

Prediction of CSO Chamber Water Levels Using Rainfall Forecasts

T. Rosin^{1,2,*}, M. Romano², K. Woodward², E. Keedwell¹, Z. Kapelan¹

¹ Centre for Water Systems, College of Engineering, Mathematics and Physical Sciences,
University of Exeter, Exeter, UK, EX4 4QJ

² United Utilities Plc, Lingley Mere Business Park, Warrington, UK, WA5 3LP
**tr319@exeter.ac.uk*

ABSTRACT

CSOs are a major source of pollution, spilling untreated wastewater directly into water bodies and/or the environment. This paper describes an investigation of the use of rainfall forecasts when predicting water level in a Combined Sewer Overflow (CSO) chamber. A prediction model based on a multi-layer feed-forward Artificial Neural Network (ANN) with a sliding time-window of inputs was developed to model the relationship between contributing rainfall and water level within a CSO chamber and predict the occurrence of spills. Three versions of the aforementioned prediction model were developed and compared: 1) without using forecast rainfall data, 2) with perfect rainfall forecasts (i.e. forecasts assuming perfect knowledge of historical rainfall into the future) and 3) with real rainfall forecasts. These models were tested on a CSO located in the Wirral area of the United Utilities network and were all able to successfully predict CSO spills 2 hours ahead. Use of forecast radar rainfall data, generated by the Met Office, was shown to demonstrably improve the accuracy of the ANN predictions and increase the prediction range of the model. Unsurprisingly, this accuracy was further improved when using perfect rainfall forecast data, due to errors inherent in real forecasts. It is envisioned that ANN models could be beneficially used by water utilities for near real-time modelling of water level in CSO chambers and generate alerts for upcoming spills events. The use of forecast data increases the detection time of CSO discharges, potentially enabling water utilities to take appropriate and timely action to address potential issues before the customers and the environment are adversely effected.

Keywords: Artificial Neural Network, Combined Sewer Overflow, Radar Rainfall Nowcasts

1 INTRODUCTION

CSOs are designed to protect downstream networks and wastewater treatment plants from hydraulic overloads and flooding during extreme rainfall events. Excess flow, containing untreated stormwater and wastewater, is discharged directly to nearby receiving watercourses such as rivers, streams, lakes, oceans and other water bodies. CSO discharges are necessary to prevent wastewater from flooding streets and backups into homes and businesses. However, spills contain untreated, or only partially treated, human, agricultural, and industrial waste, debris and stormwater which can cause significant harm to the surrounding environment, degrading water quality, threatening public health and resulting in regulatory failures. CSOs are designed to spill during heavy storm events, when pollutants in the foul water are more diluted, but damage such as blockages and siltation can cause CSOs to spill at pass forward rates lower than consented, increasing the impact on the environment.

CSO are common within the UK, where approximately 70% by length of sewer systems are combined [1]. UK Water utilities are under increasing pressure from the regulators to reduce the number of unconsented spills which occur in their network. If spills can be predicted in advance techniques are available for mitigation, such as maximising storage and adjusting set points for movable gates and pumping stations [2]. Consequently, there has been a growing interest in developing methods to accurately predict CSO performance in near real-time and provide an early warning for overflows.

Several numerical models have been developed to model storm sewer systems (CSSs). Leon et al. [3] developed a finite volume model to simulate the flow dynamics (transient and non-transient conditions) in CSSs. Politano et al. [4] designed a dynamic transient model to simulate the transition from gravity flow to pressurized flow in CSO systems in the city of Atlanta. Using a commercial urban drainage model Thorndahl & Willems [5] simulated an 18 year rainfall series to establish the relationship between system failures (including CSO discharges) and rainfall event characteristics in Denmark. Fortier & Mailhot [6], studying the possible impact of climate change, designed a simple model relating CSO occurrence to rainfall characteristics, using data from 20 CSOs in Quebec. Zimmer et al [7] developed a numerical model to predict CSO level, and successfully used this model in combination with model predictive control genetic algorithms to reduce the number of discharges in the Chicago sewer system by optimising sluice gate operation. Morales et al. [2] predicted the frequency and volume of CSOs and evaluated the operation of gates and its effect in controlling the occurrence of CSOs in the Chicago “Lawrence Avenue Tunnel” using a stochastic urban hydrologic model combined with a hydraulic model. Commercially, numerical models are also available which are capable of simulating the hydrologic and hydraulic behaviour of sewer systems, such as InfoWorks [8], EPA SWMM [9] and Mouse [10].

In recent years there has been a move towards modelling sewer systems in near real-time, however this is difficult to achieve reliably and accurately using numerical modelling. Indeed, numerical models are often difficult to build, calibrate and maintain; they have to be kept up to date and can be too computationally expensive if large networks are to be modelled. As a result, over the last decade, data-driven methods have become an increasingly attractive alternative to physical models in urban drainage simulation [11]. Artificial Neural Networks (ANNs) are one of the most popular machine learning models available, a biologically inspired computational approach, they imitate the neural systems found in the human brain [12]. ANNs do not require any physical information regarding the processes involved and are capable of learning, memorising and constructing relationships within data.

A number of ANNs have been developed to predict CSO performance. Fernando et al. [13] developed a feed-forward, back-propagation ANN to forecast wastewater overflows in an Australian CSS. When provided with antecedent flow rate and rainfall data the model was able to accurately forecast sewer overflow rates, although when only rainfall data was provided the model performed relatively poorly. Sumer et al. [14] used ANNs and time series analysis to monitor in real-time sanitary sewer overflows (SSOs) in two case studies located in Arizona, USA. Potential disruptions in the system were detected using control limit by identifying deviations between measured and expected depth and flow data. Kurth et al. [15] developed a three hidden-layer feed-forward multilayer perceptron (MLP) to model the relationship between contributing local rainfall and water level within a downstream CSO. Using level and rain gauge data the model could accurately predict the normal performance of a Yorkshire CSO 15 minutes ahead. Network inputs

consisted of five previous antecedent water depths from the CSO chamber and twelve previous rainfall values from two rain gauges. Mounce et al. [16] extended the prediction range of this ANN with the addition of rainfall intensity data provided by rainfall radar devices. A single layer feed-forward ANN model predicted CSO depth with less than 5% error 75 minutes ahead for unseen data. The model was later utilised by Mounce et al. [17] who investigated the feasibility of an online pilot to manage and reduce spills. The trained ANN model produced level predictions using level and rainfall data which were then analysed using fuzzy logic to determine if the measured and forecast level was within an acceptable range. The system was tested on a case study monitoring 50 sites and was able to produce reliable warnings of decreases in CSO performance which could potentially lead to a pollution event.

Having said the above, according to our best knowledge, short term forecast rainfall data (known as nowcasts) have not been applied previously to data-driven modelling of water level in CSO chambers. They have, however, been used successfully in many other areas of hydrology, including urban flow modelling [18], real-time control of sewer systems [19] and real-time flood forecasting [20, 21]. Sewer systems typically have fast response times and precipitation forecasts have the potential to improve the accuracy of level predictions in CSOs. Rainfall predictions can be uncertain and, at times, biased [22]. However, forecast rainfall methods have improved considerably since they were first introduced in the 1960's and, with advances in remote sensing and data assimilation systems, are becoming increasingly adept at providing high resolution predictions for short lead times.

The aim of this work is to investigate the feasibility of improving the accuracy and increasing the prediction range of ANN CSO modelling by incorporating rainfall forecast data. It is envisioned that this model could be used by water utilities to model CSO depth in near real-time and generate alerts for upcoming spills events, enabling them to take appropriate and timely action in mitigating overflows. The paper is structured as follows. After this introduction, the methodology developed for the ANN forecasting is described. The ANN models' tests on a United Utilities CSO are then detailed in the Case Study section together with a discussion of the results obtained. Finally, the conclusions are drawn.

2 METHODOLOGY

There are many different types of ANNs available for predicting and forecasting data. In this study a feed-forward single hidden-layer ANN was trained using the Levenberg-Marquardt (LM) algorithm. This is a form of MLP, a class of feed-forward networks consisting of three or more layers of units. The MLP is the most commonly used architecture for time series forecasting as it is able to robustly model non-linear phenomena. The LM algorithm was selected for training as it is the fastest method available for training moderate-sized feed-forward ANNs, although it requires a greater amount of memory than other methods [23].

Feed-forward ANNs have no internal memory to store past information and so cannot satisfactorily process time series data. Thus a sliding time window approach is used here, whereby past data is incorporated into the network by using a window of previous lagged level and rainfall inputs. However, selecting the right size of time window is an important consideration; if the window is too small it may not capture the necessary dynamics of the system; if the window is too large, unnecessary information may appear as noise and the learning time can be prolonged [24]. According to previous CSO ANN studies [15, 16], when the rainfall inputs are considered, the cross

correlation between the CSO level and rainfall intensity from each radar rainfall grid can be beneficially used for the purpose of selecting an appropriate lag, in addition to selecting the most important rainfall dataset to be used. Cross correlation is a measure of the similarity between two time series as a function of the displacement of one relative to the other. Therefore, bearing in mind that it takes time for rainfall to enter the sewer and reach the CSO, this information can be used (at least as a starting point) to help select a suitable length of past rainfall data to be used as input (see case study below). When the level inputs are considered, on the other hand, previous CSO ANN studies [15, 16] suggest that a trial and error procedure should be used. Thus, an investigation of the most suitable lag of level inputs has been performed in this study (see case study below).

To evaluate the effects of using forecast data three versions of the ANN model were considered. The first model was trained using antecedent observed rainfall and level data only. The second model included perfect rainfall forecasts (i.e. forecasts assuming perfect knowledge of historical rainfall into the future) in addition to observed rainfall and level data to investigate the maximum level of prediction improvements which could be obtained by using rainfall forecasts. In case three the model used actual forecast rainfall (obtained from the Met Office), observed rainfall and level data.

3 CASE STUDY

3.1 Description

This study focuses on the Hoylake road/Chapelhill road CSO, an urban CSO located in the Wirral area of the United Utilities network. The drainage area contains 23,184 properties, serving a population of 51,828 people. A schematic of the drainage area and the relevant sewer network is shown in Figure 1. During dry weather and under normal conditions the CSO displays a relatively stable diurnal pattern, reacting with an increase in water level during precipitation in the contributing catchment. Time-series level data (mm) was collected using an ultrasonic depth monitor located in the CSO chamber, with a uniform temporal resolution of 2 minutes. Four 1 km² Ordnance Survey (OS) grid squares contained sewers directly contributing to CSO level (highlighted in blue). Observed radar rainfall intensity data (mm/hr) from these squares was obtained with a 5 minute temporal resolution and 1 km² spatial resolution. Rainfall nowcast data (mm/hr) was obtained with a lead time of 6 hours, 2x2 km spatial resolution and 15 minute temporal resolution for the square covering the CSO. All time series data ranged from 27/4/16 to 14/4/17 and were interpolated to a resolution of 15 minutes during data pre-processing to match the resolution of the nowcast data.

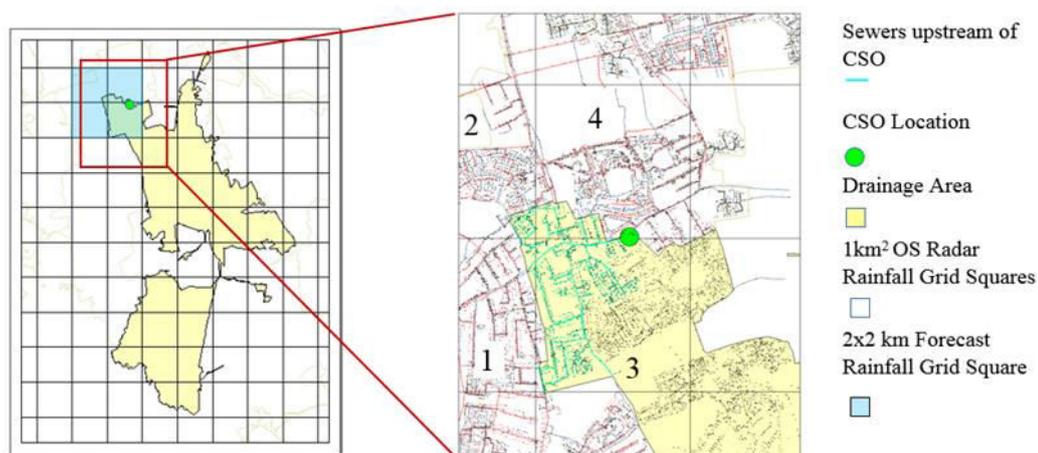


Figure 1. Case study CSO Location Showing Forecast Rainfall and Radar Rainfall Grid Squares

Both the radar rainfall datasets were obtained through the NIMROD system, an automatic radar analysis and forecast system operated by the UK Met Office. The system receives radar imagery from a network of radar stations in the UK, in this location data is provided primarily by Hameldon Hill, Lancashire. High-resolution ensemble nowcasts are then generated using the Short Term Ensemble Prediction System (STEPS) which blends extrapolated observations with the most recent high resolution numerical weather prediction (NWP) forecast.

3.2 Choice of Most Influential Ordinance Survey Grid Square

Figure 2 shows the cross correlations between the level and rainfall data calculated for a range of time lags for each rainfall radar cell over the entire dataset. All four grid squares demonstrate similar behaviour, with cross correlation increasing with increasing lag time, reaching a maximum at approximately 45 minutes before decreasing. As expected square 3, which contains a large number of upstream sewers, demonstrates the highest correlation. Based on this information, rainfall data from grid square 3 only was used as an input to the ANN models developed here. Including rainfall data from multiple grid squares was not found to noticeably improve the model results.

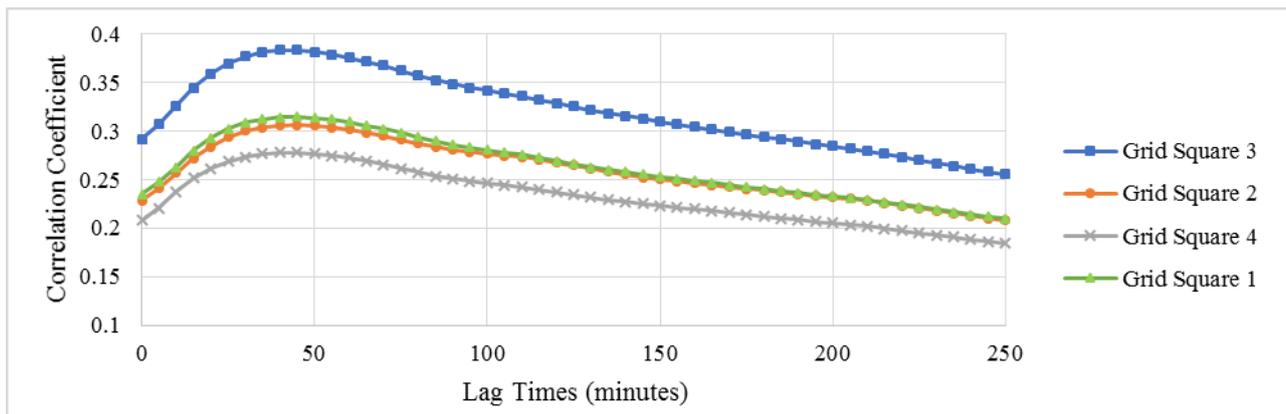


Figure 2. Cross Correlation between Historical Rainfall Intensity and CSO Level Data

3.3 ANN Structure and Training/Testing Data

No universally accepted general rule for calculating the optimum number of neurons in the hidden layer for an ANN exists, though the common recommendation is $N_{\text{HiddenNeurons}} = 2/3(N_{\text{Inputs}} + N_{\text{Outputs}})$. In this study empirical evidence showed that 8 neurons produced the best results, which is generally in line with that rule of thumb. The number of level inputs was set as 5 and the number of rainfall inputs varied depending on the forecast lead time being predicted. This is explained further in the results and discussion section below. 70% of the data was used for training the ANN, 15% for validation – ensuring that the network is generalising and to stop training before overfitting, and 15% for testing – an independent assessment of network generalisation. It was ensured that datasets contained a representative set of both dry and wet weather periods.

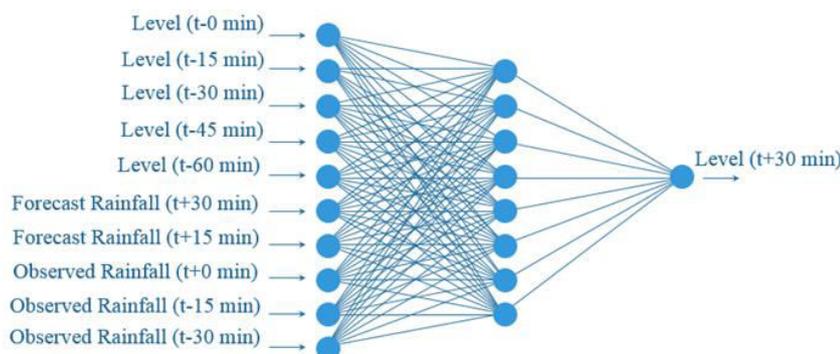


Figure 3. ANN Architecture Predicting 30 Minutes Ahead With Actual Forecast Data

As an example of a resulting ANN model, the architecture for ANN 3, modelling with actual forecast data, when predicting 30 minutes ahead is shown in Figure 3. This model contains 5 rainfall inputs; a combination of forecast rainfall data (15-30 minutes future forecasts) and past observed radar rainfall data (0 – 30 minutes past).

3.4 Results and Discussion

The three ANN-based models, as described in the above methodology, were built and applied to the case study. Predictions for future chamber depth were attempted from 15 minutes to 6 hours ahead. ANN performance was assessed by comparing the commonly used statistical indicators, mean square error (MSE) and the absolute fraction of variance (R^2) with and without forecast rainfall values.

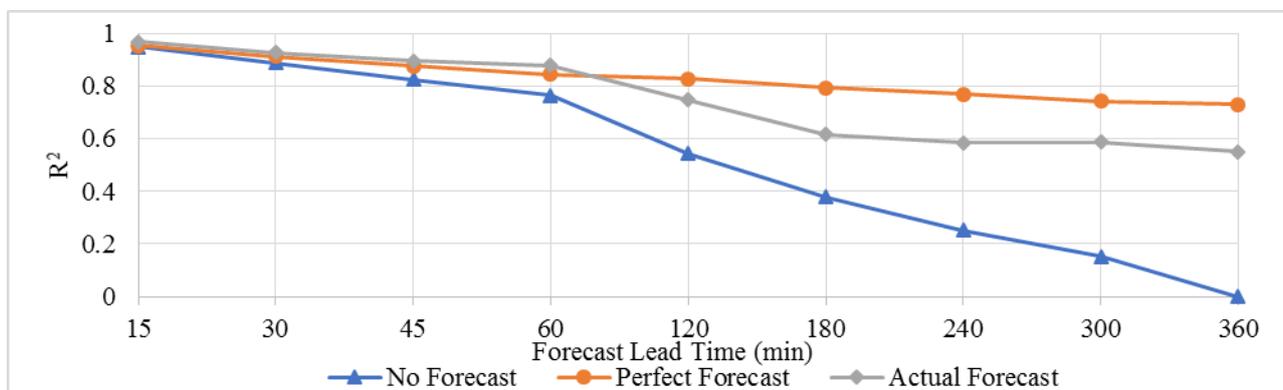
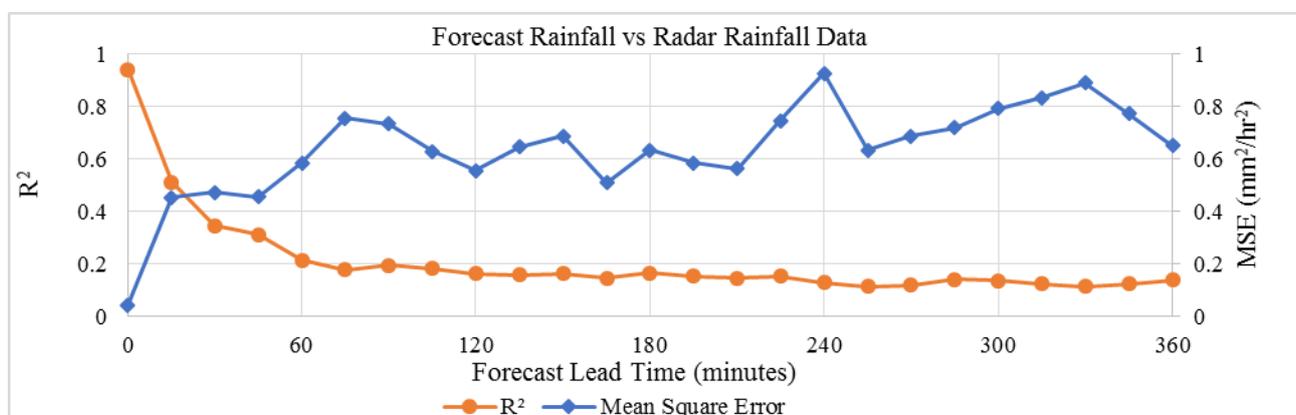


Figure 4. R^2 Value for CSO Prediction Forecast

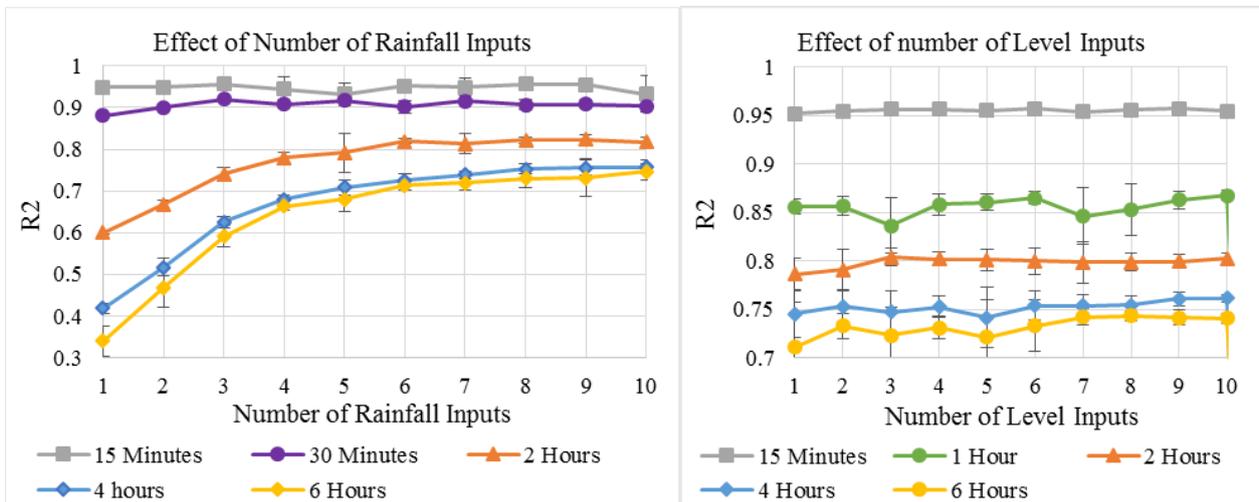
The R^2 values of the ANN models, using the optimal number of rainfall and level inputs for modelling different forecast times (as detailed below), are presented in Figure 4. The results indicate that, generally speaking, all three ANN models were able to effectively capture the relationship between rainfall and water level at the analysed CSO when predicting up to 45 minutes ahead, with R^2 values all greater than 0.85. As expected the models' accuracy decreased as the prediction horizon increased. However, including both perfect and actual forecast data enabled the



relevant ANN model to predict the CSO level values farther into the future with greater accuracy. Predicting 2 hours ahead the two forecast models were able to generate reasonable results with $R^2 > 0.75$, whilst the model using no forecast data was only able to produce an R^2 of 0.54.

Figure 5. R^2 and MSE of Forecast and Radar Rainfall Data for Different Forecast Lead Times

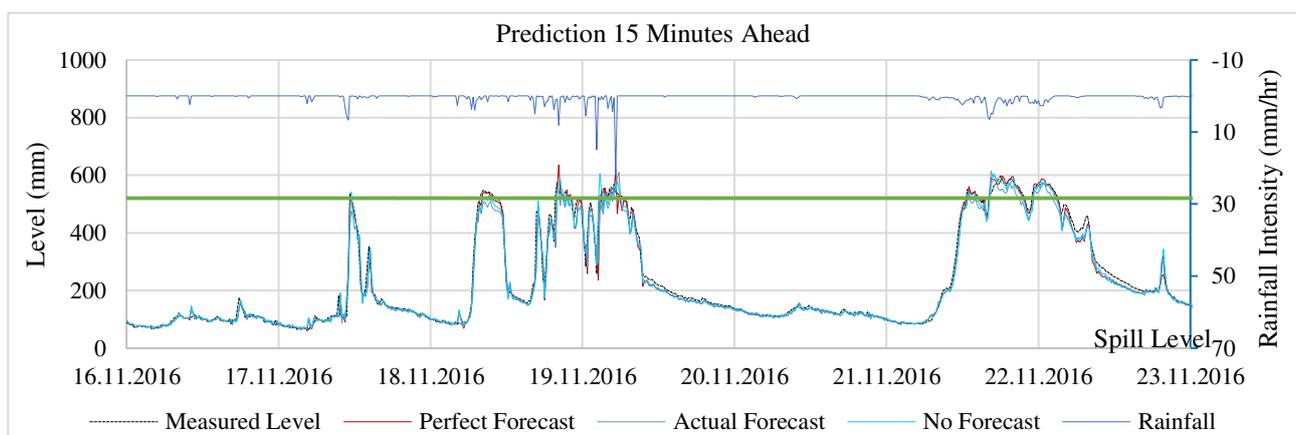
Figure 4 also demonstrates that ANN models using perfect forecast rainfall data produce better results than actual rainfall forecasts for lead times greater than 1 hour. The accuracy of forecast rainfall data was compared with observed radar rainfall data (see Figure 5). The data is in good agreement at low lead times, however, accuracy decreases rapidly for forecasts between 0 and 75 minutes, levelling off and decreasing slowly between 75 to 360 minutes. These errors therefore affect the accuracy of the ANN model predictions at larger forecast times. It should be noted, however, that even at high forecast intervals, where the R^2 value between forecast and observed radar data is very low, the ANN model with actual forecast data is still able to perform better than



the model without forecasts.

Figure 6. R^2 of Measured vs Predicted CSO Level using Perfect Forecast Rainfall Data for Different Numbers of Level and Rainfall Inputs

In order to optimise model performance, the ANNs were evaluated for different numbers of level and rainfall inputs. Each ANN was run 10 times and the average R^2 value used to assess the model, with the standard deviation shown as error bars (see Figure 6). The number of previous level inputs had little significant effect on the accuracy of the results, as did the number of rainfall inputs at low forecast times. However, when forecasting 2 or more hours ahead, including a larger number of rainfall inputs considerably improved the network performance.



For example in the case of the perfect forecast ANN model the R^2 value forecasting 6 hours ahead increased from 0.34 (1 rainfall input) to 0.75 (10 rainfall inputs). This may be because, as the model forecasts further into the future, the current and past level values have less of an effect on the future CSO level, and so forecast rainfall values become more important.

Figure 7. Modelled Predictions vs Measured Level 15 Minutes Ahead for Selected Time Period

Figure 7 shows CSO level predictions 15 minutes ahead for a time period containing three sets of spill events corresponding to periods of increased rainfall. All three ANN model predictions demonstrate very good agreement with the measured level data (dotted line), accurately predicting both the timings and magnitude of the spills.

Figure 8 shows CSO level predictions 2 hours into the future. The prediction accuracy has noticeably deteriorated, evident also in the decreased R^2 value. The ANN model without forecast data is completely unable to predict any of the overflow events and significantly under forecasts the increases in level due to rainfall. The inclusion of perfect and actual forecast rainfall enables the relevant ANN model to capture the relationship between CSO level and rainfall with greater accuracy, increasing the R^2 value from 0.54 (no forecast) to 0.74 with actual forecast data and 0.83 with perfect forecast data. The actual forecast model is able to generate a warning for the last 2 spills while the perfect forecast model predicts overflows for all 3 sets of spills.

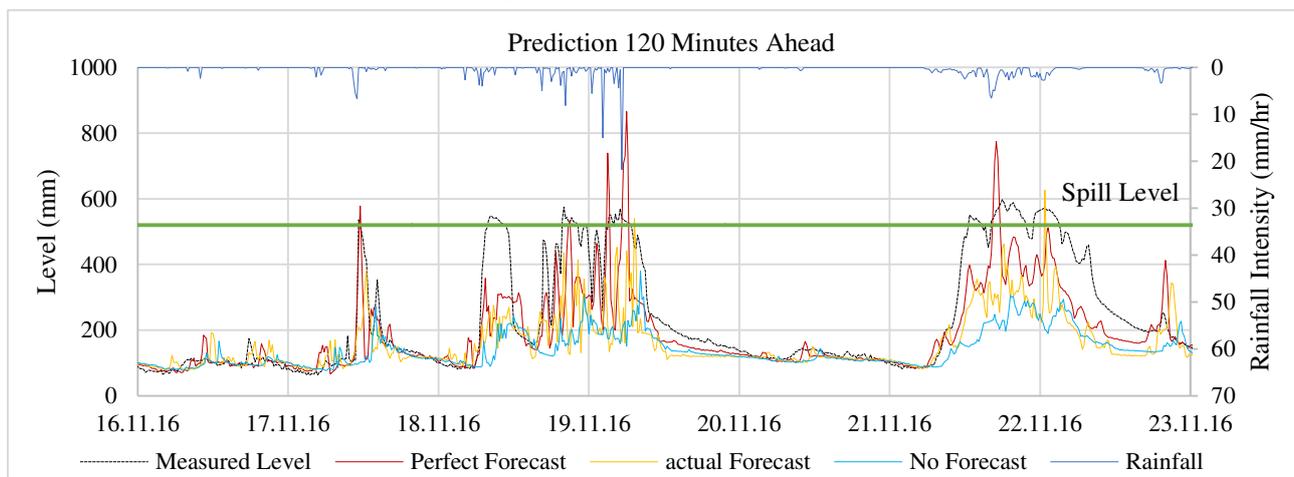


Figure 8. Modelled Predictions vs Measured Level 120 Minutes Ahead for Selected Time Period

4. CONCLUSION

It has been demonstrated in this paper that an ANN based model with a sliding time-window of inputs is able to successfully model the relationship between rainfall and water level within a CSO chamber up to 2 hours ahead for this case study. Additionally, the use of forecast rainfall, obtained from Met Office nowcasts is shown to improve the accuracy of the ANN predictions and increases its prediction range. It is envisioned that this model has the potential to be used by water utilities to model CSO depth in near real-time and generate alerts for upcoming spills events. This would enable appropriate and timely action to be taken to mitigate the effect of CSO discharges.

Acknowledgments

The authors are grateful for the access to data and funding provided by United Utilities. The work was also partially funded by the U.K. Engineering and Physical Sciences Research Council (EPSRC).

References

- [1] D. Butler and J. Davies, *Urban Drainage*, 2nd Editio. Spon Press, 2010.
- [2] V. M. Morales, J. M. Mier, and M. H. Garcia, “Innovative modeling framework for combined sewer overflows prediction,” *Urban Water J.*, vol. 14, no. 1, pp. 97–111, 2017.
- [3] M. Leon, A. Oberg, N. Schmidt, and A. Garcia, “The Illinois Transient Model: A State-of-the-Art Model for Simulating the Flow Dynamics in Combined Storm Sewer Systems,” in *Cognitive modeling urban water systems*, 2011.
- [4] M. Politano, A. J. Odgaard, and W. Klecan, “Case study: Numerical evaluation of hydraulic transients in a combined sewer overflow tunnel system,” *J. Hydraul. Eng.*, vol. 133, no. 10, pp. 1103–1110, 2007.
- [5] S. Thorndahl and P. Willems, “Probabilistic modelling of overflow, surcharge and flooding in urban drainage using the first-order reliability method and parameterization of local rain series,” *Water Res.*, vol. 42, no. 1–2, pp. 455–466, 2008.
- [6] C. Fortier and A. Mailhot, “Climate change impact on combined sewer overflows,” *Water Resour. Plann. Manag. re*, vol. 141, no. 5, 2015.
- [7] A. Zimmer, B. Minsker, A. Schmidt, and A. Ostfeld, “Evolutionary algorithm memory enhancement for real-time CSO control Challenges of Change,” *World Environ. Water Resour. Congr. 2010 Provid. Rhode Island, United States*, ., pp. 2251–2259, 2010.
- [8] Wallington Software Ltd., “Infoworks-CS.” Wallingford Software Ltd., Wallingford, UK:, 2006.
- [9] L. Rossman, “Storm water management model, User’s Manual,” Cincinnati, OH, 2010.
- [10] DHI Software, “Mouse Pipe Flow-Reference Manual,” 2003.
- [11] R. J. Abrahart, L. M. See, and D. P. Solomatine, *Practical Hydroinformatics*. 2008.
- [12] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York: Oxford University Press, 1995.
- [13] A. K. Fernando, X. Zhang, and P. F. Kinley, “Combined Sewer Overflow forecasting with Feed-forward Back-propagation Artificial Neural Network,” *Trans. Eng. Comput. Technol.*, pp. 58–64, 2006.
- [14] D. Sumer, J. Gonzalez, and K. Lansley, “Real-Time Detection of Sanitary Sewer Overflows Using Neural Networks and Time Series Analysis,” *J. Environ. Eng.*, vol. 133, no. 4, pp. 353–363, 2007.
- [15] A. Kurth, A. Saul, S. R. Mounce, W. Shepherd, and D. Hanson, “Application of Artificial Neural Networks (ANNs) for the prediction of CSO discharges,” *11th Int. Conf. Urban Drainage, Edinburgh, Scotland, UK, 2008*, pp. 1–10, 2008.
- [16] S. R. Mounce, W. Shepherd, G. Sailor, J. Shucksmith, and A. J. Saul, “Predicting combined sewer overflows chamber depth using artificial neural networks with rainfall radar data.,” *Water Sci. Technol.*, vol. 69, no. 6, pp. 1326–33, 2014.
- [17] S. R. Mounce, W. Shepherd, G. Sailor, A. J. Saul, and J. Boxall, “Application of Artificial Neural Networks to Assess CSO Performance,” no. September, pp. 7–12, 2014.
- [18] S. Liguori, M. Rico-Ramirez, A. Schellart, and A. Saul, “Using probabilistic radar rainfall nowcasts and NWP forecasts for flow prediction in urban catchments,” *Rainfall Urban Context Forecast. Risk Clim. Chang.*, vol. 103, pp. 80–95, 2012.
- [19] S. Krämer, L. Fuchs, and H. Verworn, “Aspects of radar rainfall forecasts and their effectiveness for real time control-the example of the sewer system of the City of Vienna,” *Water Pract. Technol.*, vol. 2, pp. 42–49, 2007.
- [20] M. Werner and M. Cranston, “Understanding the value of radar rainfall nowcasts in flood forecasting and warning in flashy catchments,” *Meteorol. Appl.*, vol. 16, pp. 41–55, 2009.
- [21] E. R. Vivoni, D. Entekhabi, R. L. Bras, V. Y. Ivanov, and M. P. Van Horne, “Extending the Predictability of Hydrometeorological Flood Events Using Radar Rainfall Nowcasting,” *J. Hydrometeorol.*, vol. 7, pp. 660–677, 2006.
- [22] F. Silvestro and N. Rebora, “Impact of precipitation forecast uncertainties and initial soil moisture conditions on a probabilistic flood forecasting chain,” *J. Hydrol.*, vol. 519, pp. 1052 – 1067, 2014.
- [23] C. Karul, S. Soyupak, A. F. Çilesiz, N. Akbay, and E. Germen, “Case studies on the use of neural networks in eutrophication modeling,” *Ecol. Modell.*, vol. 134, no. 2, pp. 145–152, 2000.
- [24] R. J. Frank, N. Davey, and S. P. Hunt, “Time Series Prediction and Neural Networks,” *J. Intell. Robot. Syst.*, vol. 31, no. 2, pp. 91–103, 2001.

