



 <https://doi.org/10.71573/sdhc7935>

© Authors. This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

Exploring the Potential for Machine Learning-Based Flow Predictions in Sewer Systems

Flemming Albers¹, *  <https://orcid.org/0009-0003-4420-0004>, Birgitta Hörnschemeyer¹
& Malte Henrichs¹  <https://orcid.org/0009-0001-0002-1990>

¹ Institut für Infrastruktur-Wasser-Ressourcen-Umwelt (IWARU), FH Münster, Corrensstr. 25, 48149 Münster, Germany

*Corresponding author email: flemming.albers@fh-muenster.de

Abstract

This study explores the potential of machine learning (ML) for predicting flow in sewer systems, using data generated by the Storm Water Management Model (SWMM). A Long Short-Term Memory (LSTM) network, chosen for its effective handling of time series data, was trained using both hypothetical and real rainfall data. The final model achieved a mean error of 4.6 % in predicting peak flows and demonstrated prediction times that were at least 3 times faster for individual events and up to 600 times faster in mass simulations, compared to an equivalent hydrodynamic model. Tested with 5-fold cross-validation, the model exhibited significant improvements in accuracy and speed compared to its initial version, largely due to enhanced complexity in the model architecture. However, when applied to a more complex sewer system, a decrease in accuracy was observed, underscoring the need for further validation with real-world data. These findings illustrate the promising potential of ML models to boost real-time prediction efficiency but also highlight the need for model adaptation in more complex scenarios.

Highlights

- Development of a LSTM model for predicting flow using data from a hydrodynamic model
- Final model achieved 4.6 % mean deviation from peak flow
- Prediction times were at least 3 times faster for individual events and up to 600 times faster for mass simulations compared to the hydrodynamic model

Introduction

Data has become crucial to water management, driven by advancements in sensor technology that enable more precise and comprehensive data collection. This has led to a significant increase in data volume and diversity, facilitated by extensive monitoring networks and remote sensing technologies (Grossman et al., 2015). Machine learning (ML) offers a powerful solution to leverage these vast data sets, due to its ability to identify and exploit patterns and structures within them (Cerulli, 2023).

One promising ML application is the prediction of flow events in sewer networks. Traditional hydrodynamic models are effective but often require extensive calibration and significant computational effort (Capodaglio 1994 cited in El-Din and Smith, 2002).

In contrast, ML methods offer an advantageous, data-driven alternative by using patterns in flow data to make predictions, bypassing the need to explicitly model physical processes. To explore this potential, an ML-based model was developed and evaluated for speed and accuracy using synthetic flow data generated with a hydrodynamic model.

Methodology

Concept and Model Requirements

The study focuses on developing a ML-based model to predict flow at a selected node in a sewer system using a one-hour forecast with five-minute intervals. The input horizon includes one hour of future and one hour of past precipitation data for prediction over five-minute intervals and incorporating event duration as an additional input. Other features, such as the observed flow, were also tested to potentially enhance the model's predictive capability. Training data are synthetically generated using the open source Storm Water Management Model - SWMM (Rossman, 2015). This enables the use of both artificial and recorded precipitation events for the training of the ML model.

Study Area

The selected catchment in Münster, Germany, covers 100,266 m² of which 62 % are impervious surfaces. The catchment has a low gradient ranging from 0.3 % to 1.2 %.

Event Selection

Precipitation events were filtered out of a 10-year time series, provided by the German Weather Service (DWD 2024). Of all recorded events, 100 were selected with a balanced spectrum of total precipitation and maximum intensity. These were supplemented with synthetic events from KOSTRA data (DWD, 2023), yielding a total of 226 events with total precipitation ranging from 1 mm to 100 mm and intensities up to 220 mm/h as shown in Figure 1.

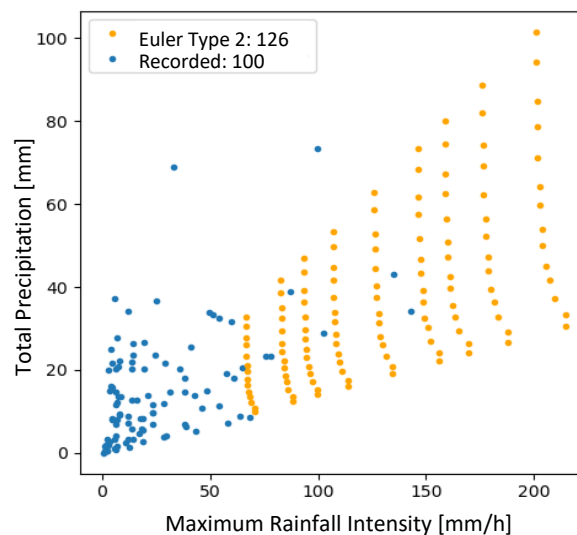


Figure 1. Selection of precipitation events based on maximum rainfall intensity and total precipitation

Machine Learning Environment

The ML environment was set up with Python using TensorFlow (Abadi et al., 2015) for model formulation and Keras (Chollet et al., 2015) for prototyping, due to their cross-platform capabilities and ease of use.

Algorithm Selection

For time series or sequence prediction, the Long Short-Term Memory (LSTM) algorithm was chosen based on its strong performance in similar applications, as highlighted by Sit et al. (2020). Developed by Hochreiter and Schmidhuber (1997), LSTM networks mitigate the limitations of traditional recurrent neural networks in handling long-term dependencies by incorporating both long-term and short-term memory capabilities. This makes LSTM particularly suitable for predicting flow patterns over time.

Data Preprocessing

Preprocessing focused on rainfall intensity and event duration as primary predictors, with additional features like observed flow and cumulative precipitation also tested. Input and output data were normalized using the "minmaxscaler" from scikit-learn (Pedregosa et al., 2011).

Model Development and Validation

Figure 2 illustrates the training routine, which consisted of two distinct phases and involved splitting the data into three subsets: training, validation, and testing. In the first phase, 5-fold cross-validation, as described by Matzka (2021), was applied to evaluate the model's reliability and reduce the risk of overfitting. In the second phase, the model was fully trained on the complete training dataset and validated on the test set to assess the final performance of each tested model.

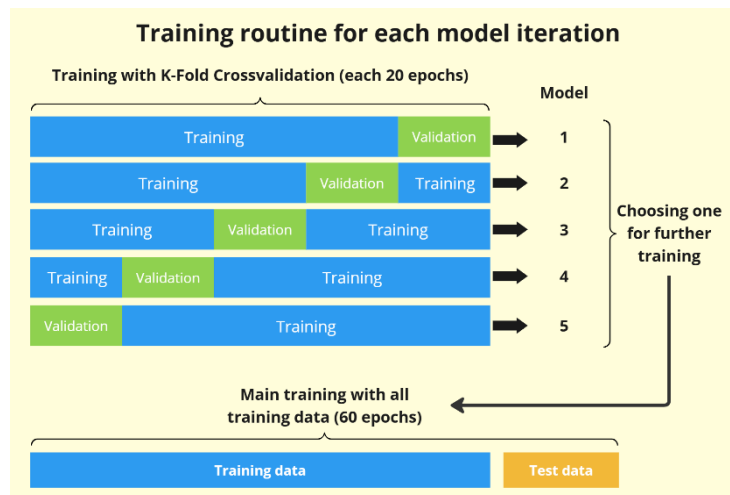


Figure 2: Training routine for each model iteration

Through iterative fine-tuning, adjustments in the model's architecture (e.g., increasing the number of layers and neurons) and input features (e.g. cumulative precipitation) were implemented to evaluate effectiveness of various configurations.

Results and Discussion

The study examined and refined various LSTM model configurations for predicting flow at the outlet of the catchment, ultimately determining that a model using an LSTM network with three layers of 128 neurons each was most effective among the tested iterations.

Accuracy

Table 1 shows the summary of the model development and the accuracy of each iteration. The final model showed significant improvements compared to the initial model, with the root mean squared error (RMSE) decreasing from 0.026 m³/s to 0.01 m³/s. Additionally, the mean deviation from peak flow improved from 9.4 % to 4.6 %. However, these enhancements were achieved predominantly through adjustments in the model's architecture, such as the loss function and the number of neurons and layers, rather than through changes in data preprocessing or feature engineering (e.g. Integration of cumulative precipitation). The observed leading flow was also tested as an input feature and demonstrated a small improvement in accuracy but was not included in the final model as it assumes that live measurements of the targeted flow are available. These findings highlight the potential for further accuracy improvements through more extensive feature engineering, such as integrating new variables or refining current inputs. To illustrate the improvements in predictions, Figure 3 depicts exemplary plots of forecasted events for both the initial and final model.

Table 1: Summary of the model development

Compared change to prior model	Type of change	Selected	RMSE	Deviation from peak flow
Initial model	-	-	0,026 m³/s	9,4 %
Loss function MAE vs. MSE	Model architecture	MSE	0,019 m ³ /s	9,9 %
Non mixed vs mixed samples	Model architecture	Mixed	0,019 m ³ /s	9,9 %
Number of neurons (32 vs. 64 vs. 128 vs. 256)	Model architecture	128	0,012 m ³ /s	4,8 %
Number of layers (1x128 vs. 2x128 vs. 3x128)	Model architecture	3x128 (final Model)	0,01 m ³ /s	4,6 %
Integration of cumulative precipitation	Input feature	-	0,01 m ³ /s	4,6 %
(Integration of observed leading flow) ¹	Input feature	-	(0,008 m ³ /s) ¹	(4,4%) ¹

¹ This configuration of the model assumed that live measurements of the observed flow are available and is not directly comparable to the other model iterations.

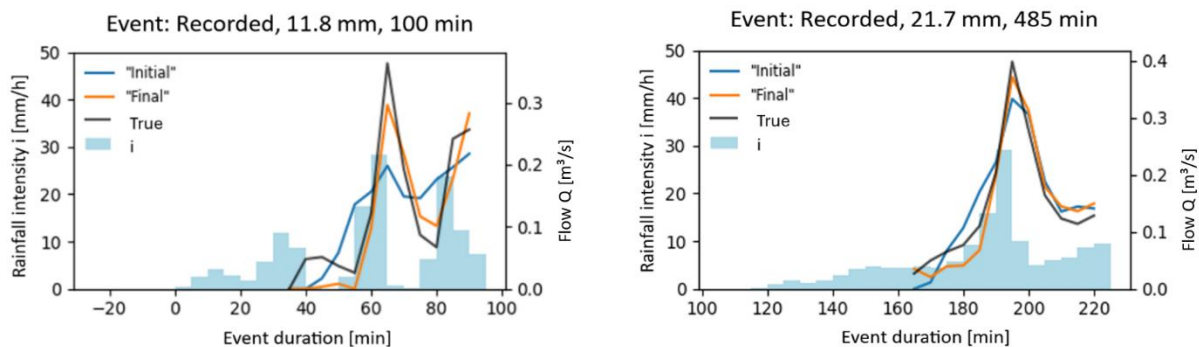


Figure 3. Comparison of predictions for two recorded events, illustrating flow predictions by the initial model, final model, and the true values.

Training and Prediction Time

The complexity of the final model increased training time from 4 to 31 minutes on a CPU (AMD Ryzen 7 5800H 3.20 GHz). The increase in training time corresponds to the model's added complexity, which included more neurons and layers.

In comparison to SWMM, the ML model's prediction time (60 milliseconds) for one event was more than three times faster. When predicting all test events at once, the ML model was, on average, up to 600 times faster, but this improvement is likely influenced by more effective utilization of parallel computing capabilities in the ML framework and is primarily relevant for mass simulations. These results confirm literature findings (Kochkov et al., 2021; Palmitessa et al., 2022) and show that ML models are highly efficient for real-time predictions and capable of accurately modelling flow in sewer systems. However, regarding this small catchment, SWMM was still effective for real-time prediction with 200 milliseconds of calculation time per event. It is important to note that these efficiencies will not be consistent for larger and more complex catchments, since complexity could substantially increase computational times for SWMM models. In contrast, the LSTM model is likely to maintain its rapid prediction capabilities, assuming the underlying model architecture remains unchanged.

Model Transfer

The ML model was adapted and trained on a more complex system, with the aim to predict overflow from a stormwater treatment basin. This change resulted in decreased accuracy, with the mean deviation from peak flow rising to 8.2 % and RMSE increasing to 0.046 m³/s. This deterioration in accuracy was anticipated due to the model's development under simpler conditions. Further research is needed to assess the accuracy in more complex systems. The transition to the new system was found to be straightforward and only required a new SWMM model, as the whole process of simulation, data preparation, and model training was automated.

Limitations and Scope

Due to its reliance on synthetic data and training within a simple catchment, the ML model may perform differently in more complex environments. It has been optimized for cases simulated with SWMM software, focusing on flow prediction at selected nodes. Future work should therefore incorporate real-world data and conduct broader applicability tests across varied system complexities.

Conclusion

AI and ML offer promising tools for anticipating the impacts of precipitation events rapidly and accurately. This study explored the potential of ML models in predicting urban sewer flows, focusing on developing a model leveraging LSTM-Networks and training data generated from the hydrodynamic simulation model SWMM.

The final LSTM-based model achieved a mean deviation from peak flow of 4.6 % for the study area and demonstrated high accuracy and efficiency in real-time prediction for sewer networks.

However, its performance was influenced by the reliance on synthetic data, underlining the need for high-quality, real-world data in future applications. Additionally, incorporating more extensive input feature selection, such as adding observed system data (e.g., observed flow, observed water levels), could further enhance the model's accuracy and adaptability. Research should address these challenges by using real-world measurement data and testing the model in varying and more complex catchments to better compare its capabilities with traditional hydrodynamic models.

Ultimately, the adaptability of these models to real-time sewer management systems could play a critical role in addressing urban water management challenges in the face of increasing climate variability.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.
- Cerulli, G., 2023. Fundamentals of Supervised Machine Learning: With Applications in Python, R, and Stata, Statistics and Computing. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-031-41337-7>
- Chollet, F. et al., 2015. Keras [WWW Document]. URL <https://keras.io/> (accessed 5.5.24).
- DWD, 2024. Wetter und Klima - Deutscher Wetterdienst - CDC (Climate Data Center) [WWW Document]. URL https://www.dwd.de/DE/klimaumwelt/cdc/cdc_node.html (accessed 5.25.24).
- El-Din, A.G., Smith, D.W., 2002. A neural network model to predict the wastewater inflow incorporating rainfall events. *Water Res.* 36, 1115–1126. [https://doi.org/10.1016/S0043-1354\(01\)00287-1](https://doi.org/10.1016/S0043-1354(01)00287-1)
- Grossman, D., Doyle, M., Buckley, N., 2015. Data Intelligence in 21st Century Water Management. The Aspen Institute, Washington, DC, USA.
- Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. *Neural Comput.* 9, 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Kochkov, D., Smith, J.A., Alieva, A., Wang, Q., Brenner, M.P., Hoyer, S., 2021. Machine learning–accelerated computational fluid dynamics. *Proc. Natl. Acad. Sci.* 118, e2101784118. <https://doi.org/10.1073/pnas.2101784118>
- Matzka, S., 2021. Künstliche Intelligenz in den Ingenieurwissenschaften: Maschinelles Lernen verstehen und bewerten. Springer Fachmedien Wiesbaden, Wiesbaden. <https://doi.org/10.1007/978-3-658-34641-6>

- Palmitessa, R., Grum, M., Engsig-Karup, A.P., Löwe, R., 2022. Accelerating hydrodynamic simulations of urban drainage systems with physics-guided machine learning. *Water Res.* 223, 118972. <https://doi.org/10.1016/j.watres.2022.118972>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Rossman, L., 2015. Storm Water Management Model User's Manual Version 5.1. EPA U.S. Environmental Protection Agency, Cincinnati, Ohio.
- Sit, M., Demiray, B.Z., Xiang, Z., Ewing, G.J., Sermet, Y., Demir, I., 2020. A comprehensive review of deep learning applications in hydrology and water resources. *Water Sci. Technol.* 82, 2635–2670. <https://doi.org/10.2166/wst.2020.369>