

Proceeding Paper

A Multivariate LSTM Model for Short-Term Water Demand Forecasting [†]

Aly K. Salem  and Ahmed A. Abokifa * 

Department of Civil, Materials, and Environmental Engineering, The University of Illinois Chicago, Chicago, IL 60607, USA; asalem22@uic.edu

* Correspondence: abokifa@uic.edu; Tel.: +1-312-413-4636

[†] 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: Accurate water demand forecasting is crucial for the effective operation and management of water distribution networks. Predicting future water demand empowers utilities to optimally operate system components. Various data-driven methodologies have been proposed for water demand forecasting, including artificial neural networks and econometric models. Recently, Long Short-Term Memory (LSTM) was shown to be particularly relevant for this application. Nevertheless, few studies have utilized multivariate-LSTM (M-LSTM) models for water demand forecasting. This study introduces an M-LSTM model incorporating historical water demands, meteorological data, and social variables to forecast short-term water demand. The proposed M-LSTM model performance was tested by applying it to the ten district metered areas (DMAs) case study of the Battle of Water Demand Forecasting (BWDF). The results demonstrated the model's ability to accurately predict the hourly water demand one week in advance. The mean absolute error of the predictions ranged between 0.5 and 2.2 l/s (2.8% to 12.9% of the average demand). The results also showed a strong correlation between the prediction error and the variability of the water demand data.

Keywords: water demand forecast; water distribution networks; machine learning; LSTM



Citation: Salem, A.K.; Abokifa, A.A. A Multivariate LSTM Model for Short-Term Water Demand Forecasting. *Eng. Proc.* **2024**, *69*, 167. <https://doi.org/10.3390/engproc2024069167>

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

Published: 25 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

The fundamental task of a water distribution network is to provide the necessary water supply under varying conditions. However, this task is typically challenged by the inherently stochastic nature of water demand patterns. Therefore, forecasting future water demand is considered a critical step in the operation and management of water distribution networks [1]. Utilities heavily rely on water demand forecasting to effectively operate essential system components such as water treatment plants, pumps, and valves [2]. This reliance is becoming even more urgent as water scarcity looms due to climate change and population growth [1]. Beyond facilitating the day-to-day operation of water supply systems, water demand forecasting plays a pivotal role in optimizing energy consumption [3] and managing system expansions [4].

Based on the desired scope, water demand forecasting can be categorized into short-term and long-term. Short-term forecasting typically extends up to three months in advance [4], while predictions with larger horizons fall under long-term forecasting. Various methodologies have been developed to address both short-term and long-term water demand forecasting. Short-term forecasting often leverages data-driven models, including artificial neural network models, support vector machines, long short-term memory (LSTM), and random forests [5]. Conversely, long-term forecasting commonly employs simulation and econometric models [1].

Despite the extensive research conducted on LSTM models for water demand forecasting [6,7], limited attention has been given to exploiting the multivariate capabilities of these models, often restricting them to single-variable forecasts. A recent contribution

by Zanfei et al. [8] introduced a Multivariate LSTM (M-LSTM) model for water demand forecasting; however, this model is constrained to one-day prediction. In this study, we present our contribution to the Battle of Water Demand Forecasting (BWDF) by advancing upon the work in the literature by developing a Multivariate LSTM (M-LSTM) model tailored for short-term demand forecasting, specifically targeting the prediction of the hourly water demand for one week in the future. The proposed M-LSTM model incorporates the historical water demand data along with meteorological data, with the aim of enabling proactive resource management several days ahead.

2. Materials and Methods

This study introduces a multivariate long short-term memory (M-LSTM) model designed to perform hourly water demand predictions for one week in advance, considering three groups of input features. The first group represents the meteorological data, encompassing air temperature, air humidity, wind speed, and rainfall depth. The second group represents the temporal characteristics, including hours, weekdays, and holidays (encoded as a binary input, where 0 denotes a standard day, and 1 denotes a holiday). The third group represents the measured water demand. The M-LSTM operates in a many-to-many fashion, wherein all inputs are in time series format, as well as the model outputs. In this study, the length of the input time series was set to one week (168 h). The workflow of the proposed M-LSTM model is illustrated in Figure 1.

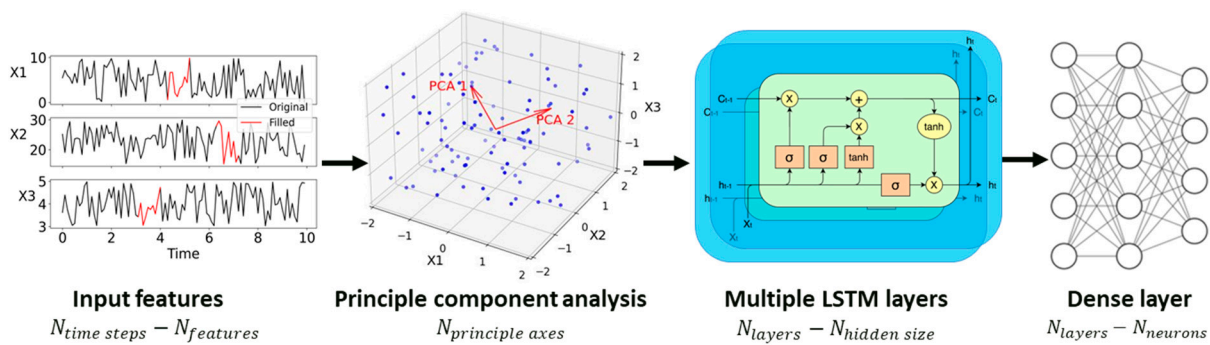


Figure 1. The workflow of the proposed M-LSTM model.

2.1. Data Preprocessing

Sensory measured data commonly experience the presence of missing values, often attributed to sensor faults. Appropriately filling in these missing values is fundamental for constructing a reliable model. In our study, the missing values were filled by employing a moving hourly basis average approach. The same approach was used for all input time series. This approach involves filling in missing values by computing the average among the specific hours over the preceding n days, where n is a user-defined parameter set to ten in our case. Subsequently, after filling in all missing data, each input feature underwent scaling. The standard and summer times were handled by removing the second instance of the duplicated hour associated with the standard time start and adding the missing hour related to the summer time start.

Before feeding the previously mentioned input features to the model, a Principal Component Analysis (PCA) was employed to transform the original inputs into more descriptive features. This process reduces the dimensionality of the forecast problem from eight to five dimensions. Furthermore, hyperparameter optimization was conducted to fine-tune the parameters of the M-LSTM model on a district-metered area (DMA) basis. The PCA, hyperparameter optimization, and training of the M-LSTM model were coded in Python, leveraging the Scikit-learn, Hyperopt, and Pytorch libraries, respectively.

2.2. Model Training, Validation, and Testing

The metrological and water demand data provided by the BWDF committee for ten district metered areas (DMAs) formed the basis for the training, validation, and testing of the M-LSTM model. Due to imposed limitations on data availability, the model development passed through the four stages detailed in Table 1. In each stage, the available data was partitioned into weeks, where 60% of the available weeks were allocated for training, 20% for validation, and 20% for testing. The testing weeks were used to assess the accuracy of the model, as represented in the results section. The subsequent week represents the BWDF target evaluation week, where the true demand is unknown.

Table 1. The training stages and their corresponding weeks.

Training Stage	Training, Validation, and Testing Weeks	Evaluation Week ¹
1	1/2021 to 29/2022	30/2022
2	Stage 1 + 31/2022 to 43/2022	44/2022
3	Stage 2 + 45/2022 to 2/2023	3/2023
4	Stage 3 + 4/2022 to 3/2023	10/2023

¹ Exact water demand data is unknown.

2.3. Loss Function

The M-LSTM model was trained to minimize the summation of three performance indicators (PIs) set by the BWDF committee, namely, $PI1$, $PI2$, and $PI3$. These PIs feature the mean and maximum absolute errors of the first day and the mean absolute error of the remaining days, respectively. The mathematical representation of these PIs is given by Equations (1)–(3).

$$PI1 = \frac{1}{24} \sum_{h=1}^{24} |D_h - \hat{D}_h| \quad (1)$$

$$PI2 = \max\{|D_1 - \hat{D}_1|, |D_2 - \hat{D}_2|, \dots, |D_{24} - \hat{D}_{24}|\} \quad (2)$$

$$PI3 = \frac{1}{144} \sum_{h=25}^{168} |D_h - \hat{D}_h| \quad (3)$$

where D_h and \hat{D}_h are the actual and predicted water demand in l/s.

3. Results and Discussion

Figure 2a depicts the performance of the M-LSTM model on testing datasets across ten distinct metered areas (DMAs). The x-axis denotes the training stages, while the y-axis denotes the values of three performance indicators (PIs), namely, $PI1$, $PI2$, and $PI3$. The governing trend indicates a progressive decrease in PIs as training stages advance, which is attributed to the expansion of available training data. Notably, the model demonstrates significantly enhanced performance starting from Stage 2. However, some fluctuations are observed between training stages, predominantly in a positive direction indicating reduced errors. Nevertheless, few instances exhibit deviations in the negative direction, potentially attributable to disparities between testing datasets across different stages.

To investigate the variations in prediction accuracies across different DMAs, a correlation analysis was conducted between the normalized Performance Indicators (PIs), denoted as the PI normalized by the mean water demand within each DMA, and the coefficient of variation of the DMA water demand time series. The correlation results, presented in Figure 2b, reveal a moderate positive correlation across all PIs. This correlation suggests that the performance of the M-LSTM model tends to improve when less variability exists in the time series training data.

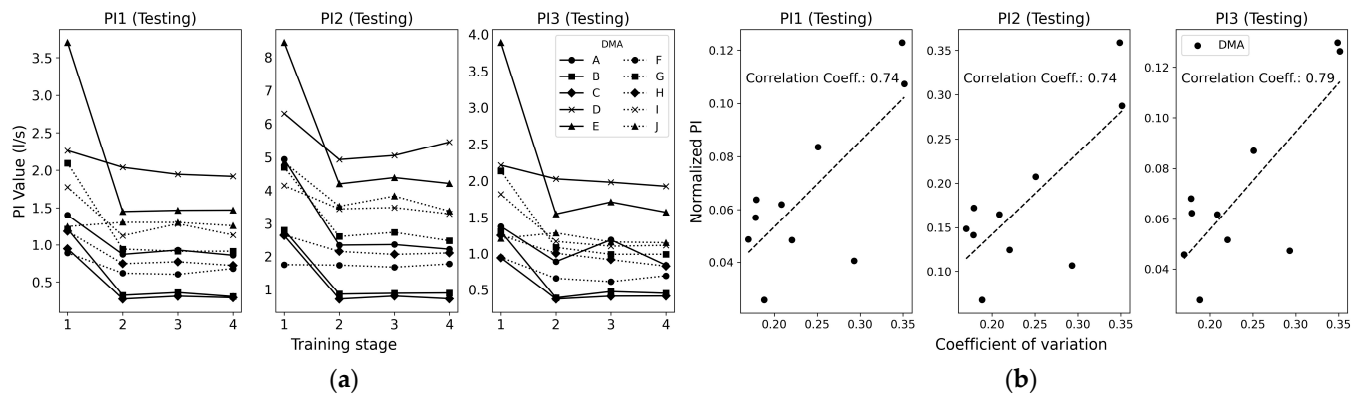


Figure 2. (a) Model results across different DMAs; (b) Correlation analysis results.

Author Contributions: Conceptualization, A.K.S.; methodology, A.K.S.; software, A.K.S.; validation, A.K.S.; formal analysis, A.K.S.; investigation, A.K.S.; resources, A.K.S.; data curation, A.K.S.; writing—original draft preparation, A.K.S.; writing—review and editing, A.A.A.; visualization, A.K.S.; supervision, A.A.A.; project administration, A.A.A.; funding acquisition, A.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Science Foundation Grant No. 2015603.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this paper are available upon reasonable request from the corresponding author.

Acknowledgments: Support from the NSF is gratefully acknowledged.

Conflicts of Interest: The authors declare no competing interests.

References

1. Ghalekhondabi, I.; Ardjmand, E.; Young, W.A., II; Weckman, G.R. Water demand forecasting: Review of soft computing methods. *Environ. Monit. Assess.* **2017**, *189*, 313. [[CrossRef](#)] [[PubMed](#)]
2. Bakker, M.; Vreeburg, J.H.G.; Van Schagen, K.M.; Rietveld, L.C. A fully adaptive forecasting model for short-term drinking water demand. *Environ. Model. Softw.* **2013**, *48*, 141–151. [[CrossRef](#)]
3. Alvisi, S.; Franchini, M.; Marinelli, A. A short-term, pattern-based model for water-demand forecasting. *J. Hydroinform.* **2007**, *9*, 39–50. [[CrossRef](#)]
4. Donkor, E.A.; Mazzuchi, T.A.; Soyer, R.; Alan Roberson, J. Urban Water Demand Forecasting: Review of Methods and Models. *J. Water Resour. Plan. Manag.* **2014**, *140*, 146–159. [[CrossRef](#)]
5. Niknam, A.; Zare, H.K.; Hosseinasab, H.; Mostafaeipour, A.; Herrera, M. A Critical Review of Short-Term Water Demand Forecasting Tools—What Method Should I Use? *Sustainability* **2022**, *14*, 5412. [[CrossRef](#)]
6. Pacchin, E.; Gagliardi, F.; Alvisi, S.; Franchini, M. A Comparison of Short-Term Water Demand Forecasting Models. *Water Resour. Manag.* **2019**, *33*, 1481–1497. [[CrossRef](#)]
7. Li, M.; 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*, 05020017.
8. 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. Hydroinform.* **2022**, *24*, 1053–1065. [[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.