engineering proceedings

MDPI

Proceeding Paper

Short-Term Urban Water Demand Forecasting Using an Improved NeuralProphet Model †

Yao Yao, Haixing Liu *, Fengrui Gao, Hongcai Guo and Jiaxuan Zou

Department of Hydraulic Engineering, Dalian University of Technology, Dalian 116000, China; yaoyaodut@mail.dlut.edu.cn (Y.Y.); 13577068797@163.com (F.G.); m19841128455@163.com (H.G.); zjxuan05@126.com (J.Z.)

* Correspondence: hliu@dlut.edu.cn; Tel.: +86-176-1518-5386

† 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: The use of machine learning models for short-term network flow prediction has become increasingly widespread in recent years. Existing data-driven models are usually able to achieve good accuracy, but machine learning models are usually weakly interpretable and cannot provide clear decision guidance to decision makers in practical applications. Determining the input data shape of the model has an important impact on improving the interpretability of the model and understanding the relationship between the input factors and the application scenarios in the case. In this study, we used an integrated model for urban water demand prediction, which is based on the NeuralProphet model, and introduced the MIC method to screen the model input factors, which led to improvements in the accuracy of the prediction model. The aim of this work is also to improve the interpretability of water demand forecasting methodologies and the applicability of this model in the context of climate change and the complexity of urban water management, in order to help water managers make optimal water resource allocation decisions under different future scenarios.

Keywords: NeuralProphet; short-term water demand forecasting; model interpretability

Citation: Yao, Y.; Liu, H.; Gao, F.; Guo, H.; Zou, J. Short-Term Urban Water Demand Forecasting Using an Improved NeuralProphet Model. *Eng. Proc.* **2024**, *69*, 175. https://10.3390/engproc2024069175

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

Published: 26 September 2024

1. Introduction

The complexities of water distribution system (WDS) management are exacerbated by urbanization and water scarcity, underscoring the necessity for precise pipeline flow predictions to optimize water allocation [1] (p. 6). Forecasting water supply and demand is essential for sustainable water resource management and addressing urban water supply–demand imbalances. Amidst escalating urbanization and water scarcity, enhancing short-term water demand prediction accuracy is pivotal for achieving efficient water supply network scheduling [2,3] (p. 1, p. 2991).

Conventional water demand forecasting methods, such as gray forecasting, water demand rating, BP neural networks, and regression analysis, though advantageous in simplicity, data requirement leniency, and result interpretability, fall short in handling highly volatile data and lack spatial–temporal dependence considerations [4,5] (p. 146, p. 339). Consequently, a more practical and comprehensive forecasting model is imperative. Enter NeuralProphet, a PyTorch-based model bridging traditional time-series analysis with deep learning, offering a promising solution to these challenges [6].

This paper discusses a method for optimizing the input factors of a model and using a neural network model for prediction, namely by using the MIC to filter the factors and inputting the NeuralProphet model [7] (p. 282) for prediction in order to improve the quality of water demand forecasts.

2. Data Processing

2.1. Data Cleansing

2.1.1. Filling of Vacancies and Outlier Detection

The data used in this study were hourly inflow data from 1 January 2021 to 15 January 2023 for ten DMAs. Due to data collection tool failures, power, and other issues, the data contained many vacant values.

- If there were 168 or more consecutive vacant values, interpolation was performed using contemporaneous data from the previous or subsequent year.
- Interpolation was also used to fill in missing data values. If there were fewer than 168 consecutive vacant values, interpolation was applied using contemporaneous data from the previous week.
- For portions of the data that did not meet the aforementioned criteria, interpolation was performed based on the average value of the current year.
- For DMAG, there was a period of time when it had more vacancies, and the historical data before the vacancies was used to train the model to fill the NaN.

2.1.2. Data Characterization

The Maximal Information Coefficient, MIC(the pipeline is shown in Figure 1), was used to measure the degree of correlation between two variables, X and Y, and to measure whether it was linear or nonlinear; this technique is commonly used for feature selection in machine learning [8] (pp. 1518–1524). Using the MIC is an excellent way of calculating the relevance of data, and when sufficient statistical samples are available, it is possible to capture a wide range of relationships without being limited to specific function types [9] (p. 9) (e.g., linear, exponential, periodic, etc.). MIC scoring of all parameters is shown in Table 1.

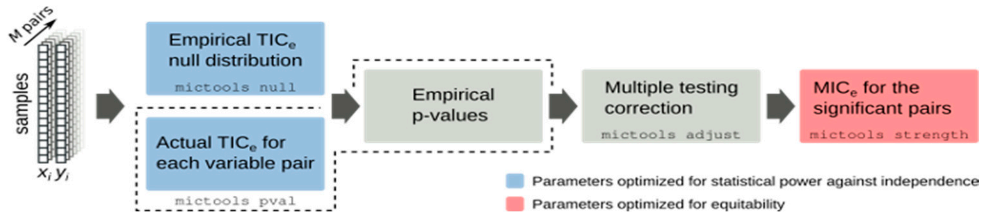

Figure 1. The MIC technique pipeline.

Table 1. MIC scoring of all parameters.

DMA Net Flow (L/s)	Rainfall Depth (mm)	Air Temperature (°C)	Air Humidity (%)	Windspeed (km/h)
A	0.424	0.691	0.601	0.203
B	0.598	0.651	0.56	0.301
C	0.49	0.621	0.58	0.314
D	0.612	0.506	0.49	0.286
E	0.384	0.584	0.63	0.276
F	0.392	0.499	0.57	0.291
G	0.216	0.326	0.62	0.3510
H	0.475	0.689	0.53	0.325
I	0.523	0.521	0.59	0.346
J	0.369	0.632	0.599	0.322

3. Model Parameter Adjustment

3.1. Model Conditioningt

The parameters chosen for the autoregression model were temperature, rainfall depth, and humidity. The best combination of parameters for model performance was determined based on the training results of the test set, as follows:

n_changepoints = 100;

trend_reg = 5;
seasonality_reg = 10;
yearly_seasonality = True;
weekly_seasonality = 20;
daily_seasonality = 10;
epochs = 80.

3.2. Model Training

We removed the last 168 data items from the historical data and divided these data into training and testing sets, with the first 95% being the training set and the last 5% being the testing set, and then used the trained model to predict the last week of the historical data [10].

4. Experimental Results

The predicted results for W1 are shown in Figure 2. Table 2 displays the model's accuracy.

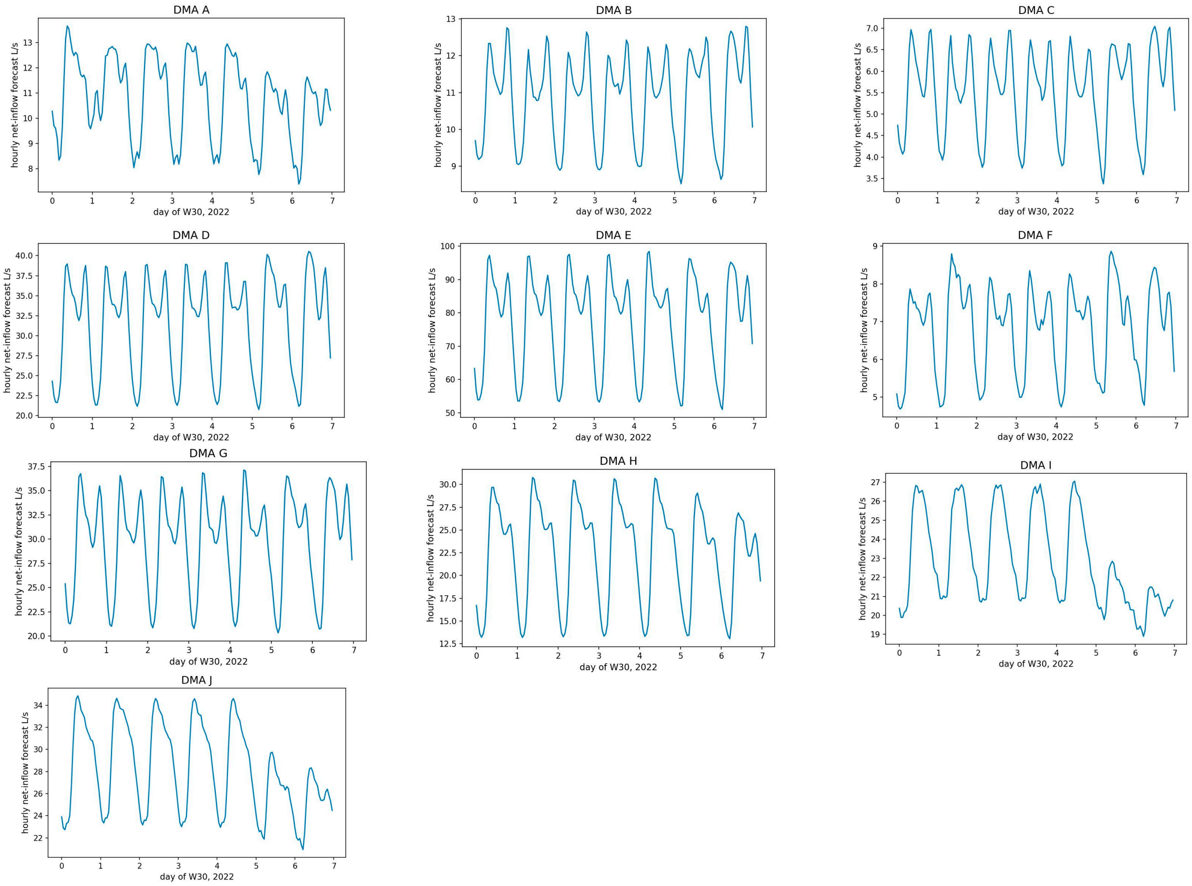

Figure 2. Experimental results (W1).

Table 2. R-squared, MAE, and MAPE values before and after data processing using the MIC.

DMA	R^2	MSE	MAE
A	0.79/0.81 0.69/0.64	0.81/0.69 4.9/3.9	1.40/1.39 0.18/0.32
B	0.69/0.53 0.72/0.776	2.36/0.86 1.82/2.34	0.098/0.058 5.14/3.92
C	0.77/0.86 0.79/0.49	4.96/4.77 0.60/0.99	3.272.66 0.72/0.664

Table 2. *Cont.*

DMA	R^2	MSE	MAE
D	0.62/0.69 0.86/0.92	3.51/2.21 0.95/1.11	0.933/1.27 1.59/2.23
E	0.886/0.78 0.930/0.81	2.66/1.03 2.828/1.43	0.814/0.87 1.187/0.92
F	0.79/0.81 0.69/0.64	0.81/0.69 4.9/3.9	1.40/1.39 0.18/0.12
G	0.69/0.53 0.72/0.776	2.36/0.86 1.82/2.34	0.098/0.058 5.14/3.92
H	0.77/0.86 0.79/0.49	4.96/4.77 0.60/0.99	3.272.66 0.72/0.664
I	0.62/0.69 0.86/0.92	3.51/2.21 0.95/1.11	0.933/1.27 1.59/2.23
J	0.886/0.78 0.930/0.81	2.66/1.03 2.828/1.43	0.814/0.87 1.187/0.92

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

Funding: This research received no external funding.

Institutional Review Board Statement: This study did not require ethical approval.

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

Data Availability Statement: The original data are available on the Battle of Water Networks | 3rd WDSA/CCWI Joint Conference (wdsa-ccwi2024.it).

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

References

1. Fu, G.; Sun, S.; Hoang, L.; Yuan, Z.; Butler, D. Artificial intelligence underpins urban water infrastructure of the future: A holistic perspective. *Camb. Prism. Water* **2023**, *1*, e14. [CrossRef]
2. Fu, G.; Jin, Y.; Sun, S.; Yuan, Z.; Butler, D. The role of deep learning in urban water management: A critical review. *Water Res.* **2022**, *223*, 118973. [CrossRef] [PubMed]
3. Liu, J.; Zhou, X.L.; Zhang, L.Q.; Xu, Y.P. Forecasting Short-term Water Demands with an Ensemble Deep Learning Model for a Water Supply System. *Water Resour. Manag.* **2023**, *37*, 2991–3012. [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. Chen, J.; Boccelli, D.L. Demand Forecasting for Water Distribution Systems. *Procedia Eng.* **2014**, *70*, 339–342. [CrossRef]
6. Miller, J.A.; Aldosari, M.; Saeed, F.; Barna, N.H.; Rana, S.; Arpinar, I.B.; Liu, N. A Survey of Deep Learning and Foundation Models for Time Series Forecasting. *arXiv* **2024**, arXiv:2401.13912.
7. Hu, S.; Gao, J.; Zhong, D.; Deng, L.; Ou, C.; Xin, P. An Innovative Hourly Water Demand Forecasting Preprocessing Framework with Local Outlier Correction and Adaptive Decomposition Techniques. *Water* **2021**, *13*, 582. [CrossRef]
8. Reshef, D.N.; Reshef, Y.A.; Finucane, H.K.; Grossman, S.R.; McVean, G.; Turnbaugh, P.J.; Sabeti, P.C. Detecting Novel Associations in Large Data Sets. *Science* **2011**, *334*, 1518–1524. [CrossRef] [PubMed]
9. Chen, K.J. An Adaptive and Fast Algorithm Based on Maximal Information Coefficient. *Microelectron. Comput.* **2016**, *33*, 70–73.
10. Triebe, O.; Hewamalage, H.; Pilyugina, P.; Laptev, N.; Bergmeir, C.; Rajagopal, R. NeuralProphet: Explainable Forecasting at Scale. *arXiv* **2021**, arXiv:2111.15397.