Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Artelt, A., Hermes, L., Strotherm, J., Hammer, B., Vrachimis, S. G., Kyriakou, M. S., Eliades, D. G., Polycarpou, M. M., Paraskevopoulos, S., Vrochidis, S., Taormina, R., Savic, D., & Koundouri, P. (2026). First AI for Drinking Water Chlorination Challenge. *Journal of Water Resources Planning and Management*, *152*(12), Article 04026059. https://doi.org/10.1061/JWRMD5.WRENG-7562

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

First AI for Drinking Water Chlorination Challenge

André Artelt, Ph.D., M.ASCE[1]; Luca Hermes[2]; Janine Strotherm[3]; Barbara Hammer, Ph.D.[4]; Stelios G. Vrachimis, Ph.D., S.M.ASCE[5]; Marios S. Kyriakou[6]; Demetrios G. Eliades, Ph.D.[7]; Marios M. Polycarpou, Ph.D.[8]; Sotirios Paraskevopoulos, Ph.D.[9]; Stefanos Vrochidis, Ph.D.[10]; Riccardo Taormina, Ph.D.[11]; Dragan Savic, Ph.D., M.ASCE[12]; and Phoebe Koundouri, Ph.D.[13]

Abstract: Water distribution systems (WDS) are vital for urban societies by providing clean and safe drinking water. Water utilities have to carefully control the injection of disinfectants such as chlorine to effectively deal with natural contaminants and accidental pollution affecting the drinking water quality. This, however, constitutes a challenging task due to sparse sensor readings, complex system dynamics, and various sources of uncertainty. Despite the importance of this problem, the field lacks realistic, publicly available benchmarks to systematically evaluate and compare automated chlorine injection control strategies. To address this gap, we introduce a comprehensive benchmark specifically designed for chlorine injection control. It features realistic operational scenarios based on the Cyprus Disinfection-By-Product (CY-DBP) network, a state-of-the-art model derived from a real-world WDS. The benchmark incorporates sparse sensing conditions, realistic water demands, fluctuating source water quality, and multiple wastewater contamination events. Standardized evaluation metrics, focusing on different desiderata, enable consistent and reproducible assessment of chlorine booster station control strategies, facilitating rigorous comparison across methods. To stimulate cross-disciplinary innovation, we also hosted the benchmark as a competition at a leading artificial intelligence (AI) conference, fostering collaboration between the AI and water engineering communities. Although participants proposed diverse data-driven control approaches, none achieved fully satisfactory performance, underscoring the complexity and open challenges of the task. By providing a realistic and demanding testbed, this benchmark establishes a foundation for advancing the research on intelligent and reliable chlorine booster station control strategies. **DOI: 10.1061/JWRMD5.WRENG-7562**. *This work is made available under the terms of the Creative Commons Attribution 4.0 International license, https://creativecommons.org/licenses/by/4.0/.*

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

Introduction

Water distribution systems are critical infrastructures responsible for delivering clean drinking water to consumers. Maintaining water quality is vital for public health because contaminants can arise from natural sources or accidental pollution, such as wastewater intrusion. To protect water quality, in many countries, water utilities commonly use chlorine as a disinfectant to prevent bacterial growth and mitigate virus infections as water flows through distribution networks (WHO 2004). In this context, the key task is to keep chlorine residuals within tight bounds (Gani et al. 2024); more importantly, it is key to keep chlorine residuals above a certain level (e.g., ≥ 50 mg/L is a common guideline) throughout the network, without exceeding some safety levels. This must be maintained across vast, sparsely sensed networks, and under substantial uncertainties, such as changing water consumption, changing water sources, changes in environmental conditions, and aging assets.

Besides complex system dynamics, rapid urban population growth and climate change further increase the complexity of water quality dynamics, posing significant challenges for conventional monitoring and control methods (Eliades et al. 2023). In this context, artificial intelligence (AI) methods, such as physics-informed deep neural networks (Cuomo et al. 2022) and reinforcement learning (RL) (Arulkumaran et al. 2017), present promising (but yet underexplored) opportunities to enhance and refine traditional chlorination control strategies. Most importantly, physics-informed deep neural networks can integrate complex physical-priors, such as water (quality) dynamics, into the AI method. In addition, RL

[1]Faculty of Technology, Bielefeld Univ., Bielefeld 33615, Germany; KIOS Research and Innovation Center of Excellence, Univ. of Cyprus, Bielefeld 33615, Cyprus (corresponding author). ORCID: https://orcid.org/0000-0002-2426-3126. Email: aartelt@techfak.uni-bielefeld.de

[2]Faculty of Technology, Bielefeld Univ., Bielefeld 33615, Germany.

[3]Faculty of Technology, Bielefeld Univ., Bielefeld 33615, Germany.

[4]Professor, Faculty of Technology, Bielefeld Univ., Bielefeld 33615, Germany. ORCID: https://orcid.org/0000-0002-0935-5591

[5]KIOS Research and Innovation Center of Excellence, Univ. of Cyprus, Nicosia, Aglantzia 2109, Cyprus. ORCID: https://orcid.org/0000-0001-8862-5205

[6]KIOS Research and Innovation Center of Excellence, Univ. of Cyprus, Nicosia, Aglantzia 2109, Cyprus.

Note. This manuscript was submitted on February 18, 2026; approved on June 18, 2026; published online on October 9, 2026. Discussion period open until March 9, 2027; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Water Resources Planning and Management***, ASCE, ISSN 0733-9496.

[7]Professor, KIOS Research and Innovation Center of Excellence, Univ. of Cyprus, Nicosia, Aglantzia 2109, Cyprus. ORCID: https://orcid.org/0000-0001-6184-6366

[8]Professor, KIOS Research and Innovation Center of Excellence, Univ. of Cyprus, Nicosia, Aglantzia 2109, Cyprus; Dept. of Electrical and Computer Engineering, Univ. of Cyprus, Nicosia, Aglantzia 2109, Cyprus.

[9]Center for Research and Technology Hellas, Thessaloniki, Thermi 57001, Greece.

[10]Center for Research and Technology Hellas, Thessaloniki, Thermi 57001. Greece. ORCID: https://orcid.org/0000-0002-2505-9178

[11]Professor, Dept. of Water Management, Technical Univ. Delft, Delft 2628, Netherlands.

[12]Professor, KWR Water Research Institute, Nieuwegein 3433, Netherlands. ORCID: https://orcid.org/0000-0001-9567-9041

[13]Professor, Athens Univ. of Economics and Business, Greece; Univ. of Cambridge, Athens 113 62, UK. ORCID: https://orcid.org/0000-0001-5979-6128

methods can learn on the fly from data and automatically adapt to changes in the system without requiring manual adjustments. The development of such AI-based methods, however, requires realistic benchmark scenarios from which large amounts of training data can be extracted (Garzón et al. 2022). Furthermore, clear evaluation metrics are required to guide the development of AI-based methods for comparison purposes. To the best of our knowledge, no such benchmark on water quality monitoring and control exists.

This work introduces a realistic and comprehensive benchmark for supporting and boosting the research and evaluation of AI-based methods for water quality control, particularly booster chlorination control algorithms (Polycarpou et al. 2002). Beyond its ability to handle deep uncertainties, the benchmark also evaluates the ability to address multiple objectives, including costs and fairness, thereby forming an ideal testbed for complex control and optimization methods. Given the huge potential of AI in this context, we hosted the proposed benchmark as a competition at the International Joint Conference on Artificial Intelligence (IJCAI) 2025, a premier AI conference, to introduce the AI community to this important problem and build a bridge with synergistic effects between the AI and water-engineering communities.

Engaging the AI community in solving water-distribution-system challenges is essential because it accelerates innovation in both fields: AI researchers gain access to rich, real-world problems, and the water sector benefits from cutting-edge methods capable of transforming system performance and resilience. Although the primary focus of the proposed benchmark (and the hosted competition) is controlling the chlorination booster stations, the benchmark can also be used out of the box for benchmarking state estimation and event diagnosis methods, both of which are critical subtasks for successful water quality control.

As evidenced by the submitted approaches, it forms a suitable testbed for diverse methods targeting complex control and optimization tasks, including reinforcement learning, mixed learning and optimization methods, as well as hybrid approaches based on more classical control methods. Finally, by providing an easy interface to the scenario generation of the proposed benchmark, we aim to establish a flexible testbed for water quality control that can be utilized beyond the particularities of the hosted competition.

The remainder of this paper is structured as follows: First, we elaborate more on the background, importance, and challenges of controlling chlorination booster stations. The competition itself is then described in full detail, whereby we dedicate a separate section to the evaluation metrics. Finally, we briefly discuss the submitted methods and their performance, before closing with a summary and conclusion.

Background and Related Work

Chlorine injection control, also referred to as chlorine disinfection control, constitutes a subfield of water quality analysis (Kang et al. 2017; Eliades et al. 2023) and is closely linked to contamination detection (Zulkifli et al. 2018) because chlorine concentrations are used as a proxy for detecting contaminations (Jonkergouw et al. 2004). Chlorination booster stations allow the injection of additional chlorine into the water network, and disinfection control (i.e., chlorine injection control) then refers to operating those booster stations for reacting to hydraulic and contamination events that threaten the water quality (Karamouz et al. 2021). Closely related is the concept of chlorination scheduling or chlorine disinfection schedules, in which the chlorine injection is done without relying on chlorine sensor measurements. In contrast, chlorine injection control refers to feedback control, in which the controller adjusts the injection in real-time based on chlorine sensor readings (Polycarpou et al. 2002).

In general, chlorine disinfection control constitutes a complex problem that can be approached with different techniques. Many different chlorination strategies have been developed over time, often specific to particular water distribution systems (Polycarpou et al. 2002; Lindmark et al. 2022; Kingham 1995; Gamiz et al. 2012; Zidan et al. 2018). The key challenges here are scalability in time and space, as well as dealing with various sources of uncertainties, such as water demands, chlorine decay rates, and so on (Boindala et al. 2025a).

There is extensive literature that uses genetic algorithms for computing (optimal) chlorine disinfection schedules (Seo et al. 2019; Maheshwari et al. 2020; Ostfeld and Salomons 2006). In addition, other methods for disinfection scheduling, such as convex optimization and heuristics (Pecci et al. 2022; Sert and Altan-Sakarya 2017), constraint-based pruning methods (Li et al. 2022), and input–output models of water quality (Vrachimis et al. 2024) have been investigated. For chlorination control, many existing approaches rely on model predictive control (MPC) (Wang et al. 2021; Xie and Brdys 2015; Polycarpou et al. 2002). Those methods, however, struggle to scale in time and space (i.e., across large networks and long time horizons), and often also struggle in handling uncertainties due to varying water flows and unknown chlorine decay dynamics, thus limiting their applicability and impact in practice.

In this context, there is a huge, but mostly unexplored, potential of AI in water quality in general, and also in chlorination disinfection control (Pérez-Beltrán et al. 2024). The few existing approaches include MPC with a simple surrogate model for predicting chlorine residuals based on support vector regression (Wang et al. 2020), novel data-driven control methods such as data-enabled-predictive-control (DeePC) (Perelman and Ostfeld 2025), and methods based on Gaussian process regression (Moeini et al. 2023). In this context, (Deep) RL (Sutton and Barto 2018) constitutes a promising, but so far unexplored, approach for learning chlorine injection control functions. In RL, the system/environment is modeled as a Markov decision process (MDP) containing a set of environment states, a set of possible control actions, a state transition function, and a real-valued reward function for representing an immediate evaluation after executing a control action in the current state. The learning method iteratively observes its current state, chooses an action that causes a transition to a new state, and observes a reward depending on the quality of the chosen action in the current state. By design, RL offers real-time responsiveness and can adapt to new emerging information, including changes in the system, without the need for reoptimization to capture future possibilities. This also makes RL adaptive to uncertainties in the system.

All existing work on water chlorination control, however, uses different, often private or proprietary, benchmarks for evaluating their proposed methods. Although there exist a large number of general benchmark networks, i.e., .inp files (Artelt et al. 2025b; Sivakumar and Prasad 2014; Haghighi et al. 2011; Haghighi and Shamloo 2011), there is still a lack of publicly available and standardized water quality control benchmarks, where not only the hydraulic dynamics are specified, but also the water quality dynamics, chemical injection/control elements, sensor locations, operational scenarios, and standard evaluation procedures. This lack of complete benchmark specifications hinders the reproducibility and comparability of reported results in this domain.

Furthermore, it puts an additional burden and responsibility regarding the creation of realistic benchmark scenarios (together with reasonable evaluation metrics) on the shoulders of the

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

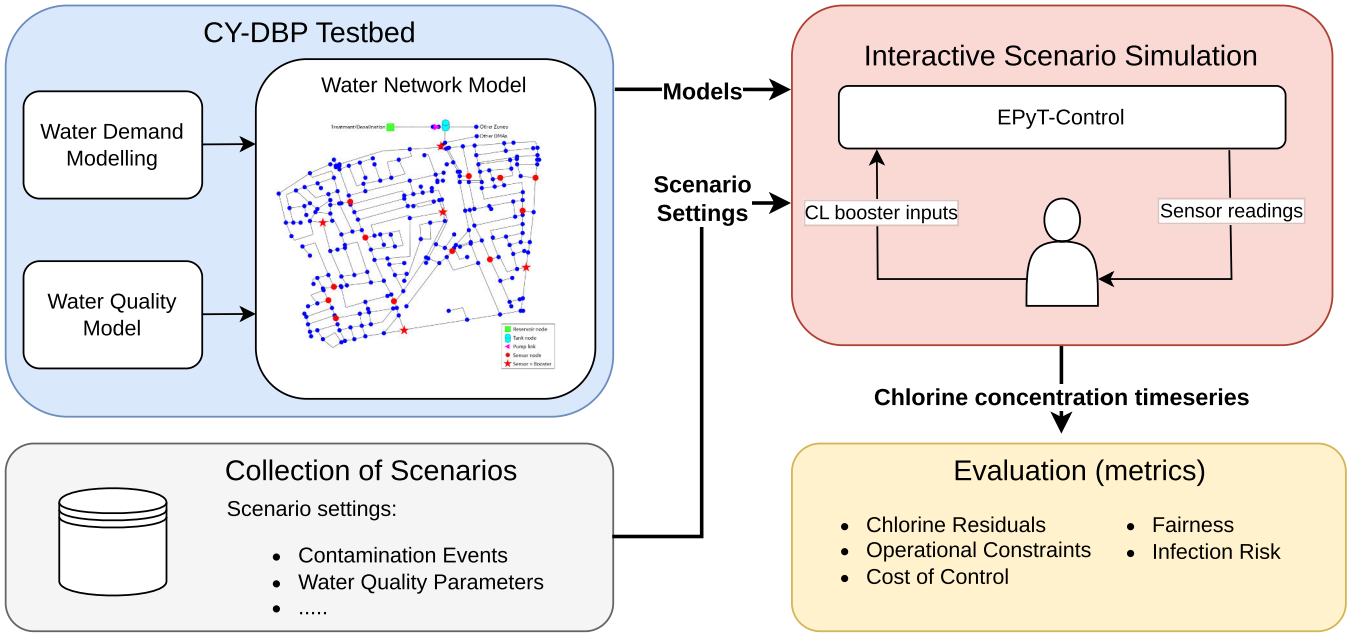


Fig. 1. Different components of the benchmark.

researchers. In this work, we address this limitation by proposing an open, easily accessible, and realistic benchmark for water chlorination control that other researchers can use for reproducible research.

The Benchmark

The proposed benchmark consists of different components (Fig. 1), including scenario modeling and generation, a stepwise and interactive scenario simulation in which control actions can be executed by the user, and an evaluation component.

Task and Materials

The benchmark focuses on regulating disinfectant concentration in drinking water while ensuring robustness against wastewater contamination events. More specifically, the challenge is about controlling the chlorination in the Cyprus Disinfection-By-Product (CY-DBP) water distribution system model (Pavlou et al. 2024) (Fig. 2 illustrates its topology), which is based on a real system in Cyprus. The task is to control the injection of chlorine (Cl) at five booster stations, based on the readings of 17 strategically placed chlorine and two flow sensors only, such that chlorine residuals stay within a tight bound. The localizations of chlorine booster stations and sensors are shown in Fig. 2 and also given in Table 1.

Scenario Generation

To ensure realistic scenarios, we utilize a model of a real-world water distribution network (i.e., the CY-DBP network), and use the stochastic modeling from Paraskevopoulos et al. (2024) for replicating real-world challenges, such as variable and realistic water demands, and variability in the chlorine decay rate due to the use of water from different sources—i.e., we varied the water quality source by introducing various levels (depending on the season) of total organic carbon (TOC) entering the water distribution system (WDS). We also adopt the wastewater contamination modeling by Paraskevopoulos et al. (2024) for modeling the contamination events. More specifically, we incorporate stochastic water demand profiles to capture daily and seasonal variability in consumption.

Regarding chlorine, we introduce both bulk- and wall-decay reactions, including the influence of natural organic matter through fast- and slow-reacting fractions of reducing agents. This allows the benchmark to reproduce realistic spatial and temporal variability in chlorine residuals across the network. We consider the intrusion of wastewater in the network as a contamination event, which includes pathogens (e.g., *Campylobacter* and *Cryptosporidium*) as well as organic matter (i.e., chlorine-reducing agent) that interacts with chlorine and thus exacerbates the challenge of maintaining adequate chlorine residuals throughout the network.

The locations of the monitoring sensors and chlorination boosters were not selected ad hoc, but were obtained via established placement methodologies that explicitly target observability and controllability objectives under realistic network dynamics. Specifically, the 17 chlorine sensor locations were computed with the S-PLACE sensor placement toolkit (Eliades et al. 2014), which formulates sensor placement as an optimization problem over multiple contamination scenarios and selects a limited set of nodes to maximize the effective monitoring coverage of the network and to improve the expected time-to-detection of contamination intrusions. This sensor configuration should be seen as a benchmark design choice derived through formal placement optimization under the selected network and scenario assumptions. It is not meant to suggest that all utilities use the same sensing density, but to provide a sparse yet informative observation setting that makes the control problem meaningful and comparable across methods.

The five chlorine booster stations were determined via a simulation: an optimization approach based on genetic algorithms, following a general booster placement methodology under extended-period simulation, as in related literature (Ohar and Ostfeld 2014). In this work, the optimization objective was to improve controllability of residual chlorine across the network, i.e., maximizing the number of demand nodes at which a controller, given a limited number of injection locations (five), can achieve at least the minimum desired residual while simultaneously ensuring that the maximum allowable residual is not exceeded at any time step. This was assessed under a nominal diurnal demand pattern and under the simplifying assumption of a single-species first-order chlorine decay model with

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

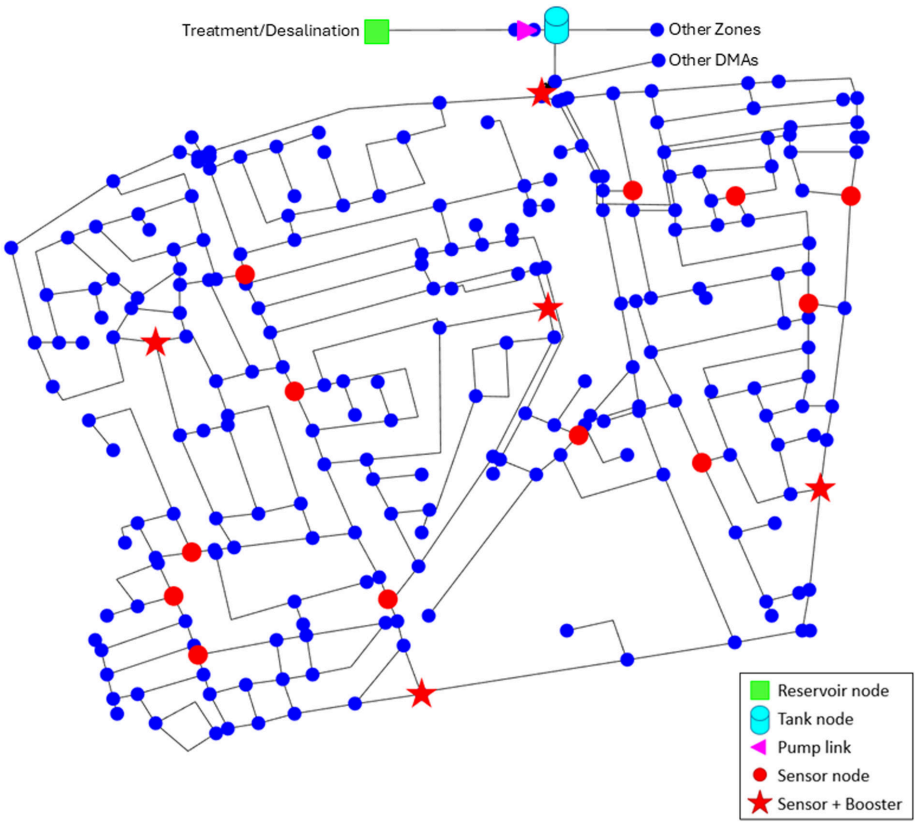


Fig. 2. CY-DBP network, containing one reservoir, one tank, 256 demand nodes, and 335 pipes, together with the location of the five chlorine booster stations, two flow sensors, and 17 chlorine sensors.

Table 1. Locations of sensors and chlorine booster stations

Type	Location
Cl sensor	dist423
Cl sensor	dist225
Cl sensor	dist989
Cl sensor	dist1283
Cl sensor	dist1931
Cl sensor	dist342
Cl sensor	dist275
Cl sensor	dist354
Cl sensor	dist885
Cl sensor	dist485
Cl sensor	dist631
Cl sensor	dist1332
Cl sensor	dist1607
Cl sensor	dist1459
Cl sensor	dist1702
Cl sensor	dist1975
Cl sensor	dist1903
Flow sensor	5
Flow sensor	p-1144
Cl booster station	dist423
Cl booster station	dist225
Cl booster station	dist989
Cl booster station	dist1283
Cl booster station	dist1931

constant chlorine decay parameters. This simplifying assumption applies only to the booster-placement design stage.

The benchmark scenarios used for controller development and evaluation include richer water-quality dynamics, including stochastic demands, bulk and wall chlorine decay, fast- and slow-reacting chlorine-reducing agents, and uncertainty in water-quality conditions. For simulating the hydraulic and water quality dynamics, we use the Python version 3.10 package EPyT-Flow version 0.14.1 (Artelt et al. 2024), which builds on top of state-of-the-art simulation software EPANET version 2.3 (Rossman 2000) and EPANET-MSX version 2.0 (Shang et al. 2008). The benchmark can be used as a realistic and reproducible testbed for developing and comparing methods, not as a full digital replica of utility operations. Although it is based on a real-world network model and realistic hydraulic and water-quality dynamics, the sensing setup, control settings, and evaluation scheme were chosen to support observability, controllability, and fair comparison across methods. Use in a specific utility would therefore require calibration and validation with utility-specific data, along with adaptation to local sensor layouts, operational practice, supervisory control and data acquisition (SCADA) constraints, and regulatory requirements.

We generated a total of 12 scenarios (Table 2), all based on the same network topology (i.e., CY-DBP) and containing the same sensors and chlorine booster stations. The sampling rate for the sensors is always 5 min, i.e., sensor readings are provided in 5-min intervals. However, the scenarios feature slightly different dynamics due to the stochastic modeling of the demands and other parameters such as the initial water quality and chemical reaction rates. The first 10 scenarios are 6 days long and contain a single wastewater intrusion event in the last 3 days; the place and time of the event are fixed but differ for each scenario.

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

Table 2. Overview of the metadata of all scenarios

Scenario ID	Duration (days)	Number of contamination events
1	6	1
2	6	1
3	6	1
4	6	1
5	6	1
6	6	1
7	6	1
8	6	1
9	6	1
10	6	1
11	365	15
12	365	15

Note: The network (i.e., CY-DBP) and the sensor locations and chlorine booster stations are the same for each scenario.

Furthermore, we also created two full-year (i.e., 365-day) scenarios, each with 15 contamination events, whereby the time and location of the contaminations have been randomized. The second of those full-year scenarios (i.e., Scenario 12) was held back and used for the evaluation of the submitted methods. In all scenarios, we use a constant chlorine injection at all booster stations in the first 3 days to let the system converge to a reasonable initial state before the participants can start observing states and interacting with the system by controlling the chlorination booster stations, i.e., the first 3 days are hidden from the participants.

After the competition, we released the well-documented code for generating the scenarios, ensuring full transparency of the competition. Furthermore, this also allows researchers to generate their own version of the benchmark, focusing on different cases; for that, we provide an easy interface for changing the sensor configuration, initial concentrations and patterns of TOC, and full control over all parameters of the contamination events (i.e., size, location, time, and so on). This aligns with the goal of establishing a flexible testbed for researching computational methods for water quality control. In addition, the benchmark was made available on the WaterBenchmarkHub platform (Artelt et al. 2025b), to make it easily findable by other researchers.

Material Provided to the Participants

Participants were provided with 10 of the 6-day scenarios and the first full-year scenario, together with a comprehensive background documentation on the scenarios and on water distribution systems in general. Additionally, a Python (Van Rossum et al. 2007) interface was provided for a step-by-step simulation of these scenario configurations, which can be used to develop a control policy for the chlorine injection at the booster stations. The provided Python interface (Fig. 3) abstracts away from all WDS details, such that participants can solely focus on developing the control policy. This was of particular importance because we expected that most participants from the AI community would have no or very little background knowledge about water distribution systems.

The Python interface (Fig. 3) is implemented utilizing the EPyT-Control version 0.4.0 package (Artelt 2025) (building on top of EPyT-Flow), which is fully compatible with the Gymnasium interface (Towers et al. 2024) and integrates the popular Stable-Baselines3 version 2.7.0 package (Raffin et al. 2021) for reinforcement learning. The evaluation metrics (section "Evaluation Criteria") were also already implemented. Participants had to implement (and submit) their control logic as a custom class in order to be able to run the implemented evaluation metrics (Figs. 4 and 5).

```python
# Load the first environment (i.e., scenario 0)
with WaterChlorinationEnv(**load_scenario(scenario_id=0)) as env:
    # Reset environment and observe initial sensor readings
    # obs: 19 dimensional array (one dimension per sensor) --
    # first 2 dimensions contain the flow sensors, and the remaining 17
    dimensions the chlorine concentration sensors
    obs, info = env.reset()

    for _ in range(20): # Run environment for up to 20 time steps
        # TODO: Your control logic goes here
        act = env.action_space.sample() # Sample random action

        # Apply action, observe new sensor readings and receive reward
        obs, reward, terminated, _, _ = env.step(act)

        if terminated is True:
            break
```

Fig. 3. Example of how to use the provided Python interface for interacting with the scenario simulation.

```python
class ChlorinationControlPolicyRandom(ChlorinationControlPolicy):
    """
    Random control policy -- picks random control actions.
    """
    def compute_action(self, observations: numpy.ndarray) -> numpy.ndarray
    :
        return self._gym_action_space.sample()


# Apply to the environment of scenario 0
with WaterChlorinationEnv(**load_scenario(scenario_id=0)) as env:
    # Create new random policy
    my_policy = ChlorinationControlPolicyRandom(env)

    # Reset environment
    obs, info = env.reset()

    # Run simulation for 20 time steps
    for _ in range(20):
        # Apply policy
        act = my_policy.compute_action(obs)
        obs, reward, terminated, _, _ = env.step(act)

        if terminated is True:
            break
```

Fig. 4. Example of how to implement the control logic in a custom class.

Challenges

Apart from the general complex reaction dynamics and long time horizons, the key challenges in the proposed benchmark stems from the limited observability due to the sparse sensor placement. Furthermore, the uncertainties in the water demands and chemical reaction parameters also add another layer of complexity to the benchmark. Finally, given that the simulation of the complex system dynamics is time-intensive, brute-force approaches, as well as data-hungry methods, are not easily applicable. Altogether, those aspects mimic the challenges that occur in the real world, making this benchmark a nontrivial but realistic testbed for developing and evaluating chlorination control methods.

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

```python
# Create & fit or load policy
my_policy = ...

# Load one environment for evaluation
env = ....

# Evaluate policy
print(evaluate(my_policy, env))
```

Fig. 5. How to run the evaluation in Python.

Evaluation Criteria

Considering the requirements and goals in practice, the primary goal and evaluation metric is to control the chlorine booster stations such that the chlorine residuals stay within tight and predefined bounds (WHO 2004). However, to mimic real-world constraints and additional desiderata, we also propose additional evaluation metrics, turning the control problem into a multiobjective problem. In practice, the stakeholders might decide to assign individual weights (i.e., importances) to those objectives, but here, we used a simple uniform weighting scheme in the competition. This choice was made for the competition to keep the ranking transparent and directly comparable across submitted methods. It should not be interpreted as implying that all objectives carry the same importance in real utility operation, where public health protection and regulatory compliance would often take priority.

In the hosted competition, the submitted control strategies and methods were evaluated on those aspects and criteria separately, and participants were ranked according to their performance on those individual aspects, whereby all aspects are weighted the same in the ranking. The evaluation was done only on the held-back full-year scenario to evaluate the generalization capabilities of the methods, i.e., a proxy for evaluating the sim-to-real gap.

In the following subsections, we describe each evaluation metric separately.

Chlorine Residuals

In order to ensure safe drinking water at all times, the chlorine concentration at junctions (i.e., demand nodes) must be within a tight bound: Too low chlorine concentrations cannot ensure safe drinking water, whereas too high chlorine concentrations and their disinfection by-products (Du et al. 2017) are known to be unhealthy for the human body (Gopal et al. 2007).

Formally, for each junction $v \in V$ in the network, the chlorine concentration c_v must be within a given interval $[a, b]$ at all discrete times $T = \{1, \ldots, T\}$

$$\frac{1}{T}\sum_{t=1}^{T} \mathbf{1}_{[a,b]}(c_v(t)) \tag{1}$$

where $\mathbf{1}_{[a,b]}(c_v(t)) = 1 \leftrightarrow c_v(t) \in [a, b]$ and zero otherwise. We aggregate the chlorine bound violations over all junctions V by taking the average of Eq. (1)

$$s_v := \frac{1}{T|V|}\sum_{v \in V}\sum_{t=1}^{T} \mathbf{1}_{[a,b]}(c_v(t)) \tag{2}$$

In the competition, we set [0.2,0.4] mg/L as the desired chlorine concentration bounds at every junction, which are inspired by regulations such as those by the World Health Organization (WHO 2011). These bounds are used here as representative operational targets for benchmark evaluation and do not reflect all utility- or jurisdiction-specific regulatory requirements.

Fairness

Besides evaluating the chlorine residuals at a particular junction, we also evaluate their spatial variations. By this, we evaluate some form of water quality fairness (Boindala et al. 2025b) where we measure the worst-case of how different all junctions (i.e., consumers) are treated; ideally, all consumers should receive the same water quality, no matter of their distance to the treatment plant

$$\max_{v1,\, v2 \in V} |s_{v1} - s_{v2}| \tag{3}$$

where s_v = mean chlorine bound violations at junction $v \in V$ [Eq. (2)], and smaller values of Eq. (3) correspond to better fairness.

Chlorine Injection Pump Operation Constraints

From an operational point of view, it is unhealthy or even unrealistic for chlorine injection pumps to have large changes/jumps in the injection rate. We therefore evaluate the smoothness of the chlorine injection over time.

The average rate of change with respect to the chlorine injection u_v at junction $v \in V$ is evaluated as follows:

$$\frac{1}{T-1}\sum_{t=1}^{T-1} |u_v(t) - u_v(t+1)| \tag{4}$$

In this benchmark, we have five chlorine injection pumps to be operated $(v_1, \ldots, v_5)$. To properly consider all of them and to avoid any averaging-out effects, we take the maximum over their average rate of Change 4 as the final evaluation, which is to be minimized

$$\max_{v \in \{v_1, \ldots, v_5\}} \frac{1}{T-1}\sum_{t=1}^{T-1} |u_v(t) - u_v(t+1)| \tag{5}$$

Cost of Control

Here, we model the cost of control, which is to be minimized, as the amount of injected chlorine mass at the chlorine booster stations $V_b \subset V$, which here is $V_b = \{v_1, \ldots, v_5\}$, using the following equation:

$$\sum_{t=1}^{T}\sum_{v_j \in V_b} u_{v_j}(t) \tag{6}$$

Infection Risk

The infection risk is an evaluation metric that quantifies the public health impact by estimating the probability of an individual getting infected after exposure to pathogens during a contamination event. Following the Quantitative Microbial Risk Assessment (QMRA) framework (Haas et al. 2014; Paraskevopoulos et al. 2024), it combines the pathogen concentration, the tap water volume, and dose-response modeling. Dose-response modeling describes how an organism responds to the exposure of a microbial contaminant.

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

Here, we want to model the risk of getting infected when consuming a specific dose of microbial contaminant.

For the purposes of this benchmark, we focus on enterovirus, a common waterborne pathogen that is known to survive in distribution systems, is infectious even at low doses, and reacts with chlorine during disinfection. The dose for each individual is calculated by multiplying the pathogen concentration in drinking water by the ingested volume (tap water) across multiple daily consumption events during a contamination event (starting at time t_0 and ending at t_1), totaling up to 1 L per person per day. The assumed daily ingestion of up to 1 L/person is distributed across multiple stochastic tap-water consumption events of various timing, consumed volume and duration, thereby preserving temporal and interindividual variability in exposure (Paraskevopoulos et al. 2024). For each junction $v \in V$, person $p \in P(v)$, and time step $t \in [t_0, t_1]$, we have

$$\text{Dose}(v,t,p) = \text{Volume}(v,t,p) \times \text{PathogenConcentration}(v,t) \quad (7)$$

The probability of infection is then obtained using the exponential dose-response model (Haas 2015) as follows:

$$\text{Risk}(v,t,p) = 1 - \exp(-r \times \text{Dose}(v,t,p)) \quad (8)$$

where $r = 0.014472$ represents the pathogen-specific infectivity (Paraskevopoulos et al. 2024). The infection risk is then defined as the ratio of expected infections to the total population as follows:

$$\text{InfectionRisk} = \frac{\sum_v \sum_p (1 - \Pi_t (1 - \text{Risk}(v,t,p)))}{\sum_v |P(v)|} \times 100 \quad (9)$$

In contrast to other water quality metrics, this indicator directly reflects the potential health consequences of a specific contamination and is therefore computed separately for each contamination event. Although we focus on enterovirus here, the framework can be readily extended to other pathogens and dose–response models. Finally, a lower infection risk indicates that the booster control strategy has been more effective at mitigating public health risks during undetected contamination events.

Submissions and Results

A total of six teams participated in the competition at IJCAI-2025. Participating teams had to submit a short written report, the source code of their proposed method, and the final (trained) controller/policy for evaluation by the organizing committee. All submissions, including reports, source code, and final evaluation results, have been published on the competition's webpage (Artelt et al. 2025a).

Submitted methods featured a wide variety of different approaches, including deep reinforcement learning, rule mining, evolutionary methods, and modern control theory. Notably, all participating teams employed some form of surrogate model to estimate the effect of control actions. In the competition, we performed a global evaluation of all submitted methods, and the results are presented in Table 3. For the purpose of better comparability, we also include a simple baseline in which we use a constant chlorine injection at the booster stations as it was done during the initialization phase (i.e., first 3 days). We also evaluate the statistical significance (by means of the Mann-Whitney U-test) with respect to the baseline of all submitted methods.

In addition, in this paper, we also performed a more detailed evaluation, in which we assess the methods' performance under different operation conditions, such as peak demand versus normal demand and no contamination versus a contamination event present. This allows us to better analyze the performance of the submitted methods in dependence on the WDS state. The results are given in Tables 4 and 5, in which we also evaluate the statistical significance of the difference by means of the Mann-Whitney U-test. We restricted this additional evaluation to the reasonable methods (including the constant baseline), i.e., we do not consider any methods that do not inject any chlorine at all.

The winning team developed a rule mining approach, guided by a long-short-term-memory (LSTM)-based surrogate model, for generating the final control rules of the chlorine booster stations. In the following, we briefly describe the approaches of the three top-performing teams. More detailed information can be found in their reports and related publications.

Table 3. Results per team

Team ID	Cost of control ↓	Control smoothness ↓	Cl bound violations ↓	Cl bound violation fairness ↓	Average infection risk ↓
1	$31{,}048{,}978.00^{***}$	5.937^{ns}	0.103^{***}	0.177^{ns}	6.656^{ns}
5	$209{,}191.72^{***}$	0^{ns}	0.171^{***}	0.199^{ns}	9.082^{ns}
3	$19{,}098{,}485.02^{***}$	50.36^{**}	0.150^{***}	0.331^{ns}	8.067^{ns}
2	0^{***}	0^{ns}	0.171^{***}	0.199^{ns}	9.085^{ns}
4	0^{***}	0^{ns}	0.171^{***}	0.199^{ns}	9.085^{ns}
6	—	—	—	—	—
Baseline	56,297,160.00	0	0.147	0.196	7.478

Note: Each evaluation metric is shown separately, and smaller numbers are better in all cases. We also show the statistical significance (with respect to the baseline) according to the Mann-Whitney U-test [not significant (ns) = p-value > 0.05; *p-value ≤ 0.05; **p-value ≤ 0.01; and ***p-value ≤ 0.001]. Team 6 did not submit any source code, and therefore, we were unable to run and evaluate their method.

Table 4. Absolute difference in performance for two different operation phases: Peak demand versus normal demand

Team ID	Average cost of control	Control smoothness	Cl bound violations	Cl bound violations fairness	Average infection risk
1	4.564^{***}	4.643^{ns}	0.009^{***}	0.029^{ns}	0.547^{ns}
5	0.004^{***}	0.001^{ns}	0.008^{***}	0.001^{ns}	0.647^{ns}
3	0.01^{ns}	0.1^{ns}	0.001^{***}	0.133^{ns}	0.616^{ns}
Baseline	0^{ns}	0^{ns}	0.008^{***}	0.007^{ns}	0.615^{ns}

Note: We show the difference together with the statistical significance according to the Mann-Whitney U-test (ns = p-value > 0.05; *p-value ≤ 0.05; **p-value ≤ 0.01; and ***p-value ≤ 0.001).

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

Table 5. Absolute difference in performance for two different operation phases: Contamination event versus no-contamination event

Team ID	Average cost of control	Control smoothness	Cl bound violations	Cl bound violations fairness
1	4.410^{**}	1.441^{ns}	0.016^{***}	0.017^{ns}
5	0.002^{***}	0^{ns}	0.008^{***}	0^{ns}
3	0.01^{ns}	0.1^{ns}	0.007^{***}	0.125^{ns}
Baseline	0^{ns}	0^{ns}	0.009^{***}	0.001^{ns}

Note: We show the difference together with the statistical significance according to the Mann-Whitney U-test ($ns = p$-value > 0.05; $*p$-value ≤ 0.05; $**p$-value ≤ 0.01; and $***p$-value ≤ 0.001).

Team 1

The winning team explored classic deep reinforcement learning algorithms (Arulkumaran et al. 2017), such as proximal-policy-optimization (PPO) (Schulman et al. 2017), as well as classic model-predictive control (Schwenzer et al. 2021). However, those methods did not perform well (probably due to their sample inefficiency), and this is why they decided to develop a rule mining approach for generating control rules. To guide the rule mining approach, they developed a simple LSTM-based surrogate model for predicting chlorine concentration states.

Team 5

This team heavily used evolutionary methods. Similar to the other teams, they also developed an LSTM-based surrogate model for predicting the effect of control actions on the chlorine concentrations; however, they used evolutionary surrogate-assisted prescription (Francon et al. 2020) for training this model. The final policy was then trained by neuroevolution of augmenting topologies (NEAT) (Stanley and Miikkulainen 2002).

Team 3

Team 3 relied on control theory methods. More specifically, they used DeePC (Coulson et al. 2019; Perelman and Ostfeld 2025) in conjunction with a fallback proportional-integral-derivative (PID) controller. For DeePC, they also built a surrogate model for approximating the system dynamics; for this, they relied on Hankel matrices for reconstructing the system dynamics from historical input–output trajectories.

Discussion

From Table 3, it becomes apparent that only the first three teams (Team 1, 5, and 3) submitted a nontrivial solution. The solutions submitted by Teams 2 and 4 do not inject any chlorine into the network for the evaluation scenario (i.e., the cost of control is zero). Although this does not constitute an interesting submission, it serves as an additional baseline that allows to quantify how much better the other submissions are compared with doing nothing. Most remarkably, none of the submissions obtained a perfect solution, which would be near-zero infection risk and chlorine bound violations.

Furthermore, only the submission by Team 1 (the winning team) outperformed the constant chlorine injection baseline. Teams 5 and 3 (i.e., evolutionary method and DeePC) performed worse than the constant chlorine injection baseline and only performed slightly better in minimizing the infection risk compared with the baselines, whereas Team 1 (rule mining + surrogate model) yielded a significant improvement over all other submitted methods. In particular, the chlorine concentration bound violations were statistically significant with respect to the baseline. This indicates that it seems to be challenging to detect the wastewater contamination events early enough to take appropriate countermeasures (i.e., increasing chlorine injection). Similar effects can be observed for the chlorine concentration bound violations, including their spatial variations. Furthermore, it seems to be difficult to satisfy the bounds at all nodes. This was also noted by Team 3, who observed that the chlorine concentrations at some nodes (mostly dead-ends) in the network are very resilient to changes in the chlorine injection rates.

Most interestingly, all teams developed and employed a data-driven surrogate model to predict chlorine concentrations over time. This highlights the potential of machine learning/data-driven approaches for substituting time-consuming simulations for efficiently learning control rules.

Finally, Tables 4 and 5 indicate that there were statistically significant differences in different operation phases (i.e., peak demand versus normal demand, and contamination event versus no-contamination event) for almost all submitted methods. In particular, the statistically significant changes in the cost of control indicate that the methods were not only able to react to changes in the demands but also able to detect and react to the presence of contamination events. However, the difference in the statistical significance and the order of magnitude in the difference itself (higher for differences in demands than for contamination events) provides further evidence that the submitted methods struggled to reliably detect contamination events. Altogether, the results indicate that the proposed benchmark is indeed nontrivial and leaves work for future research.

Conclusions

This paper introduced a realistic benchmark for controlling chlorine booster stations in a WDS under various uncertainties and random wastewater intrusion events. The benchmark contains several simulation scenarios and comes with an intuitive interface for developing and benchmarking control algorithms. Scenario simulation and evaluation methods have already been implemented to facilitate easy usage. This benchmark is the first publicly available one in the domain of drinking water quality control, addressing the urgent need for benchmarks on water quality control to boost and promote reproducible research in this area.

The presented benchmark was hosted as a competition at IJCAI-2025, a premier AI conference, for promoting the use of AI methods and building a synergistic bridge between the AI and the water system communities. Submitted solutions featured a wide variety of different methods, with the winning solution being based on a rule mining approach in conjunction with a surrogate model for replacing the time-consuming simulations for predicting the effect of control actions on the chlorine concentrations. Nevertheless, none of the submissions managed to develop a satisfactory control method that performed well in all evaluation metrics. This highlights how challenging water quality control is and opens the opportunity for follow-up work and research on this important problem.

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

A useful direction for future work is to extend the evaluation framework beyond the uniform weighting used in the competition and examine alternative prioritization settings that are closer to real operational practice, such as safety-first, compliance-first, or cost-constrained evaluation schemes. Future benchmark versions could also explore a broader set of pathogen-specific and utility-specific evaluation settings, depending on the intended operational context. The proposed benchmark also serves as a flexible testbed for water quality control research. We have made the complete, well-documented code for generating scenarios publicly available, allowing users to create customized scenarios on the CY-DBP virtual testbed that focus on specific aspects of interest. As with any simulation-based benchmark, applying it in a real utility would require further calibration, validation, and adaptation to site-specific infrastructure and operational practice. In this way, the benchmark is intended to support and facilitate future research in the water quality domain.

Data Availability Statement

All data, models, and code generated or used during the study, including submissions from the participants (implementation and report), are available in the following online repositories in accordance with funder data retention policies: https://github.com/WaterFutures/AI-for-Drinking-Water-Chlorination-Challenge-IJCAI-25; https://zenodo.org/records/20152491; https://waterfutures.github.io/WaterBenchmarkHub/benchmarks/AI4WaterChlorinationChallenge.html.

Acknowledgments

This research was supported by the Ministry of Culture and Science NRW (Germany) as part of the Lamarr Fellow Network. We also gratefully acknowledge funding from the European Research Council (ERC) under the ERC Synergy Grant Water-Futures (Grant No. 951424). This publication reflects the views of the authors only.

Author Contributions

André Artelt: Conceptualization; Data curation; Methodology; Software; Writing – original draft; Writing – review and editing. Luca Hermes: Conceptualization; Data curation; Methodology; Software; Writing – original draft; Writing – review and editing. Janine Strotherm: Conceptualization; Data curation; Investigation; Software; Writing – original draft; Writing – review and editing. Barbara Hammer: Funding acquisition; Supervision; Writing – review and editing. Stelios G. Vrachimis: Conceptualization; Data curation; Methodology; Software; Writing – original draft; Writing – review and editing. Marios S. Kyriakou: Software. Demetrios G. Eliades: Conceptualization; Supervision; Writing – review and editing. Marios M. Polycarpou: Funding acquisition; Supervision; Writing – review and editing. Sotirios Paraskevopoulos: Conceptualization; Data curation; Methodology; Software; Writing – original draft; Writing – review and editing. Stefanos Vrochidis: Supervision. Riccardo Taormina: Conceptualization; Supervision. Dragan Savic: Funding acquisition; Supervision; Writing – review and editing. Phoebe Koundouri: Funding acquisition; Supervision; Writing – review and editing.

Supplemental Materials

All data, models, and code, together with the submissions from the participants (implementation and report), are publicly available in an online repository: https://github.com/WaterFutures/AI-for-Drinking-Water-Chlorination-Challenge-IJCAI-25.

References

Artelt, A. 2025. "EPyT-control—EPANET python toolkit—Control." *GitHub Repository*. Accessed December 1, 2025. https://github.com/WaterFutures/EPyT-Control.

Artelt, A., et al. 2025a. "1st AI for drinking water chlorination challenge." Accessed December 1, 2025. https://github.com/WaterFutures/AI-for-Drinking-Water-Chlorination-Challenge-IJCAI-25.

Artelt, A., K. Giese, S. G. Vrachimis, D. G. Eliades, M. M. Polycarpou, and B. Hammer. 2025b. "The WaterBenchmarkHub: A platform for benchmarks in water distribution networks." In *Proc., 21st Computing and Control in the Water Industry Conf.* Sheffield, UK: Univ. of Sheffield.

Artelt, A., M. S. Kyriakou, S. G. Vrachimis, D. G. Eliades, B. Hammer, and M. M. Polycarpou. 2024. "EPyT-flow: A toolkit for generating water distribution network data." *J. Open Source Software* 9 (103): 7104. https://doi.org/10.21105/joss.07104.

Arulkumaran, K., M. P. Deisenroth, M. Brundage, and A. A. Bharath. 2017. "Deep reinforcement learning: A brief survey." *IEEE Signal Process Mag.* 34 (6): 26–38. https://doi.org/10.1109/MSP.2017.2743240.

Boindala, S. P., G. Jaykrishnan, and A. Ostfeld. 2025a. "Booster disinfection scheduling under uncertainty in water distribution systems: Approximate robust reformulation approach." *J. Water Resour. Plann. Manage.* 151 (2): 04024070. https://doi.org/10.1061/JWRMD5.WRENG-6555.

Boindala, S. P., S. G. Vrachimis, D. G. Eliades, and M. Polycarpou. 2025b. "Optimal chlorine dosage scheduling in drinking water distribution systems considering fairness objectives." In *Proc., World Environmental and Water Resources Congress 2025*. Reston, VA: ASCE. https://doi.org/10.1061/9780784486184.092.

Coulson, J., J. Lygeros, and F. Dorfler. 2019. "Data-enabled predictive control: In the shallows of the DeePC." In *Proc., 2019 18th European Control Conf.* New York: IEEE. https://doi.org/10.23919/ECC.2019.8795639.

Cuomo, S., V. S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli. 2022. "Scientific machine learning through physics–informed neural networks: Where we are and what's next." *J. Sci. Comput.* 92 (3): 88. https://doi.org/10.1007/s10915-022-01939-z.

Du, Y., X.-T. Lv, Q.-Y. Wu, D.-Y. Zhang, Y.-T. Zhou, L. Peng, and H.-Y. Hu. 2017. "Formation and control of disinfection byproducts and toxicity during reclaimed water chlorination: A review." *J. Environ. Sci.* 58 (Aug): 51–63. https://doi.org/10.1016/j.jes.2017.01.013.

Eliades, D. G., M. Kyriakou, and M. M. Polycarpou. 2014. "Sensor placement in water distribution systems using the S-PLACE toolkit." *Procedia Eng.* 70: 602–611. https://doi.org/10.1016/j.proeng.2014.02.066.

Eliades, D. G., S. G. Vrachimis, A. Moghaddam, I. Tzortzis, and M. M. Polycarpou. 2023. "Contamination event diagnosis in drinking water networks: A review." *Annu. Rev. Control* 55: 420–441. https://doi.org/10.1016/j.arcontrol.2023.03.011.

Francon, O., S. Gonzalez, B. Hodjat, E. Meyerson, R. Miikkulainen, X. Qiu, and H. Shahrzad. 2020. "Effective reinforcement learning through evolutionary surrogate-assisted prescription." In *Proc., 2020 Genetic and Evolutionary Computation Conf.* New York: Association for Computing Machinery. https://doi.org/10.1145/3377930.3389842.

Gamiz, J., Y. Bolea, A. Grau, J. Gamiz, F. Luque, and J. M. Vargas. 2012. "Automation of chlorination process for drinking water treatment plant: Control strategies." In *Proc., IEEE 10th Int. Conf. on Industrial Informatics*. New York: IEEE. https://doi.org/10.1109/INDIN.2012.6301355.

Gani, A., M. Singh, S. Pathak, and A. Hussain. 2024. "Recent advancements in chlorine applications for water quality control." In *Drinking water disinfection by-products*. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-031-49047-7_3.

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

Garzón, A., Z. Kapelan, J. Langeveld, and R. Taormina. 2022. "Machine learning-based surrogate modeling for urban water networks: Review and future research directions." *Water Resour. Res.* 58 (5): e2021WR031808. https://doi.org/10.1029/2021WR031808.
Gopal, K., S. S. Tripathy, J. L. Bersillon, and S. P. Dubey. 2007. "Chlorination byproducts, their toxicodynamics and removal from drinking water." *J. Hazard. Mater.* 140 (1–2): 1–6. https://doi.org/10.1016/j.jhazmat.2006.10.063.
Haas, C. N. 2015. "Microbial dose response modeling: Past, present, and future." *Environ. Sci. Technol.* 49 (3): 1245–1259. https://doi.org/10.1021/es504422q.
Haas, C. N., J. B. Rose, and C. P. Gerba. 2014. *Quantitative microbial risk assessment*. Hoboken, NJ: John Wiley & Sons.
Haghighi, A., H. M. V. Samani, and Z. M. V. Samani. 2011. "GA-ILP method for optimization of water distribution networks." *Water Resour. Manage.* 25 (7): 1791–1808. https://doi.org/10.1007/s11269-011-9775-4.
Haghighi, A., and H. Shamloo. 2011. "Transient generation in pipe networks for leak detection." *Water Manage.* 164 (6): 311–318. https://doi.org/10.1680/wama.2011.164.6.311.
Jonkergouw, P., S. Khu, and D. Savic. 2004. "Chlorine: A possible indicator of intentional chemical and biological contamination in a water distribution network." In *Proc., IWA Conf. on Automation in Water Quality Monitoring*. London: IWA Publishing.
Kang, G., J. Z. Gao, and G. Xie. 2017. "Data-driven water quality analysis and prediction: A survey." In *Proc., 2017 IEEE 3rd Int. Conf. on Big Data Computing Service and Applications (BigDataService)*. New York: IEEE. https://doi.org/10.1109/BigDataService.2017.40.
Karamouz, M., A. Yousefi, Z. Zahmatkesh, D. Mahmoodzadeh, and M. D. Pirooz. 2021. "Simulation of chlorine injection in water distribution networks in response to contaminations." *Int. J. Environ. Res.* 16 (1): 6. https://doi.org/10.1007/s41742-021-00379-1.
Kingham, T. 1995. "Chlorination control in a large water treatment works." In *Proc., IEE Colloquium on 'Application of Advanced PLC (Programmable Logic Controller) Systems with Specific Experiences from Water Treatment*. Hertfordshire, UK: Institute of Engineering and Technology. https://doi.org/10.1049/ic:19950738.
Li, J., S. Liu, F. Meng, X. Wu, and K. Smith. 2022. "An efficient constraint-based pruning method to improve chlorine dosage optimization." *J. Water Resour. Plann. Manage.* 148 (8): 04022038. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001580.
Lindmark, M., et al. 2022. "Passive in-line chlorination for drinking water disinfection: A critical review." *Environ. Sci. Technol.* 56 (13): 9164–9181. https://doi.org/10.1021/acs.est.1c08580.
Maheshwari, A., A. Abokifa, R. D. Gudi, and P. Biswas. 2020. "Optimization of disinfectant dosage for simultaneous control of lead and disinfection-byproducts in water distribution networks." *J. Environ. Manage.* 276 (Dec): 111186. https://doi.org/10.1016/j.jenvman.2020.111186.
Moeini, M., L. Sela, A. F. Taha, and A. A. Abokifa. 2023. "Bayesian optimization of booster disinfection scheduling in water distribution networks." *Water Res.* 242 (Aug): 120117. https://doi.org/10.1016/j.watres.2023.120117.
Ohar, Z., and A. Ostfeld. 2014. "Optimal design and operation of booster chlorination stations layout in water distribution systems." *Water Res.* 58 (Jul): 209–220. https://doi.org/10.1016/j.watres.2014.03.070.
Ostfeld, A., and E. Salomons. 2006. "Conjunctive optimal scheduling of pumping and booster chlorine injections in water distribution systems." *Eng. Optim.* 38 (3): 337–352. https://doi.org/10.1080/03052150500478007.
Paraskevopoulos, S., S. Vrachimis, M. Kyriakou, D. G. Eliades, P. Smeets, M. Polycarpou, and G. Medema. 2024. "Modeling the health impact of wastewater contamination events in drinking water networks." *J. Cleaner Prod.* 479 (Nov): 143997. https://doi.org/10.1016/j.jclepro.2024.143997.
Pavlou, P., M. Kyriakou, S. G. Vrachimis, and D. G. Eliades. 2024. "A comprehensive virtual testbed for modeling disinfection byproduct formation in water distribution networks." In *Proc., 3rd Int. Joint Conf. on Water Distribution Systems Analysis & Computing and Control for the Water Industry*. Basel, Switzerland: Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/engproc2024069033.
Pecci, F., I. Stoianov, and A. Ostfeld. 2022. "Convex heuristics for optimal placement and operation of valves and chlorine boosters in water networks." *J. Water Resour. Plann. Manage.* 148 (2): 04021098. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001509.
Perelman, G., and A. Ostfeld. 2025. "Data enabled predictive control for water distribution systems optimization." *Water Resour. Res.* 61 (4): e2024WR039059. https://doi.org/10.1029/2024WR039059.
Pérez-Beltrán, C., A. Robles, N. Rodriguez, F. Ortega-Gavilán, and A. Jiménez-Carvelo. 2024. "Artificial intelligence and water quality: From drinking water to wastewater." *TrAC, Trends Anal. Chem.* 172 (Mar): 117597. https://doi.org/10.1016/j.trac.2024.117597.
Polycarpou, M., J. Uber, Z. Wang, F. Shang, and M. Brdys. 2002. "Feedback control of water quality." *IEEE Control Syst.* 22 (3): 68–87. https://doi.org/10.1109/MCS.2002.1004013.
Raffin, A., A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann. 2021. "Stable-baselines3: Reliable reinforcement learning implementations." *J. Mach. Learn. Res.* 22 (268): 1–8.
Rossman, L. A. 2000. *EPANET 2: Users manual*. Rep. No. EPA/600/R-00/057. Washington, DC: USEPA.
Schulman, J., F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. 2017. "Proximal policy optimization algorithms." Preprint, submitted July 20, 2017. https://arxiv.org/abs/1707.06347.
Schwenzer, M., M. Ay, T. Bergs, and D. Abel. 2021. "Review on model predictive control: An engineering perspective." *Int. J. Adv. Manuf. Technol.* 117 (5–6): 1327–1349. https://doi.org/10.1007/s00170-021-07682-3.
Seo, J., K. Kim, J. Hyung, T. Kim, and J. Koo. 2019. "Optimization model for location and operation schedule of chlorine booster stations in water distribution networks." *Desalin. Water Treat.* 140 (Feb): 91–102. https://doi.org/10.5004/dwt.2019.23373.
Sert, C., and A. B. Altan-Sakarya. 2017. "Optimal scheduling of booster disinfection in water distribution networks." *Civ. Eng. Environ. Syst.* 34 (3–4): 278–297. https://doi.org/10.1080/10286608.2018.1463218.
Shang, F., J. G. Uber, and L. A. Rossman. 2008. "Modeling reaction and transport of multiple species in water distribution systems." *Environ. Sci. Technol.* 42 (3): 808–814. https://doi.org/10.1021/es072011z.
Sivakumar, P., and R. K. Prasad. 2014. "Simulation of water distribution network under pressure-deficient condition." *Water Resour. Manage.* 28 (10): 3271–3290. https://doi.org/10.1007/s11269-014-0677-0.
Stanley, K. O., and R. Miikkulainen. 2002. "Evolving neural networks through augmenting topologies." *Evol. Comput.* 10 (2): 99–127. https://doi.org/10.1162/106365602320169811.
Sutton, R. S., and A. G. Barto. 2018. "Reinforcement learning: An introduction." In *Adaptive computation and machine learning series*. 2nd ed. Cambridge, MA: MIT Press.
Towers, M., et al. 2024. "Gymnasium: A standard interface for reinforcement learning environments." Preprint, submitted July 24, 2024. https://arxiv.org/abs/2407.17032.
Van Rossum, G. et al. 2007. "Python programming language." In Vol. 41 of *Proc., USENIX Annual Technical Conf.*, 1–36. Berkeley, CA: USENIX Association.
Vrachimis, S. G., D. G. Eliades, and M. M. Polycarpou. 2024. "Disinfection scheduling in water distribution networks considering input time-delay uncertainty." *J. Hydroinf.* 26 (2): 386–396. https://doi.org/10.2166/hydro.2024.099.
Wang, D., J. Shen, S. Zhu, and G. Jiang. 2020. "Model predictive control for chlorine dosing of drinking water treatment based on support vector machine model." *Desalin. Water Treat.* 173 (Jan): 133–141. https://doi.org/10.5004/dwt.2020.24144.
Wang, S., A. F. Taha, and A. A. Abokifa. 2021. "How effective is model predictive control in real-time water quality regulation? State-space modeling and scalable control." *Water Resour. Res.* 57 (5): e2020WR027771. https://doi.org/10.1029/2020WR027771.
WHO (World Health Organization). 2004. Vol. 1 of *Guidelines for drinking-water quality*. Geneva: WHO.
WHO (World Health Organization). 2011. "Guidelines for drinking-water quality." *WHO Chron.* 38 (4): 104–108.

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

Xie, M., and M. Brdys. 2015. "Nonlinear model predictive control of water quality in drinking water distribution systems with dbps objectives." *Int. J. Math. Comput. Phys. Electr. Comput. Eng.* 9 (8): 354–360.

Zidan, N., M. Maree, and S. Samhan. 2018. "An IoT based monitoring and controlling system for water chlorination treatment." In *Proc., 2nd Int. Conf. on Future Networks and Distributed Systems*. New York: Association for Computing Machinery. https://doi.org/10.1145/3231053.3231084.

Zulkifli, S. N., H. A. Rahim, and W.-J. Lau. 2018. "Detection of contaminants in water supply: A review on state-of-the-art monitoring technologies and their applications." *Sens. Actuators, B* 255 (Feb): 2657–2689. https://doi.org/10.1016/j.snb.2017.09.078.

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.