*engineering proceedings*

MDPI

*Proceeding Paper*

# Sequence-to-Sequence Deep Learning for Urban Water Demand Forecasting †

**Mohammad Sina Jahangir and John Quilty ***

Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada; msjahang@uwaterloo.ca
* Correspondence: john.quilty@uwaterloo.ca

† 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 urban water demand (UWD) forecasts are key to the effective management of water distribution systems. This research explores the potential of encoder–decoder models, specifically sequence-to-sequence (S2S) deep learning models, for UWD forecasting. Two models were developed as follows: one based on long short-term memory (LSTM) networks and another using transformers. The models were trained on data from ten district metered areas (DMAs) in Northeast Italy. The results confirmed that the transformer models consistently outperformed the LSTM models across all DMAs, with an average (across all DMAs) improvement in mean absolute error of 15.3%.

**Keywords:** encoder–decoder; transformer; LSTM; urban water demand; forecasting



**Citation:** Jahangir, M.S.; Quilty, J. Sequence-to-Sequence Deep Learning for Urban Water Demand Forecasting. *Eng. Proc.* **2024**, *69*, 41. https://doi.org/10.3390/engproc2024069041

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

Published: 3 September 2024



## 1. Introduction

Accurate urban water demand (UWD) forecasts are necessary for the effective management of water distribution systems (e.g., [1,2]). Various data-driven methods have been applied for UWD forecasting, such as auto-regressive integrated moving average, artificial neural networks, and support vector regression [1]. However, deep learning (DL) models, such as long short-term memory networks (LSTMs) [3], have shown tremendous success in capturing the non-linear relationships between UWD and explanatory variables (e.g., precipitation, air temperature) that evolve through time, outperforming other machine learning and statistical methods [2].

Sequence-to-sequence (S2S) models are a specific type of encoder–decoder (ED) model, where a sequence of inputs (e.g., meteorological variables) of arbitrary length are mapped to the designated sequence target (e.g., UWD). ED frameworks excel in time series forecasting because they capture complex relationships among time series [4]. Unlike simpler models, ED frameworks separate the encoding of input data (capturing local factors) from decoding it into the desired output (e.g., UWD). This flexibility is further enhanced by the freedom to choose any neural network for the encoder and decoder, allowing researchers to tailor the model to the specific complexities of the forecasting task [5].

While S2S models have been adopted in water resources for various tasks, such as rainfall–runoff prediction [6], a thorough examination of the efficacy of such models for UWD forecasting has yet to be investigated. Furthermore, transformers that have shown outstanding performance in fields such as language modeling [7] are yet to be explored for UWD forecasting. Hence, this research explores two S2S model types, LSTM- and transform-based EDs, for multistep ahead (24 h) UWD forecasting.

The rest of the paper is structured as follows: Section 2 introduces data adopted for model development and evaluation. In Section 3, the developed models are introduced. Section 4 provides the results, followed by Section 5, where the paper is concluded.

## 2. Data

Data from the Battle of Water Demand Forecasting as part of the third WDSA-CCWI Joint Conference were used for model development and assessment. These data include UWD ($\frac{\mathrm{L}}{\mathrm{s}}$) from ten district metered areas (DMAs) in Northeast Italy (Europe), as well as meteorological variables, rainfall (mm), air temperature (°C), air humidity (%), and windspeed ($\frac{\mathrm{km}}{\mathrm{hr}}$). Missing data were replaced by the median value of the respective variable. The data were collected on an hourly basis, covering from 1 January 2021, 12 AM, to 24 July 2022, 11 PM, resulting in 13,679 data samples. The data were split into three sets: training (80%), validation (10%), and testing (10%). The training and validation sets were used for optimizing the network parameters (weights and biases) and hyperparameter tuning, respectively. The test set was solely used for model evaluation, reflecting out-of-sample performance. The results provided in Section 4 are associated with the test set, i.e., the portion of the data unseen by the model during training.

The inputs to the models included the following: lagged meteorological variables and UWD. A lookback period of 14 days was selected (covering two weeks) to produce the lagged meteorological and UWD variables. Therefore, the input data were lagged for 336 h. The target was set as the UWD for the next 24 h (i.e., a vector of length 24). Hence, for a given model (LSTM or transformer), all the forecasts for the next 24 h were obtained simultaneously.

## 3. Methodology

### *3.1. LSTM*

UWD data, with their abrupt changes, pose a challenge for traditional data-driven models. Recurrent neural networks (RNNs) were specifically designed to process temporal dynamics, making them well suited for forecasting complex, non-linear relationships among time series. However, traditional RNNs face limitations with long sequences. However, LSTMs, introduced by [3], addressed this issue by controlling information flow within the network, allowing them to model the complexities of time series effectively. In this work, two distinct LSTMs are adopted in the encoder for processing the input data sequences, one for meteorological variables and the other for processing UWD. The outputs of the LSTMs are passed to a dense network (DN). The decoder, a LSTM-DN, processes the output (encoded values) of the encoder, the meteorological variables, and the average UWD associated with the forecast dates. The outputs of the decoder are UWD forecasts for the next 24 h. More information on LSTM-based ED can be found in [4].

### *3.2. Transformer*

The transformer model had a similar structure to the LSTM-based ED models, where the only difference is that transformer encoders were used instead of LSTM blocks. The transformer architecture, introduced by [8], revolutionized sequence modeling by overcoming the limitations of RNNs in handling long sequences. Unlike RNNs, which process data sequentially, the transformer employs a self-attention mechanism. This mechanism allows the model to attend to all parts of the input sequence simultaneously, eliminating the need for recurrent layers and enabling efficient sequence processing. The self-attention mechanism generates a condensed representation capturing the entire sequence's essential information through a series of calculations involving query, key, and value vectors. This condensed representation makes the transformer adept at tasks requiring analysis of long-range dependencies, such as time series forecasting and classification tasks [9].

### *3.3. Training and Forecast Evaluation*

Two different strategies were tested for training the models. In the first strategy, a model was trained for each DMA. In the second strategy, a single model was trained for all DMAs and then fine-tuned for each DMA. Validation results confirmed that the second strategy resulted in considerably more accurate forecasts for both model types (LSTM and transformers). Consequently, the results associated with the latter strategy are presented. A dynamic learning rate was adopted, starting from an initial value of 0.001. The learning rate was halved after ten epochs if the validation loss did not decrease. A maximum of 500 epochs was considered. The Adaptive Moment Estimation (Adam) [10] was used to minimize the loss function, mean squared error. The DL model development was achieved using custom scripts in Python 3.10, leveraging the Keras 2.0 library with TensorFlow backend.

## 4. Results

For each DMA, the forecasts for each of the 24 h were evaluated using deterministic metrics, modified Kling–Gupta efficiency (KGE), Nash–Sutcliffe efficiency (NSE), and mean absolute error (MAE). The average of each metric across 24 h is presented in Figure 1. The results confirm that the transformer models outperformed the LSTM models based on all three metrics and for all DMAs. The results reveal that both models achieved good forecasting accuracy (NSE > 0.6) for all DMAs except one. Both models showed relatively poor performance in DMA 6, the suburban district. Due to the S2S structure of the models, no considerable accuracy drop was observed between the 1 and 24 h forecasts. The maximum drop in accuracy between the 1 and 24 h forecasts was associated with the transformer model in DMA 6, where a decrease in KGE of 2.8% was observed. Overall, transformers improved the MAE by 15.3% compared to LSTM across all DMAs.

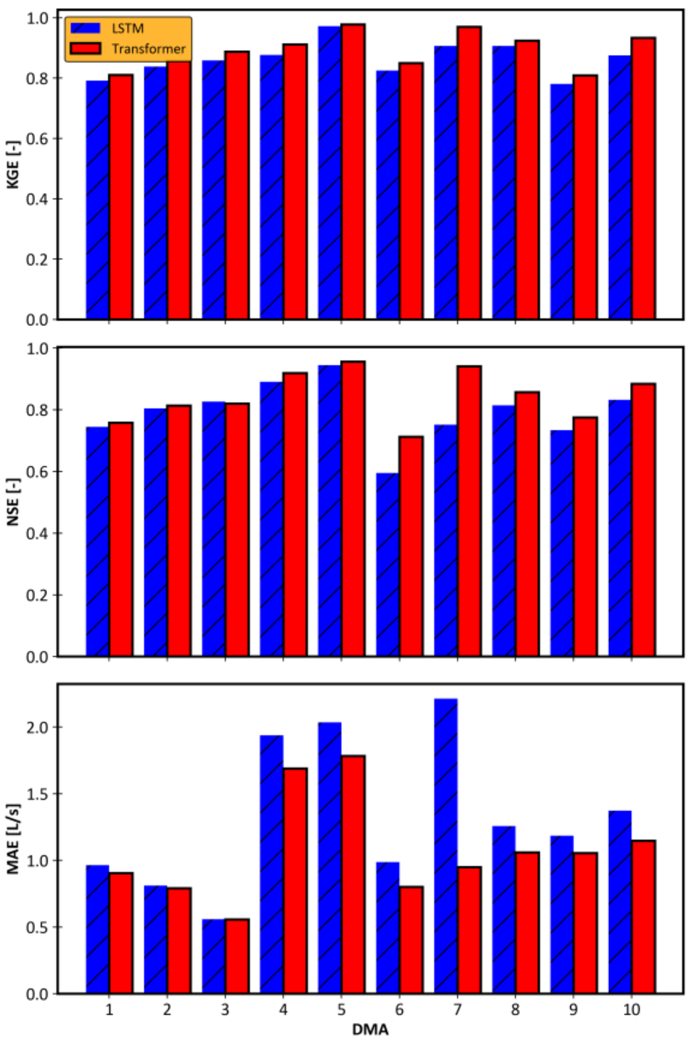


**Figure 1.** Comparison of the models' performance in the studied district metered areas (DMAs).

## 5. Conclusions

This study investigated the application of S2S DL models for 24 h ahead UWD forecasting in ten DMAs located in Northeast Italy. The following two S2S models were developed and deployed: LSTM- and transformer-based models. The transformer-based model consistently outperformed the LSTM models across all ten DMAs. These findings emphasize the significance of evaluating novel DL models, such as transformers, to ensure accurate UWD forecasting. Given the promising results presented herein, transformers are recommended when exploring DL models for forecasting UWD in other water distribution systems.

**Author Contributions:** Conceptualization, M.S.J. and J.Q.; methodology, M.S.J. and J.Q.; software, M.S.J.; writing—original draft preparation, M.S.J.; writing—review and editing, M.S.J. and J.Q. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Sinnathamby Professorship in AI for Sustainable Solutions and the Natural Sciences and Engineering Research Council of Canada Discovery Grant (RGPIN-2021-03194).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All data used in this research are public and freely accessible. The data were retrieved from https://wdsa-ccwi2024.it/battle-of-water-networks/ (accessed on 1 January 2024).

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study, in the collection, analysis, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

## References

1. Brentan, B.M.; Luvizotto Jr., E.; Herrera, M.; Izquierdo, J.; Pérez-García, R. Hybrid Regression Model for Near Real-Time Urban Water Demand Forecasting. *J. Comput. Appl. Math.* **2017**, *309*, 532–541. [CrossRef]
2. Du, B.; Zhou, Q.; Guo, J.; Guo, S.; Wang, L. Deep Learning with Long Short-Term Memory Neural Networks Combining Wavelet Transform and Principal Component Analysis for Daily Urban Water Demand Forecasting. *Expert Syst. Appl.* **2021**, *171*, 114571. [CrossRef]
3. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
4. Cui, Z.; Zhou, Y.; Guo, S.; Wang, J.; Xu, C.-Y. Effective Improvement of Multi-Step-Ahead Flood Forecasting Accuracy through Encoder-Decoder with an Exogenous Input Structure. *J. Hydrol.* **2022**, *609*, 127764. [CrossRef]
5. Jahangir, M.S.; You, J.; Quilty, J. A Quantile-Based Encoder-Decoder Framework for Multi-Step Ahead Runoff Forecasting. *J. Hydrol.* **2023**, *619*, 129269. [CrossRef]
6. Yin, H.; Zhang, X.; Wang, F.; Zhang, Y.; Xia, R.; Jin, J. Rainfall-Runoff Modeling Using LSTM-Based Multi-State-Vector Sequence-to-Sequence Model. *J. Hydrol.* **2021**, *598*, 126378. [CrossRef]
7. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019.
8. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Polosukhin, I.; Kaiser, L.; Polosukhin, I. Attention is All you Need. In Proceedings of the 31st International Conference on Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
9. Castangia, M.; Moeini, H.R.; Nourani, V.; Wößner, M.; Schaefli, B. Transformer neural networks for interpretable flood forecasting. *Environ. Model. Softw.* **2023**, *160*, 105581. [CrossRef]
10. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference for Learning Representations, San Diego, CA, USA, 7–9 May 2015.