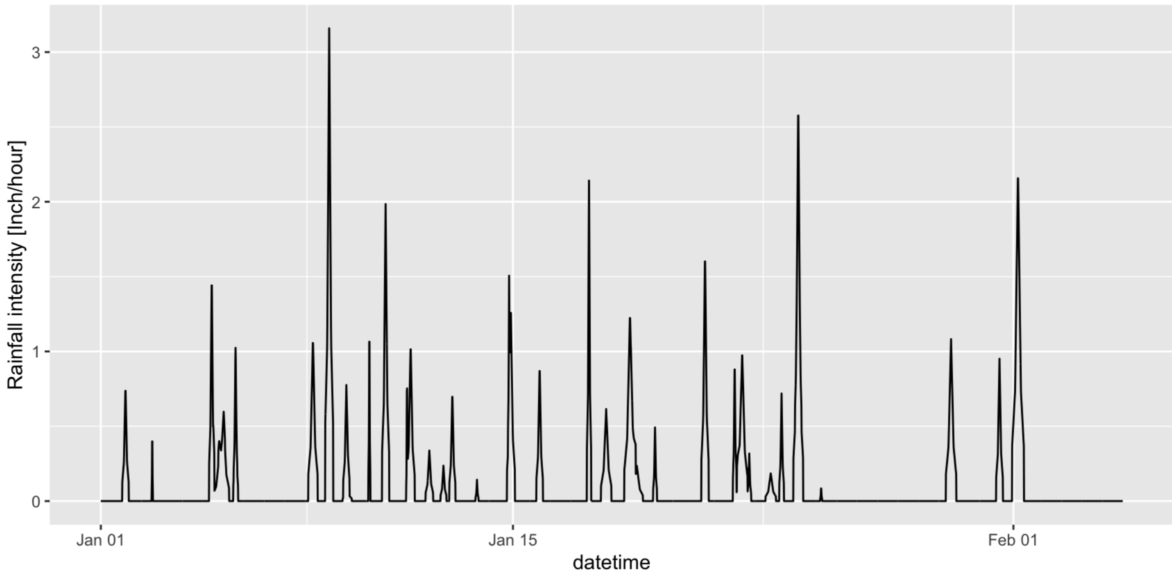
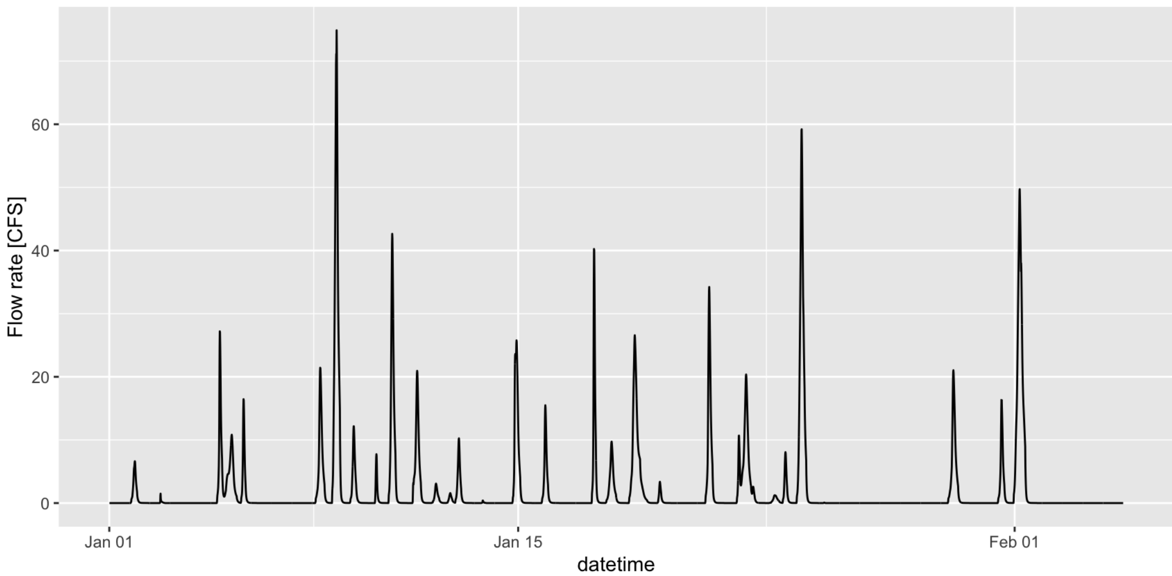
1. The rainfall time series shown in the following figure is randomly generated using the code in the "*prepare.R*", and is stored in "data/rain.data".



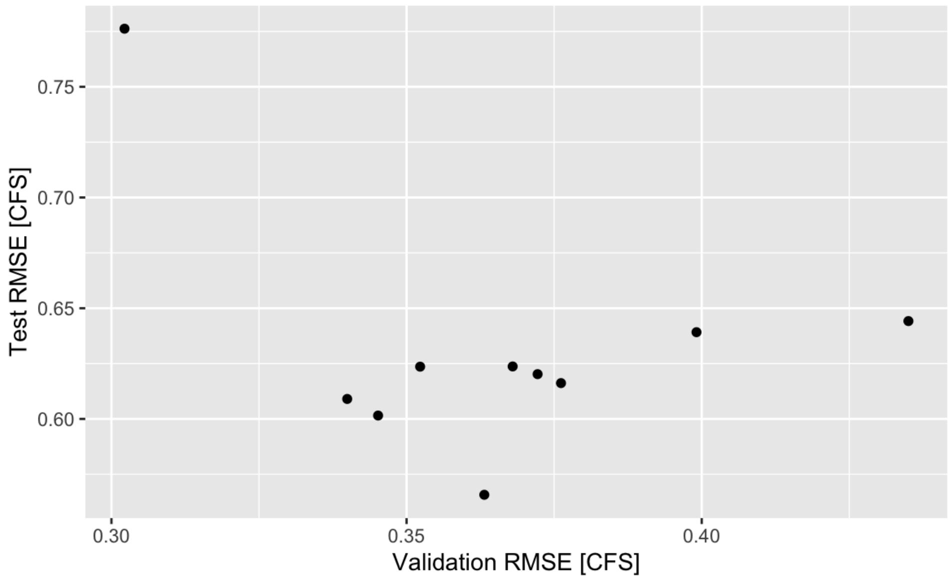
2. The rainfall time series is then used to drive the simulation of an SWMM model included in the SWMM application manual Project 4 (Gironás et al., 2010). The simulated outflow hydrograph is stored in "*data/outflow.txt*". Multiple types of sustainable urban drainage systems (SuDS) practices are installed in the catchment. The outflow hydrograph is shown as the following. The rainfall-runoff data is stored in "data/rainfall_runoff.data" and is then used for training XGBoost models.



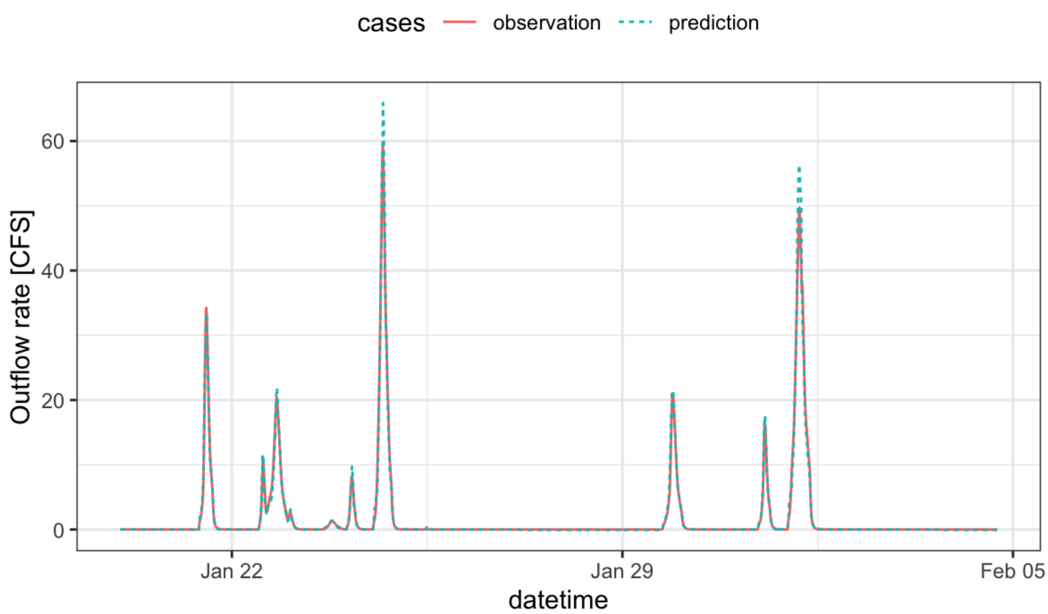
3. The dataset is divided into training, validation, and test sets. The data splitting information is stored in ` *full_event_divide_df* ` variable. The functions to generate feature engineering hypermeters and input features for training XGBoost is included in "*modeling.R*". The key functions used are "*feature_vector*", "*gen_s_e*", "*gen_feature_para*", "*gen_feature*".

4. The XGBoost models are trained using the XGBoost package (Chen and Guestrin, 2016). For each set of feature engineering hyperparameters, the effectiveness of a number of XGBoost hyperparameters in "*tune_grid*" is evaluated. The performance metrics of the models, the models, and the model predictions are stored in "*data/results*".

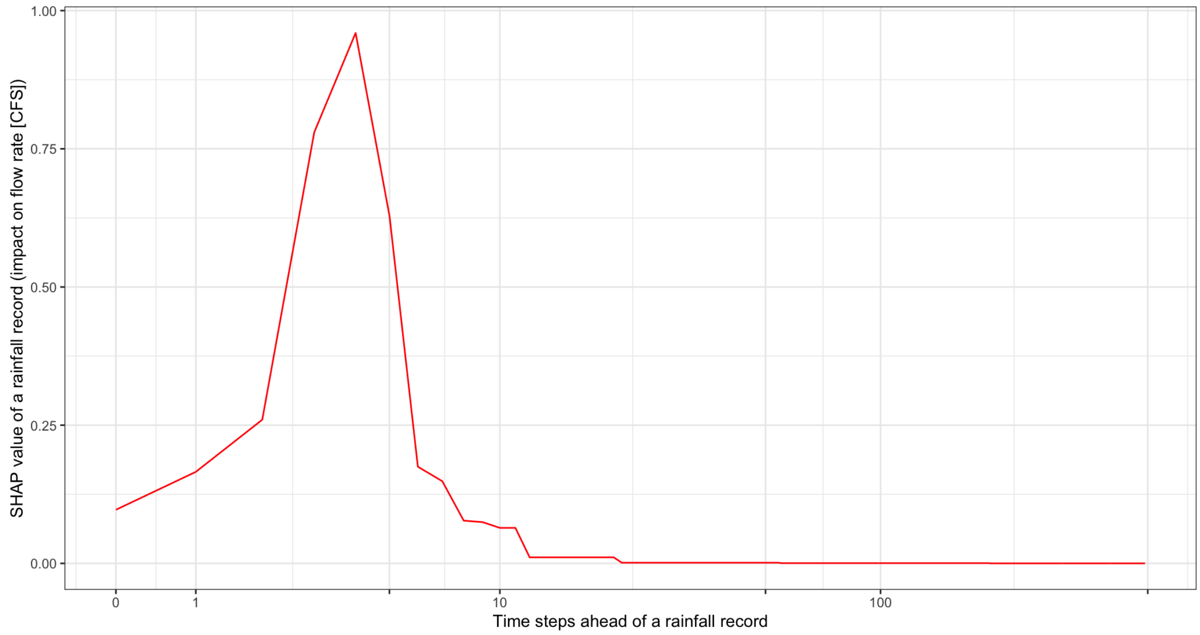
5. The validation error is compared to the test error in the following figure. The script for plotting can be found in "*plot.R*".



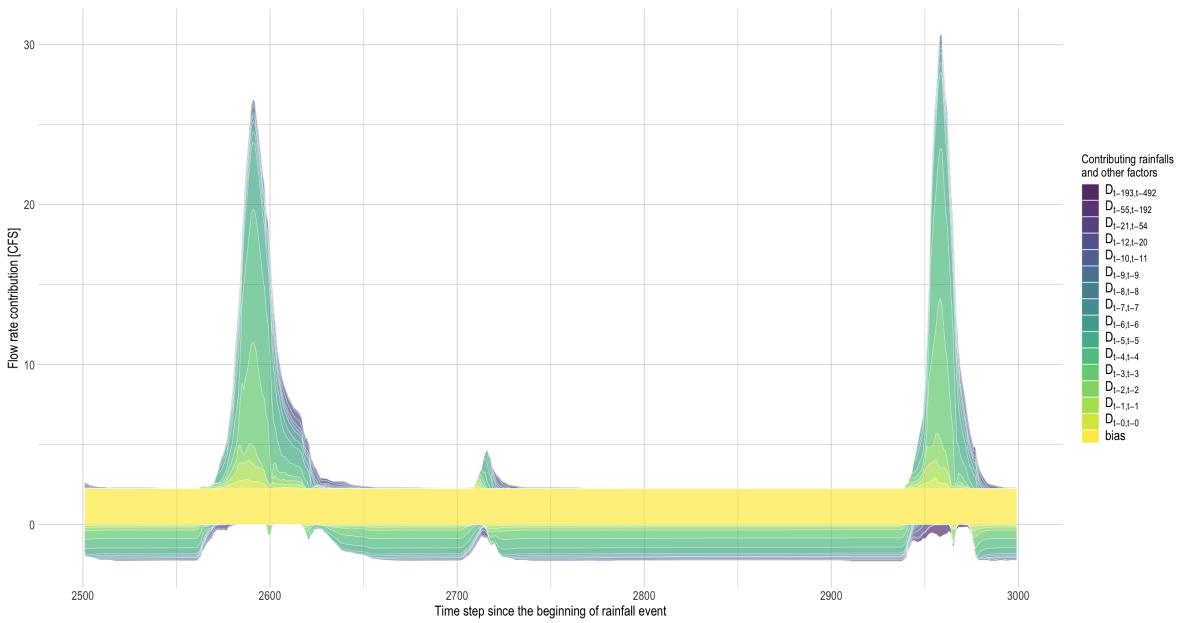
6. The following feature compares the observed and predicted hydrographs for the test period of a randomly chosen set of feature engineering hyperparameters. The predicted hydrograph matched the observed hydrograph very well.



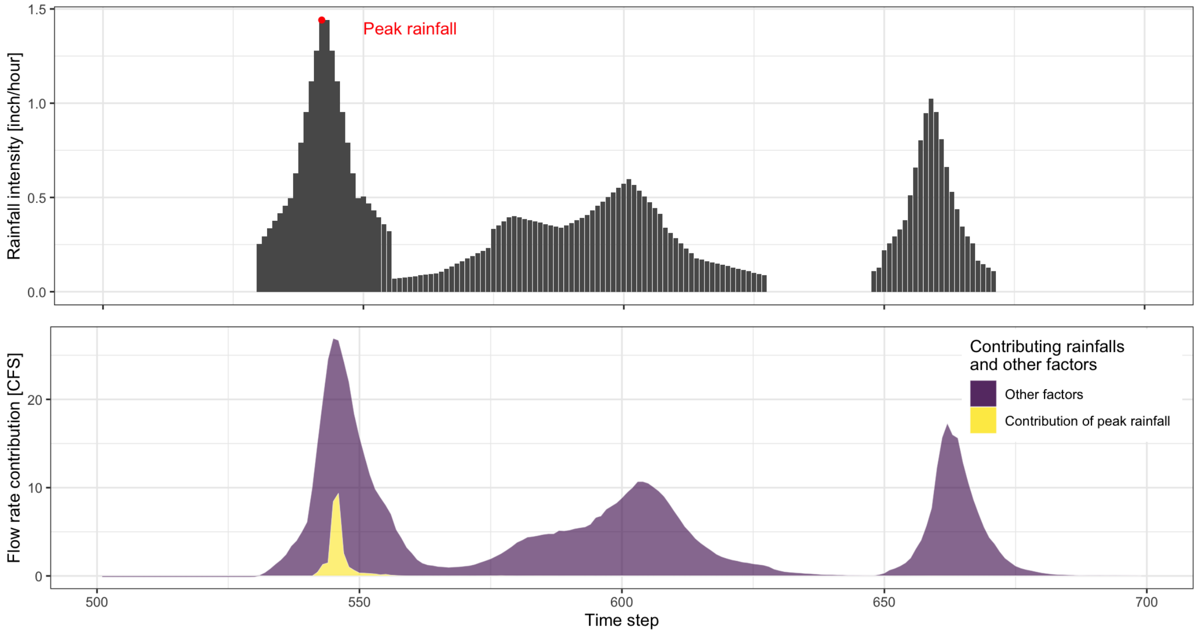
7. Contribution of rainfalls to each runoff prediction is analyzed in “SHAP.R”. The rainfall has the largest impact on runoffs at 4 time-steps ahead or 40 minutes. The script for generating the following figure is provided in code chunk “*Catchment response time*”. TreeExplainer method developed in Lundberg et al. (2020) is used to compute the contribution of each input feature to each runoff prediction.



8. The following figure shows the contribution of rainfall depth features to runoff at each time step, which is useful information for hydrograph separation. The script to generate the following figure is provided in “*hydrograph separation*” code chunk.



9. The continuing impact of a rainfall depth record is shown in the figure below. The contribution of peak rainfall is shown in yellow, and the contribution of other factors is shown in purple. The script to generate the following figure is provided in the "Continuing impact of a rainfall depth record" code chunk.



Reference:

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 785–794. https://doi.org/10.1145/2939672.2939785

Gironás, J., Roesner, L.A., Rossman, L.A., Davis, J., 2010. A new applications manual for the Storm Water Management Model (SWMM). Environ. Model. Softw. 25, 813–814. https://doi.org/10.1016/j.envsoft.2009.11.009

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9