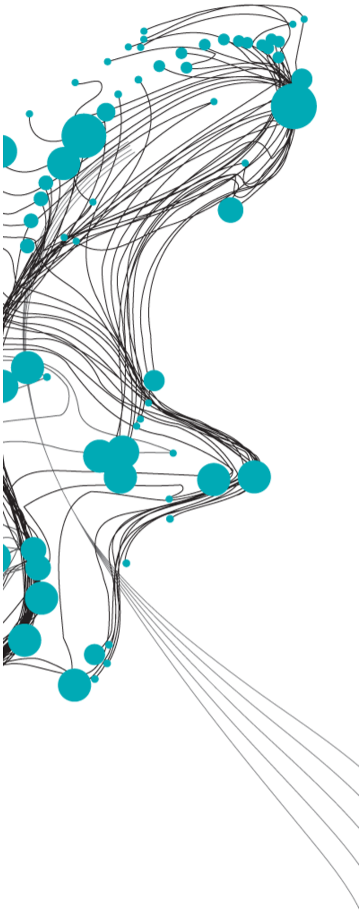
PREDICTING FLOODING DUE TO EXTREME PRECIPITATION IN AN URBAN ENVIRONMENT USING MACHINE LEARNING ALGORITHMS



Pluvial flooding in an urban environment can occur quite sudden. Therefore, flood early warning systems with a short run time are desired. A method to reduce computational load is surrogate modelling. Response surface surrogate models (machine learning (ML) algorithms) are a second level abstraction from reality. These algorithms do not emulate any internal component of the original simulation, but try to find relations between the input variables and output. They are, once trained, extremely fast in predicting the output from a given input. Therefore, the use of such ML algorithms as a flood early warning system is studied.
To limit the complexity and size of the data set, a case study area is chosen that is smaller than the whole municipality of Amersfoort. Therefore, the area of Hooglanderveen is chosen. This area has a combined sewer system, which historically has experienced frequent flooding.
Precipitation time series are used for the prediction of flooding by the ML algorithms. Precipitation statistics are used for the construction of a synthetic precipitation event data set, as there are not enough recorded data from historic flood events available that can be used to train, test and validate the ML algorithms. Maximum flood volume and flood volume time series for all manholes in the area are obtained using a sewer model.
The architecture of an artificial neural network (ANN) makes it very suitable for the simulation of a sewer system. Two ANN types are constructed and tested in the context of this research. First, the multi-layer perceptron (MLP) is constructed for both classification and regression. With classification, flooding of manholes is classified for each precipitation event, predicting only if flooding occurs at each manhole. With regression, the maximum flood volumes that occur at each manhole is predicted for each precipitation event. Second, a long short-term memory (LSTM) network is constructed for the regression of flood volume time series. Here, flood volume time series are predicted for all manholes in the studied area.
To test the LSTM algorithm on a realistic precipitation data set, 5 historic flood events from the case study area of Hooglanderveen are obtained. These are historic precipitation events that induced flooding, as reported by inhabitants of the area. The historic data is also fed into the sewer model to obtain flood volume time series for all manholes in the study area.
The algorithms are validated on a separate validation data set (20% of the data with 25 precipitation events). Evaluation scores of the three ML algorithms are shown in Tab. 1. Fig. 1 shows an example of prediction done by the LSTM on the validation data set. Fig. 2 shows an example of a prediction done by the LSTM on the historic data. High evaluation scores can be seen even on the historic data. This shows that the ML algorithms are capable of predicting flooding and can be used as flood early warning systems.

Table 1: Evaluation parameters for the ML algorithms validated on the validation data set and the LSTM regressor tested on the historic precipitation data.

Evaluation parameter	MLP Classifier	MLP Regressor	LSTM Regressor	LSTM Regressor (historic data)
Accuracy	99.24%	-	-	-
R^2	-	0.997	0.99	0.99
MAE	-	0.20 m^3	0.06 m^3	0.19 m^3
NSE	-	-	0.87	0.61

Raphaël Kilsdonk

Graduation Date:
9 April 2021

Graduation committee:
University of Twente
Prof.dr. K.M. Wijnberg
Dr.ir. A. Bomers

Hydrologic
Ir. M. van den Brink

Figure 2: Prediction of the LSTM on synthetic data.

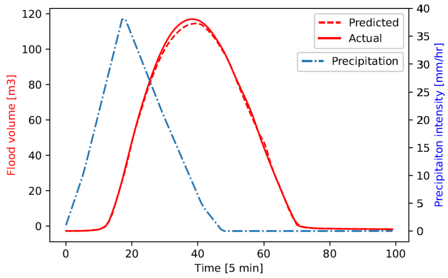
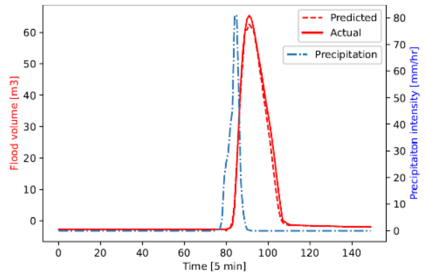


Figure 1: Prediction of the LSTM on historic data.



UNIVERSITY OF TWENTE.