EPyT-Flow: A Toolkit for Generating Water Distribution Network Data

André Artelt [1,3*¶], **Marios S. Kyriakou** [1*], **Stelios G. Vrachimis** [1,2*], **Demetrios G. Eliades** [1*], **Barbara Hammer** [3*], and **Marios M. Polycarpou** [1,2*]

1 KIOS Research and Innovation Center of Excellence, University of Cyprus, Cyprus **2** Department of Electrical and Computer Engineering, University of Cyprus, Cyprus **3** Faculty of Technology, Bielefeld University, Germany ¶ Corresponding author * These authors contributed equally.

Summary

This work introduces `EPyT-Flow`, an open-source Python package building on top of EPyT for facilitating water distribution network (WDN) simulations. `EPyT-Flow` provides a high-level interface for the easy generation of hydraulic and water quality scenario data. Additionally, it provides access to low-level functions of EPANET and EPANET-MSX for hydraulic and water quality modeling respectively. To accelerate research in WDN management, `EPyT-Flow` provides easy access to popular benchmark data sets for event detection and localization, and an environment for developing and testing feedback control algorithms.

Statement of need

Water Distribution Networks (WDNs) are designed to ensure a reliable supply of drinking water. These systems are operated and monitored by humans, supported by software tools, including basic control algorithms and event detectors that rely on a limited number of sensors within the WDN. These sensors measure hydraulic (e.g., pressure, flow) and water quality (e.g., chemical concentrations) states. However, given the rapid population growth of urban areas, WDNs are becoming more complex to manage due to the resulting time-varying system uncertainty. Consequently, key tasks such as event detection (e.g., leakage) and isolation, pump scheduling, and control are becoming more challenging. Moreover, modeling and predicting water quality in the distribution network is becoming more difficult due to changing environmental conditions. This is why water utilities are now driven to install even more sensors to gather data on their changing systems. Traditionally, model-based methods were used for planning and managing WDNs; however, due to rapid changes, these methods may no longer be sufficient. New AI and data-driven methods can now take advantage of big data and are promising tools for tackling challenges in water management.

Currently, non-water experts such as AI researchers face several challenges when devising practical solutions for water system applications, such as the unavailability of tools for easy scenario/data generation and easy access to benchmarks, which hinder the progress of applying AI to this domain. Easy-to-use toolboxes and access to benchmark data sets are extremely important for boosting and accelerating research, as well as for supporting reproducible research. This was, for instance, the case in deep learning and machine learning with toolboxes such as TensorFlow and scikit-learn.

State of the field

The modeling and simulation of hydraulic and water quality dynamics in water distribution networks have progressively advanced with the introduction of simulation software. Notably, EPANET ([Rossman, 2000](#)) and its extension EPANET-MSX ([Shang et al., 2008](#)) are foundational tools in this area. These are complemented by tools that make use of high-level programming languages, such as the EPANET-MATLAB Toolkit (EMT) ([Eliades et al., 2016](#)), the Object-Oriented Pipe Network Analyzer (OOPNET) ([Steffelbauer & Fuchs-Hanusch, 2015](#)), and the EPANET-Python Toolkit (EPyT) ([Kyriakou et al., 2023](#)). Moreover, viswaternet provides visualizations of static and time-varying attributes of EPANET-based WDNs ([Thomas et al., 2023](#)). These tools are instrumental in facilitating research in WDN resilience and response to various operational challenges.

These tools, however, do not support the creation of realistic (benchmark) scenarios with essential aspects such as realistic fault models (of leakages and sensor faults), various sensor configurations, custom control modules, and other events such as changes in water quality caused by external factors. However, such data creation mechanisms are important for supporting and enabling research on the application of Machine Learning (ML) and Artifical Intelligence (AI) in WDNs. A first step towards such software for scenario creation is the Water Network Tool for Resilience (WNTR) ([Klise et al., 2017](#)), which facilitates the performing hydraulic simulations under various scenarios such as pipe breaks, disasters such as earthquakes, power outages, and fires. However, it currently does not fully integrate (advanced) quality dynamics with scenarios such as pipe leaks and also misses other crucial modules such as sensor configurations, and considerations of industrial controls (i.e. controls that go beyond simple IF-THEN-ELSE controls as supported by EPANET).

The transition to Python-based open-source software ([Klise et al., 2017](#); [Kyriakou et al., 2023](#)), underscores a broader trend towards open-source, community-driven development in the water industry, aligning with the need for transparency, reproducibility, and innovation in the sector.

Functionality

Our Python toolbox EPyT-Flow provides a high-level interface for easy generation of WDN scenario data, but also provides access to low-level functions for maximum flexibility as needed by domain experts. By this we aim to satisfy different needs and make it a toolbox for "everybody". A special focus of EPyT-Flow is on data generation to enable and support research on the application of ML and AI in WDNs. In addition, its design and object-orientated implementation makes it easy to customize existing functionalities and implement new ones. EPyT-Flow builds upon EPyT which itself provides a Python interface to EPANET and EPANET-MSX – see Figure 1 for an illustration.
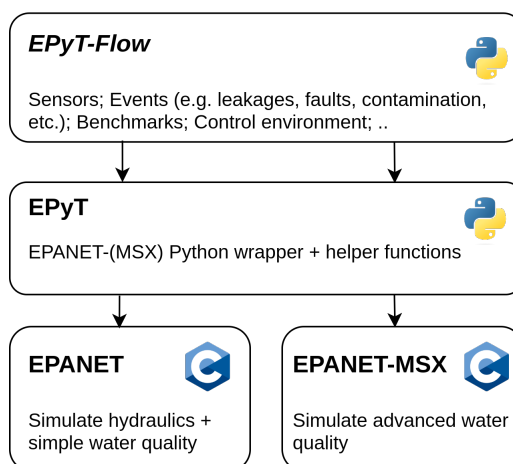
Figure 1: Illustration of the functionality of the proposed toolbox *EPyT-Flow*.

The toolbox currently includes 16 Water Distribution Networks (WDNs) that can be used for scenario generation. It goes beyond pure scenario generation by providing access to 7 popular and widely adopted benchmarks on event detection and localization (including their evaluation metrics) – ready to be utilized for building and evaluating algorithms. Furthermore, it provides an environment, inspired by the modular and extensible design of OpenAI Gym for developing and implementing (AI-based) algorithms for tasks such as energy efficient pump scheduling.

To support modeling of a wide variety of scenarios, the toolbox comes with 4 different event types and a total number of 13 pre-defined and implemented events ready to be utilized in custom scenarios: 3 different leakage types, 3 actuator events, 5 sensor fault types, and 2 communication events. All the events can be easily customized by the user.

Since the measured quantities in the real world are always subject to uncertainty, the toolbox comes with 11 pre-defined types of uncertainties ranging from classic Gaussian noise to different types of (very) deep uncertainties that can be applied to hydraulic parameters such as pipe length, diameter, and roughness, and water quality parameters, such as reaction coefficients, and sensor noise.

More information can be found in the comprehensive (200+ pages) EPyT-Flow Documentation and 14 fully working examples that demonstrate how EPyT-Flow can be utilized in different tasks.

Conclusions

In this work, we introduced a Python toolbox called EPyT-Flow for realistic scenario data generation and access to benchmarks of WDNs, that researchers can utilize to develop methods to support human WDN operators in various real-world challenges.

Our long-term vision for this toolbox is to split it into three parts to further facilitate the progress of research in this area:

1) A core part for data generation (i.e., scenario simulation).
2) A BenchmarkHub as a platform for accessing and sharing WDN benchmarks.
3) A ModelHub as a platform for accessing and sharing AI, classic models and algorithms for different tasks in WDNs.

Acknowledgments

References

Eliades, D. G., Kyriakou, M., Vrachimis, S., & Polycarpou, M. M. (2016). EPANET-MATLAB toolkit: An open-source software for interfacing EPANET with MATLAB. *Proc. 14th International Conference on Computing and Control for the Water Industry (CCWI)*, 8. https://doi.org/10.5281/zenodo.831493

Klise, K. A., Bynum, M., Moriarty, D., & Murray, R. (2017). A software framework for assessing the resilience of drinking water systems to disasters with an example earthquake case study. *Environmental Modelling & Software*, *95*, 420–431. https://doi.org/10.1016/j.envsoft.2017.06.022

Kyriakou, M. S., Demetriades, M., Vrachimis, S. G., Eliades, D. G., & Polycarpou, M. M. (2023). EPyT: An EPANET-Python toolkit for smart water network simulations. *Journal of Open Source Software*, *8*(92), 5947. https://doi.org/10.21105/joss.05947

Rossman, L. A. (2000). *EPANET 2: Users manual*. US Environmental Protection Agency. Office of Research; Development. National Risk Management Research Laboratory; US Environmental Protection Agency. Office of Research; Development. National Risk Management Research Laboratory. https://nepis.epa.gov/Adobe/PDF/P1007WWU.pdf

Shang, F., Uber, J. G., & Rossman, L. A. (2008). Modeling reaction and transport of multiple species in water distribution systems. *Environmental Science & Technology*, *42*(3), 808–814. https://doi.org/10.1021/es072011z

Steffelbauer, D., & Fuchs-Hanusch, D. (2015). OOPNET: An object-oriented EPANET in Python. *Procedia Engineering*, *119*(1), 710–718. https://doi.org/10.1016/J.PROENG.2015.08.924

Thomas, M., Trimble, T., & Sela, L. (2023). VisWaterNet: A Python package for visualization of water distribution networks. *Journal of Open Source Software*, *8*(84), 5139. https://doi.org/10.21105/JOSS.05139