*engineering proceedings*

MDPI

*Proceeding Paper*

# Short-Term Water Demand Forecasting Using Machine Learning Approaches in a Case Study of a Water Distribution Network Located in Italy †

**Qidong Que [1], Jinliang Gao [1,*], Wenyan Wu [2], Huizhe Cao [1], Kunyi Li [1], Hanshu Zhang [1], Yi He [1] and Rui Shen [1]**

[1] School of Environment, Harbin Institute of Technology, Harbin 150001, China; 18153458935@163.com (Q.Q.); caohz@hit.edu.cn (H.C.); 22b929020@stu.hit.edu.cn (K.L.); 23s029024@stu.hit.edu.cn (H.Z.); 23s129138@stu.hit.edu.cn (Y.H.); 23s136664@stu.hit.edu.cn (R.S.)

[2] School of Engineering and the Built Environment, Birmingham City University, Birmingham B4 7XG, UK; wen-yan.wu@bcu.ac.uk

* Correspondence: gjl@hit.edu.cn; Tel.: +86-13936477968

† Presented at the 3rd International Joint Conference on Water Distribution Systems Analysis & Computing and Control for the Water Industry (WDSA/CCWI 2024), Ferrara, Italy, 1–4 July 2024.

**Abstract:** Machine learning's application in short-term water demand forecasting remains a pivotal area of research in water distribution system studies. This investigation reveals a distinctive distribution pattern for the daily demand following dataset preprocessing with Random Forest and the quartile method. Inspired by the findings, this study introduces a novel Water Demand Forecast Framework (WDFF) using DMA characteristics and the CNN–Attention–LSTM architecture. By analyzing the relationship between the total and DMA-specific demand, the WDFF is found to enhance the predictions. It demonstrates expedited convergence and reduces the loss metric, demonstrating its potential to elevate the predictive precision in water demand forecasting.

**Keywords:** short-term water demand forecasting; time series; feature engineering; convolutional neural network; long short-term memory network; attention mechanism

check for updates

**Citation:** Que, Q.; Gao, J.; Wu, W.; Cao, H.; Li, K.; Zhang, H.; He, Y.; Shen, R. Short-Term Water Demand Forecasting Using Machine Learning Approaches in a Case Study of a Water Distribution Network Located in Italy. *Eng. Proc.* **2024**, *69*, 177. https://doi.org/10.3390/engproc2024069177

Academic Editors: Stefano Alvisi, Marco Franchini, Valentina Marsili and Filippo Mazzoni

Published: 29 September 2024

**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Short-term water demand forecasting plays a pivotal role in the operational, tactical, and strategic decision-making processes of water utilities. Accurate and reliable forecasts serve as a foundation for these critical decisions. Extensive research has been conducted on addressing this time series forecasting challenge, categorizing the approaches into traditional methods and learning algorithms [1,2].Traditional techniques, such as the Autoregressive Integrated Moving Average (ARIMA) model, the Seasonal ARIMA (SARIMA) model, and linear regression models, are widely applied due to their simplicity in interpretation and implementation. However, their predictive accuracy is often compromised by the inherently nonlinear nature of water demand patterns [3]. Conversely, learning algorithms, including cutting-edge machine learning or deep learning techniques, have gained significant traction in numerous fields for their adeptness in handling nonlinear issues, making them a focal point of current research in water demand forecasting.

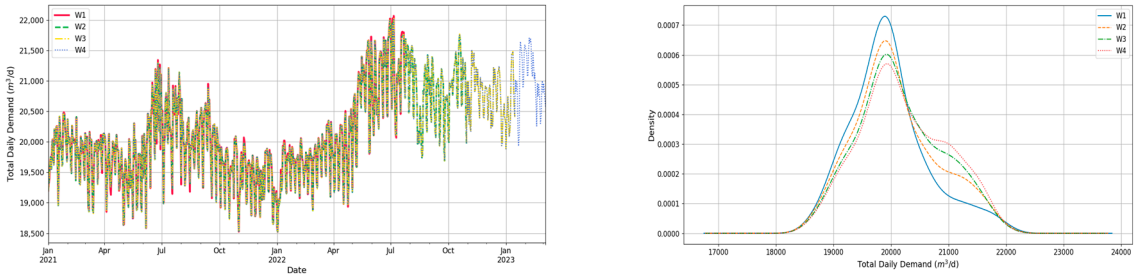
In the realm of deep learning, LSTM networks excel in predicting time series data, making them particularly effective for short-term water demand forecasting [4]. A crucial consideration in these models is the selection of the input parameters. Zubaidi et al. [5] indicate that the weather conditions, alongside water demand data, are significantly correlated with the water demand, thus serving as valuable inputs for modeling. This paper proposes a technique that employs CNN–Attention mechanisms to enhance feature extraction, subsequently integrating these features into the LSTM network for improved forecasting accuracy.

Guo et al. [6] highlighted that LSTM networks tend to accumulate errors in long-term, multi-step predictions. Therefore, considering the necessity to forecast multiple DMAs in this task, this paper introduces a novel correction technique that accounts for the relationships between individual DMAs and the total water demand. The detailed methodology is discussed later in the text.

## 2. Materials and Methods

### *2.1. Dataset Preprocessing*

This work was conducted in phases, providing four instances of net inflow data and weather data that were not post-processed and had missing data issues. Missing values in the weather data were imputed using Random Forest, and, for different DMAs, missing values were filled by averaging the water demand from temporally similar hours. Subsequent outlier removal in the flow data was performed using the quartile method. Exploratory data analysis revealed that the total daily flow volume across ten DMAs followed a certain distribution. Figure 1 presents the processed total daily flow data for the forecasting of the four evaluation weeks and its KDE function, with $\mu_{w1} = 19,964.51$, $\sigma_{w1} = 681.25$, $\mu_{w2} = 20,085.91$, $\sigma_{w2} = 732.00$, $\mu_{w3} = 20,154.49$, $\sigma_{w3} = 733.90$, $\mu_{w4} = 20,208.76$, $\sigma_{w4} = 749.95$.




(**a**) (**b**)

**Figure 1.** Exploratory data analysis on preprocessed original net inflow data (**a**) used to compare the total daily demand inputs for different evaluation weeks; (**b**) kernel density estimation (KDE) function of the daily flow data input to the model.

### *2.2. Water Demand Forecast Framework*

The Water Demand Forecast Framework (WDFF) is composed of three modules: the input module, forecast module, and correction module. Figure 2 illustrates their interconnections. The first module defines the data input into the forecast model, which include the demands of 10 DMAs at an hourly resolution, as well as historical data on the temporal and weather conditions [4]. Given the discovery that the daily total water demand conforms to a certain distribution, the model further integrates inputs reflecting the proportion of each DMA's daily demand to the total water demand and temperature features. The second module employs CNN–Attention–LSTM network architectures that are structurally identical but vary in their parameters. Historical data from DMA A-J are processed by the forecast module, which outputs uncorrected predicted demands, while the daily demand weight outputs the predicted proportion of each DMA's demand. The third module is a flexible and effective module that corrects the output of the forecast module using fully connected layers (FC). FC1 uses the weather and temporal features of the forecast day for the first correction of the results, followed by the second correction by FC2 to obtain the final results. The specific dimensions of each tensor will be presented in subsequent work.

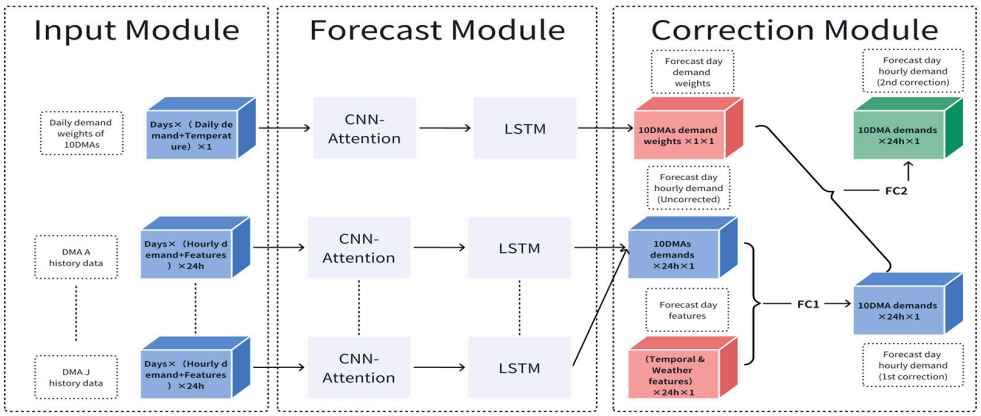


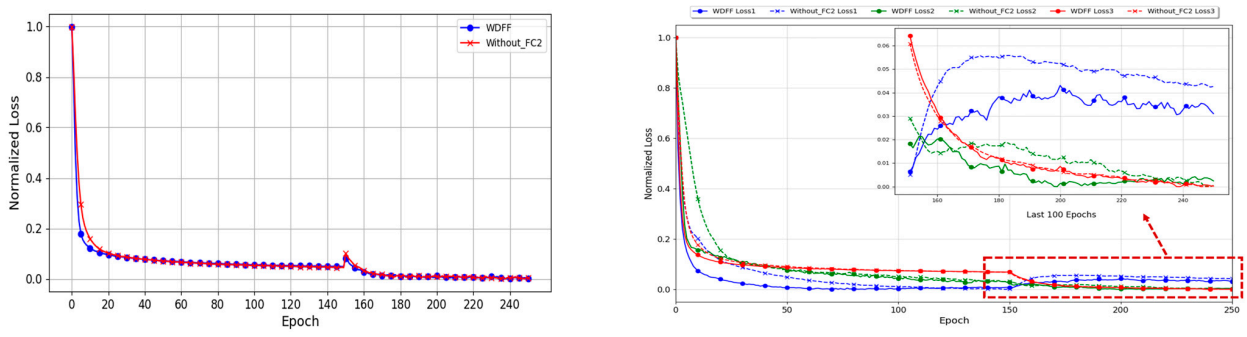

**Figure 2.** Structural schematic of the Water Demand Forecast Framework (WDFF).

The input module is divided into 11 input sections (10 DMAs and one weight), with the number of days for each input varying from 14 to 28 days, employing a sliding time series input. To ensure consistency in the number of forecast days, the prediction will start from the 29th day of the dataset. In this study, the evaluation metric employs the L1 loss during the model's training. The assessment of the predictive performance utilizes three criteria proposed by the organizing committee (PI 1–3).

## 3. Results and Discussion

This paper presents and discusses the data based on the W4 results. For the purpose of the algorithm performance comparison, this study eliminated the weight correction module depicted in Figure 2, employing a conventional weather-impacted water demand forecast framework (without FC2), wherein the results are output directly following FC1. The evaluation metrics for the test set are represented by the mean of PI 1 for all DMAs, the maximum of PI 2 for all DMAs, and the mean of PI 3 for all DMAs, which are displayed using loss 1–3, respectively.

In the original data, the range of the training loss for the WDFF and without FC2 before normalization was, respectively, [11.37, 120.19] and [12.82, 117.40]. Figure 3a shows that both the WDFF and without FC2 test sets exhibit a rapid convergence trend, yet the WDFF approaches convergence more swiftly within the first 20 epochs. For the original data's test loss, the ranges for the WDFF's loss 1–3 were [0.71, 6.24], [6.74, 99.06], and [0.99, 14.06], respectively, while, for the case without FC2, they were [0.70, 6.67], [9.13, 110.18], and [1.21, 13.40]. Figure 3b presents the comparison of the data after normalization, revealing that the WDFF's three losses were consistently lower than the others for the first 100 epochs. In the subsequent 150 epochs, both models' loss 1 slightly increased before continuing to decrease. In the convergence phase, the WDFF still outperformed the model without FC2.




(**a**) (**b**)

**Figure 3.** Performance evaluation of the training and test sets: (**a**) L1 loss of the training set after normalization; (**b**) loss 1, loss 2, and loss 3 of the test set.

In summary, the rapid convergence and superior predictive capabilities demonstrated by the WDFF in this water demand forecasting task highlight its significant potential for application in urban water demand prediction.

**Author Contributions:** Conceptualization, Q.Q.; methodology, Q.Q.; software, Q.Q.; validation, Q.Q., Y.H. and W.W.; formal analysis, Q.Q.; investigation, Q.Q.; resources, Q.Q.; data curation, Q.Q.; writing—original draft preparation, Q.Q.; writing—review and editing, J.G. and W.W.; visualization, H.C.; supervision, K.L.; project administration, H.Z. and R.S.; funding acquisition, J.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key Research and Development Program of China, grant number 2022YFC3203800; the National Natural Science Foundation of China, grant number 51978203; the Unveiling Scientific Research Program, grant number CE602022000203; and the Key Research and Development Program of Heilongjiang Province of China, grant number 2022ZX01A06.

**Institutional Review Board Statement:** Not applicable for studies not involving humans or animals.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** For access, contact the corresponding author.

**Acknowledgments:** The authors would like to thank the 3rd International Joint Conference on Water Distribution Systems Analysis & Computing and Control for the Water Industry.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Donkor, E.A.; Mazzuchi, T.A.; Soyer, R.; Roberson, J.A. Urban Water Demand Forecasting: Review of Methods and Models. *J. Water Resour. Plan. Manag.* **2014**, *140*, 146–159. [CrossRef]
2. Ghalehkhondabi, I.; Ardjmand, E.; Young, W.A.; Weckman, G.R. Water demand forecasting: Review of soft computing methods. *Environ. Monit. Assess.* **2017**, *189*, 313. [CrossRef] [PubMed]
3. Braun, M.; Bernard, T.; Piller, O.; Sedehizade, F. 24-Hours Demand Forecasting Based on SARIMA and Support Vector Machines. *Procedia Eng.* **2014**, *89*, 926–933. [CrossRef]
4. 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. *J. Hydroinformatics* **2022**, *24*, 1053–1065. [CrossRef]
5. Zubaidi, S.L.; Gharghan, S.K.; Dooley, J.; Alkhaddar, R.M.; Abdellatif, M. Short-Term Urban Water Demand Prediction Considering Weather Factors. *Water Resour. Manag.* **2018**, *32*, 4527–4542. [CrossRef]
6. Guo, G.; Liu, S.; Wu, Y.; Li, J.; Zhou, R.; Zhu, X. Short-Term Water Demand Forecast Based on Deep Learning Method. *J. Water Resour. Plan. Manag.* **2018**, *144*, 04018076. [CrossRef]

**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.