

Optimizing Short-Term Water Demand Forecasting: A Comparative Approach within the Battle of Water Demand Forecasting [†]

Bruno Ferreira ¹, Raquel Barreira ^{1,2}, João Caetano ¹, Maria Grazia Quarta ³, Nelson Carriço ^{1,*}

¹ RESILIENCE – Center for Regional Resilience and Sustainability, Escola Superior de Tecnologia do Barreiro, Instituto Politécnico de Setúbal, Portugal, bruno.s.ferreira@estbarreiro.ips.pt, raquel.barreira@estbarreiro.ips.pt, joao.caetano@estbarreiro.ips.pt

² Centro de Matemática, Aplicações Fundamentais e Investigação Operacional (CMAFclO), Universidade de Lisboa, Portugal

³ Dipartimento di Matematica e Fisica, Università del Salento, Lecce, Italy; mariagrazia.quarta@unisalento.it

* Correspondence: nelson.carriço@estbarreiro.ips.pt

[†] Presented at the 3rd WDSA/CCWI Joint Conference, Ferrara, Italy, 1-4 July 2024.

Abstract: The current paper presents a forecasting methodology for short-term water demand forecasting in the context of the Battle of Water Demand Forecasting. The methodology considers five distinct forecasting techniques, which are compared in terms of their forecasting ability for a preceding period, typically spanning a day or a week. The best-performing model is identified through error assessment between model predictions and actual measurements. This model is finally used to estimate the values for the forecasting horizon. This methodology directly considers the importance of tailoring the model to the specific case study and objectives. However, it is computationally intensive and relies on the fact that there will be not much variance between the preceding period and forecasting horizon results.

Keywords: model selection; short-term forecasting; water demand

1. Introduction

The sustainable management of urban water supply systems (UWSS) often rely on water demand forecasts for a range of purposes, namely, system design, maintenance, and operation. Water demand forecasting models can be distinguished according to the forecast horizon (i.e., the length of time for which forecasts are generated) and frequency (i.e. the time step at which water demand predicted are generated within the forecasting horizon) [1]. In broad terms, long-term models generally provide demand forecasts on a yearly or monthly basis with a forecasting horizon ranging from years to decades and are mainly used for planning and infrastructure design. Short-term models, by contrast, forecast water demand over more limited time horizons, ranging from days to months, with a time step ranging from daily to sub-hourly and are mainly used for operation and management purposes [1,2]. As noted by Ghalekhondabi *et al.* [3], neural networks and pattern-based methods are more commonly used for short-term forecasting, while econometric models are usually used for long-term forecasting.

In recent years, a plethora of research has been published related to forecasting methods for urban water demand. Niknam *et al.* [4] reviewed over 100 short-term water demand forecasting methods published between 2010 and 2022 and concluded that there is not a universal response to the question of what method one should use. In other words, the most adequate method should be selected among those showing better performance with respect both to the available data (e.g., existence of climatic data) and to the objectives of the forecast (e.g., average, or peak water demand). It was also found that a reasonable number of papers based their analysis on traditional time-series analysis (e.g., ARIMA models) and regression models (including multivariate regression, decision trees,

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Last-name

Published: date



Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

and random forests). These models have sufficient predictive ability for a range of operations in UWSS, whilst remaining interpretable and explainable to water utilities, specifically when compared with complex but less interpretable forecasting models such as the ones based on artificial neural network architectures or metaheuristics algorithms.

The current paper presents a forecasting methodology in the context of the Battle of Water Demand Forecasting (BWDF). The proposed methodology includes both pattern-based and regression techniques to improve the accuracy of the estimated values. Prior to forecasting, a pre-processing of the data and a model selection process is undertaken, wherein the best-performing model is identified through error assessment between model predictions and actual measurements from a preceding period, typically spanning a day or a week. This ensures that the optimal model, along with its parameter configuration, is chosen for forecasting considering the diverse variables such as area characteristics (e.g., proximity to maritime ports or city centres) and the importance (or not) of including weather data. The remainder of the paper is organized as follows. Section 2 presents the proposed methodology with a brief explanation of the distinct techniques and their respective parameters. Section 3 presents the main results for the four successive deliveries of the BWDF. Finally, Section 4 presents a discussion of the main benefits and drawbacks of the proposed methodology.

2. Materials and Methods

2.1. General framework

The proposed methodology is mainly driven by one of the major problems faced by water experts namely, which forecasting model should be used for a given specific case study, and which model parameters should be used. To face this problem, the assumption of no specific model will be the best in all possible situations is taken. This is done by considering five distinct forecasting techniques and by comparing their prediction ability for a preceding period, typically spanning a day or a week. The model presenting the best prediction ability for the preceding period is selected for forecasting. It is assumed that a forecasting model that presents good results for the preceding period will also present good results for the forecasting horizon.

Given the objective of the BWDF of forecasting a week of hourly water demand values for a specific case study, whilst using a given number of historical days of hourly water demand (e.g., one or two months), a general framework can be outlined into the following five steps:

1. Pre-processing: The data provided present missing values which limit the range of techniques that may be directly applied. The first steps consists on a pre-processing by seasonally decomposed missing value imputation using the R package *imputeTS* [5].
2. Model generation: The second step comprises the generation of distinct models for forecasting. This study uses five distinct forecasting techniques (as presented in 2.2). Each of these techniques contains specific parameters. Thus, a range of parameter combinations is devised for each technique. These parameter variations result in a set of candidate models for each forecasting technique.
3. Predicting of preceding period (hindcasting): Each of the candidate models generated in Step 2 is used to forecast a preceding period (e.g., a day or a week) leading up to the desired forecasting horizon. This step involves training the models on historical data preceding the target period, thereby assessing their predictive performance in a near-term context. Notably, this process may demand considerable computational resources, particularly when dealing with numerous parameter combinations or computationally intensive forecasting techniques.
4. Model evaluation and selection: The predicted values in Step 3 are compared with the real values using the mean absolute error, thus enabling ranking of each model

in terms of predictability performance in a near-term context. The model with the smallest error is selected for forecasting.

5. Forecasting: The model selected in Step 4 is finally used to estimate the values for the forecasting horizon. This step involves training the model on historical data preceding the forecasting horizon.

2.2. Forecasting techniques

Five distinct forecasting techniques are used in this study. Each technique (and its possible parameters) is briefly presented in the remainder of this section.

- The Naïve technique considers patterns either on a day-by-day basis or differentiating weekdays from weekends. For each timestamp in the forecast period, it identifies relevant historical data blocks based on the chosen pattern. Then, it calculates the predicted value using either the average, median or exponentially weighted moving average of the historical data block. The configurable parameters are the duration of historical data (e.g., one or two months), the pattern type (day-by-day basis or differentiating weekdays from weekends), and the type specified forecast type (average, median or exponentially weighted moving average).

- The support vector regression (SVR) technique uses multiple regression models to perform the forecast. Initially, it prepares the historical data by creating lagged values and by categorizing the data into weekdays, Saturdays, and Sundays/holidays. For each timestamp in the forecast period, an SVR model is trained using the relevant blocks in the prepared historical data. The trained SVR model is then used to forecast that specific timestamp. The configurable parameters are the duration of historical data, and the number of lagged values (e.g., previous 5 or 10 measurements at the same time of the day for the same type of weekday).

- The Quevedo technique firstly estimates the total daily volume for the day for which measurements are to be estimated. This is done using an ARIMA model. Then, this total daily volume is distributed to hourly values based on the average pattern for this weekday. The configurable parameters are the duration of historical data, and the pattern type (day-by-day basis or differentiating weekdays from weekends/holidays).

- Distinct from previous techniques, the XGBoost technique considers meteorological data (precipitation, air temperature, air relative humidity, wind speed). It firstly prepares the historical data with the necessary features (e.g., hour of the day, day of the week, weather variables or holiday indicators). Then, a regression model is trained using 80% of the historical data, with the remaining 20% being used to monitor training progress and prevent overfitting. The trained regression model is then used to forecast each specific timestamp (note that expected weather data is required in this phase). The configurable parameters are the duration of historical data, and multiple XGBoost-specific parameters (e.g., learning rate, number of estimators, max depth, and number of early stopping rounds).

- The Long short-term memory technique (LSTM) is a recurrent neural network that memorizes long-term dependencies of time series [6]. The configurable parameters are the duration of the historical data, some parameters related to the general neural network algorithm (e.g., dropout, batch size, number of units) and LSTM-specific parameters (e.g., window length). The LSTM technique with weather data works by using two LSTM modules: the first module deals with the historical water demand data, using past observations with a time horizon equal to the length of the window; the second module considers meteorological data (precipitation, temperature, relative humidity, wind speed) with a shorter window length horizon [7]. The configurable parameters are the same as for the LSTM, but in this case both for the water and the meteorological data.

3. Results

Table 1 presents the selected forecasting model for each of the DMA's and each of the delivering week (i.e., W1 to W4). It is possible to conclude that no unique technique presents the best solution for every situation. For instance, in W1 no technique was used in more than 3 DMA's. SVR is used in 6 out of 10 DMA's in W2, and in 5 out of 10 DMA's in W3. On the other hand, it is used only once in W4, whilst the Naïve method is used in 5 out of 10 DMA's. Furthermore, the parameters for the same DMA vary considerably (note that this data is not included on the paper due to space limitations). For instance, DMA F uses either 120 historical data days in W1 and 30 in W4. On the other hand, DMA I always resort to 30 historical data days throughout W1-W4.

Table 1. Selected forecasting model for each of the DMA's and each of the delivering weeks.

Week	DMA A	DMA B	DMA C	DMA D	DMA E	DMA F	DMA G	DMA H	DMA I	DMA J
W1	Naïve	Quevedo	Naïve	SVR	XGBoost	Naïve	Quevedo	SVR	Quevedo	XGBoost
W2	SVR	SVR	SVR	XGBoost	Quevedo	XGBoost	SVR	XGBoost	SVR	SVR
W3	XGBoost	SVR	Quevedo	Naïve	SVR	XGBoost	SVR	SVR	SVR	Naïve
W4	Naïve	Naïve	Naïve	LSTM-W	Naïve	LSTM-W	Naïve	SVR	LSTM-W	XGBoost

4. Discussion and Conclusions

The current paper presented a methodology for water demand forecasting combining multiple techniques to capture different aspects of the time series data. By hindcasting, it is possible to assess the predictive performance of distinct forecasting models in a near-term context. Finally, the model with the best predictive performance (i.e., smaller error between estimated and real values) is selected to perform the forecast. As drawbacks, it is highlighted that: 1) the computational cost can be relevant depending on the complexity of the forecasting techniques, and 2) there is no guarantee that a forecasting model that presents good results for the preceding period will also present good results for the forecasting horizon.

Author Contributions: All authors contributed equally to this study. All authors have read and agreed to the published version of the manuscript.

Funding: The authors thank the Instituto Politécnico de Setúbal for the funding granted. Maria Gra-zia Quarta was supported by a scholarship financed by the Ministerial Decree no. 351 of 9 April 2022, based on the NRRP - funded by the European Union - NextGenerationEU - Mission 4.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Available.

Data Availability Statement: Data available in the context of the BWDF.

Conflicts of Interest: There is no conflict of interest among the authors.

References

1. Pacchin, E.; Gagliardi, F.; Alvisi, S.; Franchini, M. A Comparison of Short-Term Water Demand Forecasting Models. *Water Resources Management* **2019**, *33*, 1481–1497, doi:10.1007/s11269-019-02213-y.
2. Mu, L.; Zheng, F.; Tao, R.; Zhang, Q.; Kapelan, Z. Hourly and Daily Urban Water Demand Predictions Using a Long Short-Term Memory Based Model. *J Water Resour Plan Manag* **2020**, *146*, doi:10.1061/(asce)wr.1943-5452.0001276.
3. Ghalekhondabi, I.; Ardjmand, E.; Young, W.A.; Weckman, G.R. Water Demand Forecasting: Review of Soft Computing Methods. *Environ Monit Assess* **2017**, *189*, doi:10.1007/s10661-017-6030-3.
4. Niknam, A.; Zare, H.K.; Hosseiniinasab, H.; Mostafaeipour, A.; Herrera, M. A Critical Review of Short-Term Water Demand Forecasting Tools—What Method Should I Use? *Sustainability (Switzerland)* **2022**, *14*.
5. Moritz, S.; Bartz-Beielstein, T. ImputeTS: Time Series Missing Value Imputation in R. *R J* **2017**, *9*, 207, doi:10.32614/RJ-2017-009.
6. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput* **1997**, *9*, 1735–1780, doi:10.1162/neco.1997.9.8.1735.
7. Zanfei, A.; Brentan, B.M.; Menapace, A.; Righetti, M. A Short-Term Water Demand Forecasting Model Using Multivariate Long Short-Term Memory with Meteorological Data. *Journal of Hydroinformatics* **2022**, *24*, 1053–1065, doi:10.2166/hydro.2022.055.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.