engineering proceedings

MDPI

Proceeding Paper

An Ensemble Data-Driven Approach for Enhanced Short-Term Water Demand Forecasting in Urban Areas †

Amin E. Bakhshipour [1,*], Hossein Namdari [2], Alireza Koochali [3,4], Ulrich Dittmer [1] and Ali Haghighi [1,2,5]

1 Institute of Urban Water Management, RPTU in Kaiserslautern, 67663 Kaiserslautern, Germany; ulrich.dittmer@rptu.de (U.D.); ali.haghighi@rptu.de (A.H.)
2 Faculty of Civil Engineering and Architecture, Shahid Chamran University of Ahvaz, Ahvaz 61357-83151, Iran; namdari755@gmail.com
3 German Research Center for Artificial Intelligence (DFKI), 67663 Kaiserslautern, Germany; alireza.koochali@dfki.de
4 Ingenieurgesellschaft Auto und Verkehr (IAV) GmbH, 10587 Berlin, Germany
5 Sustainable Water Infrastructure Solutions (SWIS) GmbH, 67663 Kaiserslautern, Germany
* Correspondence: amin.bakhshipour@rptu.de
† 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: This study introduces an innovative ensemble data-driven model designed for short-term water demand forecasting within urban areas. By synergistically combining three distinct machine learning approaches—NHiTS, XGBoost regression, and a multi-head 1D convolutional neural network—our model enhances forecasting accuracy and reliability. This integration not only leverages the unique strengths of each method but also compensates for their individual weaknesses, resulting in a robust solution for predicting urban water demand. Tested against the Battle of Water Demand Forecasting dataset (WDSA-CCWI-2024), our ensemble model demonstrates superior performance, offering a promising tool for efficient water resource management and decision making.

Keywords: water demand forecasting; deep learning; ensemble learning; time series analysis

check for updates

Citation: Bakhshipour, A.E.; Namdari, H.; Koochali, A.; Dittmer, U.; Haghighi, A. An Ensemble Data-Driven Approach for Enhanced Short-Term Water Demand Forecasting in Urban Areas. *Eng. Proc.* **2024**, *69*, 69. https://doi.org/10.3390/engproc2024069069

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

Published: 5 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

Water distribution networks (WDNs) are critical infrastructures that ensure the provision of water to city inhabitants at the required pressure and quantity. The optimal operation of these networks is paramount, with short-term and hourly water demand forecasting being a crucial aspect of their management [1]. Traditional forecasting methods have been categorized into linear and nonlinear approaches, with the former including techniques like exponential smoothing and ARIMA and the latter encompassing methods such as nonlinear regression and artificial neural networks (ANNs) [2,3]. Despite the application of various regression and ANN models in past studies, the challenge of accurately forecasting short-term water demand persists, primarily due to the nonlinear and multifaceted nature of water demand patterns.

Recent advancements have seen the development of deep neural networks (DNNs), including convolutional neural networks (CNNs), and recurrent neural networks (RNNs), which offer improved feature extraction and the ability to learn temporal dependencies. However, the selection of appropriate features and the risk of overfitting remain significant challenges [1].

Ensemble methods, integrating multiple machine learning models, offer a robust approach to enhance forecasting accuracy by leveraging the strengths and mitigating the weaknesses of individual models. Despite their success in various fields, their application in water demand forecasting remains underexplored. This approach presents a significant opportunity to improve the precision of short-term water demand predictions in urban settings. This study bridges these gaps by deploying an ensemble data-driven model, as detailed in Section 2, that integrates various machine learning techniques. This integration boosts the accuracy and reliability of short-term water demand forecasting.

2. Materials and Methods

This section outlines our methodology, focusing on the ensemble data-driven model that enhances short-term water demand forecasting. We detail the integration of three key machine learning models: NHiTS, XGBoost regression, and a multi-head 1D CNN.

2.1. NHiTS

N-HiTS, short for Neural Hierarchical Interpolation for Time Series forecasting [4], advances the N-BEATS model by enhancing prediction accuracy and reducing computational demands. It achieves this by sampling time series at varying rates, allowing the model to capture both immediate and overarching trends within the data. Through the process of hierarchical interpolation, N-HiTS combines forecasts from different time scales, effectively balancing short-term and long-term effects for more accurate predictions. This methodological innovation positions N-HiTS as a pivotal element in our ensemble model, aimed at refining urban water demand forecasting. For a detailed exploration of N-HiTS and its application in time series forecasting, refer to the work by [4]. We implemented N-HiTS within our ensemble model using the NeuralForecast library [5]. To optimize performance, we employed specific hyperparameter configurations. The horizon parameter was adjusted to match the input size, ensuring efficient processing. Additionally, binary co-variates were included to account for the significant impact of holidays on water demand variations. Data normalization was achieved through a standard scaler, promoting faster training and convergence. Finally, the maximum steps parameter was tailored to the forecast horizon, set at 20,000 for a 168 h horizon and 10,000 for a 24 h horizon.

2.2. XGBoost

XGBoost regression, a highly efficient and scalable variant of gradient boosting, is renowned for its performance in various machine learning competitions [6]. In this study, we adapt the time series prediction challenge into a classic supervised learning scenario, utilizing tabular data as features. This transformation allows us to incorporate a comprehensive set of features for each timestep, including zone, hour, holiday, day of the week, month, and year, thereby enriching the model's contextual understanding of water demand patterns. To fine-tune the model's hyperparameters, we employ a combination of grid search and cross-validation techniques, ensuring the selection of the most effective parameters for our specific forecasting needs. The optimal configuration identified for our test case includes n_estimators = 2000, learning_rate = 0.15, max_depth = 8, colsample_bytree = 0.9, and gamma = 0.05. This tailored approach significantly enhances the model's predictive accuracy, making XGBoost regression a pivotal component of our ensemble forecasting model.

2.3. Multi-Head 1D CNN

Two-dimensional convolutional neural networks (2D CNNs) are widely utilized in image and video recognition, classification, medical image analysis, and natural language processing. These networks, a subset of feedforward artificial neural networks (ANNs), incorporate convolutional and pooling layers. A specialized adaptation, known as 1D CNNs, has been developed for more efficient data processing. The key distinction between 2D and 1D CNNs lies in the use of 1D arrays for convolution and pooling layers, rather than 2D matrices, significantly reducing computational costs [7]. This makes 1D CNNs particularly suited for signal processing and time series data analysis.

Our multi-head 1D CNN model processes input data (24 or 168 previous timesteps) through multiple 1D CNN heads, each employing distinct filter sizes. The outputs from these heads are then concatenated and passed through flattened layers, where data are transformed into a vector before entering a fully connected layer that predicts water demand for the forthcoming 24 or 168 h. This study utilizes a multichannel approach with four 1D CNN channels to enhance forecasting accuracy. For further information, please refer to [1].

Conclusively, a final optimization step was undertaken to fine-tune the weights of each model within our ensemble, resulting in an optimal distribution of 0.3, 0.4, and 0.3 for NHiTS, XGBoost, and CNN, respectively.

3. Results and Discussion

We evaluated our methodology's performance utilizing the Battle of Water Demand Forecasting (WDSA-CCWI-2024) dataset for predicting water demand across ten zones with 24- and 168-timestep horizons. Our models were tested on data from the first week of 2023, with the period from 2021 to the end of 2022 serving as the training dataset. Table 1 presents a detailed comparison of forecasting performances across four models evaluated over ten urban zones for short-term water demand forecasting. The ensemble model emerges as the standout performer, consistently achieving the lowest mean absolute error (MAE) for both 24 h and 144 h forecast horizons and maximum error (max) for 24 h. Its superior performance is particularly notable in zones A, C, G, I, and J. Table 2 presents the average performance of each model. Here, we can see that the ensemble model outperforms other models in all criteria. This suggests the ensemble model captures the underlying trends in the data better on average.

Table 1. Comparative performance of forecasting models in each zone.

Zone	Model/Criteria	NHITS	XGB	CNN	Ensemble	Zone	Model/Criteria	NHITS	XGB	CNN	Ensemble
A	MAE 24	0.69	0.68	0.62	0.57	F	MAE 24	0.91	0.58	0.64	0.62
	Max 24	0.56	0.55	0.66	0.45		Max 24	0.80	0.63	0.70	0.62
	Max 144	1.79	1.82	1.71	1.65		Max 144	2.10	3.09	2.53	2.67
B	MAE 24	0.20	0.16	0.23	0.17	G	MAE 24	0.69	0.69	0.79	0.62
	Max 24	0.27	0.18	0.27	0.19		Max 24	0.81	0.70	1.06	0.66
	Max 144	0.78	0.59	0.62	0.65		Max 144	1.90	1.53	1.77	1.47
C	MAE 24	0.15	0.18	0.17	0.14	H	MAE 24	0.88	0.75	0.95	0.79
	Max 24	0.14	0.17	0.25	0.13		Max 24	0.89	0.81	1.10	0.80
	Max 144	0.40	0.76	0.60	0.42		Max 144	2.26	2.16	3.39	2.22
D	MAE 24	1.96	1.11	1.63	1.28	I	MAE 24	1.08	0.83	1.32	0.76
	Max 24	2.30	1.59	1.93	1.67		Max 24	1.24	0.95	1.18	0.95
	Max 144	4.78	3.04	5.19	3.99		Max 144	3.04	1.74	3.49	1.49
E	MAE 24	0.70	0.88	0.89	0.76	J	MAE 24	1.62	1.28	1.85	0.82
	Max 24	0.86	1.00	1.37	0.78		Max 24	1.24	1.11	1.48	1.13
	Max 144	2.82	2.31	3.27	2.52		Max 144	3.62	3.25	4.59	2.10

Table 2. Average performance of forecasting models.

Criteria/Model	NHITS	XGB	CNN	Ensemble
MAE 24	0.89	0.71	0.91	0.65
Max 24	0.91	0.77	1.00	0.74
Max 144	2.35	2.03	2.72	1.92

4. Conclusions

In conclusion, this study introduces a robust ensemble data-driven model that integrates the strengths of NHiTS, XGBoost regression, and multi-head 1D CNN to enhance the accuracy and reliability of short-term water demand forecasting in urban environments. Through comprehensive testing on the Battle of Water Demand Forecasting dataset, our approach demonstrated superior performance over individual models, particularly in predicting water demand across ten zones for both 24- and 168-timestep horizons. The strategic weighting of each model within the ensemble further optimized forecasting accuracy, showcasing the potential of combining diverse machine learning techniques for effective water demand prediction. Our findings contribute valuable insights to the field of urban water management, offering a promising tool for decision makers to ensure sustainable and efficient water distribution.

Author Contributions: Conceptualization, A.K.; methodology, A.E.B. and H.N.; validation, A.K.; data curation, H.N.; writing—original draft preparation, A.E.B.; writing—review and editing, U.D. and A.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available from the corresponding author by request.

Conflicts of Interest: Author Alireza Koochali was employed by the company Ingenieurgesellschaft Auto und Verkehr (IAV) GmbH. Author Ali Haghighi was employed by the company Sustainable Water Infrastructure Solutions (SWIS) GmbH. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Namdari, H.; Ashrafi, S.M.; Haghighi, A. Deep learning–based short-term water demand forecasting in urban areas: A hybrid multichannel model. *AQUA—Water Infrastruct. Ecosyst. Soc.* **2024**, *73*, 380–395. [CrossRef]
2. Ristow, D.C.M.; Henning, E.; Kalbusch, A.; Petersen, C.E. Models for forecasting water demand using time series analysis: A case study in Southern Brazil. *J. Water Sanit. Hyg. Dev.* **2021**, *11*, 231–240. [CrossRef]
3. 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. Plann. Manag.* **2020**, *146*, 05020017. [CrossRef]
4. Challu, C.; Olivares, K.G.; Oreshkin, B.N.; Garza Ramirez, F.; Mergenthaler Canseco, M.; Dubrawski, A. NHITS: Neural Hi-erarchical Interpolation for Time Series Forecasting. *AAAI* **2023**, *37*, 6989–6997. [CrossRef]
5. Olivares, K.G.; Challú, C.; Garza, F.; Canseco, M.M.; Dubrawski, A. NeuralForecast: User Friendly State-of-the-Art Neural Forecasting Models. 2022. Available online: https://github.com/Nixtla/neuralforecast (accessed on 31 March 2024).
6. Chen, T.; Guestrin, C. XGBoost. In Proceedings of the KDD '16: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Krishnapuram, B., Shah, M., Smola, A., Aggarwal, C., Shen, D., Rastogi, R., Eds.; ACM: New York, NY, USA, 2016; pp. 785–794.
7. Kiranyaz, S.; Avci, O.; Abdeljaber, O.; Ince, T.; Gabbouj, M.; Inman, D.J. 1D convolutional neural networks and applications: A survey. *Mech. Syst. Signal Process.* **2021**, *151*, 107398. [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.