Using Optimisation Method for Input Data Selection in AI-Based Time-Series Flood Forecasting Models

School of Computing and Engineering

IIoT research group

UNIVERSITY OF WEST LONDON

Farzad Piadeh, Kourosh Behzadian, Amir Alani

School of Computing and Engineering, University of West London, St Mary's Rd, London, W5 5RF, UK
Contact e-mail: 21452390@student.uwl.ac.uk

Introduction

Data-driven models for real-time flood forecasting use time-series input data by using specific range of input data for each iteration of both training and validation processes. However, little is known about the preparation and classification of these input datasets to achieve the best model performance.

Aim and Objectives

Proposing an input variable selection method based on nature-based optimisation model to obtain the best dataset of input data for AI-Based time-series flood forecasting models.

Methodology

- The methodology comprises three components: "data collection and preparation", "model development" and "performance assessment" (Figure 1) and applied for real case study, meaning Ruislip urban drainage system (Figure 2).
- The nonlinear autoregressive network with exogenous inputs (NARX) models is selected, which is recommended as most powerful model is suitable for multivariate time series hydrological and hydraulics problems (Piadeh *et al.*, 2022).
- Three conventional models first are defined: (1) Built model with one rainfall monitoring station (RMS), (2) Built model with three more correlated RMS, (3) Built model with 6 RMS to show the importance of data integration in model performance.
- Range of input data are selected based on other studies' recommendations, i.e. using lag time in which best cross-correlation or cross-covariance coefficient is obtained between rainfall station and water level data) (Snieder *et al.,* 2020).
- Performance are assessed for 15-minute, 1-hour, 2-hour and 3-hour steps ahead forecasting in conventional models and only 3-hour lead time for optimised model.
- Indicators of performance assessment are set to (1) Root mean square error (RMSE), (2) Normalised Nash–Sutcliffe model efficiency coefficient (NNSE), and (3) Accuracy of overflow detection (ACC) for identified flood events only inspired by Piadeh *et al.* (2021).
- Best conventional model is optimised by the shuffled frog leaping algorithm to show difference between conventional approach and optimised approach in input data selection, each trial includes 4 and shuffle sample for exploration and exploitation step, respectively. Objectives were RMSE, NSE, and ACC enhancement. Stopping criteria is set to improvement less than 0.01% (Bui *et al.,* 2020).

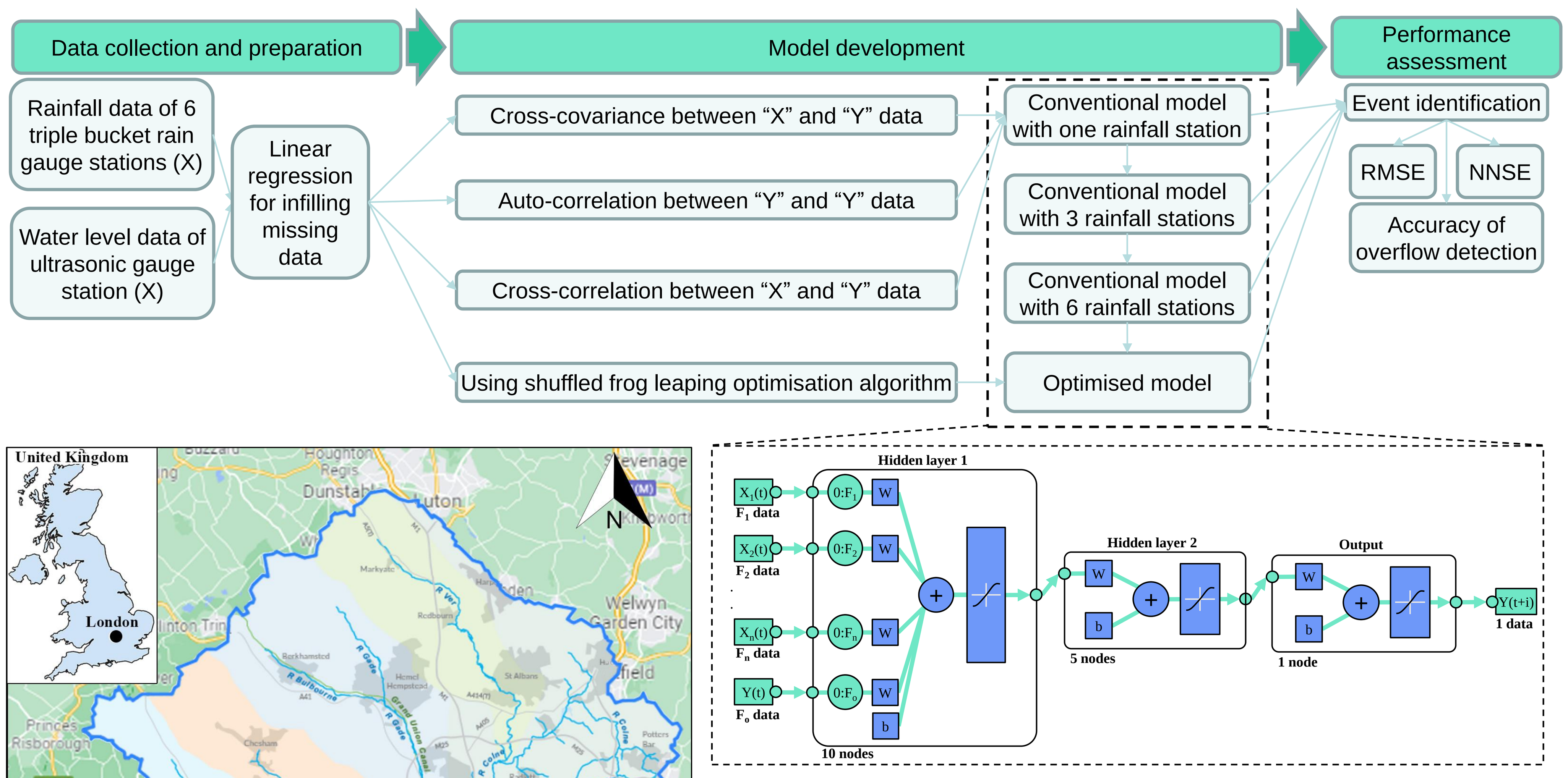

Figure 1. The methodology applied for developing different urban flood forecasting models

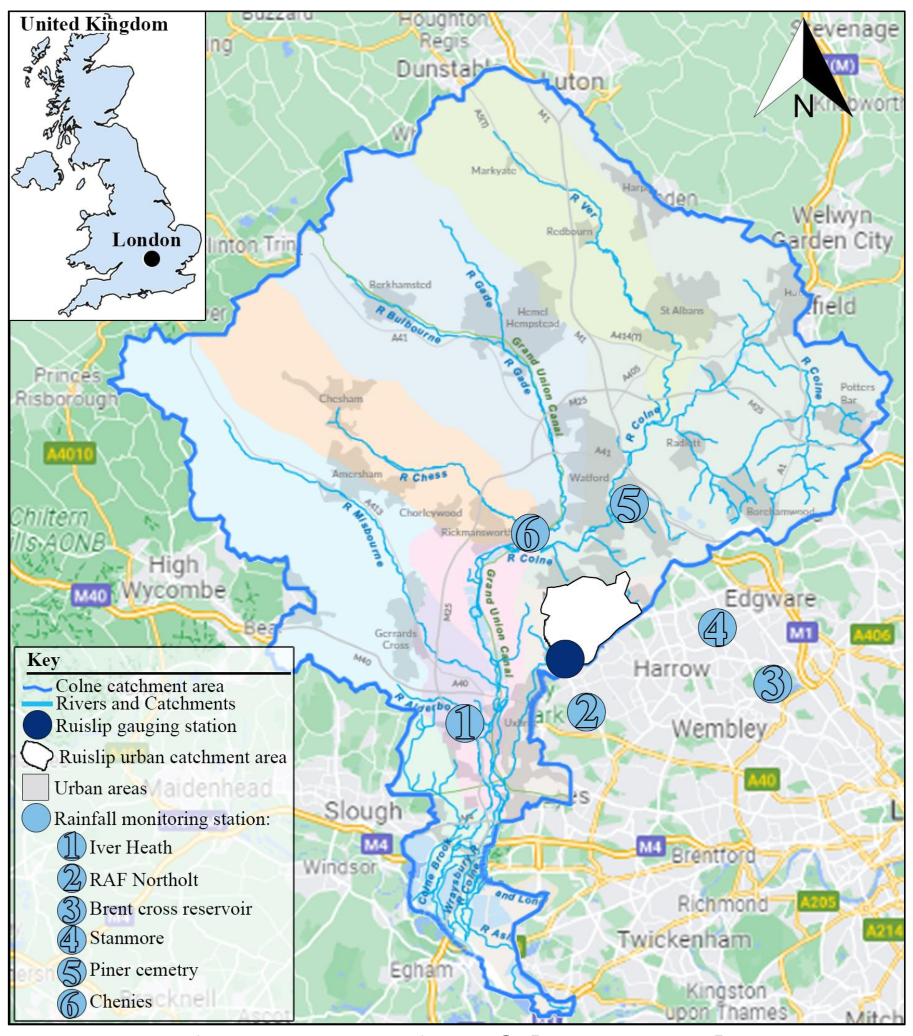

Figure 2. Location of the case study

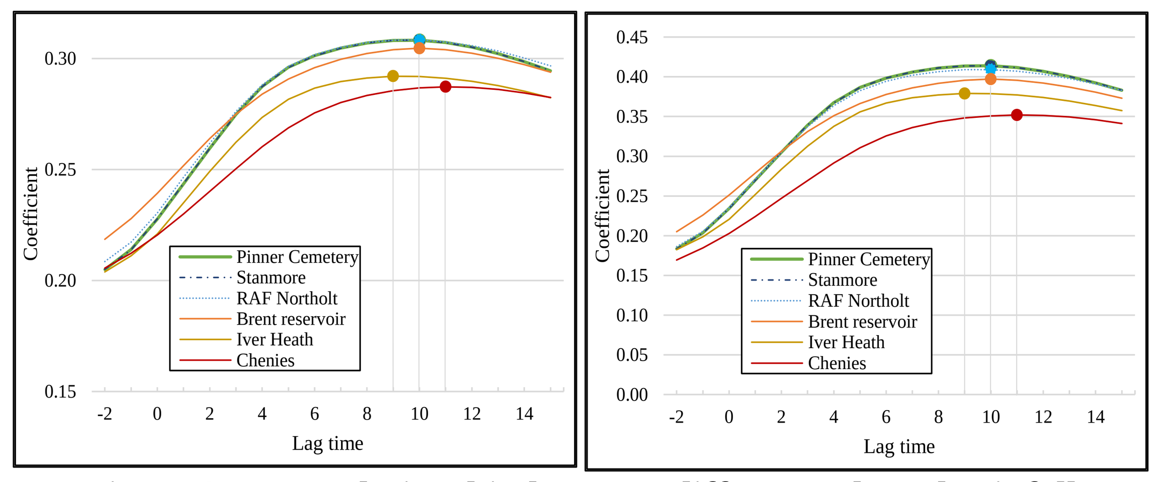

Figure 3. Data relationship between different selected rainfall stations and water level: (Left): Cross-correlation, (Right): Cross-covariance

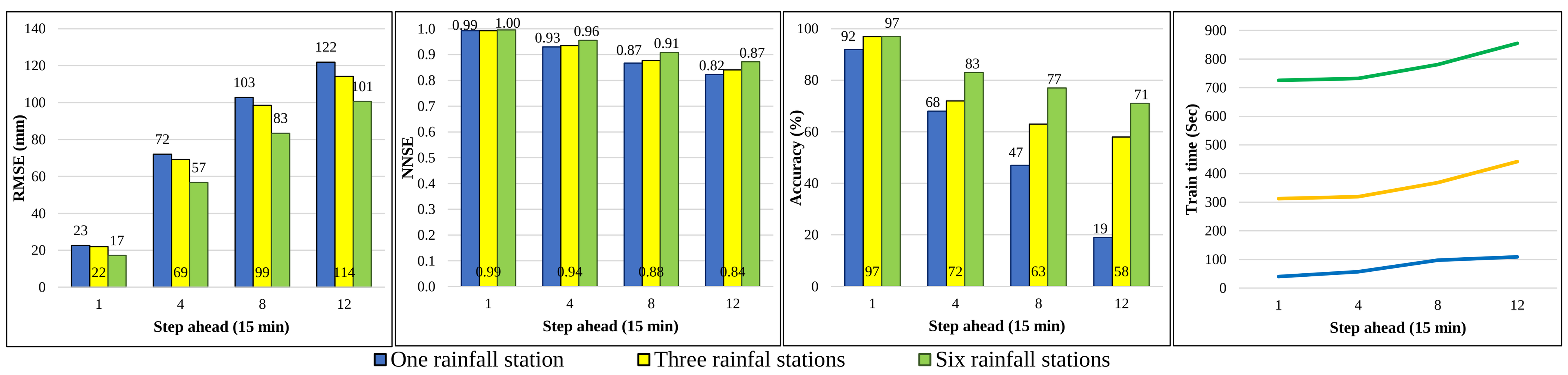

Figure 4. Performance indicators for conventional built models

Trial	Lag times of dataset						
	R1	R2	R3	R4	R5	R6	Water level
0 (Conventional approach)	9	10	10	10	10	11	1
1	7	7	3	4	8	8	6
2	4	5	9	10	2	11	3
3	4	4	4	9	5	7	2
4	3	3	9	8	5	4	8
5	7	9	6	7	2	5	9
6	2	3	5	3	5	7	2
7	5	5	2	6	8	7	5
8	4	8	8	5	7	7	2
9	3	5	9	6	7	7	5
10	4	5	9	6	7	7	5

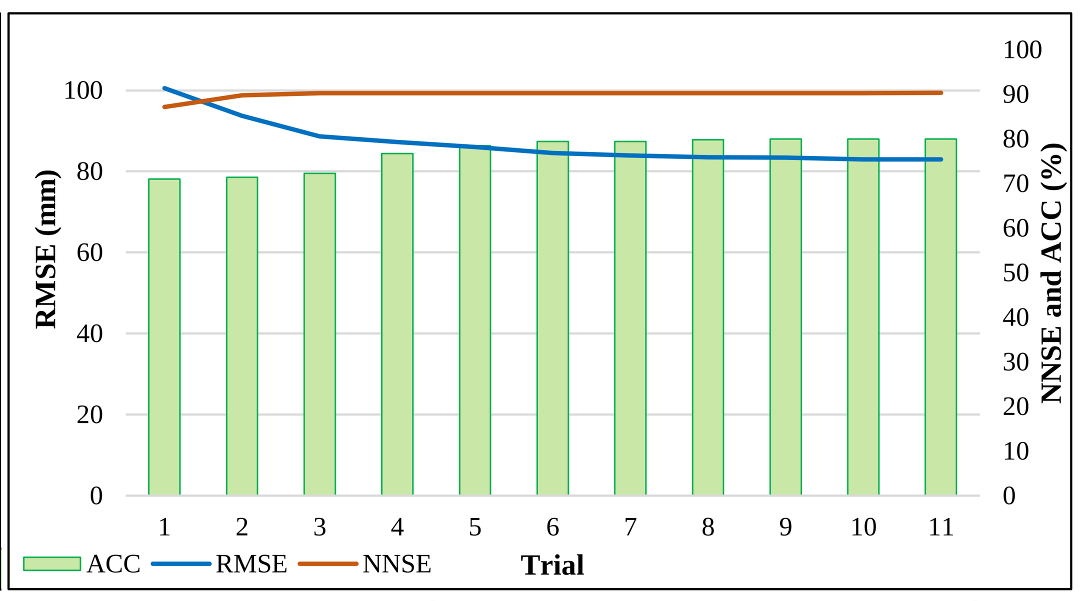

Figure 5. Each trial's model improvement using optimisation model (Left) Decision variables, (Right) Performance

Results

- Conventional models are improved with including more rainfall data stations, where RMSE improved 20% and ACC increased form 19% to 71% for 12-step ahead for instance. This result shows that data integration can be helpfully increased model performance of AI models, which also is compatible by result obtained from Zounemat-Kermani *et al.* (2020). However, computational time increased significantly for best model (See Figure 4-Right).
- Testing all different feedback delays (range of data input) required around 280 years (each run model needs around 15 min. for training and validation), Therefore, optimised model can help noticeably.
- Results shows that input variables for optimised model are not same as conventional approaches were previously recommended (Figure 5-Left).
- Optimised model was built after 20 hr. (10 trial and 80 run model), in which ACC is raised from 71% to 80%, NNSE slightly improved and more importantly RMSE decreased from 101 to 83 mm (Figure 5-Right).
- Optimised model could reduced error in especially overflow condition and high depth uprising (Figure 6).

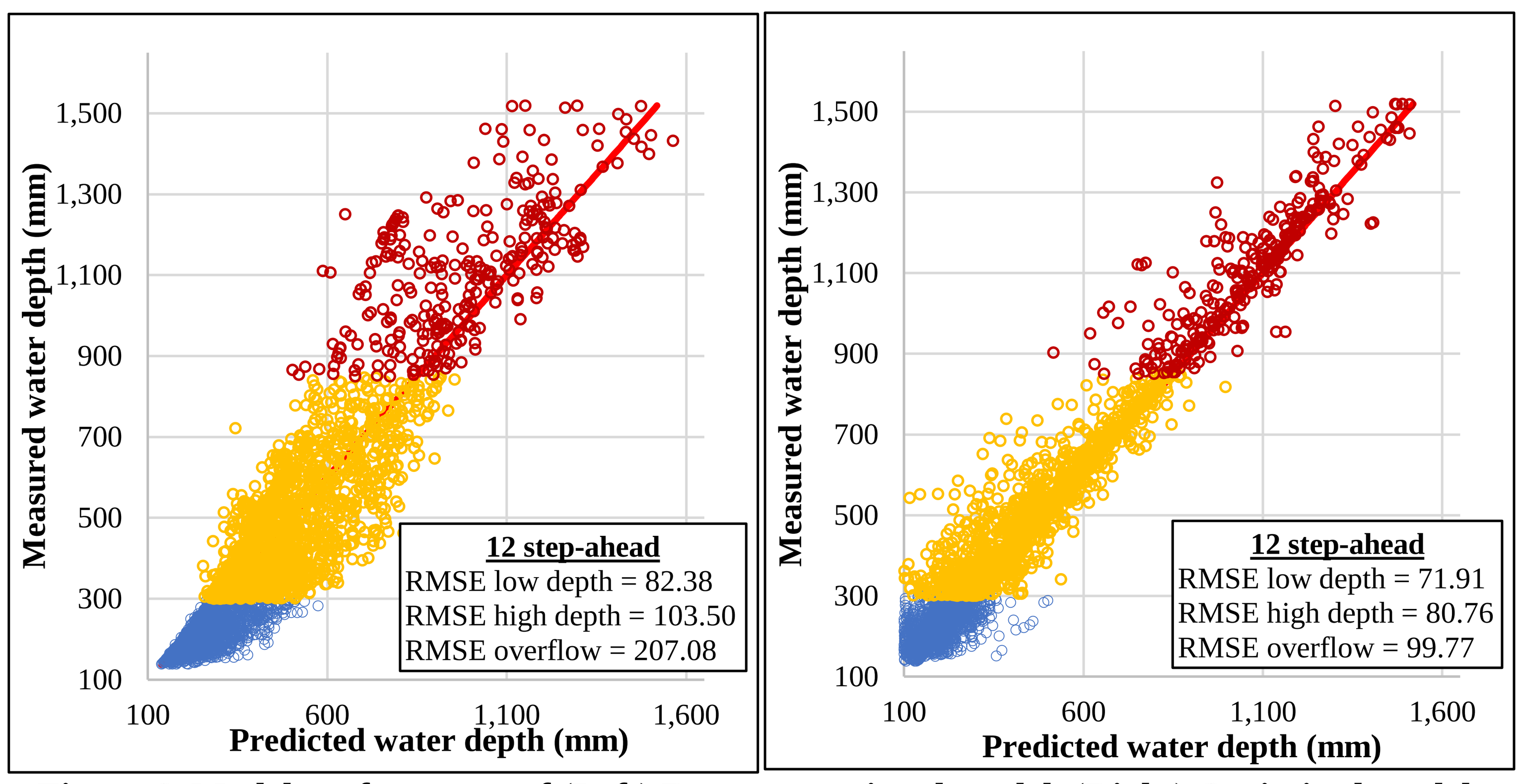

Figure 6. Model performance of (Left) Best conventional model, (Right) Optimised model

References

Bui, Q., Nguyen, Q., Nguyen, X., Pham, V., Nguyen, H., Pham, V. (2020). Verification of novel integrations of swarm intelligence algorithms into deep learning neural network for flood susceptibility mapping. *Journal of Hydrology*, 581: 124379.

Piadeh F., Behzadian K. Alani A. (2021). The Role of Event Identification in Translating Performance Assessment of Time-Series Urban Flood Forecasting. *UWL annual conference*, July 2021, London UK.

Piadeh F., Behzadian K. Alani A. (2022). A Critical Review of Real-Time Modelling of Flood Forecasting in Urban Drainage Systems. *Journal of Hydrology*, 127476.

Snieder, E., Shakir, R., Khan, U.T. (2020). A comprehensive comparison of four input variable selection methods for artificial neural network flood forecasting models. *Journal of Hydrology*, 583:124299.

Zounemat-Kermani, M., Matta, E., Cominola, A., Xia, X., Zhang, Q., Liang, Q., Hinkelmann, R. (2020). Neurocomputing in surface water hydrology and hydraulics: A review of two decades retrospective, current status and future prospects. *Journal of Hydrology*, 588, 125085.