Proceeding Paper

Leveraging Potentials of Local and Global Models for Water Demand Forecasting †

Matthias Groß * ![ORCID] **and Lukas Hans** ![ORCID]

IDE+A Institute, TH Köln, 51643 Gummersbach, Germany; lukas.hans@th-koeln.de

* Correspondence: matthias.gross2@th-koeln.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 paper examines the effectiveness of local and global models in predicting water demand, employing data from the Battle of Water Demand Forecasting. Utilizing LightGBM models under local, semi-global, and global settings, we analyze the performance of these models across different configurations. The results suggest that inadequately optimized hyperparameters do not always enhance model performance, but well performing hyperparameters can be appropriate for different model types inside the domain of water demand forecasting. Semi-global and global models frequently outperformed local models, underscoring the benefits of contextual information. Our findings indicate that while semi-global approaches offer promising results, extensive tuning and a strategic selection of a time series for modeling are imperative for forecasting accuracy.

Keywords: water demand forecasting; LightGBM; predictive modeling; global and local modeling; time series analysis

1. Introduction

Accurate predictive models are essential in hydrological forecasting, especially for predicting water demand. With the increasing worldwide need for water and the challenges posed by climate change, advanced and dependable forecasting methods are increasingly important. This paper explores the differences and effectiveness of global and local forecasting strategies in predicting water demand.

Central to this analysis are global forecasting models. These models stand out due to their simultaneous use of multiple time series data during training. The primary objective is to augment the quantity of training samples, thereby enhancing the accuracy of future predictions [1]. Traditionally, these data groups were homogeneous in nature to optimize forecasting accuracy. However, recent research has demonstrated the effectiveness of this approach, even with heterogeneous datasets [2].

In contrast, local forecasting models prioritize specific, individual time series. They are often favored for their precision and locality-specific accuracy, yet they may fall short in encompassing the broader context that global models capture. This paper sets out to compare these modeling strategies within the domain of water demand forecasting, aiming to ascertain which approach is more effective and under what circumstances.

The paper is methodically structured into four main sections. Section 2 outlines the materials and methods detailing the technical approach and the data utilized. Section 3 reveals the findings of our study. Finally, in Section 4, we conclude by discussing the primary findings and exploring avenues for future research.

2. Materials and Methods

Our study employed data sourced from the Battle of Water Demand Forecasting, comprising three distinct datasets. The primary dataset of net inflow data encompasses a

series of ten time series, each representing water demand in various district types including residential, hospital, and industrial areas. The objective of the battle was to forecast these time series over four distinct timeframes, each spanning over 168 time steps—equivalent to one week—with hourly data updates. The time series covered a period from 1 January 2021 at 12 a.m. to 5 March 2023 at 11 p.m.

Additionally, the dataset included supplementary data detailing holiday schedules and weather conditions, extending from the same start date to 12 March 2023 at 11 p.m. Holiday data were encoded in binary format, while the weather data comprised four columns with floating-point values. Both the inflow data and weather datasets exhibited missing values, which we addressed through various imputation techniques. For weather data, Multiple Imputation by Chained Equations (MICE) [3] was employed. In contrast, the inflow data gaps were filled using linear interpolation for single-point gaps and Partial Sequence Matching (PSM) [4] for larger gaps. Prior to analysis, all datasets underwent a standardization process to ensure consistency in the modeling phase.

Our analytical approach centered on the LightGBM model, with variations in predictive strategies across three distinct settings:

- Local Setting: Each of the ten time series were paired with a dedicated model, resulting in ten distinct models. This approach focuses on the unique characteristics of each district's water demand.
- Semi-Global Setting: Time series were grouped into four clusters. Four models in total were trained, with each model responsible for forecasting the time series within its respective cluster.
- Global Setting: A single model was employed to predict all ten time series. This setting evaluates the model's capability to generalize across diverse datasets.

Alongside these settings, we also investigated the effects of optimized hyperparameters on our models' performance. The evaluation of these varied approaches was conducted through backtesting, which involved the following:

- Starting from the initial point in the test dataset, we forecasted the subsequent 168 time steps.
- These forecasts were compared against the actual values to determine the mean absolute error (MAE) for that period.
- The process was then repeated, moving forward in 24 h strides, until the end of the test dataset was reached. This results in 192 separate evaluations.

This iterative method of evaluation provided us with multiple data points to ensure the numerical stability and accuracy of our results.

3. Results and Discussion

The depicted workflow yielded MAE values for all 192 evaluations of each time series using each model. The MAE values were averaged per time series to produce a single value to assess the performance of each model on every time series. The error values are based on standardized values to provide more comparability between series of different scales.

Figure 1 shows the results of these evaluations in tabular form, augmented with color coding for better perceptibility. The first finding resulting from these values is that the models trained with tuned hyperparameters (middle three rows) performed worse than their untuned counterparts in most cases. Although surprising at first, this might be attributed to several factors.

Due to time and computational constraints, the conducted hyperparameter tuning only covered 103 evaluation trials for each model. This can already lead to promising results—cluster 2 in the semi-global setting (DMAs D, E, G, and H) performed better using the tuned parameters—but a more exhausting optimization can potentially further improve the models' performance.

Model Type	DMA A	DMA B	DMA C	DMA D	DMA E	DMA F	DMA G	DMA H	DMA I	DMA J	Average
Local **no tuning**	0.4591 (±0.2758)	0.1989 (±0.0823)	0.2137 (±0.0657)	0.3075 (±0.0486)	0.1199 (±0.0405)	**0.3492** **(±0.0282)**	0.1607 (±0.0307)	0.2427 (±0.0527)	0.5465 (±0.1807)	0.3128 (±0.1690)	0.2911
Semi-Global **no tuning**	0.4022 (±0.2297)	0.2084 (±0.0695)	0.2040 (±0.0770)	0.3041 (±0.0390)	0.1141 (±0.0362)	0.3547 (±0.0276)	0.1618 (±0.0306)	**0.2003** **(±0.0500)**	0.4980 (±0.1961)	0.3149 (±0.1768)	0.2762
Global **no tuning**	0.4045 (±0.2382)	0.1956 (±0.0721)	0.1979 (±0.0845)	0.3009 (±0.0393)	0.1234 (±0.0348)	0.3636 (±0.0407)	0.1651 (±0.0291)	0.2202 (±0.0448)	0.5318 (±0.2072)	0.3133 (±0.1847)	0.2816
Local **tuned**	0.4540 (±0.1855)	0.1999 (±0.0733)	0.7141 (±0.1352)	0.4940 (±0.0655)	0.3606 (±0.0277)	0.5408 (±0.1825)	0.2534 (±0.0383)	0.2852 (±0.0716)	0.9333 (±0.2505)	0.4573 (±0.1635)	0.4693
Semi-Global **tuned**	0.6595 (±0.1207)	0.3312 (±0.0611)	0.6994 (±0.1283)	**0.3001** **(±0.0396)**	**0.1137** **(±0.0372)**	0.6878 (±0.2907)	**0.1495** **(±0.0315)**	0.2166 (±0.0524)	0.8837 (±0.2492)	0.4665 (±0.1638)	0.4508
Global **tuned**	0.4168 (±0.2037)	0.1874 (±0.0761)	0.2526 (±0.0728)	0.3258 (±0.0428)	0.1629 (±0.0313)	0.3819 (±0.0614)	0.1811 (±0.0317)	0.2775 (±0.0501)	0.5855 (±0.2107)	0.3267 (±0.1534)	0.3098
Local **tuned with best parameters**[1]	**0.3709** **(±0.1935)**	0.1839 (±0.0771)	0.2324 (±0.0617)	0.3067 (±0.0493)	0.1263 (±0.0425)	0.3497 (±0.0334)	0.1507 (±0.0279)	0.2887 (±0.0716)	0.5015 (±0.1989)	0.2989 (±0.1615)	0.2810
Semi-Global **tuned with best parameters**[1]	0.3764 (±0.1978)	0.1843 (±0.0730)	0.2106 (±0.0735)	**0.3001** **(±0.0396)**	**0.1137** **(±0.0372)**	0.3497 (±0.0331)	**0.1495** **(±0.0315)**	0.2166 (±0.0524)	**0.4834** **(±0.2097)**	**0.2961** **(±0.1581)**	**0.2680**
Global **tuned with best parameters**[1]	0.3866 (±0.2144)	**0.1799** **(±0.0751)**	**0.1918** **(±0.0820)**	0.3013 (±0.0392)	0.1248 (±0.0361)	0.3574 (±0.0356)	0.1566 (±0.0289)	0.2267 (±0.0502)	0.5142 (±0.2003)	0.2982 (±0.1562)	0.2738

[1] Best parameters found for any model (the cluster containing DMAs D, E, G and H in semi-global setting).

Figure 1. Tabular results of model backtesting: averaged MAE (± standard deviation), rounded, color coded per series (red—high MAE, green—low MAE). The lowest average MAE per time series is highlighted in bold.

Furthermore, the validation sets account for 10% of the total datasets. Due to the limited timeframe covered by the data, this validation set represents a specific time of the year—from June to August. The test sets cover the last 25% of the data. Coupled with the subpar performance of the tuned models, this could indicate that the tuned hyperparameters are especially fit for picking up the specific patterns in the summer months but lack this fitness for the rest of the year. This can be further assessed using larger datasets, e.g., allowing the evaluation set to cover an entire year of data.

For the second finding, using the hyperparameters found for the best performing tuned model (the model for cluster 2 in the semi-global setting) on all the remaining models surprisingly yields the best performance for most time series, as depicted in the last three rows of the table depicted in Figure 1. This could suggest that searching hyperparameters over a range of different settings can to some amount mitigate the need for an extensive number of optimization cycles.

As for the last finding, focusing on the untuned models, the semi-global approach produced the lowest average MAE over all time series. Their performance is generally on par with their local and global counterparts, surpassing them notably for some series. The local models only surpassed the other settings for a single time series (DMA F). This indicates that the additional contextual information from the other time series does help the LightGBM model with producing more accurate forecasts than local models. Clustering the time series so that the most similar series add context to one another further improved the performance of the forecasting models.

4. Conclusions

From the evaluations, it can be concluded that adding contextual information to a time series can aid the forecasting performance, as on average the global and semi-global models outperformed the local models in all settings. The semi-global cluster models were able to outperform the global models in most cases, indicating the importance of providing context using the most homogeneous series as possible. Especially when tuned by reasonable hyperparameters, the semi-global models produced the lowest MAE values on average.

Computational and time constraints limited the amount of possible hyperparameter tuning cycles. Nevertheless, although the hyperparameters used did emerge from a tuning process on a specific cluster model, most other models still gained a performance boost using these parameters. This suggests a high importance of a functioning combination

of hyperparameters for the application of water demand forecasting. Inside this domain, there seem to be no large differences in performance when applying hyperparameters that proved to perform well on one model onto another model.

In addition to the computational and time constraints, the exceptionally high MAE values for most tuned models might also partly be attributed to the specific structure of the validation set and the structural differences between training, validation, and testing sets. The limited timeframe of the total datasets (2 years and 9 weeks) restrains the possibilities to provide adequate amounts of representative data to all the sets.

Beyond that, even without hyperparameter tuning, semi-global and global methods performed well, again showing the improvement from adding context to forecasting models. However, adding contextual time series indiscriminately—even from the same domain—does not always yield a better performance, as can be seen from the superior performance of semi-global cluster models over the global model in most cases. Thus, at least for the water demand forecasting domain, "more" does not always equal "better".

In conclusion, grouping similar time series can boost the performance of forecasting models for a negligible computational cost.

Author Contributions: Conceptualization, M.G. and L.H.; methodology, M.G. and L.H.; software, M.G.; validation, M.G. and L.H.; formal analysis, M.G. and L.H.; investigation, L.H.; writing—original draft preparation, M.G and L.H.; writing—review and editing, M.G. and L.H.; visualization, M.G. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this research were provided by the organizing committee of the 3rd WDSA/CCWI as part of the Battle of Water Demand Forecasting through https://wdsa-ccwi2024.it/battle-of-water-networks/ (accessed on 5 March 2024).

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

References

1. Januschowski, T.; Gasthaus, J.; Wang, Y.; Salinas, D.; Flunkert, V.; Bohlke-Schneider, M.; Callot, L. Criteria for Classifying Forecasting Methods. *Int. J. Forecast.* **2020**, *36*, 167–177. [CrossRef]
2. Montero-Manso, P.; Hyndman, R.J. Principles and Algorithms for Forecasting Groups of Time Series: Locality and Globality. *Int. J. Forecast.* **2021**, *37*, 1632–1653. [CrossRef]
3. van Buuren, S.; Groothuis-Oudshoorn, K. Mice: Multivariate Imputation by Chained Equations in R. *J. Stat. Softw.* **2011**, *45*, 1–67. [CrossRef]
4. Khampuengson, T.; Wang, W. Novel Methods for Imputing Missing Values in Water Level Monitoring Data. *Water Resour. Manag.* **2023**, *37*, 851–878. [CrossRef]