Water Practice & Technology

Water Practice & Technology Vol 19 No 7, 2693 doi: 10.2166/wpt.2024.166

A large-scale evaluation of machine learning algorithms in mid-term water demand forecasting

Christos Michalopoulos [a,b,*], Panagiotis Dimas [a], Panagiotis Kossieris [a], Nikos Pelekanos [a] and Christos Makropoulos [a,b]

[a] Urban Water Management and Hydroinformatics Group, Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens (NTUA), Iroon Politechniou 5, Zografou, Athens 157 80, Greece
[b] KWR Water Research Institute, Groningenhaven 7, 3433 PE Nieuwegein, Netherlands
*Corresponding author. E-mail: christos.michalopoulos@kwrwater.nl

CM, 0009-0004-5322-1509; PD, 0000-0002-5605-0846; PK, 0000-0002-5201-3079; NP, 0000-0001-8050-7917; CMak, 0000-0003-0308-4265

ABSTRACT

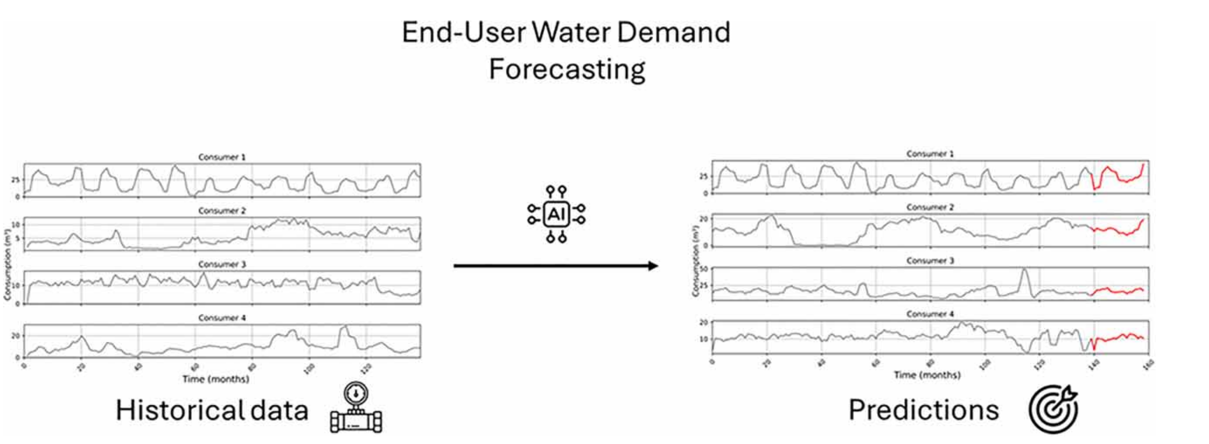
Water utilities currently face challenges in accurately measuring user-end consumption as they have to visit every conventional meter that is installed in their service area. This study explores the use of accessible machine learning models to forecast water demand at two crucial temporal scales, addressing practical needs such as customer billing. To draw meaningful conclusions, these models were tested and trained from actual user-end water consumption data from over 2.1 million customers over a 10-year period. These data were provided by the Water and Sewage Company of Greece (EYDAP). The large-scale experiment included the use of statistical models (ARIMA and SARIMA), deep learning [long short-term memory (LSTM)], and clustering techniques (k-NN algorithm). The model with the best performance was ARIMA, outperforming all the other models including the more complex ones. The LSTM model did not perform as expected in predicting water consumption, it excelled only in predicting the total amount of water consumed. These forecasts can be valuable tools for water utility companies, aiding in tasks such as customer billing and water balance calculations. This paper covers all the necessary steps that must be taken to achieve meaningful user-end water consumption forecasts, from raw data preprocessing to hyperparameter tuning for machine learning algorithms.

Key words: demand forecasting, long short-term memory, machine learning, water consumption

HIGHLIGHTS

- A massive dataset comprising actual user-end water consumption data from over 2.1 million customers.
- Variety of statistical and machine learning models.
- Forecast for accurate billing of customers.
- Explanation of the entire process for generating forecasts, encompassing data preprocessing through hyperparameter tuning.

GRAPHICAL ABSTRACT



INTRODUCTION

The provision of accurate and reliable water demand forecasts is of high importance for operational, tactical, and strategic decisions within drinking water utilities (Gardiner & Herrington 1986). These predictions enable utilities to effectively meet short-term objectives, such as determining the volume of water to be processed in treatment plants and optimally controlling and scheduling water pumping, as well as long-term goals, like quantifying distribution losses (Donkor *et al.* 2014) and setting medium- and long-term leakage reduction goals.

Water meters, by providing consumption data at the end-user level, play a key role in the development of methods and tools for water demand forecasting. Historically, water utilities have relied on mechanical meters to measure residential water use. Data from these meters, often referred to as 'conventional meters', are collected periodically, typically on a quarter or semi-annual basis, from water utilities for customer billing purposes. On the other hand, conventional water meters are characterized by their limitations in supporting comprehensive water management and planning efforts, mainly due to their inability to store data related to consumption patterns and seasonal fluctuations at fine scales (i.e., daily and finer ones).

A remedy to the above is offered by smart water meters, which monitor consumption at granular time intervals (even down to 1 s), enabling also frequent automatic meter readings and the real-time transmission of data to utility providers (Boyle *et al.* 2013; Randall & Koech 2019). The high-resolution data, delivered from such smart meters, have spurred the development of numerous models and methods to forecast (Pesantez *et al.* 2020; Rahim *et al.* 2020) and model water demand at fine time scales (even down to 1 s and at appliance level) (Buchberger & Wu 1995; Alvisi *et al.* 2003; Blokker *et al.* 2010; Kossieris *et al.* 2019). Currently, there is an increasing trend in the installation of such smart metering systems (Di Mauro *et al.* 2021), mainly as pilot cases, while their wider deployment faces various challenges, both technical and financial (Cominola *et al.* 2015).

Currently, most installed meters remain mechanical (single-jet, multi-jet, volumetric) (Li & Gao 2021), and despite the fact that manual meter reading is a labor-intensive process, they are the most popular choice for water utilities with respect to reliable customer billing. Furthermore, they have a proven record of accurately measuring consumption over spans of multiple years and – quite often – decades, making them a trusted option for utility billing. Their simple functioning principles and durability contribute to their extended lifespan, resulting in less maintenance and calibration than their electronic equivalents. Moreover, they are more affordable in advance, making them an ideal choice for utilities, which most of the time seek robust and cost-effective solutions requiring less specialized personnel (Johnson 2007; Stoker *et al.* 2012).

As discussed above, a significant drawback of traditional water meters is the requirement for physical presence to collect the readings, which is a labor-intensive and time-consuming process (Randall & Koech 2019). This often results in water companies struggling to maintain a feasible meter reading schedule due to resource constraints. To tackle the latter predicament, water utilities are attempting to decrease the cost of large-scale deployment of measuring crews by incorporating water demand forecasting methodologies in the billing process (House-Peters & Chang 2011; Ghalehkhondabi *et al.* 2017; Duerr *et al.* 2018; Niknam *et al.* 2022).

In light of the above, our work focuses on a mid-term timeframe, specifically the monthly and quarterly, aligning with typical billing intervals for water companies and the resolution of conventional water meters. Water demand forecasting models can be broadly categorized into two main types: stochastic and deterministic (Donkor *et al.* 2014). Deterministic models consider all factors that influence the predictions and aim to identify causal relationships between these factors and the target variable. On the other hand, stochastic models are often developed based on statistical models adapted to previous time-series (Box *et al.* 2015). In the available literature, extensive research on short-term models has been conducted, while it is very limited in the case of mid- and long-term. Below, we showcase the most common methods delineated in the literature, providing insight into the operational mechanisms of each one. Common stochastic models include the autoregressive (AR), the moving average (MA), the combination of those two with an integration step, the autoregressive integrated moving average (ARIMA), and the seasonal autoregressive integrated moving average (SARIMA) (Box *et al.* 2015; Hyndman & Athanasopoulos 2018). The forecasted values of these models are derived from a function of the previous observations. Due to its simplicity and accessibility via numerical programming languages, the ARIMA model is widely used as a baseline model for comparative purposes (Adamowski *et al.* 2012; Chen & Boccelli 2018; Liu *et al.* 2023).

Another very common model that has shown great potential in time-series forecasting is neural networks (Herrera *et al.* 2010; Chen & Boccelli 2018; Liu *et al.* 2023). Several studies have found that neural network models exhibit the best forecasting performance for water demand time-series, as a large number of parameters are ideal in understanding patterns in the time-series (Chen & Boccelli 2018; Kontopoulos *et al.* 2023; Liu *et al.* 2023). Pu *et al.* (2022) proposed a state-of-the-art method using a hybrid long short-term memory (LSTM)-Convolutional Neural Network (CNN) model on time-series that has been first separated with wavelet decomposition to high- and low-frequency components. With this structure, it outperformed regular machine learning models. Another approach to water demand forecasting is to use exogenous parameters apart from the previous time-series values. For example, as described in Firat *et al.* (2009), other parameters such as total precipitation, average humidity, and inflation rate have been used as input values to machine learning models.

Another frequently employed machine learning tool in the domain of time-series forecasting is the support vector machines (SVM). There are a lot of papers in the available literature that use SVM algorithms to predict future states. A comparative study conducted by Msiza *et al.* (2008) showed that neural network-based architectures have better generalization abilities than SVM models and overall achieved lower performance errors. The timeframe where the models were trained and tested was daily and is comprised of the total daily consumption of the Gauteng Province in South Africa.

An alternative approach available in the literature involves the application of customer clustering for forecasting purposes, whereby a singular predictive model is trained for a group of customers (Huang *et al.* 2019; Kontopoulos *et al.* 2023). Recent empirical evidence indicates that clustering models tend to have an inferior performance compared to conventional forecasting methodologies (i.e., using the previous values of the time-series as predictors) when dealing with analogous data structures (Kontopoulos *et al.* 2023).

Following recent trends in water demand forecasting and embracing the requirements of water utilities, in this work, we provide evidence on the performance of several machine learning approaches with respect to their accuracy and computational efficiency in predicting water demand at monthly and quarter-year scales. Particularly, we train and assess five forecasting models: naïve, ARIMA, SARIMA, LSTM, and k-NN. To obtain concrete results, we conduct a large-scale experiment, exploiting a one-of-its-kind big dataset composed of water meter readings from 2,107,555 water meters with an irregular time resolution that spans from 2 to 4 months, the time-series span from the first month of 2010 to the end of 2020, 10 years in total. The data were collected from EYDAP which is the water utility company that is responsible for watering the whole region of Attica the most populated area in Greece, making it the largest water utility in Greece. This is a key innovative element of the present work, since, to the best of our knowledge, such a large-scale comparative study has never been conducted in the past.

Furthermore, this work addresses another key practical complication which is associated with the irregularity in meter readings collected from water utilities. Specifically, in large service areas (like Athens, Greece), it is technically infeasible for water companies to conduct regular measurements with a constant time step. This causes irregularities in time-series, hampering the direct use of datasets from the statistical and machine learning models. Going beyond the usual practice that entails discarding time-series with missing values or with unusual timesteps, here we develop and test novel techniques to address these issues, ranging from simple interpolation to more complex kernel-based models (Rehfeld *et al.* 2011). Missing values are filled in using an interpolation technique that accounts for seasonality, using as a proxy for seasonality the total exported water from the water treatment facilities each month, as proposed by Billings & Jones (2011).

The paper is structured as follows: The initial section discusses various methodologies employed to address the water demand issue. The methodology section is divided into three parts. Firstly, the problem formulation scrutinizes the nature of the problem at hand. Secondly, it delves into the forecasting models utilized, and lastly, it addresses the evaluation functions employed for method comparison. The subsequent chapter focuses on data preprocessing, elucidating the process of transforming raw time-series with irregular time steps into consistent, meaningful the term timeseries has been adress correctly in the hole documen without missing data. Lastly, the paper concludes with a comprehensive discussion, providing insights into the superior performance of certain methodologies and the reasons for the shortcomings of others. The findings of this paper, derived from the large-scale experiment, will support future analysts in addressing forthcoming challenges pertaining to mid-term water demand forecasting. The results provide valuable insights for advancing research in mid-term forecasting.

METHODOLOGY

Problem formulation

This work studies the water demand forecasting problem in two settings that are differentiated with respect to the temporal step and time horizon of predictions. The two problem settings are illustrated graphically in Figure 1.

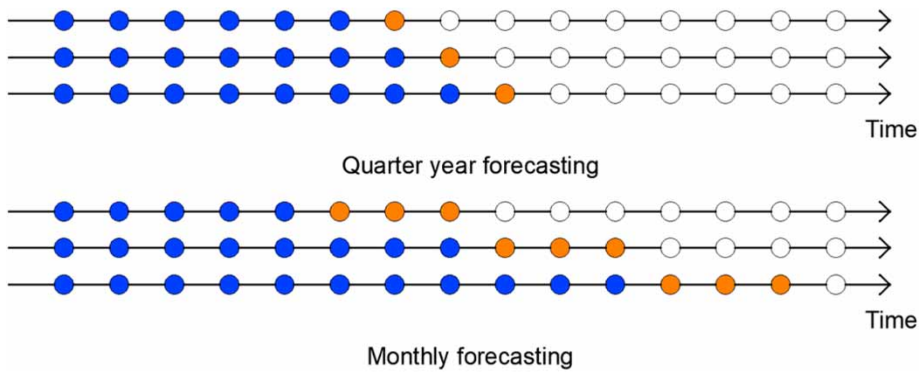


Figure 1 | Training and testing data in a time-series are shown in blue and orange colors, respectively. In the first schematic, each point represents a 3-month consumption, whereas in the below schematics, each point refers to monthly consumption.

Specifically, we provide forecasts on the water demand of each individual user on quarter and monthly time scales. The selection of these time scales, and particularly of the former one, stems from the fact that customer billing is usually based on accumulated consumptions belonging in a quarter time period (~90 days).

In the quarter forecast, the models are developed based on time-series with a 3-month time step to forecast the consumption of the next quarter, and hence the forecast horizon is equal to one step. On the other hand, in the monthly forecast, the models are trained using monthly time-series and provide forecasts for three consecutive months, and hence the forecasting horizon is three steps. The monthly forecasts are then aggregated to the quarter level to allow the comparison of two different forecast settings. In this work, by comparing the two forecast settings on the basis of a large dataset of customers, we attempt to provide empirical evidence on the advantages and disadvantages of the two approaches, and particularly whether there is any added value by first transforming the quarter data into monthly series, and perform predictions, or forecast water consumption directly on the quarter scale. In general, we can argue that the quarter scale entails a simpler forecast problem with a horizon of one step, in contrast to the monthly scale where the predictions required are three, which are then summed up to provide quarter consumption estimates.

The methods discussed in the next section, particularly naïve, ARIMA-SARIMA, and LSTM methods, were used to provide predictions at the single customer scale using only endogenous variables as predictors, and specifically, past consumption measurements of the same end-user. Moreover, here we also develop and assess an alternative forecasting approach that exploits information from end-users of similar characteristics, whose consumption is assumed to be known at the time step of interest, to provide predictions at unmeasured users at the same time step. A schematic representation of this approach is provided in Figure 2.

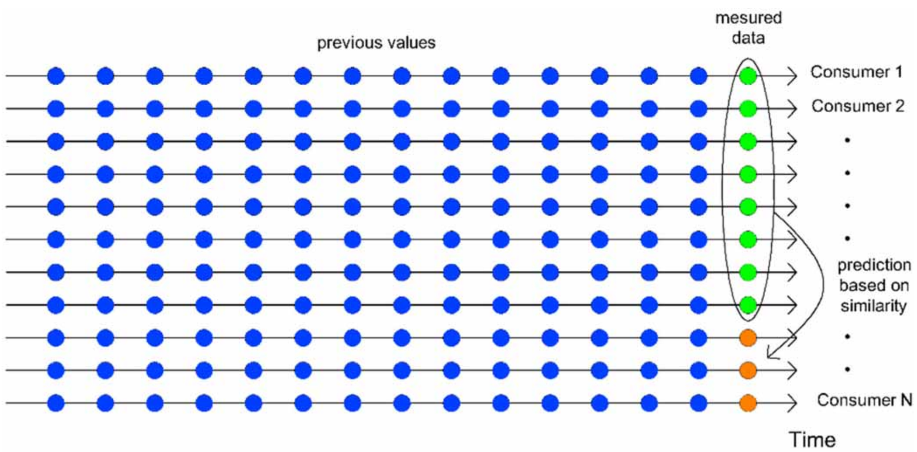


Figure 2 | Collaborative method flowchart. Green corresponds to the customers that have been measured in time, while orange is the value that we want to predict. The similarity is calculated with the previous values of the time-series, which are showcased as blue.

Forecasting models

In this study, naïve, ARIMA-SARIMA, LSTM, and k-NN time-series forecasting methods were selected for further customization and comparison. These methods were chosen based on the belief that they can provide better solutions to the problem, as indicated by the available literature. Each of these models represents a distinct approach to making predictions: ARIMA for a probabilistic approach, LSTM for a deterministic approach, k-NN for a clustering approach and naïve as a baseline model for reference. The selection of models was based on specific criteria. The ARIMA family of models was chosen due to its documented effectiveness in mid-term forecasting as evidenced in the literature, while additionally no normalization is required as ARIMA is a scale-free model meaning that it will be able to capture the behavior for both low- and high-level consumer without any problems. The decision to incorporate LSTM stemmed from its notable performance in short-term forecasting studies. Its intricate architecture enables it to discern complex patterns and mitigate noise through mechanisms like the forget gate. As for the clustering approach, the k-NN approach was selected for its simplicity and suitability in assessing clustering performance. In the context of this work, we have identified and tailored the naïve, ARIMA-SARIMA, LSTM, and k-NN time-series forecasting methods for in-depth investigation and comparative analysis. The selection of these methodologies was underpinned by the literature, which strengthens the belief that they provide superior solutions to the problem at hand. The models were trained and tested in Python 3.10.5.

Seasonal naïve approach

In the field of time-series forecasting, naïve models are typically used as baseline forecasting scenarios, serving as benchmarks for more complex and sophisticated forecasting methodologies. Naïve models, typically use historical values to make predictions, and hence are quite straightforward and rely on minimum data assumptions, providing fast initial estimates. Their main drawback is their limited predictive capacity and extendability of predictions out of the domain of observations. The most frequently utilized naïve process is the *naïve forecast* or *persistence model*, which is grounded on the assumption that the future values will not change in relation to the most recent observed value. Other popular naïve models include the seasonal naïve (assuming that the forecasted value corresponds to the value from the same season in the year that precedes the forecasted time frame) and the mean naïve (assuming that the forecasted value is derived from the average of previous values). Hyndman & Athanasopoulos (2018) have argued that naïve methods are working remarkably well in the fields of economic and financial time-series forecasting. They also suggest that naïve models produce optimal results when the time-series data are following a random walk, and consecutively, they are characterized as random walk forecasts. However, the seasonal naïve method is more useful in the case of seasonal data as analyzed below. Another useful variation of the naïve model (drift method) allows the forecasts to increase or decrease over time. The amount of the aforementioned change (the so-called drift) is set equal to the average change as observed in the historical data. Here, due to the seasonality exhibited in water consumption, we select the seasonal naïve model to develop benchmark forecast scenarios. Sun *et al.* (2011) have also presented a methodology for short-term load forecasting for distribution companies, where they utilized a naïve Bayesian method as their baseline forecasting technique. Kabbilawsh *et al.* (2022) have also explored in detail the applicability of seasonal naïve methods in forecasting seasonal data (i.e., rainfall time-series) by comparing them with three other univariate time-series forecasting techniques, namely Hyndman Khandakar-Seasonal Autoregressive Integrated Moving Average (HK-SARIMA), Non-Stationary Thomas-Fiering (NSTF) and Yeo-Johnson Transformed Non-Stationary Thomas-Fiering (YJNSTF). Furthermore, Donkor *et al.* (2014) have described in detail the capabilities of forecasting through univariate time-series analysis (or what is technically known as extrapolation forecasts in the water industry) noting that they do not account for the effect of exogenous variables such as weather or price (Gardiner & Herrington 1986; Billings & Jones 2011).

The seasonal naïve forecasting method sets each forecast to be equal to the last observed value from the corresponding season. For example, the prediction for a winter quarter-year will be equal to the previous value of last year's winter quarter-year. The equation describing the naïve approach is as follows:

$$\hat{y}_{T+h|T} = y_{T+h-km} \tag{1}$$

where $\hat{y}$ is the forecast, y_{T+h-km} is the observed value, m is the seasonal period, k is an integer value given by $(h-1)/m+1$ with h being the timesteps from the last known value (Hyndman & Athanasopoulos 2018). It is highlighted that this approach stands as the prevailing choice of water utilities for customer billing in cases where measurement data are unavailable, primarily because of its inherent practicality.

ARIMA-SARIMA

These models are extensions of the basic ARMA (AutoRegressive Moving Average) models (Hyndman & Athanasopoulos 2018) and incorporate differencing to handle non-stationary data and seasonal components for time-series with inherent seasonality. In the field of water consumption forecasting such models have been widely used in the short-term scale. Indicative successful past implementations of such models are (i) the study of Maidment & Miaou (1986), who studied the performance of the ARIMA model in daily demand forecasting of nine cities in the USA, (ii) Smith (1988), who developed a time-series model of daily municipal water use and produced a water use forecast through an updating step in which a revised estimate of current mean water use is computed, and (iii) Wong *et al.* (2010), who attempted to forecast daily urban water consumption in Hong Kong by deploying Fourier series, which can adequately represent seasonal cycles. On the other hand, more recent works (Pandey *et al.* 2021) argue that such models encompass inherent limitations in water demand forecasting and, hence, suggest hybrid methods to improve their accuracy. Nonetheless, since ARIMA and SARIMA models can capture the seasonality of time-series and are showcasing relative convenience and simplicity in their parameterization, they are still holding up as a stronghold of comparison with other innovative methods.

ARIMA(p,d,q) model is the combination of autoregressive (AR), integrated (I), and MA models. The autoregressive component AR(p) represents the regression terms, where p is the number of lagged observations used. The moving average component MA(q) represents the MA terms, where q is the number of lagged error terms included. The Integrated component is denoted by d, which represents the differencing order required to make the data stationary.

There are numerous ways to calculate the coefficients and the white noise variance of the ARIMA(p,d,q) model in the literature like maximum likelihood estimator, least squares method, and moments method (Shumway & Stoffer 2017). The *statsmodels* library by default utilizes the maximum likelihood estimator to find the optimal parameters (Perktold *et al.* 2023).

To capture seasonal patterns, the SARIMA model has been developed, which takes into consideration the seasonality of the problem. The additional parameters of the SARIMA model are $(P,D,Q)_m$ which are the seasonal order terms, while m denotes the seasonality. The model is expressed as SARIMA$(p,d,q)(P,D,Q)_m$. For this research, the seasonal order was $m = 4$ for quarter-year analysis while when performing monthly analysis, $m = 12$. Statistical models encompass a degree of uncertainty, and this inherent attribute can introduce a complicating factor in the assessment of model effectiveness. To address this challenge, numerous statistical methodologies have been devised, with one of the most prevalent being the utilization of the Akaike Information Criterion (AIC) (Stoica & Selen 2004). AIC can be calculated as follows:

$$\mathrm{AIC} = 2k - 2\ln(\hat{L}) \tag{2}$$

where k is the number of model parameters and $\hat{L}$ is the maximized value of the likelihood function. For the ARIMA and SARIMA, the *statsmodels* library was used while for AutoARIMA, the *pmarima* module was utilized (Perktold *et al.* 2023; Pmdarima 2023).

Long short-term memory

LSTM models are a type of recurrent neural network (RNN) that can capture long-term dependencies in time-series data and are specifically designed to address the challenges faced by standard RNNs (Hochreiter & Schmidhuber 1997). They have been widely used in a variety of time-series forecasting tasks, including water consumption forecasting. LSTM models are particularly powerful when dealing with temporal patterns characterized by increased complexity and sequences with irregular or non-linear trends. Pu *et al.* (2022) have proposed a new method which combines the time-frequency decomposition of wavelet multi-resolution analysis (MRA) with an advanced deep-learning model. Their approach exhibits improved performance than traditional models in its predicting capability, demonstrating improved accuracy and stability in both single and multi-step predictions. Hu *et al.* (2019) constructed a hybrid model which incorporates a convolutional neural network and a bidirectional long- and short-term memory network to evaluate the effects that climatic factors have on the usage of urban water. To quantify the results of their models they used a multitude of baseline models, including among others, the typical LSTM architecture. Kühnert *et al.* (2021) examined the applicability of LSTM models in water demand predictions as well as control decisions (i.e., optimal pump control) and performed an extensive comparison against other popular methods used by water utilities. They manage to outperform other methods by integrating additional exogenous information (i.e., the day of the week or national holidays), while they show exceptional performance in replicating or even improving the current operational policies of the utilities examined.

Each LSTM cell contains three different gates: the input gate, the forget gate, and the output gate. This architecture enables the handling of longer sequences without encountering issues such as gradient vanishing or exploding. A detailed presentation of the LSTM cell and an in-depth explanation of its components can be found in recent machine learning books (e.g., Salem 2022), and due to this, it is not given here.

Tuning LSTM models can be challenging due to the numerous hyperparameters that need to be optimized (Abbasimehr *et al.* 2020). In this study, the lag (number of previous data points fed to the LSTM) depends on the research timescale of choice. For the monthly scale, the input vector consists of 12 previous values to gather adequate data from 1 year of data, while in the quarter-year scale, the input vector contains 8 values which correspond to the last 2 years of data. For the other parameters, there is a lot of available guidance from other researchers (Song *et al.* 2020) suggesting that by increasing the number of stacked layers smaller errors can be achieved with the cost of execution time. Furthermore, it states that the error slowly decreases until 700 epochs, after which the error plateaus. The hyperparameter with the most significant influence on the error reduction is the number of neurons, and finding a good balance is crucial to avoid underfitting or overfitting issues. Finally, it must be noted that each customer's consumption is normalized to the interval [0,1] to avoid instability issues. By reducing the scale of the given data, the number of corresponding weights that have to be optimized is also expected to be smaller making the training process more efficient (Song *et al.* 2020). The Python module that was used here was *TensorFlow* 2.15.0 (TensorFlow 2023).

Matrix completion (k-nearest neighbors)

k-Nearest Neighbors (k-NN) is a widely known and applied supervised learning method that was first proposed by (Cover & Hart 1967). Its main functioning mechanism considers the labels or the values of the k nearest points, with the most assigned label being the output of the algorithm, where k is the hyperparameter of the model. Because of the nature of the model, k-NN is primarily employed in classification problems. Oliveira & Boccelli (2017) implement a variation of the algorithm in short-term water demand forecasting. They used the method of delays to transform the time-series data into vectors. However, the method was not able to outperform the seasonal autoregressive model. Another approach to using k-NN methodology for time-series forecasting has been performed by Antunes *et al.* (2018). In this work, the time-series were converted into vectors containing four consecutive values, and the last three values were used to make a prediction for the fourth value.

Collaborative filtering methods, including matrix completion algorithms, have been widely adopted by researchers to address missing data in various domains, such as recommendation systems and time-series data correction (Schafer *et al.* 2007; Ma *et al.* 2019). Because of the nature of the data available (water consumption of individual multitype users), the unmeasured customers could be forecasted through matrix completion algorithms, as in the work of Ma *et al.* (2019). Shortly, the main context is to treat unmeasured customers by having their last value missing and the matrix completion algorithm detects the most suitable value as a forecast. Numerous matrix completion algorithms have been detailed in the literature, as comprehensively reviewed by Ramlatchan *et al.* (2018). However, for the sake of simplicity, this study focuses exclusively on the k-NN algorithm, widely recognized as the most popular choice for this specific application. Within the context of this algorithm, the forecasting process is executed as follows:

$$\hat{y} = \frac{\sum_{i=1}^{k} y_{\text{known}|i}}{k} \tag{3}$$

where $y_{\text{known}|i}$ is the value of the i nearest neighbor from the set of customers with measured consumption, and k is the number of nearest neighbors. The value of k, which is the only hyperparameter in this model, is selected through exhaustive search. The performance of the model is evaluated for various numbers of neighbors, and the number of neighbors that outputs the lowest error is selected. A flowchart of this process is described in Figure 2. The problem with this procedure is that it treats every neighbor in the same manner, despite its similarity to other users. To overcome this limitation, it is wise to introduce some type of similarity coefficient in the equation. One popular example that is extensively used in recommendation systems is the Pearson correlation (Schafer *et al.* 2007), but for our application, the existence of customers with very different base consumption may cause instability issues. For that reason, we adopt the cosine similarity as it has scale-free properties (Cui 2017). The cosine similarity is expressed in Equation (5), where u_i and v_i denote the two vectors that we want to find the similarity. In the context of this problem, the vector $\vec{u_\iota}$ corresponds to customers with known last consumption, while the vector $\vec{v_\iota}$ pertains to the customers for whom we seek to predict the consumption.

$$\hat{y} = \frac{\sum_{i=1}^{k} \mathrm{sim}(u_i, v_i) * y_{\mathrm{known}|i}}{\sum_{i=1}^{k} \mathrm{sim}(u_i, v_i)} \tag{4}$$

$$\mathrm{sim}(u_i, v_i) = \cos(u_i, v_i) = \frac{\vec{u_\iota} \cdot \vec{v_\iota}}{|\vec{u_\iota}| * |\vec{v_\iota}|} \tag{5}$$

To simulate the problem that water utilities must deal with, the data are split into two subsets: The measured, where the working crew has managed to record the consumption, and the unmeasured, where the crew did not perform a measurement. For the Water and Sewage Company of Greece (EYDAP) the measured meters account for 90% with the unmeasured meters accounting for 10%.

Evaluation protocol

The assessment of model performance is performed through the utilization of the following metrics:

1. Mean absolute percentage error (MAPE): This metric was chosen because it allows a fair comparison among consumers with relatively high and low consumption due to the property of being scale-independent. MAPE is expressed in Equation (6), where y_t is the actual value and $\hat{y}_t$ is the forecasted value.
2. Forecast bias: This metric assesses the model's tendency to overestimate or underestimate the forecasts in total. The closer this metric is to zero, the more accurate the model is in predicting the total consumption. Forecast bias is expressed in Equation (7).

$$\mathrm{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100 \tag{6}$$

$$\mathrm{Bias} = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t) \tag{7}$$

These metrics are intended to showcase the performance of each model using different approaches. If the water utility company, for example, needs to make a prediction for billing their customers, they should consider models that perform well on the MAPE metric. On the other hand, if the company aims to calculate the yearly total balance of exported and imported water, the models that performed better on the bias metric should be considered.

DATA AND TEMPORAL REGULATION

To assess and compare the performance of different forecasting algorithms, we employ a one-of-a-kind large dataset comprising water consumption records from 2,107,637 customers, located in the greater metropolitan area of Athens (Attica, Greece). The dataset was provided by the Water and Sewage Company of Greece (EYDAP), the water utility serving the city of Athens, and the records are 10 years long. This dataset includes consumption data from all EYDAP's customers, including both domestic and industrial sectors. The records comprise water volumes measured from conventional water meters and logged manually by the staff of EYDAP, for billing purposes, with an approximate frequency of a 3-month period (i.e., quarter time step). The extensive geographical coverage of the city of Athens and logistic constraints make the compliance with a regular recording schedule practically impossible, and due to this, the data records exhibit irregular time intervals, ranging in average between 80 and 110 days.

This poses extra difficulties in the use of such data records and their insertion in forecasting methods and tools. For this reason, the raw dataset needs to be preprocessed so as to establish common regular time steps for all records. Literature provides several approaches for smoothening irregular time-series, with the simplest one being the interpolation-type methods (Rehfeld *et al.* 2011). In this work, for the sake of simplicity, we use such an interpolation method, with an intuitive variation incorporated to account for seasonality.

Particularly, we develop a new temporal regularization method that transforms the irregular, approximately quarter time step, raw records into records with regular monthly time steps. A visual representation, along with an example, of the whole regularization process for a single measurement can be seen in Figure 3.

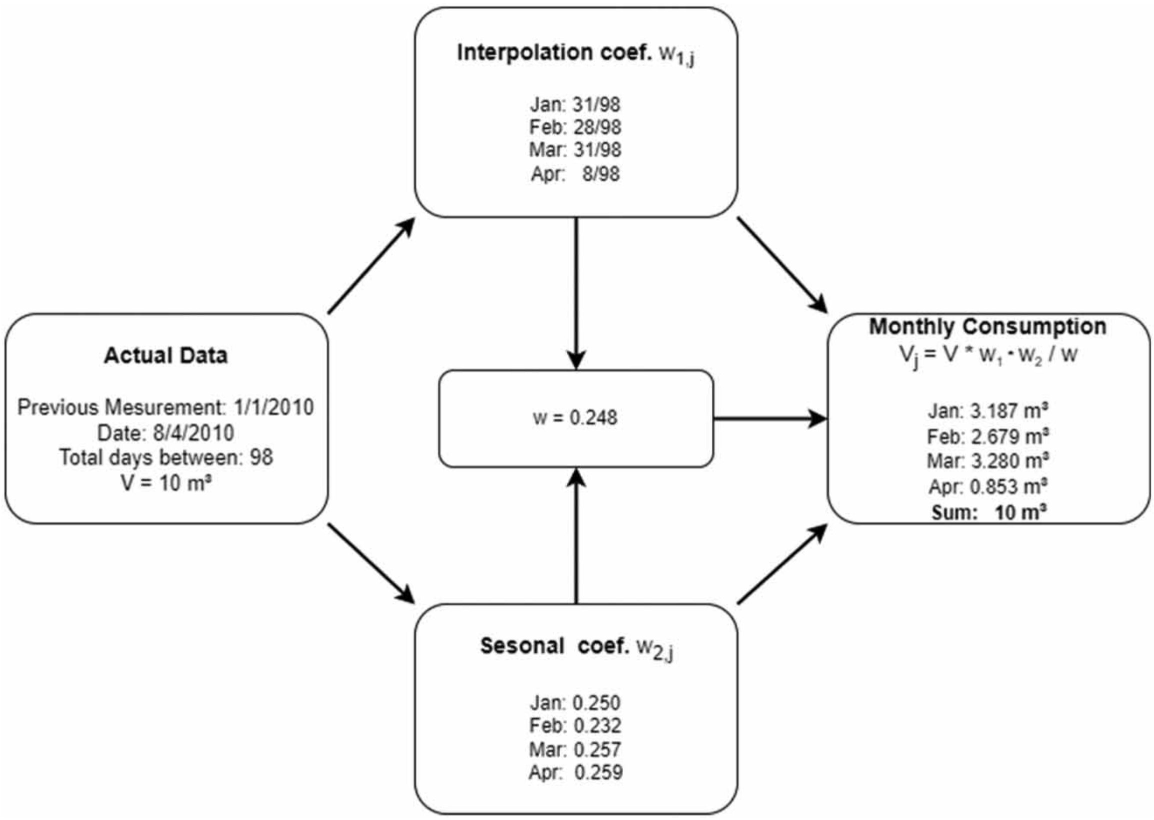


Figure 3 | Data preprocessing flowchart for a single measurement.

Specifically, the target consumption value of month j, V_j, can be obtained from the quarterly consumption value V according to the formula:

$$V_j = \frac{V * w_{1j} * w_{2j}}{w} \tag{8}$$

$$w = \sum_{j=1}^{J} w_{1j} * w_{2j} \tag{9}$$

where w_{1j} and w_{2j} are the interpolation and seasonality coefficient, respectively, while w is a normalizing constant that ensures that the sum of V_j for the j months of a single measurement adds to V. The interpolation coefficient, w_{1j}, is given by the ratio between the number of days of month j and the number of days of the measuring period. Furthermore, the seasonality coefficient, w_{2j}, expresses the month-by-month variation in water consumption. At the water system level, this can be easily obtained from the monthly pattern of water supplied from the water treatment plants (Billings & Jones 2011). Specifically, w_{2j} can be obtained as the ratio of the water supply of month j, c_j, over the total annual supply

$$w_{2j} = \frac{c_j}{\sum_{j=1}^{12} c_j} \tag{10}$$

The values of w_{2j} for the case of the under-study region are given in Figure 4, depicting the significant seasonal variability of water consumption in Athens (Greece).

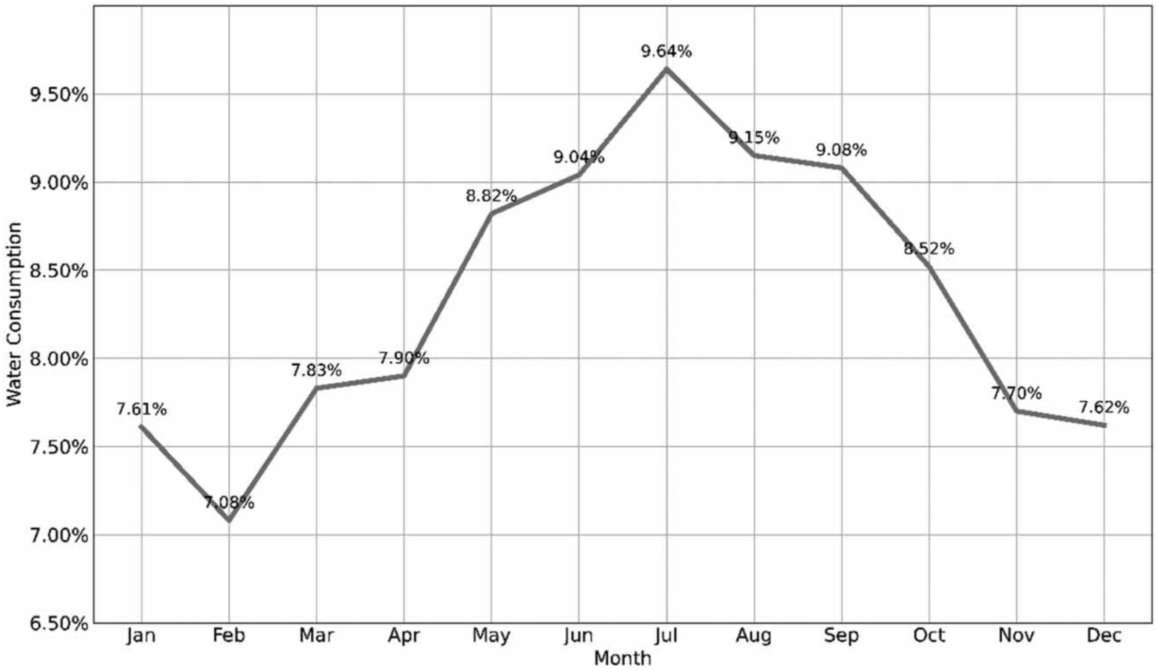


Figure 4 | The values of the seasonality coefficient w_{2j} for the under-study region.

RESULTS AND DISCUSSION

To assess and compare the performance of different algorithms, we use the dataset composed of 2,107,637 time-series of water consumption with monthly regular timestep, as obtained after the preprocessing that is presented in the above section. The dataset was further analyzed to exclude the time-series with nonrealistic values or the strong presence of zero values, which indicates the absence of water consumption. Specifically, all time-series that include zero values at their 12 last values were excluded from the analysis, leading to a dataset composed of a total of 1,535,813 records. This choice was made not only to prevent the calculation of a useless prediction, but also to avoid instability problems that the MAPE metric demonstrates when the actual value is equal to zero.

For this research, the ratio between the measured and unmeasured consumers is set to 80% of consumers that we know their last value while 20% are the consumers that we want to predict. Pure time-series models are trained on the unmeasured subset. Furthermore, the unmeasured time-series data are subdivided into distinct train and test sets to further facilitate the assessment of each model's performance. All models were trained and tested with the same data for a fair comparison. The whole process is described in Figure 5.

The models that operate without requiring any parameter tuning comprise the naïve and AutoARIMA models. The optimization of parameters for each model is accomplished through an iterative process of trial and error. Specifically, in the case of the ARIMA and SARIMA models, exhaustive exploration of parameter combinations is undertaken. Since the time-series becomes stationary after a single integration step ($d = 1$), parameter tuning is exclusively focused on the remaining parameters. For the ARIMA model, the parameters p and q are tested, while for the SARIMA model, the seasonal parameters P and Q undergo similar examination. In contrast, the LSTM model's optimization process entails the exploration of a different number of epochs and neurons within the

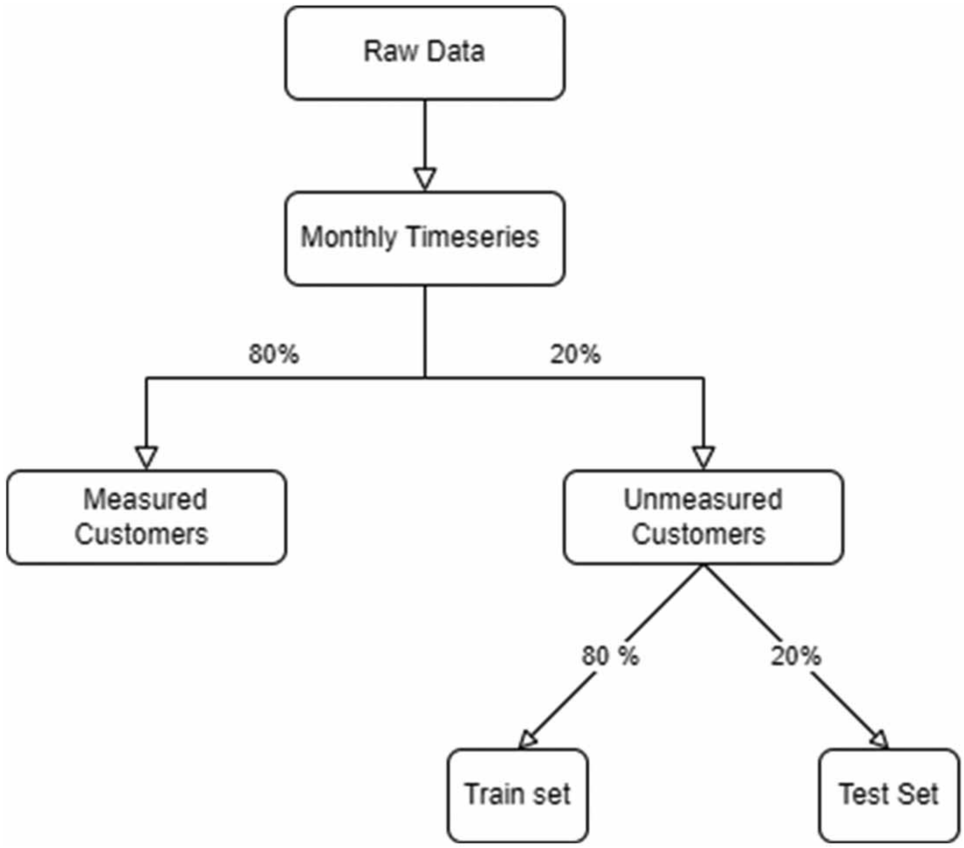


Figure 5 | Every step of data preprocessing from raw data.

LSTM cell. Finally, the sole hyperparameter within the k-NN method is the number of neighbors (k), which is optimized by assessing model performance across different values. The error in the test set of each model with respect to each parameter is summarized in Table 1 and Figure 6.

Table 1 | MAPE score for every combination of parameters for ARIMA, SARIMA, and LSTM models. The optimal performance is indicated with Bold

ARIMA					SARIMA				
p \ q	0	1	2	3	P \ Q	0	1	2	3
0	0.214	0.215	0.225	0.238	0	0.220	0.225	0.232	0.240
1	**0.213**	0.219	0.230	0.238	1	0.221	**0.219**	0.230	0.235
2	0.218	0.220	0.221	0.239	2	0.230	0.224	0.235	0.231
3	0.219	0.219	0.225	0.237	3	0.225	0.225	0.233	0.237

LSTM		Number of Neurons			
		25	50	75	100
Number of epochs	50	0.370	0.360	0.332	0.370
	100	0.340	0.415	0.386	0.356
	200	0.331	0.302	0.295	0.311
	300	0.320	0.295	**0.283**	0.320

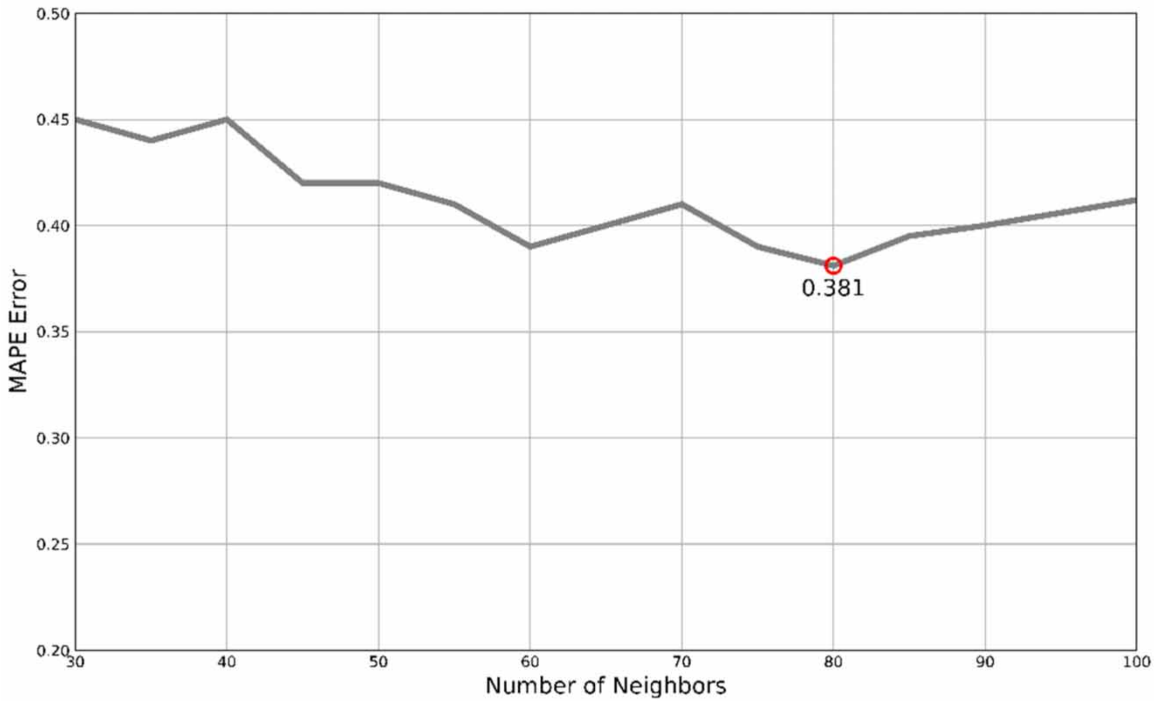


Figure 6 | MAPE score against the number of neighbors in the case of k-NN method.

It is observed that the ARIMA and SARIMA models are relatively insensitive to different parameter combinations, since the error stays practically the same for different parameters. The same pattern is observed by Liu *et al.* (2023). Due to the limited performance of ARIMA models when having high parameter values, the maximum of the values in this study is set to 3 and no higher combinations are examined, so that the total run time is reduced and convergence problems are avoided. The optimal parameter combination is determined to be ARIMA(1,1,0), scoring the lower percentage error among the other combinations. The SARIMA model performs slightly worse than the ARIMA with the best parameter set being (0,1,0)(1,0,1). From this result, it can be concluded that the seasonal part of the SARIMA model did not improve the overall performance of the forecast. One possible reason for this behavior could be the existence of data during the COVID-19 pandemic, which may influence the seasonality patterns as derived from the previous years, since water usage has been remarkably impacted during this period (Kalbusch *et al.* 2020).

For the LSTM approach, it seems that the number of epochs increases the performance of the model for the whole set that we are investigating. This means that the model does not demonstrate overfitting problems. However, with this approach the computational time also increases significantly, making the total run time technically infeasible. For that reason, the maximum number of epochs that are used is 300, which achieves optimal performance. The other parameter that was investigated was the size of the inner LSTM cell varying from 25 to 100. Higher values were not considered due to the limiting length of the time-series. The forecasting is performed with the help of a fully connected dense layer with three neurons, each corresponding to the monthly forecast of each quarter-year. An iteration is performed for each quarter-year and four iterations are required to cover the whole test set. The iterations are performed with an activated cell state for enhanced information flow between the LSTM cells. The number of neurons that achieve the highest performance is 75. In every instance, the models are trained with the Adam optimizer. Also, the dropout percentage was set constant and equal to 10%. The Loss function that the models were trained upon is the mean squared error, as it's the most widely used and established baseline for regression problems.

For the k-NN model, finding the optimal number of neighbors is a remarkably simpler process. The optimal number can be found by finding the global minimum in the corresponding graph that illustrates the error with respect to the number of neighbors. The minimum error, as shown in Figure 6, is achieved with 80 neighbors ($k = 80$). When considering larger numbers of neighbors, the model is prompted to overfitting problems.

To better visualize the results of each model, boxplot (box-and-whisker) diagrams are used for MAPE on a quarterly and monthly time scale Figure 7. As for the bias of the models, a bar plot is used to illustrate the results as shown in Figure 9. The only model for which the architecture had to be shifted for a quarter-year prediction was the LSTM. The sole modification involved reducing the size of the fully connected neuron layer to one, resulting in a single forecast per quarter-year.

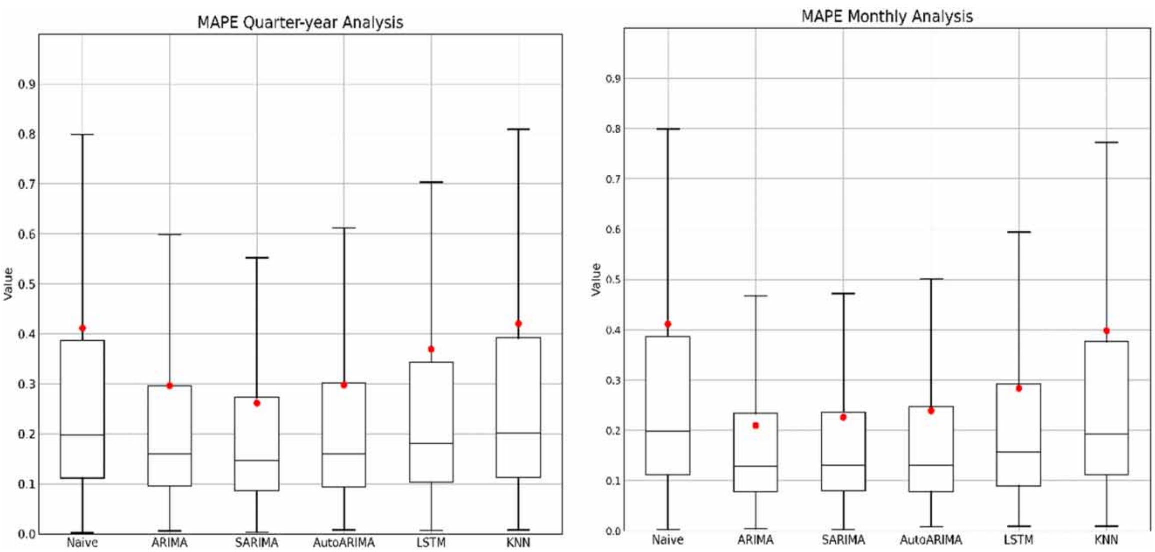


Figure 7 | MAPE performance of each model for a quarter-year and a monthly analysis.

The results of the large-scale analysis can be seen in Figure 7. All the models performed better on a monthly time scale in contrast to the quarter-year analysis. This indicates that the models demonstrated improved adaptability to the monthly data compared to the findings presented by Kofinas *et al.* (2014). The finer timescale results in lengthier time-series in contrast to working with a quarter-yearly scale. The larger number of values in the time-series enables the models to optimize their parameters more effectively. Liu & Lin (1991) also suggest that a monthly timeframe results in more accurate predictions. For the naïve method, it was expected that the performance would remain consistent for both monthly- and quarter-year analysis as the timestep does not influence the model performance. For the monthly scale, the best performing model was the ARIMA with an error close to 21% followed very closely by SARIMA. These two models have similar variance ranges as can be seen in Figure 7. A very interesting observation is that the AutoARIMA model does not outperform ARIMA and SARIMA models. This can be attributed to how AutoARIMA selects optimal parameters. Unlike ARIMA and SARIMA, which optimize parameters by achieving lower mean absolute percentage error (MAPE), AutoARIMA chooses optimal parameters by minimizing the AIC. The mean error for the LSTM model was 28%, placing it right after the statistical models in terms of performance. k-NN model performance was very close to that of the naïve method achieving a mean error of 40 and 41%, respectively.

For the quarter-year scale, the SARIMA model outperformed the rest, with an error of 26%. It seems that the model is able to capture yearly seasonality effectively, outperforming both ARIMA and AutoARIMA, which achieved the same error of 30%. LSTM showcased a 37% error, while k-NN was outperformed by the naïve method with a mean error of 42%. The models with the most significant error reduction due to the finer scale were the LSTM and the ARIMA with a difference of over 8%. This was expected, especially for LSTM, due to the substantial number of parameters it requires in order to be tuned. Longer time-series provide more training data, resulting in a better forecasting model.

LSTM and ARIMA demonstrated a greater reduction in error with respect to the finer time scale, with a difference of over 8%. This was expected, especially for LSTM, due to the substantial number of parameters it requires in order to be tuned. The longer the time-series are, the more training data are available for training resulting in a better forecasting model. All models exhibit a significant difference between the average and the median value in their error distribution. This discrepancy is mainly attributed to the presence of customers experiencing significant temporal fluctuations, which can be caused by various factors, including changes in residence, or changes in the consumption type (i.e., a small industrial customer replacing a domestic one), with the accompanying varying water usage requirements. An example of those time-series is illustrated in Figure 8. By excluding these customers from the dataset, the average performance of each model will increase, while more symmetrical error distributions will be formed. The average and median performance of each model is shown in Table 2 and Table 3.

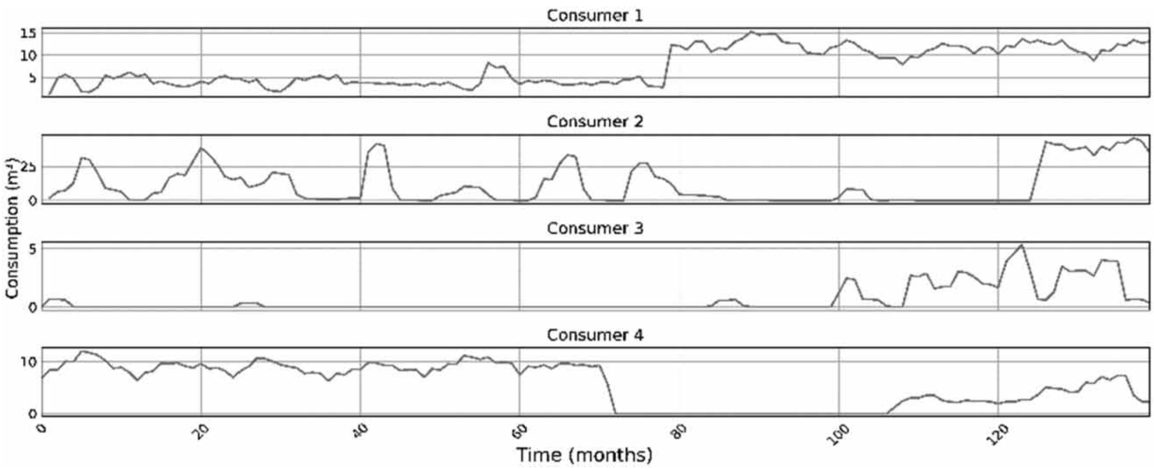


Figure 8 | Examples of abnormalities among consumers.

Table 2 | Average MAPE performance for monthly and quarter-year analysis

Models		Naïve	ARIMA	SARIMA	AutoARIMA	LSTM	k-NN
MAPE	**Quarter**	0.411	0.296	0.261	0.297	0.369	0.420
	Monthly	0.411	0.209	0.226	0.239	0.283	0.398

Table 3 | Median MAPE performance for monthly and quarter-year analysis

Models		Naïve	ARIMA	SARIMA	AutoARIMA	LSTM	k-NN
MAPE	**Quarter**	0.197	0.160	0.146	0.160	0.180	0.201
	Monthly	0.197	0.128	0.131	0.130	0.157	0.192

Although the MAPE metric favors the monthly time scale analysis, the bias metric tends to favor the quarterly scale analysis, with all models achieving values closer to zero. Similar to the MAPE error, the bias metric is the same across the two timeframes in the naïve model for the same reasons, achieving a bias of 0.683 m^3 as can be seen in Figure 9. For the monthly timescale, the model that achieved the best performance is the AutoARIMA with an average bias of 0.498 m^3. Surprisingly, the naïve model achieved the second-best performance with a bias of 0.683 m^3, followed by LSTM, SARIMA, and ARIMA with almost similar performances. The k-NN model performed the poorest by a significant margin, outscaling all the performances of other models. For the quarter-year scale, the model with the best performance is the LSTM scoring 0.018 m^3, closely followed by ARIMA and SARIMA. The k-NN model shows a much better performance on the monthly scale compared to the quarterly scale, indicating a difference in scale factors. Nonetheless, k-NN still falls short of matching the performance of other models.

It is observed that all models underestimate the consumption, with the best one being LSTM. This outcome was unexpected, considering its poor performance on the MAPE metric. Considering that, we can conclude that the forecasts are symmetrically oscillating around the correct value, effectively balancing the bias error. The worst-performing model for both instances is the k-NN model, achieving the lowest score by a significant margin. All the models seem to overestimate the consumption except for the SARIMA model, which exhibits a negative

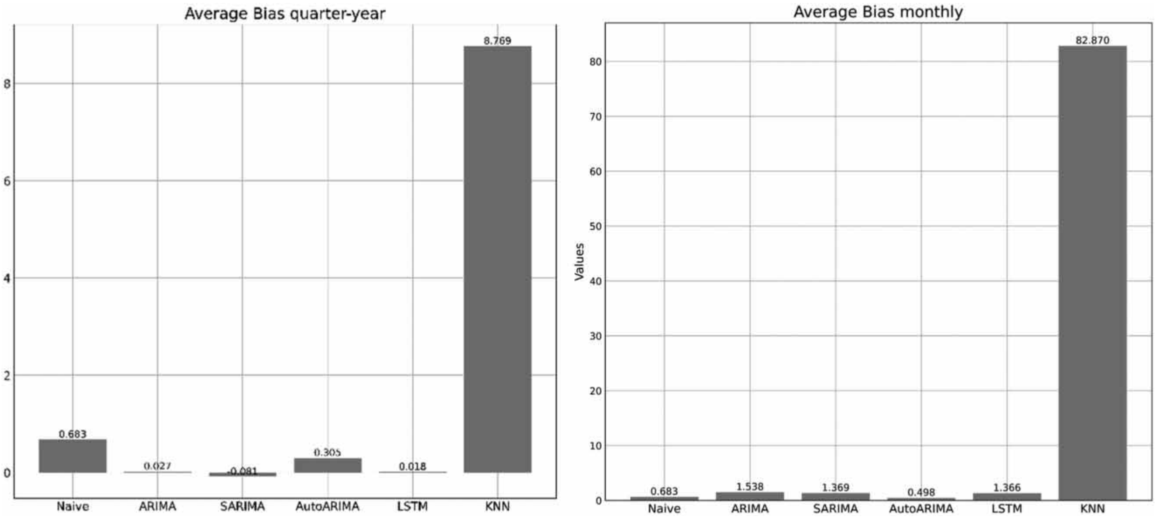


Figure 9 | Average bias performance for a quarter-year and a monthly analysis.

bias error. This behavior holds significant value as it can provide a precise estimate of the total water that has been consumed by the customers but not recorded, making it a credible tool for calculating the total water balance on both monthly and quarterly year scales.

Finally, for the monthly scale, we measure the performance of each model on a quarterly scale. This allows us to detect if the models exhibit any preference in a specific season or if certain seasons exhibit more predictable data patterns than others. The results shown in Figure 10 confirm the results of Figure 7, illustrating that the best model through every season is the ARIMA. All the models exhibit the same pattern in terms of performance in

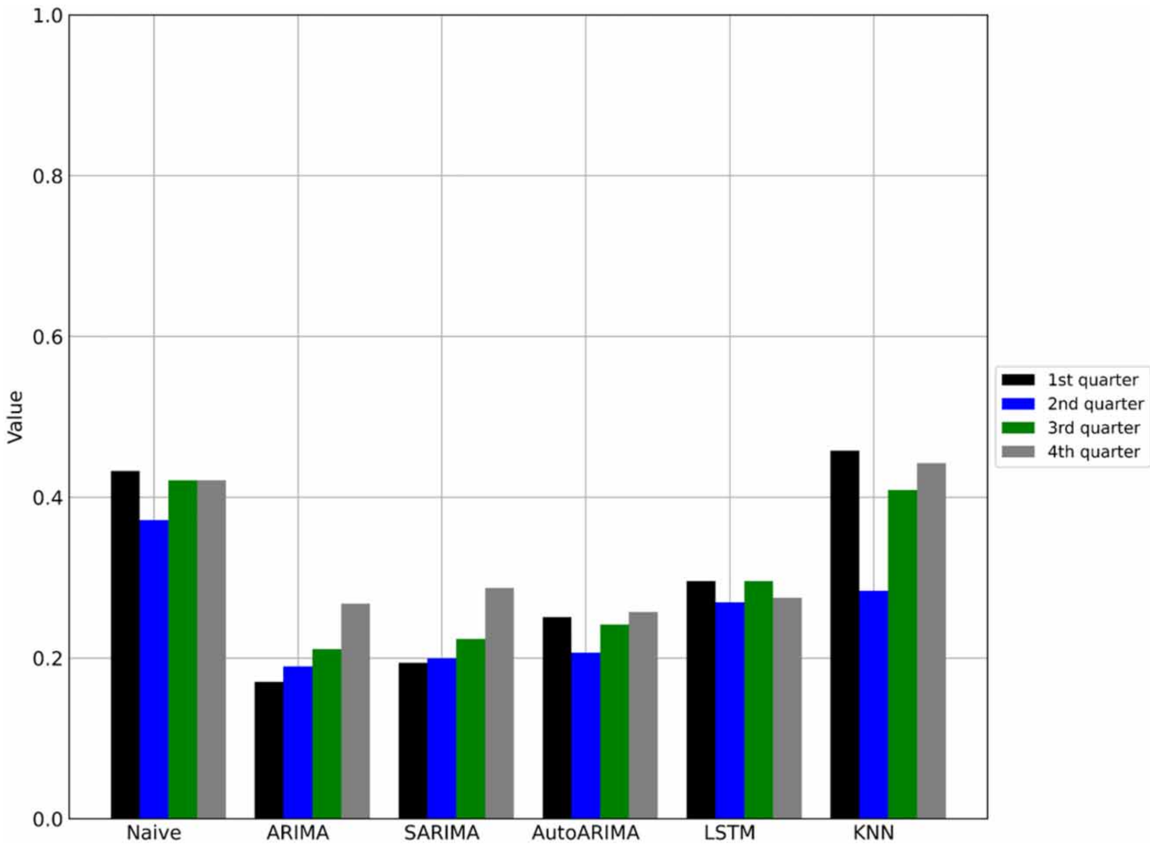


Figure 10 | Model performance per season.

the quarterly time scale. The most predictable period was the second quarter of the year, when almost all models are showcasing the best performance. Despite the poor performance of k-NN in previous metrics, it performed equally good as LSTM in the second quarter, leading to the conclusion that the second quarter is the most predictable. All statistical models exhibited nearly identical patterns, except for AutoARIMA's first quarter. Conversely, the most challenging quarter to predict was the fourth, with almost all models recording their poorest performance. The clustering method did not outperform the baseline naïve method. The same issue was addressed in the study of Kontopoulos *et al.* (2023) where the clustering approach did not outperform the regular models. One explanation for the poor performance of the clustering method is the large amount of measured data. The k-NN algorithm struggles to find meaningful neighbors from the large amount of data used in the training procedure. Another problem that this algorithm must deal with is the curse of dimensionality. The latter is a well-documented problem that k-NN models suffer from when dealing with high dimensions of data (Kouiroukidis & Evangelidis 2011). One possible way to deal with it is by reducing the dimensions of the vectors with a proper function without losing valuable information. There are numerous available methodologies each with its own advantages and disadvantages (Carreira-Perpiñán 1997).

CONCLUSIONS

This study presents a comparative evaluation of different machine learning-based models to forecast water consumption at the end-user level at monthly and quarter-temporal scales. These two scales of forecast are of high practical interest for water utilities, since they are the time periods for which customer bills are typically issued. To provide concrete results and empirical evidence on the performance of different algorithms and approaches, we use a one-of-a-kind large dataset of water consumption records collected from 2,107,555 customers in Athens (Greece) with a total length of 10 years and a large-scale experiment is performed. The raw quarterly-based data, which typically becomes available with irregular time steps, were first transformed into a regular monthly time-series, following a new methodology, tailored to the peculiarities of the water consumption process. The transformed time-series served as inputs to train the ML models. The models were evaluated on both monthly and quarterly timescales. The findings revealed that, for the MAPE metric, all models achieved lower errors for the monthly timescale. The model that outperformed all the others was the ARIMA(1,1,0), with the other statistical models not being far behind. On the other hand, the bias metric seems to favor the quarter-year scale, with all models performing better in this timeframe. While the ARIMA model could not outperform the LSTM, the difference was marginal, making ARIMA the model with the best overall performance. Exhaustive parameter testing revealed that, for the statistical models, parameter selection does not have a meaningful impact on model performance. The LSTM model did not perform as expected, scoring higher errors than every stochastic model. The poor performance in each individual forecasting can be attributed to the existence of a small training set relative to the number of parameters that need to be trained, making it unsuitable for mid-term forecasting. In contrast, models with simpler architecture seemed to benefit from the smaller size of the time-series. The k-NN proved to be the least effective, showing errors nearly equivalent to the benchmark, suggesting it is not an appropriate method for mid-term time-series forecasting. Ultimately, the second quarter of the year emerged as the most predictable quarter, with even the worst model managing to perform meaningful predictions. This work provides valuable insights into the application of time-series forecasting algorithms for water demand forecasting, highlighting the benefits and limitations of each method. It underscores the potential of these models in enhancing operational efficiency, providing key insights for water service providers. These findings lay the groundwork for future research and practical applications in big data analysis for water utilities.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Abbasimehr, H., Shabani, M. & Yousefi, M. 2020 An optimized model using LSTM network for demand forecasting. *Comput. Ind. Eng.* **143**, 106435. https://doi.org/10.1016/j.cie.2020.106435.

Adamowski, J., Fung Chan, H., Prasher, S. O., Ozga-Zielinski, B. & Sliusarieva, A. 2012 Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resour. Res.* **48**. https://doi.org/10.1029/2010WR009945.

Alvisi, S., Franchini, M. & Marinelli, A. 2003 A stochastic model for representing drinking water demand at residential level. *Water Resour. Manage.* **17**, 197–222. https://doi.org/10.1023/A:1024100518186.

Antunes, A., Andrade-Campos, A., Sardinha-Lourenço, A. & Oliveira, M. S. 2018 Short-term water demand forecasting using machine learning techniques. *J. Hydroinf.* **20**, 1343–1366. https://doi.org/10.2166/hydro.2018.163.

Billings, R. B. & Jones, C. V. 2011 *Forecasting Urban Water Demand*. American Water Works Association, Denver, Colorado, USA.

Blokker, E. J. M., Vreeburg, J. H. G. & van Dijk, J. C. 2010 Simulating residential water demand with a stochastic End-Use model. *J. Water Resour. Plann. Manage.* **136**, 19–26. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000002.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C. & Ljung, G. M. 2015 *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, Hoboken, New Jersey, USA.

Boyle, T., Giurco, D., Mukheibir, P., Liu, A., Moy, C., White, S. & Stewart, R. 2013 Intelligent metering for urban water: A review. *Water* **5**, 1052–1081. https://doi.org/10.3390/w5031052.

Buchberger, S. G. & Wu, L. 1995 Model for instantaneous residential water demands. *J. Hydraul. Eng.* **121**, 232–246. https://doi.org/10.1061/(ASCE)0733-9429(1995)121:3(232).

Carreira-Perpiñán, M. Á. 1997 *A Review of Dimension Reduction Techniques*.

Chen, J. & Boccelli, D. L. 2018 Forecasting hourly water demands with seasonal autoregressive models for real-time application. *Water Resour. Res.* **54**, 879–894. https://doi.org/10.1002/2017WR022007.

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

Cover, T. & Hart, P. 1967 Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **13**, 21–27. https://doi.org/10.1109/TIT.1967.1053964.

Cui, B.-B. 2017 Design and implementation of movie recommendation system based on kNN collaborative filtering algorithm. *ITM Web Conf.* **12**, 04008. https://doi.org/10.1051/itmconf/20171204008.

Di Mauro, A., Cominola, A., Castelletti, A. & Di Nardo, A. 2021 Urban water consumption at multiple spatial and temporal scales. A review of existing datasets. *Water* **13**, 36. https://doi.org/10.3390/w13010036.

Donkor, E. A., Mazzuchi, T. A., Soyer, R. & Alan Roberson, J. 2014 Urban water demand forecasting: Review of methods and models. *J. Water Resour. Plann. Manage.* **140**, 146–159. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000314.

Duerr, I., Merrill, H. R., Wang, C., Bai, R., Boyer, M., Dukes, M. D. & Bliznyuk, N. 2018 Forecasting urban household water demand with statistical and machine learning methods using large space-time data: A comparative study. *Environ. Modell. Software* **102**, 29–38. https://doi.org/10.1016/j.envsoft.2018.01.002.

Firat, M., Yurdusev, M. A. & Turan, M. E. 2009 Evaluation of artificial neural network techniques for municipal water consumption modeling. *Water Resour. Manag.* **23**, 617–632. https://doi.org/10.1007/s11269-008-9291-3.

Gardiner, V. & Herrington, P. 1986 *Water Demand Forecasting*. CRC Press, London, United Kingdom.

Ghalehkhondabi, I., Ardjmand, E., Young, W. A. & Weckman, G. R. 2017 Water demand forecasting: Review of soft computing methods. *Environ. Monit. Assess.* **189**, 313. https://doi.org/10.1007/s10661-017-6030-3.

Herrera, M., Torgo, L., Izquierdo, J. & Pérez-García, R. 2010 Predictive models for forecasting hourly urban water demand. *J. Hydrol.* **387**, 141–150. https://doi.org/10.1016/j.jhydrol.2010.04.005.

Hochreiter, S. & Schmidhuber, J. 1997 Long Short-Term Memory. *Neural Comput.* **9**, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

House-Peters, L. A. & Chang, H. 2011 Urban water demand modeling: Review of concepts, methods, and organizing principles. *Water Resour. Res.* **47**. https://doi.org/10.1029/2010WR009624.

Hu, P., Tong, J., Wang, J., Yang, Y. & Oliveira Turci, L. d. 2019 A hybrid model based on CNN and Bi-LSTM for urban water demand prediction. In *2019 IEEE Congress on Evolutionary Computation (CEC). Presented at the 2019 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1088–1094. https://doi.org/10.1109/CEC.2019.8790060.

Huang, M., Bao, Q., Zhang, Y. & Feng, W. 2019 A hybrid algorithm for forecasting financial time series data based on DBSCAN and SVR. *Information* **10**, 103. https://doi.org/10.3390/info10030103.

Hyndman, R. J. & Athanasopoulos, G. 2018 *Forecasting: Principles and Practice*. OTexts, Melbourne, Australia.

Johnson, E. H. 2007 Estimating uncertainty of flow measurement for the in-situ calibration of large water meters. *Australas. J. Water Resour.* **11**, 41–52. https://doi.org/10.1080/13241583.2007.11465310.

Kabbilawsh, P., Kumar, D. S. & Chithra, N. R. 2022 Performance evaluation of univariate time-series techniques for forecasting monthly rainfall data. *J. Water Clim. Change* **13**, 4151–4176. https://doi.org/10.2166/wcc.2022.107.

Kalbusch, A., Henning, E., Brikalski, M. P., Luca, F. V. d. & Konrath, A. C. 2020 Impact of coronavirus (COVID-19) spread-prevention actions on urban water consumption. *Resour. Conserv. Recycl.* **163**, 105098. https://doi.org/10.1016/j.resconrec.2020.105098.

Kofinas, D., Mellios, N., Papageorgiou, E. & Laspidou, C. 2014 Urban water demand forecasting for the island of Skiathos. *Procedia Eng.* **89**, 1023–1030. 16th Water Distribution System Analysis Conference, WDSA2014. https://doi.org/10.1016/j.proeng.2014.11.220.

Kontopoulos, I., Makris, A., Tserpes, K. & Varvarigou, T. 2023 *An evaluation of time series forecasting models on water consumption data: A case study of Greece*.

Kossieris, P., Tsoukalas, I., Makropoulos, C. & Savic, D. 2019 Simulating marginal and dependence behaviour of water demand processes at any fine time scale. *Water* **11**, 885. https://doi.org/10.3390/w11050885.

Kouiroukidis, N. & Evangelidis, G. 2011 The effects of dimensionality curse in high dimensional kNN search. In *2011 15th Panhellenic Conference on Informatics*. Presented at the 2011 15th Panhellenic Conference on Informatics, pp. 41–45. https://doi.org/10.1109/PCI.2011.45.

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**, 644. https://doi.org/10.3390/w13050644.

Li, X. J. & Gao, Y. 2021 Digitisation of conventional water meters using automated number recognition. In *TENCON 2021 – 2021 IEEE Region 10 Conference (TENCON)*. Presented at the TENCON 2021. 2021 IEEE Region 10 Conference (TENCON), pp. 470–475. https://doi.org/10.1109/TENCON54134.2021.9707425.

Liu, L.-M. & Lin, M.-W. 1991 Forecasting residential consumption of natural gas using monthly and quarterly time series. *Int. J. Forecasting* **7**, 3–16. https://doi.org/10.1016/0169-2070(91)90028-T.

Liu, G., Savic, D. & Fu, G. 2023 Short-term water demand forecasting using data-centric machine learning approaches. *J. Hydroinf.* **25**, 895–911. https://doi.org/10.2166/hydro.2023.163.

Ma, W., Nowocin, K., Marathe, N. & Chen, G. H. 2019 An interpretable produce price forecasting system for small and marginal farmers in India using collaborative filtering and adaptive nearest neighbors. In *Proceedings of the Tenth International Conference on Information and Communication Technologies and Development, ICTD '19*. Association for Computing Machinery, New York, NY, USA, pp. 1–11. https://doi.org/10.1145/3287098.3287100.

Maidment, D. R. & Miaou, S.-P. 1986 Daily water use in nine cities. *Water Resour. Res.* **22**, 845–851. https://doi.org/10.1029/WR022i006p00845.

Msiza, I., Nelwamondo, F. & Marwala, T. 2008 Water demand prediction using artificial neural networks and support vector regression. *J. Comput.* **3**. https://doi.org/10.4304/jcp.3.11.1-8.

Niknam, A., Zare, H. K., Hosseininasab, H., Mostafaeipour, A. & Herrera, M. 2022 A critical review of short-term water demand forecasting tools – what method should I use? *Sustainability* **14**, 5412. https://doi.org/10.3390/su14095412.

Oliveira, P. J. A. & Boccelli, D. L. 2017 k-Nearest neighbor for short term water demand forecasting, In World Environmental and Water Resources Congress 2017, pp. 501–510. https://doi.org/10.1061/9780784480625.046.

Pandey, P., Bokde, N. D., Dongre, S. & Gupta, R. 2021 Hybrid models for water demand forecasting. *J. Water Resour. Plann. Manage.* **147**, 04020106. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001331.

Perktold, J., Seabold, S., , Sheppard, K., ChadFulton, Shedden, K., Jbrockmendel, J-Grana6, Quackenbush, P., Arel-Bundock, V., McKinney, W., Langmore, I., Baker, B., Gommers, R., Yogabonito, S-Scherrer, Zhurko, E., Brett, M., Giampieri, E., Liu, Y., Millman, J., Hobson, P., Vincent, Roy, P., Augspurger, T., Tvanzyl, Alexbrc, Hartley, T., Perez, F., Tamiya, Y. & Halchenko, Y. 2023 statsmodels/statsmodels: Release 0.14.0. https://doi.org/10.5281/ZENODO.593847.

Pesantez, J. E., Berglund, E. Z. & Kaza, N. 2020 Smart meters data for modeling and forecasting water demand at the user-level. *Environ. Modell. Software* **125**, 104633. https://doi.org/10.1016/j.envsoft.2020.104633.

pmdarima 2023 Available from: https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.AutoARIMA.html#pmdarima.arima.AutoARIMA (accessed 4.25.24).

Pu, Z., Yan, J., Chen, L., Li, Z., Tian, W., Tao, T. & Xin, K. 2022 A hybrid wavelet-CNN-LSTM deep learning model for short-term urban water demand forecasting. *Front. Environ. Sci. Eng.* **17**, 22. https://doi.org/10.1007/s11783-023-1622-3.

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**, 294. https://doi.org/10.3390/w12010294.

Ramlatchan, A., Yang, M., Liu, Q., Li, M., Wang, J. & Li, Y. 2018 A survey of matrix completion methods for recommendation systems. *Big Data Min. Anal.* **1**, 308–323. https://doi.org/10.26599/BDMA.2018.9020008.

Randall, T. & Koech, R. 2019 Smart water metering technology for water management in urban areas. *Water E-J.* **4**, 1–14. https://doi.org/10.21139/wej.2019.001.

Rehfeld, K., Marwan, N., Heitzig, J. & Kurths, J. 2011 Comparison of correlation analysis techniques for irregularly sampled time series. *Nonlinear Processes Geophys.* **18**, 389–404. https://doi.org/10.5194/npg-18-389-2011.

Salem, F. M. 2022 *Recurrent Neural Networks: From Simple to Gated Architectures*. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-89929-5.

Schafer, J. B., Frankowski, D., Herlocker, J., Sen, S., 2007 Collaborative filtering recommender systems. In: *The Adaptive Web: Methods and Strategies of Web Personalization, Lecture Notes in Computer Science* (Brusilovsky, P., Kobsa, A. & Nejdl, W., eds.). Springer, Berlin, Heidelberg, pp. 291–324. https://doi.org/10.1007/978-3-540-72079-9_9.

Shumway, R. H. & Stoffer, D. S. 2017 *Time Series Analysis and Its Applications: With R Examples, Springer Texts in Statistics*. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-52452-8.

Smith, J. A. 1988 A model of daily municipal water use for short-term forecasting. *Water Resour. Res.* **24**, 201–206. https://doi.org/10.1029/WR024i002p00201.

Song, X., Liu, Y., Xue, L., Wang, J., Zhang, J., Wang, J., Jiang, L. & Cheng, Z. 2020 Time-series well performance prediction based on long short-term memory (LSTM) neural network model. *J. Pet. Sci. Eng.* **186**, 106682. https://doi.org/10.1016/j.petrol.2019.106682.

Stoica, P. & Selen, Y. 2004 Model-order selection: A review of information criterion rules. *IEEE Signal Process Mag.* **21**, 36–47. https://doi.org/10.1109/MSP.2004.1311138.

Stoker, D. M., Barfuss, S. L. & Johnson, M. C. 2012 Flow measurement accuracies of in-service residential water meters. *J. AWWA* **104**, E637–E642. https://doi.org/10.5942/jawwa.2012.104.0145.

Sun, Y., Bansal, R., Bhardwaj, A. K. & Srivastava, A. K. 2011 Electricity market price forecasting using support vector machines. *Int J Comput. Aided Eng. Technol.* **3**, 1–18. https://doi.org/10.1504/IJCAET.2011.037865.

TensorFlow 2023 GitHub. Available from: https://github.com/tensorflow/tensorflow/releases/tag/v2.15.0 (accessed 4.25.24).

Wong, J. S., Zhang, Q. & Chen, Y. D. 2010 Statistical modeling of daily urban water consumption in Hong Kong: Trend, changing patterns, and forecast. *Water Resour. Res.* **46**. https://doi.org/10.1029/2009WR008147.

First received 19 January 2024; accepted in revised form 9 June 2024. Available online 25 June 2024