Water Supply

IWA PUBLISHING

Water Supply Vol 24 No 4, 1352 doi: 10.2166/ws.2024.055

AI-Forecast: an innovative and practical tool for short-term water demand forecasting

Ariele Zanfei [a,*], Andrea Lombardi[a], Alberto De Luca[a] and Andrea Menapace[b]

[a] AIAQUA S.r.l., Via Volta 13/A, Bolzano, Italy
[b] Faculty of Engineering, Free University of Bozen-Bolzano, Piazza Domenicani 3, Bolzano, Italy
*Corresponding author. E-mail: a.zanfei@aiaqua.tech

AZ, 0000-0002-3759-6421

ABSTRACT

Water management is a major contemporary and future challenge. In an increasing water demand scenario related to climate change, a water distribution system must ensure equal access to water for all users. In this context, a reliable short-term water demand forecasting system is crucial for reliable water management. However, despite the abundance of studies in the scientific literature, few examples highlight complete tools for providing such models to real water utilities and water managers. This study presents AI-Forecast, an innovative tool developed to predict water demand with state-of-the-art models. Such tool is based on the data-driven logic, and it is designed to provide a complete data-driven chain that starts from the data and arrives to the short-term water demand prediction. AI-Forecast can import data, properly manage them, and assess tasks like outlier detection and missing data imputation. Eventually, it can implement state-of-the-art forecasting models and provide the forecasts. The prediction is shown through an intuitive web interface, which is designed to highlight the major information related to the prediction accuracy. Although this tool does not provide a new prediction algorithm, it proposes a complete data-driven chain that is practically designed to take such models in practice to real water utilities.

Key words: artificial neural network, deep learning, innovation, water demand forecasting, water distribution systems

HIGHLIGHTS

- This study proposes a complete data-driven chain that is practically designed to consider such models in practice to real water utilities.
- The results highlight how such practical tool can deliver reliable prediction.
- AI-Forecast is designed to host state-of-the-art methods based on the data-driven formulation.

INTRODUCTION

Currently, efficient water distribution systems (WDSs) are critical for the conservation of drinking water. This response requires careful monitoring in the future. In the world of water management, the ever-increasing demand for water from agriculture, humans and industries due to climate change and socio-economic factors forces researchers, practitioners and utilities to meet the complex request for as efficient management of water resources as possible shall be addressed (Ramos *et al.* 2020).

In addition to the same characteristics as the power grid, the water distribution system has embarked on a major renewal in recent years to change the de facto paradigm of the current grid and shift towards the smart grid concept (Lee *et al.* 2015; Antzoulatos *et al.* 2020; Menapace *et al.* 2020a). This important transition from green-smart systems is even more important for water supply systems today, with climate change increasing global water demand and putting WDS under stress (Wang *et al.* 2016). Climate change is forcing utilities to take actions in many parts of the world (e.g. Mynett & Vojinovic 2009; Yu *et al.* 2014; Elkiran *et al.* 2021), where the water resource availability is changing, leaving a sense of uncertainty for the future. Today is even more fundamental than in the past to avoid wastes and leakages through a proper water management and in order to ensure water safety for all. New technologies will play a crucial role in this target (Menapace *et al.* 2020b).

Generally, short-term water demand forecasts are needed to support the operation management of water systems (Zanfei *et al.* 2022a). The forecasts are designed to support decision-makers who have to deal with the ordinary operations of the systems but also to make strategic decisions for future investment in WDSs (Donkor *et al.* 2014). It is also worth mentioning

that, within the operational control framework, the water prediction should include not just the amount necessary for consumers but also the water lost from the distribution network due to leakage, as it is the total volume that is supplied. For effective management of water distribution networks, it's commonly necessary to predict demand to programme pumping schedules for the upcoming 24 h, leveraging the structure of electricity tariffs (Alvisi *et al.* 2007).

Nowadays, we live in the era of Big Data and artificial intelligence methods that are the state-of-art for multiple tools and algorithms that support sustainable water use (Sit *et al.* 2020; Oberascher *et al.* 2022). These new techniques can now be reliable and comprehensive tools that can effectively help improve water management efficiency. In recent decades, many studies have attempted to develop these instruments for many tasks. There are many important examples in scientific literature, such as metamodels for state estimation (Zanfei *et al.* 2023), data analysis techniques for smart water metering systems (Rahim *et al.* 2020), for water demand modelling (House-Peters & Chang 2011), detection of intrusion (Mboweni *et al.* 2021), or leakage detection (Wu & Liu 2017; Zanfei *et al.* 2022c), and for much more different and important tasks. Arguably, one of the most important of the many possible ways that these data-driven techniques can be accomplished is forecasting (e.g. Herrera *et al.* 2010; Brentan *et al.* 2017). All studies demonstrate how forecasting can improve water management efficiency and optimise it. For example, for hydropower systems (Avesani *et al.* 2022), or for improving the pump efficiency and reducing operating costs for irrigation systems (Pulido-Calvo & Gutierrez-Estrada 2009), and, of course, WDSs. This latest case study (Bakker *et al.* 2013) highlighted that the use of a water demand forecasting model allowed significant savings in a WDS in the Netherlands. Moreover, using a reliable water demand forecasting model allows water to be managed more efficiently, i.e., water and energy are saved, and costs are reduced.

In such data-driven context, the quality and volume of accessible data are critical for the majority of data-driven methods and their applications. It becomes even more significant for a forecasting model, which directly depends on the data to generate trustworthy predictions (Xenochristou & Kapelan 2020). Consequently, the data-processing operation assumes a fundamental role for these methodologies. Among all problems, the issue of missing data can greatly hinder the efficiency of data-driven approaches. A potential resolution in many scientific disciplines might be to simply ignore the missing data. Nonetheless, this is not a viable approach for data-driven models involving water demand data, given the nature of such time-series data. In reality, water demand exhibits a typical daily pattern and a seasonal element. This implies that excluding certain data might create complications for the data-driven model that needs to learn the behaviour of water demand over time. To overcome this problem and tackle the issue of missing data, the data imputation strategy is commonly employed. In practical terms, there exist various techniques for imputing missing data, ranging from conventional and straightforward approaches (such as deletion or imputation using mean values) to more sophisticated methods (such as imputation with machine learning algorithms) (Osman *et al.* 2018).This present study proposes a practical tool for water demand forecasting named AI-Forecast. It is worth noting that this article aims to showcase a comprehensive tool that utilises established methods from existing literature to forecast short-term water demand for water utilities. AI-Forecast is a tool that, starting from the raw data, allows for short-term water demand forecasting for water utilities using well-consolidated methods of the literature. In essence, the objective is not to present a new methodology, but rather to demonstrate how the findings from extensive research can be applied to develop a practical tool that assists water utilities in their daily activities. Therefore, this paper aims to present the complete data-driven chain, starting from the overall architecture of AI-Forecast. Afterwards, the data-processing module that is designed to process the data and detect eventual outliers in the historical data needed for training the data-driven models will be presented, as well as the missing data imputation algorithm that is equipped within such module. Then, it presents the three algorithms that are already implemented, which are a support vector regression (SVR), a multilayer perceptron (MLP) and a K-nearest neighbour (KNN). Finally, a proof-of-concept case study is presented alongside with the system performances and functionalities.

MATERIALS AND METHODS

This paper proposes an innovative and practical tool to support water utilities providing water demand forecasting. This section provides a description of such tool, showing the overall architecture of the system, the algorithms that are incorporated in the tool and some functionalities. The overall architecture of the system is shown in Figure 1.

Figure 1 highlights the main idea behind AI-Forecast. In fact, the tool is designed to provide water demand forecasting to water utilities adopting some well-established methodologies. Furthermore, this tool has been designed to be adaptable, allowing for the incorporation of different methods. The architecture of the tool follows a modular concept, enabling

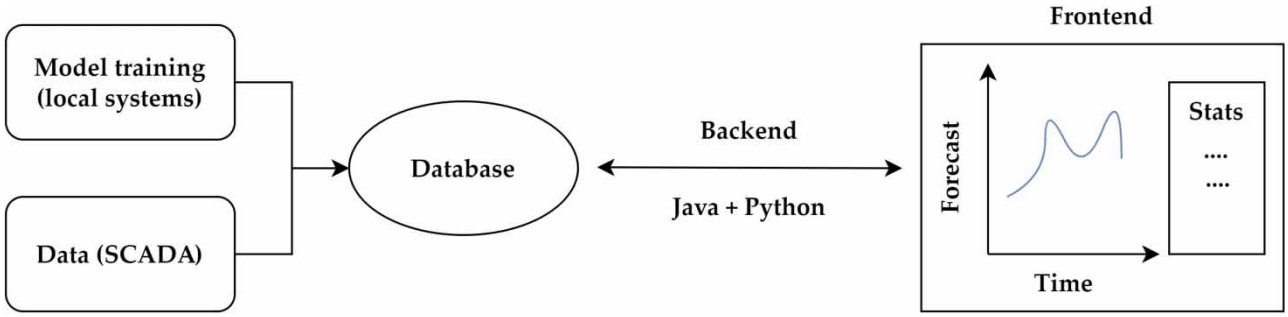

Figure 1 | Diagram of the AI-Forecast framework.

modifications to various modules such as data pre-processing, data imputation, and model tuning, and also the module that has to build the model for actually provide the forecast. In order to accomplish such task, the data-driven models are first trained and prepared on local systems. During this stage, the historical data for each case study are prepared and used to build the models. In addition, a data pre-processing module is incorporated to grant an optimal model creation. In particular, this module permits to deal with missing data using a KNN imputer algorithm (Zanfei *et al.* 2022b), and allows to detect and clean outliers in the data with some simple filtering techniques. It also allows the creation of the samples for the model training phase, which are also used afterwards to validate the models. This module is also equipped with an optimal model training system, which is based on the dynamic grid search proposed in Menapace *et al.* (2021). Then, the model is uploaded automatically to the database, which is also designed to be connected with a SCADA system in order to acquire and update the data in almost real-time. Furthermore, a backend written using both Python and Java programming languages connects the database with a web application in order to provide the forecasts requested by the user. The frontend–backend architecture allows to access multiple functionalities, such as choosing a different period for the forecast, selecting the model to use, the station to be forecasted, and also visualising some metrics to understand the prediction performances.

In the following sections, all the models that are now implemented in the system will be presented, along with some functionalities of the data pre-processing part and the modules designed for outlier detection and missing data imputation. Furthermore, the details of the software architecture and also a proof-of-concept case study used to demonstrate the capability of AI-Forecast will be proposed.

Forecasting methodology

The AI-Forecast tool is designed to host multiple different data-driven model architectures. At the moment, an MLP, a SVR and KNN methods are already implemented and can be used as forecasting models. In the following sections, these well-established methods are described.

MLP

Artificial neural networks (ANNs) have been one of the most used methods for short-term water demand forecasting in recent decades. In the field of ANNs, the MLP (also known as feedforward neural network) is probably the most adopted architecture. The design of this kind of network was influenced by biological neural networks. Specifically, it is an extremely quick and simple architecture in which data flows through every layer from the input to the output. It is comprised of several interconnected and layered computing units known as neurons. First, the input data and the input layer are directly connected. Secondly, the data are moved from the hidden layers to the output layer, which is the final layer. The MLP for the study is shown in Figure 2.

Figure 2 shows the MLP architecture. Regarding the number of neurons and layers, such hyperparameters are chosen case-by-case, meaning that its decision is demanded by the tuning module (Menapace *et al.* 2021). Nevertheless, the input layer is designed to process the past observation of the water demand time-series to be forecasted, and it can host the different number of past observations based on the case studies. Differently, the output layer is composed by 24 units, which are needed to provide the simultaneous forecast of 24 h (i.e., 1 day ahead with hourly aggregation).

Such models are built using the Keras (Chollet *et al.* 2015) and TensorFlow (Abadi *et al.* 2016) libraries in Python.

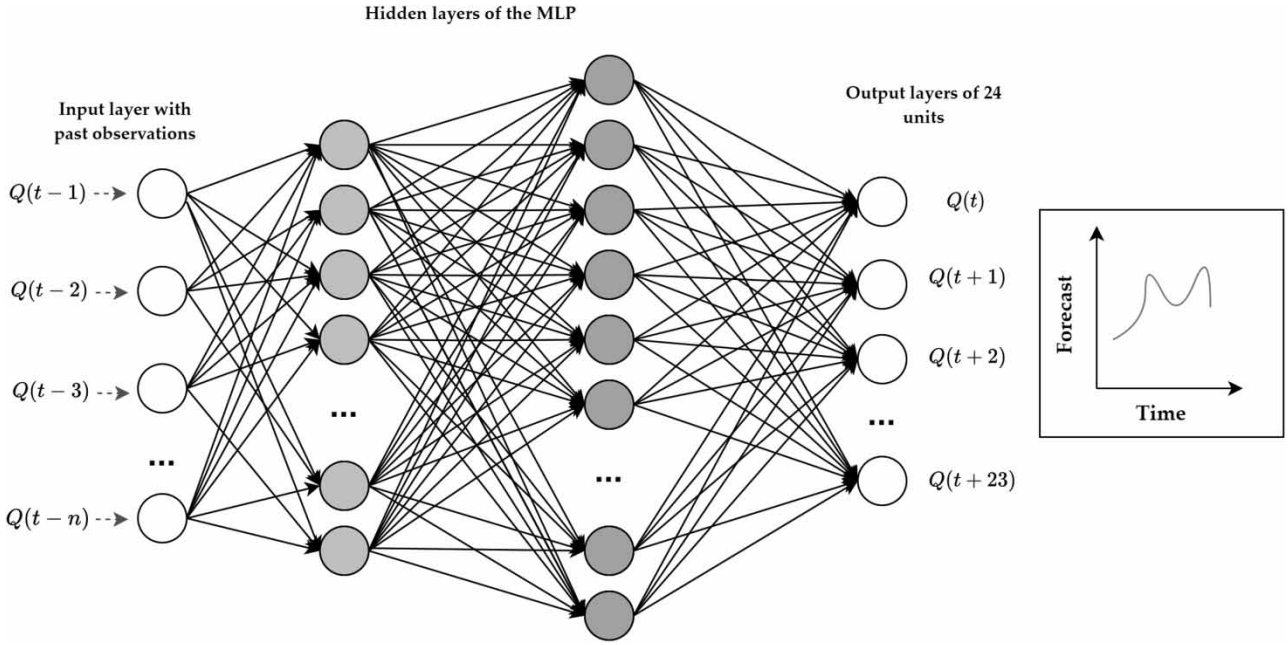

Figure 2 | MLP architecture adopted in the AI-Forecast tool.

SVR

SVR extends the principles of support vector machines (SVMs) to address regression problems. In particular, the SVMs' classification techniques are based on identifying an optimal hyperplane that maximises the separation distance between two data point groups by minimising the margin error. In the case of the SVR, the definition of the regression hyperplane that provides the optimal fit to the data points implies building a mapping (ϕ) of the data points exploiting a kernel function. The SVR method considers a margin of error, which means that the possibility of some predicted points being outside the region formed by $\pm\varepsilon$ is tolerated because it doesn't affect the quality of the predictions. The kernel function, given the input data x_i, is:

$$\hat{f}(x) = \langle w, \phi(x) \rangle + b \tag{1}$$

where w represents the weight vector, b the bias term, and $\phi(x)$ the mapping into a higher-dimensional space (i.e., feature space) of the input data x. The aim is to define a function that deviates at most ε from the observed data point y_i associated to x_i and that, at the same time, minimises the complexity of the model. This corresponds to a convex optimisation problem that can be written as follows:

$$minimize\ \frac{1}{2}\|w^2\|\ subject\ to\ |y_i - \langle w, \phi(x) \rangle - b| \leq \varepsilon \tag{2}$$

The main assumptions are the existence of $\hat{f}(x)$ for all observations and its precision of ε. In cases where a solution to the problem might not exist, the method introduces slack variables to penalise deviations beyond ε from the actual observation. These slack variables, denoted as ξ^+, and ξ^-, are defined as:

$$\xi^+ = \hat{f}(x) - y(x_i) > \varepsilon \tag{3}$$

$$\xi^- = y(x_i) - \hat{f}(x) > \varepsilon \tag{4}$$

Hence, considering the size of the training observation n and the regularisation parameter C which strikes a balance between the model's tolerated error and its complexity, the optimisation problem can be written as:

$$minimize\ \frac{1}{2}\|w^2\| + C\frac{1}{n}\sum_{i=1}^{n}(\xi^+ + \xi^-) \tag{5}$$

The main parameters of the entire formulation are ε and the regularisation factor C (Smets *et al.* 2007) and to select the most appropriate hyperparameters a grid search procedure is used, and it is automatically made for each case study. As for the MLP, also the SVR model is designed to take as input the past observation of the water demand time-series to be forecasted.

KNN

The KNN is an algorithm that can be exploited for classification and regression problems. In this work, the KNN is used for regression problems such as water demand data forecasting and imputation. In both cases the input dataset is represented by a collection of training objects characterised by specific properties. Given objects whose properties are unknown, the KNN algorithm allows to find the k most similar objects in the training dataset according to a distance metric and consequently the unknown properties as an average of the KNNs (Zhang 2012). The most used distance metrics are the Euclidean distance and the Manhattan distance, which both derives from the general expression of the Minkowski distance:

$$d(x, z) = \left(\sum_{i=1}^{n}|x_i - z_i|^p\right)^{1/p} \tag{6}$$

where x_i and z_i belong, respectively, to points $X = (x_1, \dots, x_n)$ and $Z = (z_1, \dots, z_n)$ between which the distance is calculated. The Minkowski coefficient $p = 2$ leads to the Euclidean distance expression, instead, with $p = 1$ the Manhattan distance expression is obtained. In the AI-Forecast tool, the KNN algorithm is adopted using the Euclidean distance.

It is worth mentioning that the KNN algorithm employed in AI-Forecast is used both as a regressor for forecasting and also as the algorithm for imputation. Regarding the forecasting application, the mechanism is the same of the others, which means that the KNN takes as input the past observation of the demand and performs the day-ahead forecast. Regarding the imputation, the KNN takes as input only few past observations (the previous three values) but also the calendar variables that represent the month, the day and the hour of the period that has to be imputed. Such calendar variables are provided to the imputer as simple numbers.

Data pre-processing module

In data-driven models, the input data play a crucial role; for this reason, they need to be accurately prepared to guarantee a high-quality prediction. Data pre-processing is conducted following several steps, including identifying outliers, missing data statistics, and imputation.

Outliers refer to data points within the dataset whose values deviate significantly from others due to errors of measurement or other ambiguities. In the case of water demand, outliers can be identified in negative values, typically related to sensor malfunctions, and in extremely hourly high or low values compared to the others contained in the dataset. For negative values, no additional considerations are needed, but for the other types of outliers, it is fundamental to have a knowledge of the water supply system and its history. In fact, for example, there can be periods (hours or days) characterised by water consumption restrictions imposed by the water utility company (Shah *et al.* 2018), which are important to know to conduct proper data analysis.

The procedure for the identification and treatment of the outliers is reported in the flowchart of Figure 3. For each timestep, the identified outliers are substituted with a Not-a-Number (NaN).

Figure 3 shows the process that allows AI-Forecast to deal with the outliers. The whole dataset is analysed to detect and delete eventual outliers. Then, the dataset is updated, and it is analysed to evaluate the statistics of NaNs, which is a fundamental step before proceeding with the imputation. These latter comprise the missing data count, the maximum gap dimension and the Nans distribution along the dataset with a mask. Indeed, this analysis is necessary to understand if the

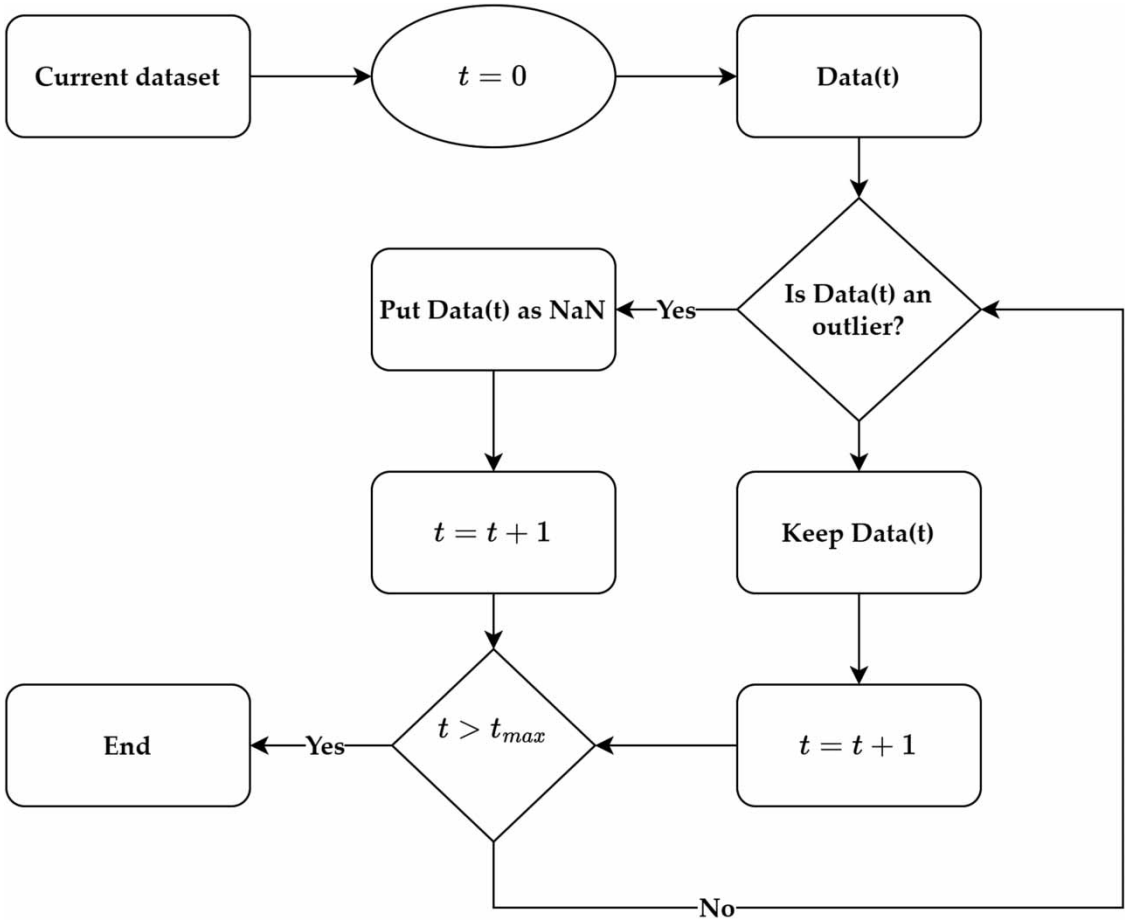

Figure 3 | Algorithm flowchart for outliers' detection and process.

data availability is sufficient. Once this condition is verified, the KNN method described in the previous section can be exploited for the imputation of the missing data.

Software architecture

The system architecture leverages the well-known client-server model. The server is in charge of receiving requests from a client, processing them, performing the necessary calculations and modelling tasks, and sending back the results to the client itself. In this particular case, the client is a web browser, therefore, it is possible to address this application as a 'web application'. The frontend and the backend communicate over the Internet.

The backend is made of two three main components: a relational database, the modelling software, and a web server. Firstly, the database stores the data that are to be used as inputs by the modelling software. These data are mainly time-series of recorded water demand data (i.e. measurements coming from flow meters), and can be updated in real-time if needed. These time-series are used both to train the models and as inputs to run the actual simulations. In future iterations, the database could also be used to store the simulation outputs for performance analysis. Secondly, the modelling software is composed of a series of Python modules packaged in a Docker container. Thirdly, the web server is also a containerised application written in Java using the Spring Boot framework. Its role is to receive requests from clients, launch the modelling software, read the results, and serve them back to the client.

Finally, the frontend is a simple web application written using state-of-the-art technologies such as ReactJS for the user interface and GraphQL query language for the Application Programming Interfaces (APIs). Additional libraries were used to display charts (Apache ECharts) and maps (OpenLayers). All the software was written using open-source tools and libraries.

Proof-of-concept case study

To test the proposed approach, the well-known Modena network (Bragalli *et al.* 2008) is used as synthetic case study. In particular, the data adopted in this case study are the ones generated by Zanfei *et al.* (2022c), which are open and available online. Three weeks of these data are shown in Figure 4.

Figure 4 highlights the data, which were generated to account from the typical seasonality of the water demand, with also an important random component due to the stochastic behaviour of the domestic consumptions. Nevertheless, it shows the 4 water signals of the case study, which can all be forecasted by the AI-Forecast tool and can then be chosen by the use of the web application. The dataset consists of 4 years of generated data with hourly time steps.

Performance evaluation metrics

To assess the performance of predictions, the AI-Forecast tool is equipped with the most conventional metrics, which are the mean absolute percentage error (MAPE), the determination coefficient (R^2), and the mean absolute error (MAE). The equations are shown through the following equations:

$$\text{MAPE} = \frac{100}{Z}\sum_{i=1}^{Z}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \tag{7}$$

$$\text{MAE} = \frac{1}{Z}\sum_{i=1}^{Z}|Y_i - \hat{Y}_i|. \tag{8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{Z}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{Z}(Y_i - \bar{Y}_i)^2} \tag{9}$$

where term Z represents the number of predicted values, while y_i and $\hat{Y}_i$ are the observed and the forecasted values at the i hourly time step, respectively. The term $\bar{Y}_i$ is the average of the observed values. Furthermore, low MAE and MAPE values

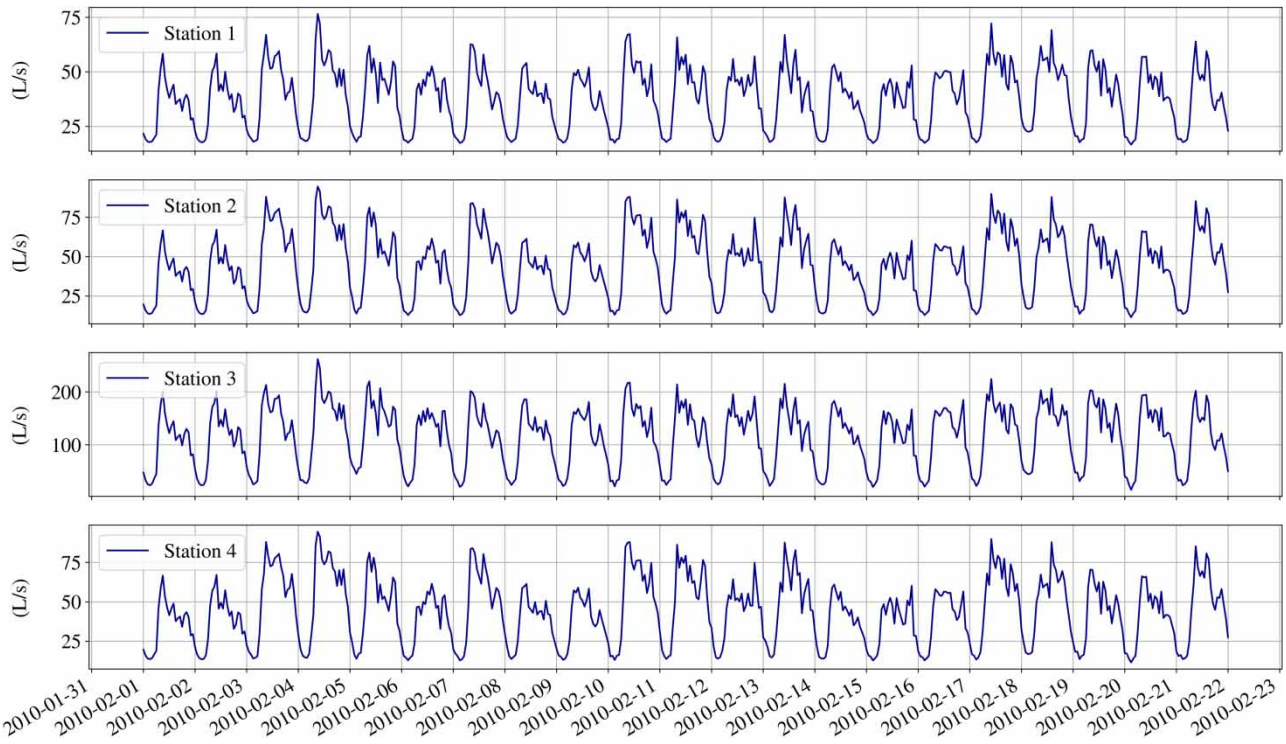

Figure 4 | Portion of 1 month of data of the time-series used for the case study.

are indicators of better model performances. Differently, R^2 metrics indicate better performance when its value is as close as possible to one.

RESULTS AND DISCUSSION

This section presents and shows the results of the tool, highlights some of the functionalities and shows the performances of the implemented methods over the case study adopted. Figure 5 shows the frontend of the application during its usage.

Figure 5 highlights the forecast produced by the algorithm selected for the selected station, and it compare it with the historical data. As previously mentioned, the overall system is built over some artificially generated data that allows to show how the system performs and its functionalities. In fact, this allows to highlight that the system permits to choose among different models, different periods, different stations to forecast but also the number of days to simulate (in case of availability of the historical values). This figure shows also that the system generates a small report of the forecast performance, and this is very important for a user to understand how much he can trust the prediction algorithm.

Nevertheless, AI-Forecast is now equipped with three different models, and it is designed to host as many algorithms as desired. Of course, each algorithm must be optimally tuned to perform well over the problem. Nevertheless, the system is already equipped with the module to tune the algorithm that are already in it. To show their performances and highlight that the system works for all the methods, Figure 6 shows the prediction performed by the SVR model implemented.

As can be noticed, the model performs well with a MAPE of 6.1%, an R^2 of 0.93 and a MAE of 4.4 L/s. The overall error is distributed along the day, especially during the demand peaks that are stochastic. Nevertheless, the error is normal distributed, which highlights the fact that the model is well-fitted and well-designed for the problem.

Figure 7 shows the prediction performed by the MLP model implemented.

Figure 7 highlights the prediction performed by the model. It is worth to mention that the architecture chosen by the tuning process is an MLP with two hidden layers of 48 and 92 neurons, respectively. Although the MLP show worse performances compared to the SVR model, it is still a good model with a MAPE of 8.5%, an R^2 of 0.91 and a MAE of 5.7 L/s. The daily seasonality of the water demand is well captured by the data-driven model, which has the residuals normal distributed. The error fluctuations are around the peaks, which are random and complex to be captured.

Figure 8 shows the prediction performed by the KNN model implemented.

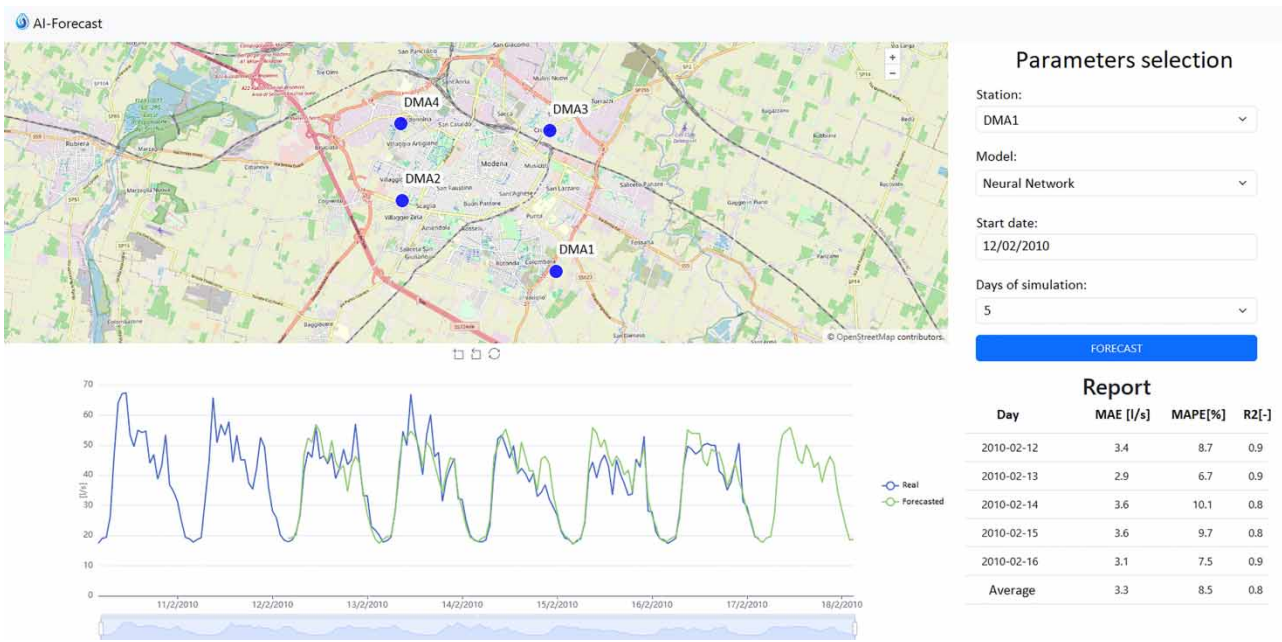

Figure 5 | Screen-shot of the AI-Forecast frontend output.

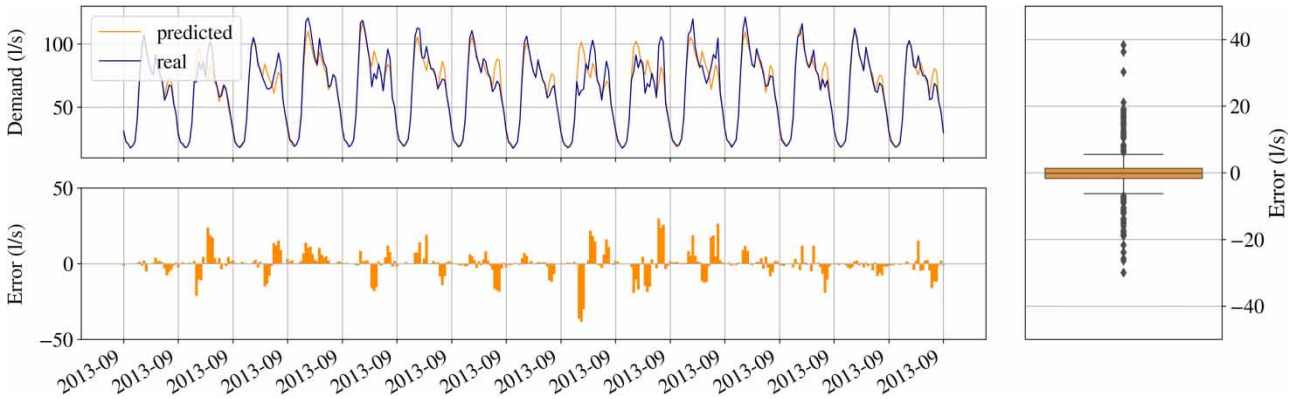

Figure 6 | Performance of the SVR model on 15 days of prediction of the testing dataset.

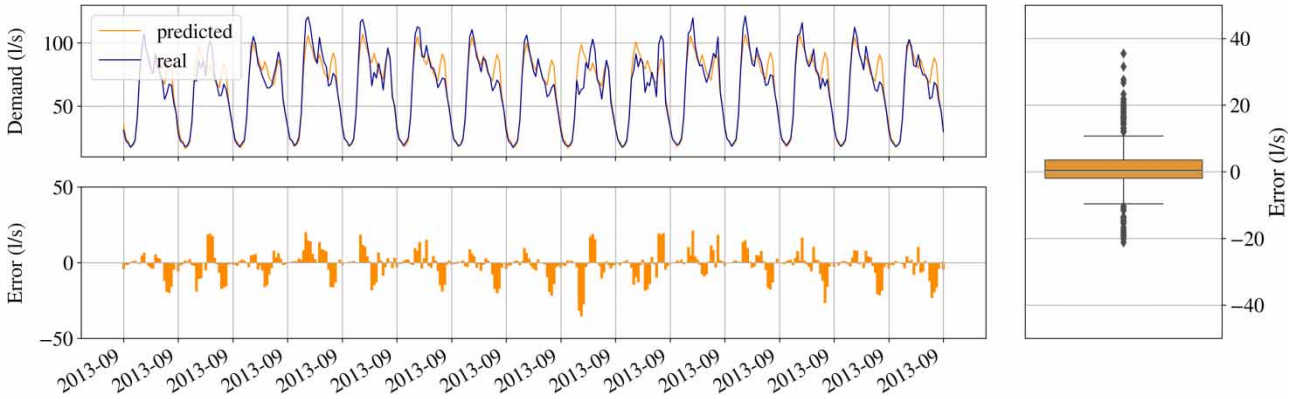

Figure 7 | Performance of the MLP model on 15 days of prediction of the testing dataset.

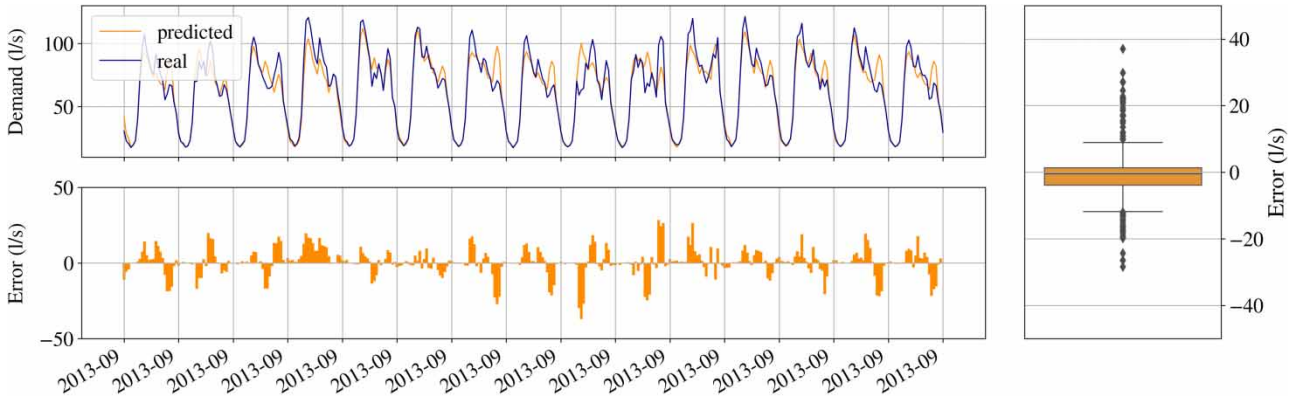

Figure 8 | Performance of the KNN model on 15 days of prediction of the testing dataset.

Once again, the model highlights good prediction capability. Also, the KNN is well-fitted over the data, with the errors that are normally distributed around the zero. This time, the KNN performs with a MAPE of 8.1%, and R^2 of 0.90 and a MAE of 5.7 L/s.

Overall, all three models implemented perform well with the prediction task. This highlights that the complete system behind AI-Forecast is capable of building well-designed models and providing reliable short-term water demand forecasting.

Functionalities and limitations

The AI-Forecast tool has demonstrated promising results, indicating its ability to effectively address the proposed forecasting problem. It is worth to remind that the main purpose of this study is to present a comprehensive tool that utilizes established techniques from existing research to forecast short-term water demand for water utilities. The objective is not to introduce a new methodology, but rather to demonstrate how the insights gained from extensive research can be applied to develop a practical tool that supports the daily operations of water utilities. With this aim, AI-Forecast shown his capability to perform the prediction of short-term water demand for water utilities by utilizing well-established methods found in the literature, starting from the raw data. Furthermore, this tool has been designed with flexibility in mind, allowing for the incorporation of numerous methods. The architecture of the tool follows a modular concept, enabling modifications to various modules such as data pre-processing, data imputation, and model tuning, and so on. However, it is essential to acknowledge that the tool does have inherent limitations. It is widely recognised that there is no universally applicable method for forecasting water demand. There may be instances where the proposed methods are unable to adequately handle prediction problems, particularly when dealing with real data. Real data often exhibit greater complexity, with numerous anomalies and challenges. Due to the AI-Forecast tool's reliance on data-driven approaches, it is highly likely that anomalies in the data stream will result in incorrect predictions and unrealistic outcomes. It is important to note that this issue is unavoidable when utilising data-driven methods. Nonetheless, our future plans involve expanding the algorithm pool underlying AI-Forecast, incorporating pattern-based methods that are less reliant on near real-time observations. These methods have the potential to mitigate issues associated with anomalies in the data stream.

CONCLUSION

This study proposes AI-Forecast, a tool for providing short-term water demand forecasting to water utilities by means of a complete data-driven chain. The main idea behind AI-Forecast is to build a reliable and practical system that can support water managers with accurate 24 h of forecasts. In order to do so, AI-Forecast is equipped with multiple modules that allow building a machine learning model according to the state-of-the-art data-driven approach. This means that AI-Forecast is designed with multiple modules that can manage the data, import them, detect outliers, and perform missing data imputation. The frameworks allow the host of various models. At the moment, the SVR, the MLP and the KNN models are already implemented alongside the needed tuning procedures.The application of AI-Forecast is shown in this study by means of an artificially generated case study. The results allowed to highlight the mechanism behind AI-Forecast, as well as its performance on this case study. Furthermore, the interface of AI-Forecast is shown to enhance how AI-Forecast can provide simple information to the users. In fact, the mission of AI-Forecast is to show how it is possible to build a complete tool for water demand forecasting that starts from the raw data till the creation of a web application that provides the prediction for the water utilities.

In the future, we plan to keep working on AI-Forecast to add more functionalities, such as choosing the prediction horizon for the forecast, but also to include more models in the system.

ACKNOWLEDGEMENT

This study was founded by the project 'Applicativi di intelligenza artificiale per la gestione delle reti idriche' founded by the autonomous province of Bozen-Bolzano (act 43/22 of year 2022).

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories (https://github.com/ArieleZanfei/generated-datasets-for-burst-detection-in-water-distribution-systems).

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y. & Zheng, X. 2016 "Tensorflow: A system for large-scale machine learning." *12th Symp. Oper. Syst. Des. Implement. 16*, 265–283.

Alvisi, S., Franchini, M. & Marinelli, A. 2007 A short-term, pattern-based model for water-demand forecasting. *J. Hydroinf.* **9** (1), 39–50. https://doi.org/10.2166/hydro.2006.016.

Antzoulatos, G., Mourtzios, C., Stournara, P., Kouloglou, I.-O., Papadimitriou, N., Spyrou, D., Mentes, A., Nikolaidis, E., Karakostas, A., Kourtesis, D., Vrochidis, S. & Kompatsiaris, I. 2020 Making urban water smart: The SMART-WATER solution. *Water Sci. Technol.* **82** (12), 2691–2710. https://doi.org/10.2166/wst.2020.391.

Avesani, D., Zanfei, A., Di Marco, N., Galletti, A., Ravazzolo, F., Righetti, M. & Majone, B. 2022 Short-term hydropower optimization driven by innovative time-adapting econometric model. *Appl. Energy* **310**, 118510. https://doi.org/10.1016/j.apenergy.2021.118510.

Bakker, M., Vreeburg, J., Palmen, L., Sperber, V., Bakker, G. & Rietveld, L. 2013 Better water quality and higher energy efficiency by using model predictive flow control at water supply systems. *J. Water Supply Res. Technol.* **62** (1), 1–13. IWA Publishing.

Bragalli, C., D'Ambrosio, C., Lee, J., Lodi, A. & Toth, P. 2008 'Water network design by MINLP.' *Rep No RC24495 IBM Res. Yorktown Heights NY*.

Brentan, B. M., Luvizotto Jr, E., Herrera, M., Izquierdo, J. & Pérez-García, R. 2017 Hybrid regression model for near real-time urban water demand forecasting. *J. Comput. Appl. Math.* **309**, 532–541. https://doi.org/10.1016/j.cam.2016.02.009.

Chollet, F. *et al.* 2015 'Keras'. GitHub. https://keras.io

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** (2), 146–159. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000314.

Elkiran, G., Nourani, V., Elvis, O. & Abdullahi, J. 2021 Impact of climate change on hydro-climatological parameters in North Cyprus: Application of artificial intelligence-based statistical downscaling models. *J. Hydroinf.* **23** (6), 1395–1415. https://doi.org/10.2166/hydro.2021.091.

Herrera, M., Torgo, L., Izquierdo, J. & Pérez-García, R. 2010 Predictive models for forecasting hourly urban water demand. *J. Hydrol.* **387** (1–2), 141–150. https://doi.org/10.1016/j.jhydrol.2010.04.005.

House-Peters, L. A. & Chang, H. 2011 Urban water demand modeling: Review of concepts, methods, and organizing principles: REVIEW. *Water Resour. Res.* **47**, 5. https://doi.org/10.1029/2010WR009624.

Lee, S. W., Sarp, S., Jeon, D. J. & Kim, J. H. 2015 Smart water grid: The future water management platform. *Desalin. Water Treat.* **55** (2), 339–346. https://doi.org/10.1080/19443994.2014.917887.

Mboweni, I. V., Abu-Mahfouz, A. M. & Ramotsoela, D. T. 2021 A machine learning approach to intrusion detection in water distribution systems – A review. In: *IECON 2021 – 47th Annu. Conf. IEEE Ind. Electron. Soc.*, Toronto, ON, Canada. IEEE, pp. 1–7.

Menapace, A., Boscheri, W., Baratieri, M. & Righetti, M. 2020a An efficient numerical scheme for the thermo-hydraulic simulations of thermal grids. *Int. J. Heat Mass Transfer* **161**, 120304. Elsevier.

Menapace, A., Zanfei, A., Felicetti, M., Avesani, D., Righetti, M. & Gargano, R. 2020b Burst detection in water distribution systems: The issue of dataset collection. *Appl. Sci.* **10** (22), 8219. https://doi.org/10.3390/app10228219.

Menapace, A., Zanfei, A. & Righetti, M. 2021 Tuning ANN hyperparameters for forecasting drinking water demand. *Appl. Sci.* **11** (9), 4290. https://doi.org/10.3390/app11094290.

Mynett, A. E. & Vojinovic, Z. 2009 Hydroinformatics in multi-colours – part red: Urban flood and disaster management. *J. Hydroinf.* **11** (3–4), 166–180. https://doi.org/10.2166/hydro.2009.027.

Oberascher, M., Kinzel, C., Kastlunger, U., Schöpf, M., Grimm, K., Plaiasu, D., Rauch, W. & Sitzenfrei, R. 2022 Smart water campus – a testbed for smart water applications. *Water Sci. Technol.* **86** (11), 2834–2847. https://doi.org/10.2166/wst.2022.369.

Osman, M. S., Abu-Mahfouz, A. M. & Page, P. R. 2018 A survey on data imputation techniques: Water distribution system as a use case. *IEEE Access* **6**, 63279–63291. https://doi.org/10.1109/ACCESS.2018.2877269.

Pulido-Calvo, I. & Gutierrez-Estrada, J. C. 2009 Improved irrigation water demand forecasting using a soft-computing hybrid model. *Biosyst. Eng.* **102** (2), 202–218. Elsevier.

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. https://doi.org/10.3390/w12010294.

Ramos, H. M., Carravetta, A. & Nabola, A. M. 2020 New challenges in water systems. *Water* **12** (9), 2340. https://doi.org/10.3390/w12092340.

Shah, S., Ben Miled, Z., Schaefer, R. & Berube, S. 2018 Differential learning for outliers: A case study of water demand prediction. *Appl. Sci.* **8** (11), 2018. https://doi.org/10.3390/app8112018.

Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y. & Demir, I. 2020 A comprehensive review of deep learning applications in hydrology and water resources. *Water Sci. Technol.* **82** (12), 2635–2670. https://doi.org/10.2166/wst.2020.369.

Smets, K., Verdonk, B. & Jordaan, E. M. 2007 Evaluation of performance measures for SVR hyperparameter selection. In *2007 Int. Jt. Conf. Neural Netw.*, Orlando, FL, USA. IEEE, pp. 637–642.

Wang, X., Zhang, J., Shahid, S., Guan, E., Wu, Y., Gao, J. & He, R. 2016 Adaptation to climate change impacts on water demand. *Mitig. Adapt. Strateg. Glob. Change* **21** (1), 81–99. https://doi.org/10.1007/s11027-014-9571-6.

Wu, Y. & Liu, S. 2017 A review of data-driven approaches for burst detection in water distribution systems. *Urban Water J.* **14** (9), 972–983. https://doi.org/10.1080/1573062X.2017.1279191.

Xenochristou, M. & Kapelan, Z. 2020 An ensemble stacked model with bias correction for improved water demand forecasting. *Urban Water J.* **17** (3), 212–223. https://doi.org/10.1080/1573062X.2020.1758164.

Yu, P.-S., Yang, T.-C., Kuo, C.-M. & Chen, S.-T. 2014 Development of an integrated computational tool to assess climate change impacts on water supply–demand and flood inundation. *J. Hydroinf.* **16** (3), 710–730. https://doi.org/10.2166/hydro.2013.018.

Zanfei, A., Brentan, B. M., Menapace, A. & Righetti, M. 2022a A short-term water demand forecasting model using multivariate long short-term memory with meteorological data. *J. Hydroinf.* **24** (5), 1053–1065. https://doi.org/10.2166/hydro.2022.055.

Zanfei, A., Menapace, A., Brentan, B. M. & Righetti, M. 2022b How does missing data imputation affect the forecasting of urban water demand? *J. Water Resour. Plann. Manage.* **148** (11), 04022060. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001624.

Zanfei, A., Menapace, A., Brentan, B. M., Righetti, M. & Herrera, M. 2022c Novel approach for burst detection in water distribution systems based on graph neural networks. *Sustainable Cities Soc.* 104090. https://doi.org/10.1016/j.scs.2022.104090.

Zanfei, A., Menapace, A., Brentan, B. M., Sitzenfrei, R. & Herrera, M. 2023 Shall we always use hydraulic models? A graph neural network metamodel for water system calibration and uncertainty assessment. *Water Res.* **242**, 120264. https://doi.org/10.1016/j.watres.2023.120264.

Zhang, S. 2012 Nearest neighbor selection for iteratively kNN imputation. *J. Syst. Software* **85** (11), 2541–2552. https://doi.org/10.1016/j.jss.2012.05.073.

First received 4 January 2024; accepted in revised form 6 March 2024. Available online 19 March 2024