Leveraging Generative Adversarial Networks (GANs) to Improve the Accuracy of Data-Driven Combined Sewer Flow Prediction Models

A. E. Bakhshipour, Ph.D.[1], A. Koochali[1,2,3*], M. Bakhshizadeh[1,2], U. Dittmer, Prof.[2], A. Dengel, Prof.[1,2], S. Ahmed, Ph.D.[2], A. Haghighi, Prof.[3]

[1]*TU Kaiserslautern, Kaiserslautern, Germany*
[2]*German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany*
[3]*Shahid Chamran University of Ahvaz, Ahvaz, Iran*
[4]*Ingenieurgesellschaft Auto und Verkehr(IAV) GmbH, Berlin,Germany*
**Corresponding author email: Alireza.Koochali@dfki.de*

Highlights

- We investigated the potential of GANs in the field of urban water management.
- As proof of concept, GAN is used to generate synthetic time series to improve the accuracy of data-driven combined sewer flow prediction models
- The time series generated by GAN look quite genuine even on the first try, and still, significant improvement can be achieved after reaching a more proper GAN architecture. Besides, further evaluations are needed to draw general conclusions.

Introduction

Data analysis and mining would facilitate smart operation and management of interconnected urban water systems with environmental challenges. Working with big data is especially relevant in a changing world where demands change faster than the lifespan of critical infrastructure elements. In recent years, we have witnessed numerous breakthroughs of machine learning techniques in a wide variety of domains, such as computer vision, language processing, and reinforcement learning (Eggimann et al. 2017). However, still, several key shortcomings hinder progress for environmental management. They include (1) lack of freely available data (e.g., due to data privacy or high expenses of data gathering) for an extended period as well as lack of rare or extreme events in the training data, (2) lack of robust methods for anomaly detection, particularly for drift detection, (3) absence of probabilistic time series forecasting data-driven methods to consider different sources of uncertainty for optimal and robust operation of critical urban water infrastructures. The Generative Adversarial Networks (GANs) might help to overcome the abovementioned challenges in water resources and environmental management problems. The reasons are represented in Figure 1, summarizing the capabilities of GANs. To the best of our knowledge, GANs application's potentials in water resources and environmental management are not yet discovered. In this study, we want to test one application of GANs to urban drainage systems. We use GANs to generate synthetic time series to balance our data and improve the accuracy of data-driven combined sewer flow prediction models as explained in the next section.

Methodology

Generative Adversarial Networks (GANs)

In machine learning, generative models are a class of methods that are capable of learning data distribution to generate artificial data. Recently GANs (Goodfellow et al. 2014) are introduced as neural network-based generative models which demonstrate outstanding performance in learning complex probability distribution and generating very realistic data samples. A GAN architecture consists of two neural networks namely generator and discriminator. During training, these two networks are engaged in a two-player minimax game using an adversarial loss function.

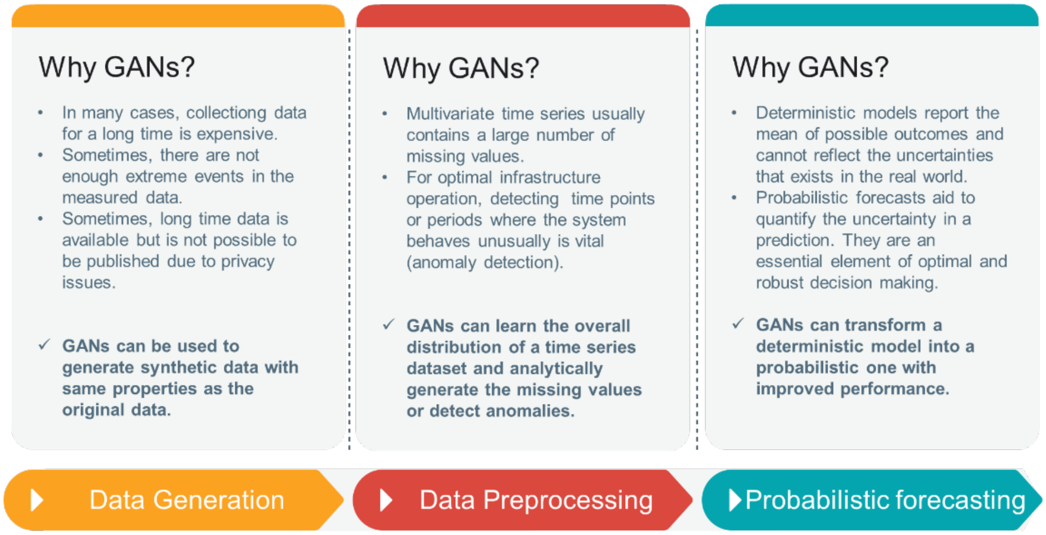

Figure 1. Potential applications of GANs in different stages of time series data modeling

So far, GANs are applied widely to domains in which their results are intuitively assessable, e.g., images and acquired prominent results. However, their application on other domains such as time-series is limited due to difficulties in the assessment of GANs in these domains. Nevertheless, there are successful applications of GANs in the time-series domain for various tasks such as generate time series sequences in the health care, finance, and energy industry. Moreover, they have been fruitfully utilized for anomaly detection and the probabilistic prediction of time series (Koochali et al. 2020).

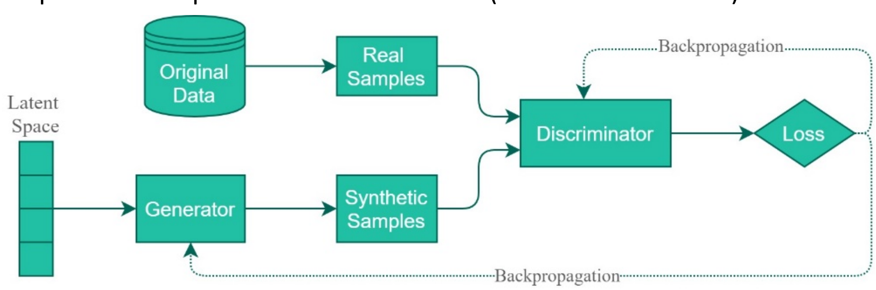

Figure 2. GAN architecture

Application to Combined Sewer Systems

Many approaches have been developed to improve combined sewer systems management and operation using real-time control (RTC). The remarkable amount of measured data coupled with RTC would aid in managing combined sewer overflow (CSO) based on the potential impact on the water bodies (Eggimann et al. 2017). As relevant events are relatively rare in the historical data, pure data-driven rainfall-runoff models often underestimate the runoff when predicting these events. The optimal operation of the sewer system during these events, which result in the most critical states for the urban area and environment, depends highly on accurate flow predictions. To overcome these challenges, we used GANs to generate synthetic time series from approximately the same statistical distribution of our data set. The generator model enables us to balance our training data with extreme synthetic events. To evaluate the performance of the proposed approach, we train a specific deep neural network model with and without synthetic data and test their performance using a similar test dataset.

Study Area and Data

In this study, the sewer system of a small town in Baden-Württemberg state in Germany is considered to evaluate the performance of the proposed model. Precipitation (mm), temperature (°C), and inflow to the storage tanks are the data set used for the model development. All data are measured from the beginning of July to the end of December 2017, at a 5-min time resolution (Ayazpour et al. 2018). Figure 3 summarizes our data set. It also depicts an example of our measured time series. The aim is to build a black-box simulator to predict combine sewer flow in the system using historical and synthetic data. The model then can be employed for optimal operation (e.g., to minimize volume and duration of CSO) of the system using RTC.

Month	Precipitation(mm)	Average of T(∞C)	Min of T(∞C)	Max of T(∞C)	Average of Q(l/s)	Max of Q(l/s)	Average of V(m/s)	Max of V(m/s)
July	22,90	20,36	15,50	23,60	14,98	890,00	0,32	1,37
December	17,10	10,19	5,40	13,30	11,28	403,00	0,38	1,00
September	14,20	19,03	0,00	21,40	18,46	855,00	0,32	1,39
November	11,50	12,19	6,40	16,80	15,04	295,00	0,34	0,91
October	9,30	16,46	11,90	18,50	9,51	781,00	0,33	1,05
August	9,20	20,40	16,80	23,20	8,95	617,00	0,33	1,09
January	8,00	10,38	6,90	12,40	17,08	565,00	0,35	1,05
June	1,70	19,87	18,20	21,00	10,02	78,80	0,31	0,78
Total	**93,90**	**15,78**	**0,00**	**23,60**	**12,92**	**890,00**	**0,34**	**1,39**

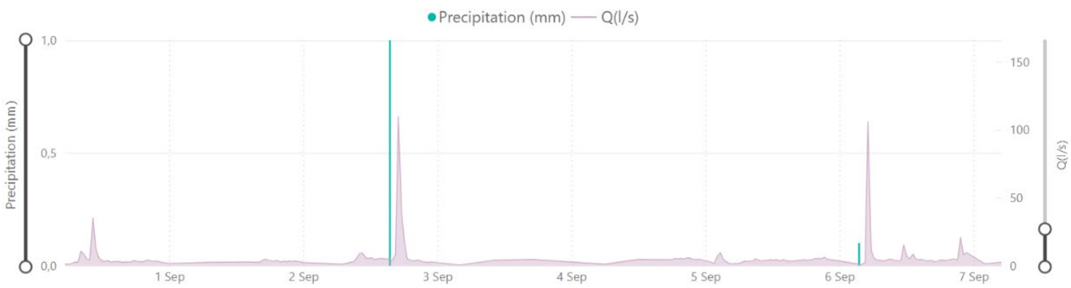

Figure 3. Case Study and data set summary

Results and future works

Figure 4 shows a part of the generated data using a GAN model. We are still working on various GANs architectures to get more realistic results. The next step would be to compare the accuracy of different data-driven sewer flow prediction models with (the developed model in Ayazpour et al. 2018) and without synthetically generated data to evaluate the influence of leveraging GANs on the performance of such models. These two models will be also compared by developing an optimization framework to minimize the volume and duration of CSOs in the system using RTC. We also plan to employ GANs toward moving from deterministic data-driven models to probabilistic models leading to the capability of optimizing the operation of CSO tanks under uncertainties.

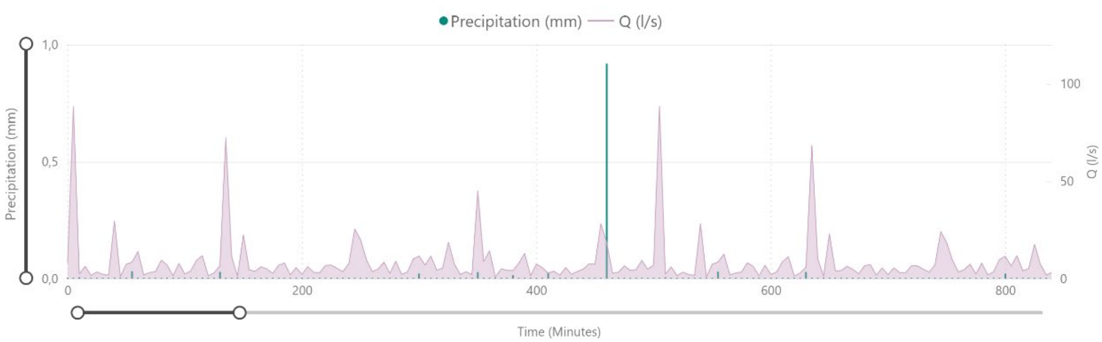

Figure 4. Examples of synthetic events generated by the developed time series GAN

References

Eggimann, S.; Mutzner, L.; Wani, O.; Schneider, M.Y.; Spuhler, D.; Moy de Vitry, M.; Beutler, P.; Maurer, M. The Potential of Knowing More: A Review of Data-Driven Urban Water Management. Environ. Sci. Technol. 2017, 51, 2538–2553.

Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks, 2014.

Koochali, A.; Dengel, A.; Ahmed, S. If You Like It, GAN It—Probabilistic Multivariate Times Series Forecast with GAN. Eng. Proc. 2021, 5, 40. https://doi.org/10.3390/engproc2021005040

Ayazpour Z., Bakhshipour A.E., Dittmer U. (2019) Combined Sewer Flow Prediction Using Hybrid Wavelet Artificial Neural Network Model. In: Mannina G. (eds) New Trends in Urban Drainage Modelling. UDM 2018.