Proceeding Paper

A Methodology for Forecasting Demands in a Water Distribution Network Based on the Classical and Neural Networks Approach †

Yesid Coy [1], Laura González [1], Laura Basto [1], Valeria Rodríguez [1], Santiago Gómez [1], Juan Perafán [1], Simón Cardona [1], Alejandra Tabares [2] and Juan Saldarriaga [3,*]

1   Water Distribution and Sewerage Systems Research Center, Universidad de los Andes, Bogotá 111711, Colombia; b.coy@uniandes.edu.co (Y.C.); lk.gonzalez@uniandes.edu.co (L.G.); ls.basto@uniandes.edu.co (L.B.); v.rodriguezh@uniandes.edu.co (V.R.); s.gomezm3@uniandes.edu.co (S.G.); j.perafan@uniandes.edu.co (J.P.); s.cardonao@uniandes.edu.co (S.C.)
2   Industrial Engineering Department, Universidad de los Andes, Bogotá 111711, Colombia; a.tabaresp@uniandes.edu.co
3   Civil and Environmental Engineering Department, Universidad de los Andes, Bogotá 111711, Colombia
*   Correspondence: jsaldarr@uniandes.edu.co; Tel.: +57-1-339-49-49 (ext. 2805)
†   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 paper proposes a three (3)-step methodology to forecast the future water demands of a water distribution network (WDN) composed of ten (10) district metered areas (DMAs). First, pre-processing of the time-series data was performed through outlier elimination, imputation by K-Nearest Neighbors (KNN), and statistical data scaling. Second, the model hyperparameters were calibrated using Bayesian optimization. Third, Long Short-Term Memory (LSTM) coded as a Multi-Step Multivariate Time-Series forecasting model was implemented. Our results indicate that the proposed model produces accurate future water demands, suggesting that feasible short-term water demand forecasting models require combining engineering judgment and computational tools to achieve reliability.

Keywords: water demand; forecasting; DMA; KNN; LSTM; Bayesian optimization

1. Introduction

In an era where population growth and climate change intensify the imperative for sustainable water resource management, the accuracy of water demand forecasts has become crucial for the operational and strategic decisions of drinking water utilities [1]. This paper presents a pioneering methodology that enhances short-term water demand forecasting within water distribution networks by integrating traditional engineering insights with advanced neural network models. Focusing on ten district metered areas, the study employs a three-step process involving meticulous preprocessing of time-series data, calibration of model hyperparameters using Bayesian optimization, and the deployment of a Long Short-Term Memory model tailored for multivariate time-series forecasting. This work not only responds to the critical need for reliable water demand predictions to optimize water distribution systems but also demonstrates the potential of combining engineering judgment with computational tools, setting a new benchmark in water resource management amidst the challenges of urbanization and climate variability.

2. Methodology

This study introduces an innovative approach to improve the short-term forecasting of water demand within distribution networks, leveraging Long Short-Term Memory (LSTM) models. Our method, designed around ten district metered areas (DMAs), consists of three

pivotal stages, data preprocessing, hyperparameter tuning via Bayesian optimization, and LSTM model deployment, for multivariate time-series forecasting. This tripartite approach integrates engineering expertise with neural network models, aiming to set new standards in water resource management amidst the challenges of climate change and urbanization.

2.1. Preprocessing

The initial stage focuses on data preparation, crucial for the effectiveness of the LSTM algorithm. This process involves outlier detection, missing value imputation, and data normalization. Outliers are identified and removed based on domain-specific knowledge, ensuring that the data accurately reflect stable consumption patterns typical within DMAs [2]. We employed K-Nearest Neighbors (KNN) for imputation [3], leveraging its simplicity and resilience to noise, thereby maintaining the integrity and coherence of our dataset. Data normalization is achieved through the min–max scaler technique, standardizing the input features to scale the data within a specified range.

2.2. LSTM for Water Demand Forecasting

LSTM, a variant of Recurrent Neural Networks (RNNs), is employed for its superior ability to process and forecast complex multivariate time-series data, a capability crucial for accurately predicting water demand in district metered areas (DMAs) [4]. Unlike traditional RNNs, LSTM is designed to overcome the limitations associated with the vanishing and exploding gradient problems through its unique architecture, enabling it to retain information over longer sequences effectively [5].

To enrich the forecasting model, we integrated a Multi-Step Multivariate Time-Series approach, leveraging weather data and other relevant variables over the observed period as dependent factors. Utilizing the Fourier transform, we dissected time variables to highlight both short- and long-term patterns. This methodology enables the segmentation of consumption patterns into distinct signals, factoring in holidays and outlier consumption behaviors, with a keen focus on DMAs exhibiting non-seasonal or abrupt pattern changes.

2.3. Hyperparameter Optimization and Model Enhancement

Our LSTM model's accuracy is enhanced through two key methods: incorporating synthetic data and applying Bayesian optimization for fine-tuning [6]. We enhance the model with synthetic data that simulate both typical and rare water usage events, broadening its understanding of possible demand patterns. This is crucial for adjusting to unusual data, like the 2021 pandemic's impact, ensuring reliable predictions across various scenarios [7]. By using probability distributions that match observed trends, our synthetic data accurately represent real-world demand patterns, from regular to exceptional.

Moving beyond traditional tuning, we used Bayesian optimization to efficiently find the best model parameters. This method improves upon past results to hone in on the most effective parameter sets, enhancing model performance while reducing the need for extensive computations. It uses a smart balance of past performance analysis and new parameter exploration to streamline the optimization process. Together, these approaches not only sharpen the LSTM model's forecasting accuracy but also equip it with the ability to navigate the complex dynamics of urban water demand forecasting with greater precision and less computational demand.

3. Results

Our results, derived from applying the proposed methodology over four evaluation weeks (W1–W4), illustrate significant strides in water demand forecasting accuracy for ten district-metered areas (DMAs). By custom-tuning the parameters for each DMA, we tailored our approach to enhance prediction precision for each area's unique net inflow (Q_{net}) time series.

Table 1 encapsulates the forecasting performance across all DMAs, showcasing a notable trend: as the Determination Coefficient (R^2) for each DMA increases, indicating

better model fit, the Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) concurrently decrease. This inverse relationship underscores our model's reliability, accurately mirroring the intricate behavior of water demand across diverse urban contexts.

Table 1. General metrics of the forecasted water demands along evaluation weeks W1–W4: Determination Coefficient (R^2), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE).

DMA	A	B	C	D	E	F	G	H	I	J
R^2 (—)	0.78	0.74	0.82	0.83	0.92	0.60	0.87	0.92	0.66	0.81
MAPE (%)	12.26	4.34	10.73	6.93	3.54	8.03	5.67	10.91	5.64	6.27
RMSE (—)	0.92	0.51	0.39	2.75	4.24	1.31	1.92	3.50	1.73	1.70

The Determination Coefficient (R^2) consistently exceeded 0.70 for most DMAs, affirming the model's capability to closely predict actual water demand patterns. Exceptions were noted in DMAs F and I, attributed to their unique geographic and functional characteristics— DMA F's suburban makeup and DMA I's proximity to commercial and industrial activities near the port. The slight discrepancies in forecasting accuracy for DMAs F and I highlight the model's sensitivity to the varied water usage patterns characteristic of non-residential areas. These findings signal the need for further model refinement to better account for the diverse factors influencing water demand in such distinct settings.

Figure 1 shows the water demand time series and the correlation curves of the best-performing DMAs along evaluation week W4. The water demand time series and the correlation curves consist of the predicted net inflows (Q_{net}) obtained for week W4 and the net inflows (Q_{net}) of the week previous to W4 ($W4_0$).
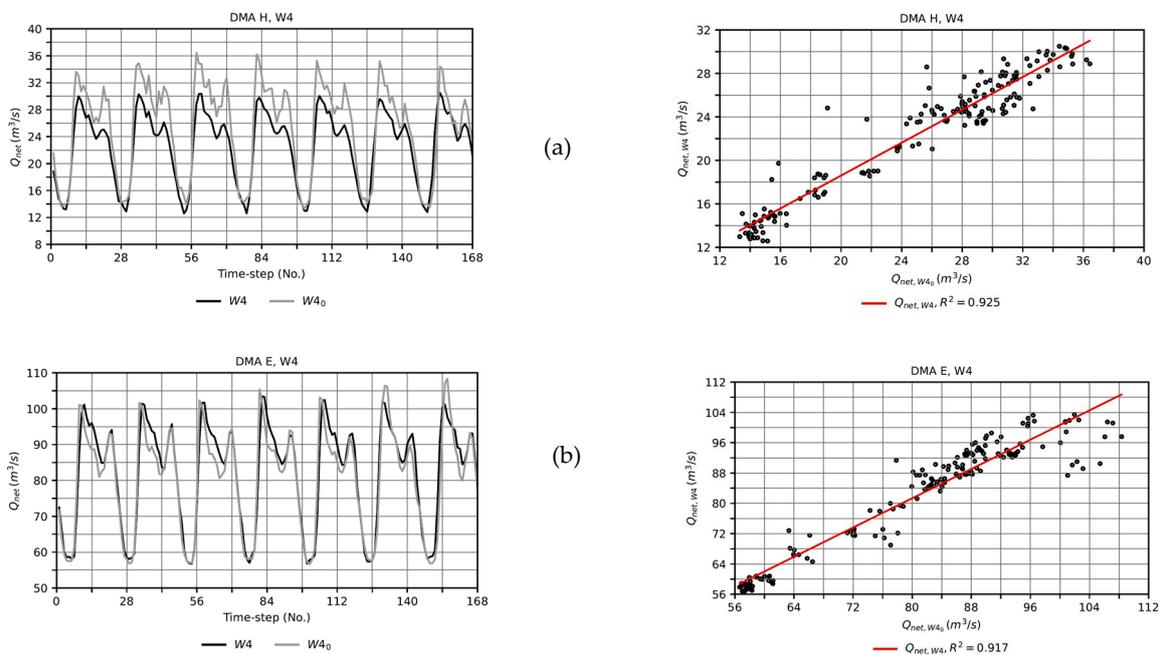
Figure 1. Water demand time-series and correlation curves of the best-performing DMAs in evaluation week W4: (**a**) DMA H, $R^2 = 0.925$; (**b**) DMA E, $R^2 = 0.917$.

From Figure 1, note that all previous affirmations hold since the predictive outputs of the model were able to capture the general demand patterns of the shown DMAs while a strong correlation between net inflows (Q_{net}) was preserved along any evaluation week of this study. Indeed, the model is sensible enough to capture the studied water

demand patterns, i.e., water demand peaks, the week/weekend transition, and the general curve characteristics.

4. Conclusions

This study proposes the use of an AI model to represent and predict the water demands of the DMAs of a water network considering seasonal patterns, including daily consumption characteristics and non-conventional days such as weekends or atypical consumption. The model obtained a prediction of four non-continuous weeks throughout the years 2022 and 2023. The performance metrics reported a MAPE between 3.54 and 12.26% and an RMSE between 0.39 and 4.24. Since the raw weather data (temperature, rainfall, depth, air humidity, and wind speed) were not sufficient for training the neural network, a seasonal decomposition was performed to provide additional layers in the model. For future research, the Prophet algorithm can be used as it is an additive model that describes the long-term behavior and seasonality of the series and provides signals about outliers.

References

1. 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]
2. Avni, N.; Fishbain, B.; Shamir, U. Water Consumption Patterns as a Basis for Water Demand Modeling. *Water Resour. Res.* **2015**, *51*, 8165–8181. [CrossRef]
3. Xiao, Y.; Kong, W.; Liang, Z. Short-Term Demand Forecasting of Urban Online Car-Hailing Based on the K-Nearest Neighbor Model. *Sensors* **2022**, *22*, 9456. [CrossRef] [PubMed]
4. Kühnert, C.; Gonuguntla, N.M.; Krieg, H.; Nowak, D.; Thomas, J.A. Application of LSTM Networks for Water Demand Prediction in Optimal Pump Control. *Water* **2021**, *13*, 644. [CrossRef]
5. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
6. Yang, K.; Liu, L.; Wen, Y. The Impact of Bayesian Optimization on Feature Selection. *Sci. Rep.* **2024**, *14*, 3948. [CrossRef] [PubMed]
7. Ortiz, C.; Salcedo, C.; Saldarriaga, J. Assessment of the Effects of COVID-19 Pandemic Stay-at-Home Measures on Potable Water Consumption Patterns, Location, and Financial Impacts for Water Utilities in Colombian Cities. *Water* **2022**, *14*, 3004. [CrossRef]