1	1
N12	0	1	1	0	0	1	1	1	1
N13	1	1	1	0	0	1	1	0	1
R1	1	1	1	0	0	1	0	1	1
R2	1	0	1	0	0	0	1	0	1
R3	1	1	1	0	0	0	1	0	0
R4	1	0	1	0	0	1	0	1	1
R5	1	1	0	0	0	0	1	0	0
R6	0	1	1	0	0	0	1	0	0
R7	1	1	1	0	0	1	1	0	1
S1	0	1	1	0	0	1	1	1	0
S2	1	1	0	1	0	0	0	1	1
S3	1	1	0	1	0	0	1	0	1
S4	1	0	1	0	0	1	1	0	1
H1	1	0	1	0	0	1	1	0	1
H2	1	1	1	0	0	1	0	0	1
H3	1	0	1	1	0	1	0	0	1
H4	1	1	1	0	1	0	0	1	1
H5	1	0	1	0	0	0	0	1	1
H6	1	1	1	0	0	1	0	0	1
H7	1	1	1	1	1	1	0	1	1
H8	1	1	1	1	0	1	0	1	1
H9	1	0	0	1	0	1	1	0	1
H10	1	1	1	0	0	0	1	1	1
H11	1	1	1	0	0	1	1	1	1
H12	1	1	1	1	0	1	0	1	1
H13	1	1	1	1	0	1	0	1	1

Table 6: Analysis criteria to define gaps

Based on the above analysis, it can be observed that the majority of the studies used neural network models (N) for forecasting urban water demand. This suggests that neural networks are a popular choice among researchers in this field, perhaps due to their ability to handle complex, non-linear relationships in data. Regression-based models (R) were also used in a few studies, indicating that these models can still provide valuable insights into urban water demand forecasting. Additionally, hybrid models (H), which combine multiple modeling techniques, were used in a couple of studies, showing that these models may have some advantages over using a single model type. Overall, the choice of modeling technique may depend on various factors such as the nature of the data, the research question, and the availability of computational resources. Therefore, it is essential to carefully evaluate different modeling techniques before selecting the most appropriate one for urban water demand forecasting.
Additional findings and conclusions that can be drawn from analyzing the relevant studies based on the defined criteria are as follows:

- Most studies (33 out of 37, or 89.1%) validated their models using a separate validation dataset, which is a good practice of model evaluation.
- Based on the fact that 70% of the analyzed studies used multiple inputs in their models, we can infer that using multiple factors to forecast water demand is likely to result in more accurate and reliable predictions. Additionally, researchers can gain a deeper understanding of the determinants and impact of water demand by analyzing the individual contributions of each factor and their interrelationships.
- A significant proportion of the reviewed studies, specifically 32 out of 37 (or 86.5%), tested their models using real-world data, which suggests that researchers are aware of the importance of testing their models in real-world settings.

Moreover, this emphasizes the importance of bridging the gap between theoretical research and practical application and highlights the need for researchers to consider the practical implications of their work. Additionally, the utilization of real-world data in the majority of the studies reviewed underscores the commitment of researchers to test their models in practical settings. For instance, Xenochristou and Kapelan (2020) collaborated with water authorities in the UK to collect data from 1793 properties, while Pesantez et al. (2018) incorporated hourly water consumption records from smart meters. This trend of engagement with real-world data not only demonstrates researchers' awareness of the importance of practical testing but also emphasizes the need to bridge the gap between theoretical research and practical application. These examples highlight the importance of considering the practical implications of research endeavors, thus enhancing the relevance and utility of the findings for real water companies and municipalities.

- The fact that 38% of the studies present a multi-horizon water demand forecast indicates that there is growing recognition of the importance of considering multiple time horizons in water demand forecasting. This approach allows decision makers to plan and allocate resources more effectively, as they can take into account short-term fluctuations in demand as well as longer-term trends and patterns. Multi-horizon water demand forecasting can also help identify potential water shortages or surpluses in the future and inform decisions around water allocation and infrastructure investments. Overall, incorporating multiple time horizons in water demand forecasting can lead to more accurate and reliable predictions of water demand, which is crucial for effective water resource management.
- The results show that that a relatively small proportion (8%) of the studies that were analyzed used forecasting models with multiple periodicities to forecast water demand. The low percentage indicates that this approach has not yet been widely adopted in the field of water demand forecasting. However, it suggests that there is potential for this approach to improve the accuracy of water demand predictions and a growing interest in modeling water demand at different temporal resolutions. By accounting for multiple periodicities, such as weekly and monthly cycles, a model may be better equipped to capture the various factors that affect water demand at different time scales. For example, weekly patterns may be influenced by factors such as day of the week or weather conditions, while monthly patterns may be influenced by factors such as holidays or seasonal changes. Incorporating multiple periodicities in water demand forecasting models may lead to more accurate predictions, which can help water utilities better manage their resources and plan for future demand. This can ultimately lead to a more efficient use of water resources, which is particularly important in regions where water scarcity is a concern.
- Out of the 37 articles reviewed, 9 did not incorporate benchmarking. Hence, it can be concluded that there is a relatively high number of articles that include benchmarking. Therefore, this suggests that benchmarking is a commonly used practice in water demand forecasting literature. However, the fact that only 75.6% of the articles included benchmarking indicates that there is still a need for more studies that compare different models and their performance and there is still room for improvement in terms of the accuracy and effectiveness of forecasting models. Further research could focus on comparing different models and their performance to identify the most effective approaches for forecasting water demand.
- Behavioral changes in forecasting water demand were considered in approximately 30% of the 37 studies examined. This suggests that behavioral changes are not always taken into account when forecasting water demand. Therefore, it is essential to continue investigating the role of behavioral changes in water demand forecasting and to develop more comprehensive and accurate models that consider these changes. By incorporating behavioral changes into forecasting models, policymakers and water managers can make informed decisions about water resource management, identify opportunities for water conservation and efficiency, and ensure the sustainable use of water resources.
- At least three criteria were used to assess the performance of the proposed models in 22 out of the 37 studies reviewed. This indicates that researchers are using a comprehensive approach to evaluate the accuracy of water demand forecasting models and they recognize the importance of accuracy and reliability in water demand forecasting. Overall, the use of multiple evaluation criteria in assessing water demand forecasting models is a positive step towards more accurate and reliable forecasting. By improving the accuracy of these models, policymakers and water managers can make more informed decisions about water resource management, leading to a more sustainable and efficient use of water resources.

4. Conclusions and Future Work

The objective of this study was to review the methods and models used for forecasting short-term water demand. Using a systematic review process, 37 articles retrieved from reputable databases and top-tier journals were selected and carefully analyzed to identify the gaps and limitations in the current research. The goal of this analysis is to provide insights that can be used by researchers to address these gaps and limitations in future studies.

The insights provided by this review may inspire new research initiatives, and the following recommendations are proposed:

1. Developing machine learning operations (MLOps) approaches for water demand forecasting. MLOps can be used to develop a robust and scalable machine learning solution that can accurately predict water demand.
2. Developing on top of MLOps for different countries and populations. Water resource management authorities can ensure the sustainable management of water resources by customizing the MLOps approach to suit the unique needs of different countries and populations. This tailored approach can result in a more efficient and effective solution for water demand forecasting.
3. Designing automated machine learning (AutoML) method- ologies for predicting water demand. AutoML method- ologies can aid in the creation of predictive models for water demand forecasting by automating the selection of the most suitable machine learning algorithm and hyperparameters, based on the data.
4. Developing forecasting methods for emerging water demand sectors: Multi-horizon water demand forecasting can be extended to emerging water demand sectors, such as the water-energy nexus, water-food nexus, and water-recreation nexus. Future research can investigate the development of customized forecasting methods for these sectors, taking into account their unique characteristics and interactions with other sectors.
5. Exploring the optimal combination of different time horizons: Future research can investigate the optimal combination of different time horizons for water demand forecasting. This can involve identifying the most relevant time horizons for specific regions, sectors, or water uses, as well as evaluating the trade- offs between accuracy and computational complexity.
6. Developing integrated forecasting models: multi-horizon water demand forecasting can be further im- proved by integrating different forecasting models and techniques to leverage the strengths of each method and enhance the accuracy of water demand forecasts.
7. Further investigating the benefits of using multiple periodicities in water demand forecasting models. This could involve comparing the performance of models that incorporate different combinations of periodicities, as well as exploring the impact of incorporating additional variables and factors.
8. Developing user-friendly software or tools that make it easier for water utilities and municipalities to incorporate multiple periodicities in their forecasting models. This can help to promote the adoption of this approach and encourage a more accurate and efficient use of water resources.
9. Exploring and creating new evaluation criteria that take into account additional factors like the effects of extreme weather events, so water demand forecasting models can become more comprehensive and produce precise predictions that can be utilized effectively in managing water resources.
10. Further evaluating the models' performance considering different datasets from different countries to generalize the results. Assessing how well machine learning models perform on various datasets from different countries can enhance the accuracy and usability of water demand forecasting models. Consequently, this can facilitate more sustainable water resource management practices.
11. Several studies included in this review have emphasized the effectiveness of attention mechanisms and convolutional neural networks (CNNs) in assigning higher significance to key features. This enhanced understanding of the importance of key factors and their influence on future water consumption patterns in the studied area. The capability of these techniques highlights their effectiveness and warrants further exploration in the context of water forecasting research.

Availability of data and materials

Not applicable.

Statements and Declarations

Funding

"*No funding was received for conducting this study.*"

Competing Interests

"The authors have no relevant financial or non-financial interests to disclose."

Author Contributions

Safa Ghannam*: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Writing – original draft, Writing – review &editing.*

Farookh Hussain*: Supervision*

Ethics Approval: Not applicable.

Consent to Participate: All authors consent to participate.

Consent for Publication: All authors consent for publications.

References

Abadi, M. L., Samé, A., Oukhellou, L., Cheifetz, N., Mandel, P., Féliers, C., & Chesneau, O. (2017). Predictive classification of water consumption time series using non-homogeneous markov mod- els. *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 323–331.

Al-Zahrani, M. A., & Abo-Monasar, A. (2015). Urban residential water demand prediction based on artificial neural networks and time series models. *Water Resources Management*, *29*(10), 3651–3662.

Bata, M., Carriveau, R., & Ting, D. S.-K. (2020). Short-term water demand forecasting using hybrid supervised and unsupervised machine learning model. Smart Water, 5(1), 1–18.

Bata, M. H., Carriveau, R., & Ting, D. S.-K. (2020). Short-term wa- ter demand forecasting using nonlinear autoregressive artificial neural networks. Journal of Water Resources Planning and Management, 146(3), 04020008.

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.

Blokker, E., Vreeburg, J., & Van Dijk, J. (2010). Simulating residential water demand with a stochastic end-use model. Journal of Water Resources Planning and Management, 136(1), 19–26.

Caiado, J. (2010). Performance of combined double seasonal univariate time series models for forecasting water demand. Journal of Hydrologic Engineering, 15(3), 215–222.

Candelieri, A. (2017). Clustering and support vector regression for water demand forecasting and anomaly detection. Water, 9(3), 224.

Chen, G., Long, T., Bai, Y., & Zhang, J. (2019). A forecasting framework based on kalman filter integrated multivariate local polynomial regression: Application to urban water demand. *Neural Processing Letters*, *50*(1), 497–513.

Coelho, B., & Andrade-Campos, A. G. (2019). Short-term forecasting of hourly water demands-a portuguese case study. *International Journal of Water*, *13*(2), 173–207.

Du, B., Huang, S., Guo, J., Tang, H., Wang, L., & Zhou, S. (2022). Interval forecasting for urban water demand using pso optimized kde distribution and lstm neural networks. *Applied Soft Computing*, *122*, 108875.

Du, B., Zhou, Q., Guo, J., Guo, S., & Wang, L. (2021). Deep learning with long short-term memory neural networks combining wavelet transform and principal component analysis for daily urban water demand forecasting. *Expert Systems with Applications*, *171*, 114571.

Gagliardi, F., Alvisi, S., Kapelan, Z., & Franchini, M. (2017). A probabilistic short-term water demand forecasting model based on the markov chain. *Water*, *9*(7), 507.

Guo, G., Liu, S., Wu, Y., Li, J., Zhou, R., & Zhu, X. (2018). Short-term water demand forecast based on deep learning method. *Journal of Water Resources Planning and Management*, *144*(12), 04018076.

Herrera, M., Torgo, L., Izquierdo, J., & Pérez-Garcıa, R. (2010). Predictive models for forecasting hourly urban water demand. *Journal of hydrology*, *387*(1-2), 141–150.

Kühnert, C., Gonuguntla, N. M., Krieg, H., Nowak, D., & Thomas, J. A. (2021). Application of lstm networks for water demand prediction in optimal pump control. *Water*, *13*(5), 644.

Liu, J., Zhao, L., & Mao, Y. (2019). Bayesian regularized nar neural network based short-term prediction method of water consumption. *E3S Web of Conferences*, *118*, 03024.

Makki, A. A., Stewart, R. A., Beal, C. D., & Panuwatwanich, K. (2015). Novel bottom-up urban water demand forecasting model: Revealing the determinants, drivers and predictors of residential indoor end-use consumption. *Resources, Conservation and Recycling*, *95*, 15–37.

Mouatadid, S., & Adamowski, J. (2017). Using extreme learning machines for short-term urban water demand forecasting. *Urban water journal*, *14*(6), 630–638.

Mu, L., Zheng, F., Tao, R., Zhang, Q., & Kapelan, Z. (2020). Hourly and daily urban water demand predictions using a long short-term memory based model. *Journal of Water Resources Planning and Management*, *146*(9), 05020017.

Nguyen, B. N., Prevedello, C., Cools, M., & Teller, J. (2021). Factors influencing residential water consumption in wallonia, belgium. *Utilities Policy*.

Nunes Carvalho, T. M., de Souza Filho, F. d. A., & Porto, V. C. (2021). Urban water demand modeling using machine learning techniques: Case study of fortaleza, brazil. *Journal of Water Resources Planning and Management*, *147*(1), 05020026.

Odan, F. K., & Reis, L. F. R. (2012). Hybrid water demand forecasting model associating artificial neural network with fourier series. *Journal of Water Resources Planning and Management*, *138*(3), 245–256.

Pesantez, J. E., Berglund, E. Z., & Kaza, N. (2018). Modeling and forecast- ing short-term water demand reported by smart meters:(187). *WDSA/CCWI Joint Conference Proceedings*, *1*.

Rahim, M. S., Nguyen, K. A., Stewart, R. A., Giurco, D., & Blumenstein, M. (2020). Machine learning and data analytic techniques in digital water metering: A review. *Water*, *12*(1), 294.

Rathnayaka, K., Malano, H., Arora, M., George, B., Maheepala, S., & Nawarathna, B. (2017). Prediction of urban residential end- use water demands by integrating known and unknown water demand drivers at multiple scales ii: Model application and validation. *Resources, Conservation and Recycling*, *118*, 1–12.

Romano, M., & Kapelan, Z. (2014). Adaptive water demand forecasting for near real-time management of smart water distribution systems. Environmental Modelling & Software, 60, 265–276.

Sardinha-Lourenço, A., Andrade-Campos, A., Antunes, A., & Oliveira, M. (2018). Increased performance in the short-term water demand forecasting through the use of a parallel adaptive weighting strategy. Journal of Hydrology, 558, 392–404.

Stańczyk, J., Kajewska-Szkudlarek, J., Lipiński, P., & Rychlikowski, P. (2022). Improving short-term water demand forecasting using evolutionary algorithms. Scientific Reports, 12(1), 1–25.

Timotewos, M. T., Barjenbruch, M., & Behailu, B. M. (2022). The as- sessment of climate variables and geographical distribution on residential drinking water demand in ethiopia. Water, 14(11), 1722.

Tiwari, M. K., & Adamowski, J. (2013). Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap- neural network models. Water Resources Research, 49(10), 6486–6507.

Tiwari, M. K., & Adamowski, J. F. (2017). An ensemble wavelet bootstrap machine learning approach to water demand forecasting: A case study in the city of calgary, canada. Urban Water Journal, 14(2), 185–201.

Vijai, P., & Sivakumar, P. B. (2018). Performance comparison of techniques for water demand forecasting. Procedia computer science, 143, 258–266.

Wu, S., Han, H., Hou, B., & Diao, K. (2020). Hybrid model for short- term water demand forecasting based on error correction using chaotic time series. Water, 12(6), 1683.

Xenochristou, M., & Kapelan, Z. (2020). An ensemble stacked model with bias correction for improved water demand forecasting. Urban Water Journal, 17(3), 212–223.

Xu, Y., Zhang, J., Long, Z., & Lv, M. (2019). Daily urban water demand forecasting based on chaotic theory and continuous deep belief neural network. Neural Processing Letters, 50(2), 1173–1189.

Xu, Y., Zhang, J., Long, Z., Tang, H., & Zhang, X. (2019). Hourly urban water demand forecasting using the continuous deep belief echo state network. Water, 11(2), 351.

Zhou, S., Guo, S., Du, B., Huang, S., & Guo, J. (2022). A hybrid framework for multivariate time series forecasting of daily urban water demand using attention-based convolutional neural network and long short-term memory network. Sustainability, 14(17), 11086.

Zhou, S. L., McMahon, T. A., Walton, A., & Lewis, J. (2002). Forecasting operational demand for an urban water supply zone. Journal of hydrology, 259(1-4), 189-202.

Zounemat-Kermani, M., Matta, E., Cominola, A., Xia, X., Zhang, Q., Liang, Q., & Hinkelmann, R. (2020). Neurocomputing in sur- face water hydrology and hydraulics: A review of two decades retrospective, current status and future prospects. Journal of Hydrology, 588, 125085.

Zubaidi, S. L., Gharghan, S. K., Dooley, J., Alkhaddar, R. M., & Abdellatif,

M. (2018). Short-term urban water demand prediction consid- ering weather factors. Water resources management, 32(14), 4527–4542.