



EPiC Series in Engineering

Volume 3, 2018, Pages 1787–1795

HIC 2018. 13th International
Conference on Hydroinformatics



Prediction of CSO Chamber Level Using Evolutionary Artificial Neural Networks

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

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

²United Utilities Plc, Lingley Mere Business Park, Warrington, UK.
tal.ia.rosin@uuplc.co.uk

Abstract

Combined Sewer Overflows (CSOs) are a major source of pollution, spilling untreated wastewater directly into water bodies and/or the environment. If spills can be predicted in advance then interventions are available for mitigation. This paper presents Evolutionary Artificial Neural Network (EANN) models designed to predict water level in a CSO chamber up to 6 hours ahead using inputs of past CSO level, radar rainfall and rainfall forecast data. An evolutionary strategy algorithm is used to automatically select the optimal ANN input structure and parameters, allowing the ANN models to be constructed specifically for different CSO locations and forecast horizons. The methodology has been tested on a real world case study CSO and the EANN models were found to be superior to ANN models constructed using the trial and error method. This methodology can be easily applied to any CSO in a sewer network without substantial human input. It is envisioned that the EANN models could be beneficially used by water utilities for near real-time modelling of the water level in multiple CSOs and the generation of alerts for upcoming spills events.

1 Introduction

Combined Sewer Overflows (CSOs) are a necessary part of sewage infrastructure, designed to prevent overloading during heavy rainfall events which could otherwise lead to flooding of properties and sewage treatment works. Excess flow is discharged directly to a nearby waterbody. However, CSOs contain untreated human, agricultural, and industrial waste and are therefore a significant pollution concern, potentially harming the surrounding environment, degrading water quality, threatening public health and causing regulatory failures. CSOs are designed to operate only during periods of intense rainfall, when pollutants are dilute. However, events such as blockages and siltation can cause CSOs to spill at pass forward rates lower than consented, increasing environmental impacts. Water utilities in the UK are under growing pressure from the regulators to reduce the number of unconsented spills

which occur in their network. Consequently there has been increasing interest in developing methods to accurately predict CSO performance in near real-time and generate early warnings for spills. By predicting overflows in advance preventative measures can be implemented, such as maximizing storage and adjusting set points for movable gates and pumping stations, which mitigate their effects [1].

Traditionally, utilities have constructed physically based numerical models of the sewer system. However, these can be difficult to build, calibrate and maintain and are often computationally expensive. Data-driven models provide an exciting alternative; the Environment Agency announced in 2014 that they will require monitoring at the majority of CSOs by 2020 and, as a result, water utilities have begun to install large quantities of increasingly accurate level sensors in the sewer systems, routinely monitoring their networks in near real-time and collecting large volumes of sewer level data. This detailed data provides opportunities for in-depth analysis and data-driven modelling. One of the most popular data-driven techniques is Artificial Neural Networks (ANNs), biologically inspired computational models inspired by neural systems in the human brain. ANNs do not require a physical understanding of the processes involved in a system and are proficient at learning, memorising and constructing relationships within data. Several authors have demonstrated the potential of ANNs to successfully predict CSO performance. Fernando et al. [2] developed one of the earliest CSO level models, a feed-forward, back-propagation ANN designed to forecast spills in a CSO using flow rate and rainfall data. Kurth et al. [3] presented a feed-forward multilayer perceptron to predict CSO level 15 minutes ahead using level and rain gauge rainfall data. Sumer et al. [4] monitored sanitary sewer overflows using ANNs in near real-time. By identifying deviations between the measured and expected depth and flow data potential disruptions in the system were detected. This ANN was further developed by Mounce et al. [5] who used rainfall intensity data to predict CSO depth 75 minutes ahead. The model was then utilised by Mounce et al. [6], investigating the feasibility of an online system to reduce spills using fuzzy logic. Rosin et al [7] demonstrated that forecast rainfall data measurably improves the ANN prediction accuracy and prediction range. Forecasting 2 hours ahead the model accurately predicted CSO levels, providing alerts for upcoming spill events.

When utilising ANNs it is necessary to initially establish several parameters, such as the number of hidden neurons, the selection of input features and the structure of the network, to ensure good performance. This is a time consuming empirical process, often involving heuristic rules or searches of the parameter space, and, additionally, requires a human expert. One solution is to use Evolutionary Artificial Neural Networks (EANNs), whereby good ANNs are designed through the use of an Evolutionary Algorithm (EA) optimisation routine. An EANN model is able to automatically search for the best network architecture, parameters and input structure. Yao [8,9] produced two comprehensive reviews, examining the different combinations between ANNs and EAs, exploring the use of EAs when evolving connection weights, architectures, learning rules and input features, and reviewing the different search operators used in various EAs.

There has been increasing interest in applying EAs and EANNs to hydraulic systems in recent years. Chang [10] demonstrated that an EANN can forecast 10 day reservoir flow, with an accuracy greater than autoregressive models. Moradi & Dariane [11] proposed an EANN to derive operating policies for complex reservoir operations. A number of EAs have been developed for use in sewer networks, largely to help identify optimum design and rehabilitation strategies of sewer systems to reduce spill events [12,13] and for operational control of networks to minimise overflows [14,15]. EANNs have not yet, however, been applied to CSO level forecasting.

The main objective of the work presented here is to develop data-driven CSO level prediction models by employing an EANN approach to automatically select the optimal ANN input structure and

parameters. In this study, the resulting EANN models are compared to corresponding ANN models developed manually, i.e. as in Rosin et al. [7]. The rationale for this is to assess if the use of an EA optimisation routine enables the development of an ANN model that, while being automatically tailored to a specific CSO site and forecast horizon, also shows improved predictive performance. The EANN methodology used here is based on the work by Romano & Kapelan [16] who developed an adaptive water demand forecasting model to predict water demand up to 24 hours ahead.

The paper is structured as follows: after this introduction the methodology is presented and the selection of input structure and parameters process for the EANN and trial and error ANN models is described. In the Case Study section the various ANN/EANN models are applied to forecasting level in a CSO chamber and the EANN models' results analysed and compared with the trial and error ANN models' results to assess performance. Finally, the conclusions are drawn.

2 Methodology

In this study feed forward ANNs with a hyperbolic tangent transfer function for the neuron in the single hidden layer and a linear transfer function for the neuron in the output layer are trained using the back propagation method. This setup was identified by Romano & Kapelan [16] as producing accurate results with fast training times. The resulting ANNs are then used to produce level forecasts from 15 minutes to 6 hours ahead (24 forecast horizons at 15 minute intervals). The CSO levels for the next 6 hours are obtained by running the 24 EANNs in parallel. Inputs to the ANNs consist of antecedent level, antecedent rainfall and forecast rainfall data. The day of the week and the time of day associated with the forecast horizon, converted into a field type form (i.e. ones and zeros), are also used as potential inputs (determined during the input structure and parameters selection process). 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 employed here, whereby, past or future (in the case of forecast rainfall data) time series data is incorporated into the ANN using windows of antecedent lagged level, rainfall and forecast rainfall inputs. The number of past/future timesteps (i.e., window size) is selected during the input structure and parameters selection process. The ANN input structure and parameters to be selected in the aforementioned selection process are presented in Table 1 together with their relevant range of values. These are selected either by trial and error or by an EA-based optimisation process. The selection of input structure and parameters to achieve good performance is an important process and can have a significant impact on the model predictions. As different CSOs may exhibit very different behaviours and responses to rainfall events using pre-defined ANN input architectures and parameters may result in poor predictions.

When determining the window size of level and rainfall (both antecedent and forecast) inputs, for example, the selection of too few inputs can result in a model unable to capture the necessary dynamics of the system, however, too many inputs prolongs the learning time and may cause unnecessary information to appear as noise. Regarding the ANN architecture, a suitable number of hidden neurons have to be selected. A network with too few hidden neurons can have poor accuracy, while an excessive number may decrease the generalisation ability of the model due to overfitting. Finally, suitable numbers for the training cycles and the coefficient of weight decay regularisation have to be selected as well. These parameters are designed to prevent overfitting, by controlling the number of training cycles and applying a penalisation coefficient to the weights of an ANN model.

Selection of the trial and error ANN input structure and parameters is accomplished here by constructing models with different configurations of parameters and input structure and computing the

results for the different forecast horizons. The ANN with the best generalisation capacity is selected by identifying, for each parameter, which value produces the best overall performance for all 24 forecast horizons. Overall, 15 different ANN model configurations were constructed, examining the effect of varying different parameters and input structures (results not shown here), with each model configuration trained and tested on the 24 forecast horizons (15 minutes to 6 hours ahead). This was a very time consuming process that, additionally, considered only a small number of possible model configurations.

Decision variables	Range of values used in the input structure and parameters selection process
Level data lag size (number of time steps)	2 – 72
Radar rainfall data lag size (number of time steps)	2 – 72
Forecast rainfall data lag size (number of time steps)	1 - 24
Time of day	Use / Do not use
Day of week	Use / Do not use
Number of hidden neurons	10 - 100
Number of training cycles	50 - 500
Coefficient of weight decay regularisation (α)	10-5 – 103

Table 1: Decision variables and associated ranges of variability

The EA-based selection process, on the other hand, uses an evolutionary strategy algorithm [17] to automatically select the optimum ANN input structure and parameter set for the specific training data and forecast horizon considered. The network parameters are initially randomly selected. For each cycle of the evolutionary strategy algorithm the ANN model prediction error on the testing dataset is computed using the Nash-Sutcliffe model efficiency coefficient [18] and used in the objective function of the evolutionary strategy. This process is repeated until a pre-defined termination criterion, Nf.f.e., is reached (defined as the maximum number of generations in the case study presented in this paper). This results in the selection of the combination of parameters which produce the lowest error during testing [9]. Thus, 24 separate EANN models, optimised for a specific forecast horizon and CSO location, are created. The parameters of the evolutionary strategy algorithm, and the values selected and used in the case study are presented in Table 2. As noted by Romano & Kapelan [16] most EAs are quite insensitive to the exact parameter settings and therefore extensive sensitivity analysis is not required to determine satisfactory values.

Parameter	Value
Number of Parents per generation - μ	10
Number of offspring per generation – λ	20
Number of fitness function evaluations – Nf.f.e.	210
Probability of a parameter being perturbed – Pmut.	0.75
Mutation strength – σ	0.6
Selection operator	+

Table 2: Values of the evolutionary strategy algorithm's parameters

3 Case Study

The trial and error ANN and EANN models were applied to 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. This CSO exhibits a relatively stable diurnal pattern during normal conditions, with increasing 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 resolution of 2 minutes. Observed radar rainfall intensity data (mm/hr) and forecast rainfall intensity data (mm/hr) were obtained from the UK Met Office, observed data with a 5 minute temporal resolution and 1×1 km spatial resolution and forecast data with a 15 minute temporal resolution and 2×2 km spatial resolution. All time series were interpolated to a common, uniform resolution of 15 minutes during data pre-processing. Data was collected from April 2016 to February 2017 and divided into 60% training, 25 % testing and 15% validation datasets.

Separate ANN/EANN models were constructed and optimised using: 1) actual forecast rainfall data, 2) perfect forecast data (i.e. forecasts assuming perfect knowledge of historical rainfall into the future) and 3) no forecast rainfall data. When used in near real-time by a water utility actual forecast data will be used. However, Rosin et al. [7] demonstrated that the accuracy of forecast rainfall data significantly decreases at higher forecast horizons which, in turn, affects the accuracy of CSO level predictions. Therefore, in order to rigorously assess the performance of the EANN in comparison to the trial and error ANN, models using perfect and no forecast data were also constructed and assessed. Bearing this in mind, for each “forecast rainfall type”, 24 separate EANN models, with forecast horizons from 15 minutes to 6 hours ahead at 15 minute increments were constructed.

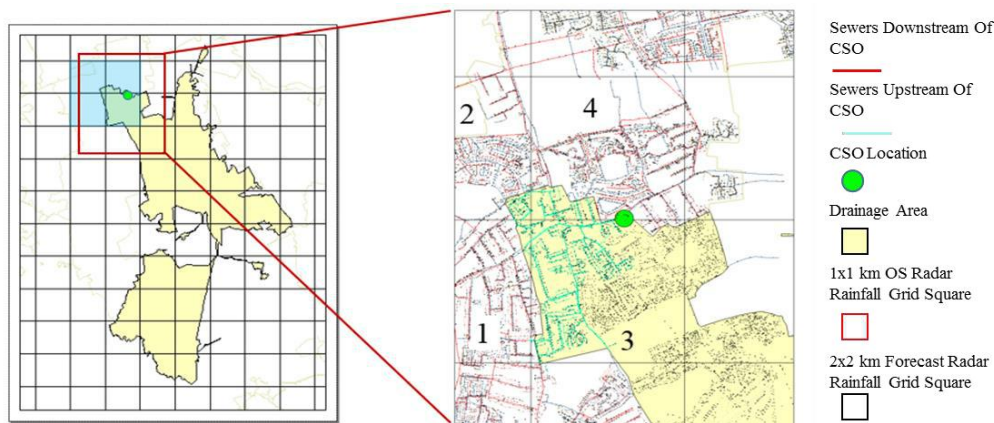


Figure 1: Case Study Location

4 Results and Discussion

Figure 2 presents the Nash-Sutcliffe indices, computed by comparing the measured level with forecast data over the validation dataset, for each case considered. As can be seen from this figure, for almost all forecast horizons, the EANN models outperform the trial and error ANNs, with improvements becoming more significant when predicting farther into the future. It is posited that this

is because, when predicting at short forecast horizons, future CSO level is highly dependent on past level data and so this parameter has the strongest influence on the model output. Thus, the trial and error ANNs, which are “optimised” generally for all the forecast horizons, enable achieving good performance. When predicting further into the future, however, past and future level are less directly correlated and other model inputs and parameters become more significant. Thus, selecting the optimal parameters and inputs via the EANN methodology, i.e. tailoring the ANN model specifically to a particular forecast horizon, results in a greater improvement to the model results.

An example of the forecasts obtained by the trial and error ANN and the EANN when predicting 15 minutes ahead using actual forecast rainfall data is presented in Figure 3a. There is excellent agreement between the measured and forecast data. Both models accurately forecast CSO level during dry weather and a rainfall event, predicting accurately the times and magnitudes of two overflows caused by precipitation. Figure 3b shows the CSO level forecast 4 hours ahead using actual forecast rainfall data and demonstrates the improvements attained using the EANN model; the trial and error ANN completely fails to capture any level changes, forecasting an almost flat line for all timesteps. Conversely the EANN is able to accurately forecast the diurnal pattern in CSO Level.

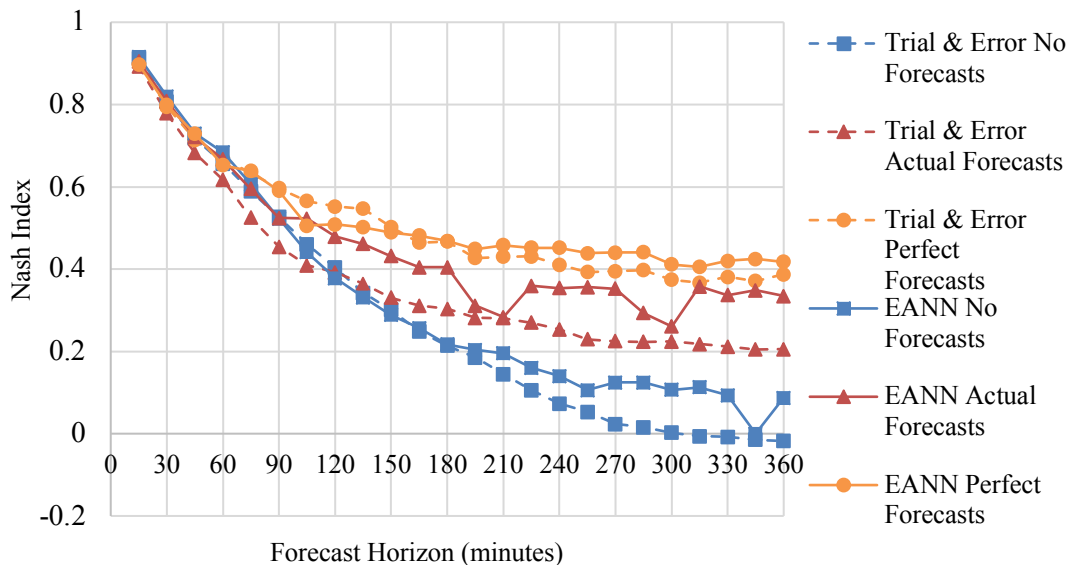


Figure 2: Validation dataset Nash-Sutcliffe Indices for each case considered

Major rainfall events are more difficult to be modelled precisely at higher forecast horizon values, as illustrated in Figure 4 which shows forecasts 4 hours ahead using actual forecast rainfall data during precipitation. The EANN and trial and error ANN model both forecast a rise in CSO level but are unable to predict the full extent of the increase. The EANN model quickly recovers from the rainfall event and accurately forecasts the subsequent dry weather period, however, the trial and error model remains high, predicting a second peak during dry weather. This pattern is found consistently after rainfall events throughout the dataset.

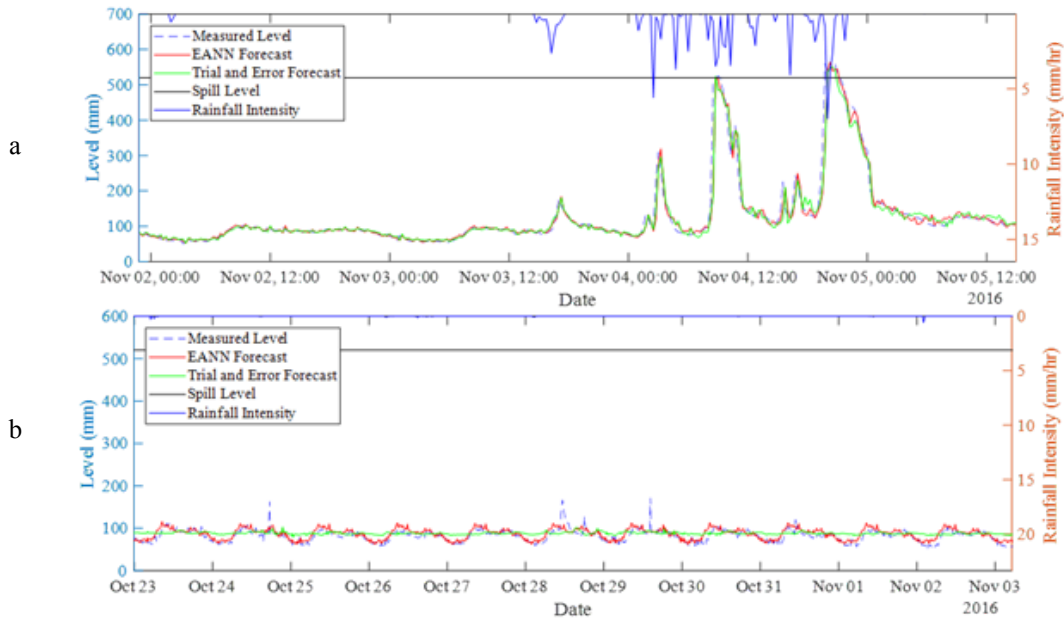


Figure 3: EANN vs trial & error ANN level forecast (a) 15 minutes ahead during wet weather and (b) 4 hours ahead during dry weather

The reduction in model performance at high forecast horizons during peak rainfall events is in part due to the inaccuracy of forecast rainfall data when predicting this far ahead. However, it is also caused by the imbalanced nature of the dataset; a significant majority of rainfall data is under 1 mm/hr and, as the ANN models are designed to reduce overall error and maximise overall accuracy, the models are optimised for dry weather and fail to properly capture the rarer peak rainfall events. Further research will be carried out to improve the models’ performance during these events, employing techniques to mitigate data imbalance such as undersampling and oversampling [19].

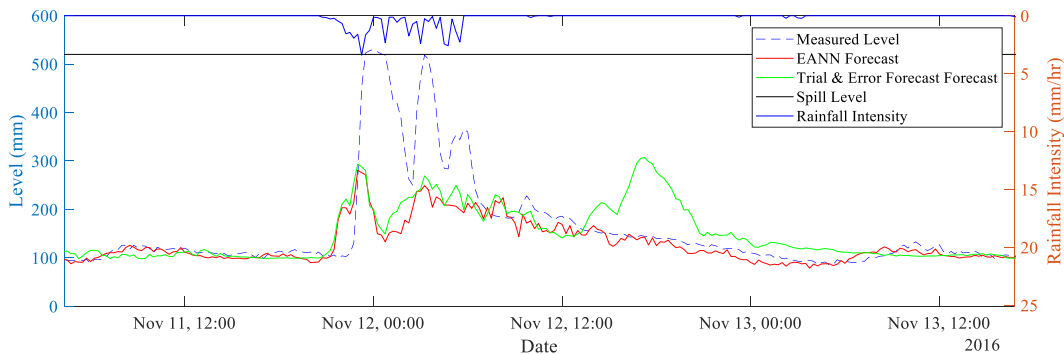


Figure 4: EANN vs trial & error ANN level forecast 4 hours ahead during wet weather

The parameters and input structure selected by the EA for different forecast horizons when using perfect forecasts, actual and no forecast rainfall data were also analysed to better understand the effect they have on the performance of the various models. As an example, table 3 shows the EANN selected values for the two forecast horizons presented above and the trial and error ANN for all forecast horizons when using actual forecast rainfall data. Detailed results of all ANN models are not shown due to space

limitations. The results obtained show that, as expected, when using forecast rainfall data the number of selected inputs increases when predicting further into the future. Generally, both the day of the week and time of day are used only at higher forecast horizons for all 3 models. The coefficient of weight decay regularisation is low for all forecast horizons. The other parameters do not exhibit a strong trend related to the forecast horizon or the forecast rainfall type. As the methodology is generic it has the potential to include additional model parameters in the future, for example the antecedent precipitation index or soil moisture index, which could potentially improve the results.

Decision variables	EANN <i>15 minutes ahead</i>	EANN <i>4 hours ahead</i>	Trial & Error ANN <i>All forecast horizons</i>
Level data lag size (number of time steps)	48	36	72
Radar rainfall data lag size (number of time steps)	12	12	72
Forecast rainfall data lag size (number of time steps)	1	22	24
Time of day	Use	Use	Use
Day of week	Do not use	Do not use	Do not use
Number of hidden neurons	10	80	100
Number of training cycles	350	350	500
Coefficient of weight decay regularisation (α)	0.0001	0	0.1

Table 2: Example of automatically selected ANN parameters/input structures using actual forecast rainfall

5 Conclusion

EANN models are proposed here to automatically select the optimal ANN input structures and parameters for forecasting water level in a CSO chamber. The methodology consists of a feed-forward single hidden layer ANN, with structure and parameters optimised using an evolutionary strategy algorithm. The methodology was tested on a real world case study CSO. Comparison with ANN models developed manually through trial and error indicates that the EANN models produce more accurate CSO level predictions whilst also requiring substantially less human time and effort. The EANN methodology is generic, and so can automatically be applied to different catchments and optimised for different forecast horizons. Thus it is envisioned that the EANN model can be used by water utilities to model CSO levels in the wastewater network in near real-time in order to proactively manage overflow events.

References

- [1] 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.
- [2] 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.
- [3] 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.

- [4] 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.
- [5] 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.
- [6] 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.
- [7] T. Rosin, M. Romano, K. Woodward, E. Keedwell, and Z. Kapelan, "Prediction of CSO chamber water levels using rainfall forecasts," in *Computing and Control for the Water Industry*, 2017.
- [8] X. Yao, "A review of Evolutionary Artificial Neural Networks," *Int. J. Intell. Syst.*, vol. 8, pp. 539–567, 1993.
- [9] X. Yao, "Evolving Artificial Neural Networks," *Proc. IEEE*, vol. 87, no. 9, pp. 1432–1447, 1999.
- [10] Y. Chen and F. Chang, "Evolutionary artificial neural networks for hydrological systems forecasting," *J. Hydrol.*, vol. 367, no. 1, pp. 1525–1540, 2009.
- [11] M. Moradi, A. and B. Dariane, A., "Evolving neural networks and fuzzy clustering for multireservoir operations," *Neural Comput. Appl.*, vol. 28, no. 5, pp. 1149–1162, 2017.
- [12] B. Ward and D. A. Savic, "A multi-objective optimisation model for sewer rehabilitation considering critical risk of failure," *Water Sci. Technol.*, vol. 66, no. 11, pp. 2410–2417, 2012.
- [13] B. Itaquy, O. Ogidan, and M. Giacomoni, "Application of a Multiobjective Genetic Algorithm to Reduce Wet Weather Sanitary Sewer Overflows and Surge," *J. Sustain. Water Built Environ.*, vol. 3, no. 3, 2017.
- [14] B. Joseph-duran, C. Ocampo-martinez, and G. Cembrano, "A Control-Oriented Hybrid Modelling Approach for Sewer Net- works : Barcelona Case Study," 11th IWA Conf. Instrum. Control Autom., 2013.
- [15] U. S. Rathnayake and T. T. Tanyimboh, "Evolutionary multiobjective optimal control of combined sewer overflows," *Water Resour. Manag.*, vol. 29, no. 8, pp. 2715–2731, 2015.
- [16] M. Romano and Z. Kapelan, "Adaptive water demand forecasting for near real time management of smart water distribution systems," *Environ. Model. Softw.*, vol. 60, pp. 265–276, 2014.
- [17] H. Schwefel and P., *Evolution and Optimum Seeking*. New York: Wiley, 1995.
- [18] R. H. McHuen, Z. Knight, and A. G. Cutter, "Evaluation of the Nash-Sutcliffe Efficiency Index," *J. Hydraul. Eng.*, vol. 11, no. 6, 2006.
- [19] H. He and E. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. adata Eng.*, vol. 21, no. 9, pp. 1263–1284, 2009.