




Disinfection scheduling in water distribution networks considering input time-delay uncertainty

Stelios G. Vrachimis ^{a,*}, Demetrios G. Eliades ^a and Marios M. Polycarpou ^{a,b}

^a KIOS Research and Innovation Center of Excellence, University of Cyprus, Nicosia, Cyprus

^b Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus

*Corresponding author. E-mail: vrachimis.stelios@ucy.ac.cy

 SGV, 0000-0001-8862-5205; DGE, 0000-0001-6184-6366; MMP, 0000-0001-6495-9171

ABSTRACT

A significant challenge when attempting to regulate the spatial-temporal concentration of a disinfectant in a water distribution network is the large and uncertain delay between the time that the chemical is injected at the input node and the time that the concentration is measured at the monitoring output nodes. Uncertain time delays are due to varying water flows, which depend mainly on consumer water demands. Existing approaches cannot guarantee that the concentration of the disinfectant will remain within a specified range at the output, even though bounds on time-delay uncertainty may be known. In this work, given bounded water-flow uncertainty, we use the input–output modeling approach to develop a disinfectant scheduling methodology that guarantees a bounded output disinfectant concentration. The proposed methodology creates an input–output model uncertainty characterization by utilizing estimated bounds on water-quality states using the backtracking approach. An optimization problem is formulated and solved to find an input schedule that keeps the disinfectant concentration within predefined bounds for a specified time horizon. Simulation results in two case studies where water demands varied between $\pm 20\%$ of their nominal value show that the proposed scheduler is able to avoid lower bound violations of disinfectant concentration.

Key words: modeling, prediction, uncertainty quantification, water quality

HIGHLIGHTS

- A methodology for calculating the disinfection chemical input in water distribution networks is presented that considers the uncertainty in water flows.
- The methodology guarantees bounded disinfectant concentration at monitored locations, given bounds on water flows for a time horizon.
- Results on an example network show robust performance under high demand uncertainty, in scenarios with abrupt changes, and flow reversals.

1. INTRODUCTION

There is currently significant research interest in real-time monitoring and regulation of drinking water quality, especially after a recent EU Directive (European Union 2020), which encourages continuous monitoring of water-quality parameters in water distribution networks (WDN). According to the World Health Organization (WHO 2017), a disinfectant residual needs to be sustained throughout drinking water networks, such that it is sufficient to deactivate waterborne pathogens and, at the same time, small enough to reduce the formation of harmful disinfection byproducts (DBPs), such as trihalo-methanes (THMs) and haloacetic acids (HAA) (Mouly *et al.* 2010).

The effective regulation of water quality in WDN is based on the use of mathematical models that facilitate the estimation of disinfectant concentration in the water at different parts of the network (Elsherif *et al.* 2022). This is a challenging task, since water-quality models are inaccurate and complex, especially in large-scale networks (Frankel *et al.* 2023). This is why heuristic algorithms (Li *et al.* 2021) and machine learning models (Sun *et al.* 2019; Li *et al.* 2023) have been used for the detection and source identification of contaminants or other water-quality issues in water networks. A significant challenge when regulating the concentration of a disinfectant in a WDN is the significant delay between the time that a

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

chemical is injected at the input node and the time that the concentration is measured at the monitoring output nodes. Moreover, the time delays are time-varying and uncertain due to varying pipe flows, which depend mainly on consumer demands that are an uncontrolled and typically unknown input to the system (Eliades *et al.* 2023). In addition, the disinfectant reaction dynamics are usually only partially known and the reaction parameters may vary due to the use of water originating from diverse sources or due to fluctuations in temperature (Monteiro *et al.* 2017).

One of the first studies addressing the challenge of large and varying time delays when controlling water quality is by Polycarpou *et al.* (2002), where the proposed input–output (I–O) model of disinfectant concentration is used to adaptively learn periodic parameters to characterize the time delay. The input–output approach can be thought of as a model-reduction technique that overcomes the challenge of using a small number of measurements compared to the system states, to design a water-quality control algorithm (Wang *et al.* 2006). More recently, Wang *et al.* (2021) derived a state-space model of chlorine concentration evolution in WDN, which traces over time all water-quality states; i.e., the concentration at all the links and nodes of the network. To do this, they solve the hydraulic equations to obtain the steady-state flow solution for a hydraulic-step, and then create the state-space equations, which are time-varying due to the varying flows. The drawback of this approach in practice is the lack of observability for estimating the complete state vector, and the need for using model-reduction techniques to calculate a control law. A simulation-based model is used in Xie & Brdys (2015) to compute hydraulics and water quality, while nonlinear model predictive control (MPC) within a hierarchical control framework is then used to control disinfectant levels. The authors in Xie *et al.* (2018) use a multi-input multi-output modeling approach, which considers also the formation of DBPs.

The aforementioned approaches, however, cannot guarantee that the concentration of a chemical will remain within a specified range at the output, even though bounds on water flows uncertainty, and, by extension, the water-quality input–output time-delay uncertainty, may be known. The quantification of water-quality state uncertainty is typically achieved through Monte Carlo Simulations (MCS) of hydraulic scenarios, due to the model complexity in large-scale networks (Hart *et al.* 2019). However, this approach is limited by computational constraints in large systems. Thus, researchers have looked into alternative methods, such as spectral propagation (Braun *et al.* 2020). Recently, an algorithmic approach has been proposed by the authors that guarantee the validity of calculated water-quality bounds, as opposed to MCS (Vrachimis *et al.* 2021). Given these developments, a water-quality regulation approach is needed that is not reliant on MCS to maintain water-quality states within predefined value sets.

In this work, we use the input–output modelling approach to develop a disinfectant injection methodology that guarantees a bounded output disinfectant concentration, given bounds on water-flow uncertainty. The proposed methodology creates an input–output model which considers time-delay uncertainty by utilizing recent results on estimating bounds on water-quality states using the backtracking approach (Vrachimis *et al.* 2021) Following the model predictive scheduling framework, an optimization problem is formulated and solved to find an input schedule that keeps the disinfectant concentration within predefined bounds for a specified time horizon. Simulation results illustrate the effectiveness of this approach in an example network case study.

2. PROBLEM FORMULATION

The following notation is followed throughout this work: matrices are denoted with capital letters, vectors with lower bold letters, and scalars with italic letters. Sets and graphs are denoted by calligraphic capital letters. Estimated parameters are denoted with a hat; i.e., $\hat{\boldsymbol{w}}$ denotes the estimated values of the parameter vector \boldsymbol{w} . Uncertain parameters are represented by a continuous interval of values defined by a lower and upper bound. Intervals are accompanied by a tilde as follows: $\tilde{\boldsymbol{v}} = [\boldsymbol{v}^l, \boldsymbol{v}^u]$, where \boldsymbol{v}^l is the lower bound vector and \boldsymbol{v}^u is the upper bound vector, such that: $\tilde{\boldsymbol{v}} = \{\boldsymbol{v} \in \mathbb{R}^n : v_i^l \leq v_i \leq v_i^u, \forall i = \{1, \dots, n\}\}$, and n is the size of the vector.

Consider a WDN whose topology is modeled by a directed graph denoted as $\mathcal{G} = (\mathcal{N}, \mathcal{L})$. Let $\mathcal{N} = \{1, \dots, n_n\}$ be the set of all nodes, where $|\mathcal{N}| = n_n$ is the total number of nodes. These represent junctions of pipes, consumer water demand locations, reservoirs, and tanks. Moreover, let $\mathcal{N}_t \subset \mathcal{N}$ represent the tank nodes, where $|\mathcal{N}_t| = n_t$ is the total number of tanks. Finally, let $\mathcal{L} = \{1, \dots, n_l\}$ be the set of links, where $|\mathcal{L}| = n_l$ is the total number of links. These represent network pipes, water pumps, and pipe valves.

The hydraulic state associated with a node j at each time instant $t \in \mathbb{R}$ is the *hydraulic head*, denoted by $h_j(t)$, $j \in \mathcal{N}$. The hydraulic state associated with a link $i \in \mathcal{L}$ is the *water flow*, denoted by $q_i(t)$. In general, the complete set of hydraulic states

can be expressed as $\mathbf{x}_H(t) = [\mathbf{h}^\top(t) \mathbf{q}^\top(t)]^\top$. Each node j is associated with a water consumer demand at the node location, denoted by $d_j(t)$. Water demands drive the hydraulic dynamics of a WDN and in this work are considered as an unknown and uncontrolled input to the system. It is noted that some demands $d_j(t)$ may be zero if the node is not associated with any consumers. Background leakages that may exist in the network are modeled as part of the unknown water demands. Pumps and valves are the main hydraulic actuators in a WDN and are modeled at the network links. Their control settings are indicated using the hydraulic input vector $\mathbf{u}_H(t)$.

Specific hydraulic states are measured at regular time intervals Δt_H using hydraulic sensors installed in the network. The interval Δt_H is in the range of minutes, and hydraulic analysis is typically performed in discrete time. The measured states may be flow, pressure, or tank level and are indicated by the hydraulic output vector $\mathbf{y}_H(k)$, where $k \in \mathbb{N}$ indicates the discrete hydraulic time-step associated with time intervals Δt_H .

Water quality refers to the concentration of various chemical substances and biological species in drinking water that may affect human health. In general, the concentration of these substances in the water should be within predefined bounds, defined by regulating authorities. Disinfectants are the only controllable substance in the water. The existence of a sufficient concentration level of disinfectant in the water indicates that any infectious substances have reacted with the disinfectant and have been neutralized. In this work, water-quality state refers to the concentration of a disinfectant in the water, specifically at the nodes of the network. The water-quality state vector at each time instant $t \in \mathbb{R}$ is denoted by $\mathbf{x}(t) \in \mathbb{R}^{n_s}$.

Water quality is regulated using chemical dosing pumps, designed to inject specific amounts of disinfectant into the water at a set of network nodes $\mathcal{N}_u \subset \mathcal{N}$, where $|\mathcal{N}_u| = n_u$ is the total number of controlled inputs. Water quality is monitored using sensors (e.g., chlorine sensors) installed at nodes $\mathcal{N}_s \subset \mathcal{N}$, where $|\mathcal{N}_s| = n_s$ is the total number of quality sensors. Control and measurement signals are transmitted at discrete time-steps Δt . For simplicity, in this work, we assume that Δt is equal to the hydraulic step, i.e., $\Delta t = \Delta t_H$; to ensure the stability of the solver, the water-quality time step must be sufficiently small compared to the water transport times in the network pipes (Rossman 2000). The water-quality control input vector is denoted by $\mathbf{u}(k) \in \mathbb{R}^{n_u}$, while the measurement vector of quality states is denoted by $\mathbf{y}(k) \in \mathbb{R}^{n_s}$.

An illustration of the inputs, outputs, and states of a water distribution network considering hydraulic and water-quality dynamics is given in Figure 1. In this work, we assume a single-input single-output (SISO) water-quality system, i.e., $n_u = n_s = 1$. The goal is, given a water-quality model and bounds on water flows, to choose the input $\mathbf{u}(k)$ such that the output $\mathbf{y}(k)$ stays within predefined output reference bounds $\tilde{r}(k) = [r^l(k), r^u(k)]$, such that $r^l(k) \leq y(k) \leq r^u(k), \forall k \geq 0$. Moreover, input constraints should be satisfied, such that $u^l \leq u(k) \leq u^u, \forall k \geq 0$.

3. STATE ESTIMATION METHODOLOGY

3.1. Hydraulic-state estimators

A *hydraulic model* is a set of equations that describe the hydraulic dynamics. These equations are formulated considering the conservation of energy and mass laws (Vrachimis *et al.* 2019). The relationship between the network topology \mathcal{G} and the physical parameters denoted by the vector $\boldsymbol{\theta}_H$ (pipe roughness, length, diameter, node elevations, tank dimensions) are also

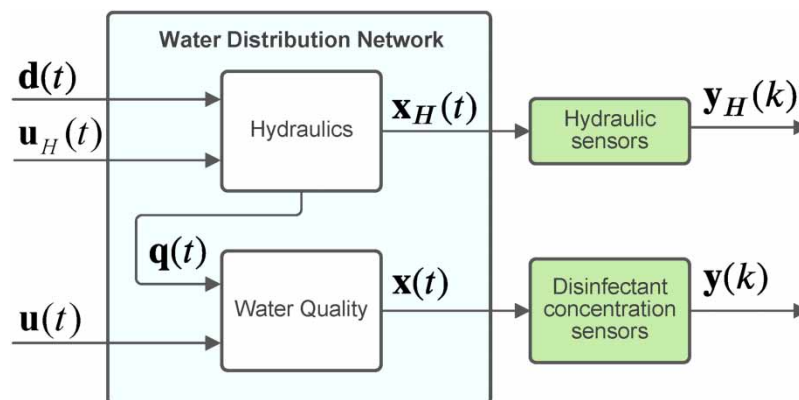


Figure 1 | Diagram illustrating the inputs, outputs, and states of a water distribution network.

described by the model. Hydraulic parameters are typically time-invariant for the time scales considered in this work. The hydraulic model used in this work is indicated by $\mathcal{M}_H(\mathcal{G}; \theta_H)$.

The installed hydraulic sensors measuring flow and pressure in the network typically do not satisfy hydraulic-state observability conditions (Díaz *et al.* 2017). This is because water demands are typically unknown and only an approximate pattern of consumption at nodes is available *a priori*. The approximate consumption patterns are used as pseudo-measurements to achieve hydraulic-state observability. Due to the high uncertainty of pseudo-measurements, an interval of possible demand values for each node is used at each time step instead, indicated by $\tilde{\mathbf{d}}(k)$. Moreover, the network's physical parameters are not exactly known. An estimate of these parameters is used instead, indicated by $\hat{\theta}_H$. The uncertainty of these parameters is considered bounded in this work, thus the range of values they may take is indicated by the interval vector $\tilde{\theta}_H$, with $\hat{\theta}_H \in \tilde{\theta}_H$.

Hydraulic-state estimation is enabled using hydraulic sensors in an observable configuration that sends measurements at a discrete time-step k . A hydraulic-state estimator (HSE) is a function that uses the hydraulic model $\mathcal{M}_H(\mathcal{G}; \hat{\theta}_H)$, the known hydraulic inputs $\mathbf{u}_H(k)$, demand pseudo-measurements $\tilde{\mathbf{d}}(k)$, as well as the current measurements $\mathbf{y}(k)$ to calculate the most suitable hydraulic-state estimate $\hat{\mathbf{x}}_H(k) \in \mathbb{R}^{n+n_t}$ as follows:

$$\hat{\mathbf{x}}_H(k) = f_{\text{HSE}}(\mathbf{y}_H(k), \mathbf{u}_H(k), \tilde{\mathbf{d}}(k); \mathcal{M}_H(\mathcal{G}; \hat{\theta}_H)) \quad (1)$$

In this work, the bounds on hydraulic states are calculated using a *hydraulic-state interval estimator* (HSIE) (Vrachimis *et al.* 2019). This is a function that uses the hydraulic model $\mathcal{M}_H(\mathcal{G}; \theta_H)$ containing parametric uncertainty $\tilde{\theta}_H$, and demand uncertainty $\tilde{\mathbf{d}}(k)$, to calculate an interval of values for the hydraulic state as follows:

$$\tilde{\mathbf{x}}_H(k) = f_{\text{HSIE}}(\mathbf{u}_H(k), \tilde{\mathbf{d}}(k); \mathcal{M}_H(\mathcal{G}; \tilde{\theta}_H)) \quad (2)$$

3.2. Water-quality estimator considering flow uncertainty

A water-quality model describes the dynamics of species transport and reaction in the water. Transport refers to the movement of substances in water due to the water flow in pipes. Substances can be tracked throughout the network when water flows are known for a number of time-steps, referred to as the *memory* of the system m . This is defined as the maximum number of discrete time-steps needed for any input node to affect any output and can be computed given the minimum flow rate of pipes during the system operation.

Many substances, such as various disinfectants, react with other substances in the water and materials attached to pipe walls. Thus, reaction dynamics need to be taken into account when estimating their concentration in the water. Reaction rates are typically unknown due to the multiple species existing in the water and the different characteristics of the water inserted in the network. We assume the availability of bounds on the uncertain estimates of disinfectant reaction rates in the water and with pipe walls. The range of disinfectant reaction rates is indicated by the interval vector $\tilde{\theta}$. Finally, the water-quality model is indicated by $\mathcal{M}(\mathcal{G}; \tilde{\theta}, m)$.

In this work, we assume the availability of a disinfectant concentration bounding estimator, such as the backtracking uncertainty bounding algorithm (BUBA) described in Vrachimis *et al.* (2021), which provides bounds on disinfectant concentration at monitored sensor nodes. For convenience, we denote the set of previous time-steps included in the memory of the system as $k \in k_m = \{k, k-1, \dots, k-m\}$. The BUBA calculates disinfectant concentration bounds $\tilde{y}(k)$, given the vectors of past inputs $\mathbf{u}(k_m) = [\mathbf{u}(k), \mathbf{u}(k-1), \dots, \mathbf{u}(k-m)]^\top$ and the vectors of uncertain water flows $\tilde{Q}(k_m) = [\tilde{\mathbf{q}}(k), \tilde{\mathbf{q}}(k-1), \dots, \tilde{\mathbf{q}}(k-m)]$, as follows:

$$\tilde{y}(k) = f_{\text{BUBA}}(\mathbf{u}(k_m), \tilde{Q}(k_m); \mathcal{M}(\mathcal{G}; \tilde{\theta}, m)) \quad (3)$$

4. DISINFECTION SCHEDULING METHODOLOGY

4.1. Past input effect calculation considering flow uncertainty

The first part of the proposed methodology considers the effect of inputs $u(k)$ that have already been applied to the system. An illustration of the calculation of past input effects is given in Figure 2. The estimation of the effect of past inputs on the output

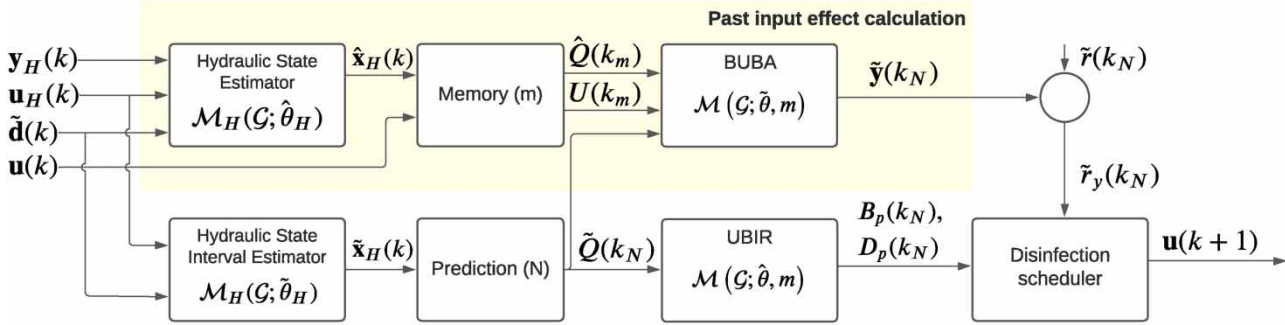


Figure 2 | Description of the proposed disinfection scheduling methodology.

is done by utilizing available hydraulic measurements from the system and using the hydraulic state estimator of (1). The flow estimates for all the network links for the memory of the system m are combined into the matrix $\hat{Q}(k_m)$.

Let us define the prediction horizon $N \in \mathbb{N}^+$ for which we want to estimate the output $y(k) \forall k \in k_N, k_N = \{k + 1, \dots, k + N\}$. The horizon of interest is typically chosen to be 24 h, to capture the periodicity of the system dynamics inherited by the periodicity of water demands. Intervals of the uncertain future flow of the system are calculated using (2). The predicted interval of output values $\tilde{y}(k_N)$ due to the past inputs $u(k_m)$, and considering that future inputs $u(k_N) = 0$, is then calculated using the BUBA algorithm of (3). The output due to past inputs $\tilde{y}(k_N)$ is used to adjust, using interval arithmetic (Moore *et al.* 2009), the reference bounds $\tilde{r}(k_N)$ as follows:

$$\tilde{r}_y(k_N) = \tilde{r}(k_N) - \tilde{y}(k_N) \quad (4)$$

4.2. Input-output model of water quality

Let the Input-Output (I-O) water-quality model for a SISO system, as described in Polycarpou *et al.* (2002), be written as follows:

$$y(k) = \sum_{i=1}^{n_t} \alpha_i(k)y(k-i) + \sum_{l=1}^m \beta_l(k)u(k-l) + \eta(k) \quad (5)$$

where $y(k)$ is the disinfectant concentration at a monitored node at sampling instant k , $u(k)$ is the forced disinfectant concentration at the input node, n_t is the number of tanks in the water network, m is the maximum water transport delay from input to output node, $\eta(k)$ is the modeling and time-discretization uncertainty. Parameters $\alpha_i(k)$ and $\beta_l(k)$ are time-varying coefficients corresponding to $y(k-i)$ and $u(k-l)$, respectively, and depend on the hydraulic dynamics of the network as well as the disinfectant decay rates.

For simplicity, in this work, we assume a network without water tanks, thus $\alpha_i = 0$; this assumption is typically applicable to district metered areas (DMAs) of many water systems. Note that, tanks may be used as chlorine injection locations, since they are easily accessible locations in the network. Moreover, we assume the availability of the hydraulic estimator of (1). Given the network flows, the backtracking approach (Vrachimis *et al.* 2021) can be utilized to calculate the set of delays \mathcal{D} by which the input affects the output. The following relationship then describes the input-output water-quality model of a WDN without tanks:

$$y(k) = \sum_{l \in \mathcal{D}(k)} \beta_l(k)u(k-l) + \eta(k) \quad (6)$$

4.3. Impulse response

A simulation-based approach to calculate the coefficients of the I-O model of (6) is the Impulse Response (IR) method, that has been used in WDN for modeling and detecting leaks (Kim 2005; Gao *et al.* 2006). The IR method calculates, using a

simulation model of the hydraulics and/or water-quality dynamics of the water system, the output $y(k)$ when $u(k-l) = 1$ and $u(k-\tau) = 0$, $\tau \neq l$ for $l, \tau \in k_m$. The I-O model coefficient $\beta_l(k)$ is then equal to the calculated output $y(k)$. The complete I-O model of (6) is derived by calculating the coefficients $\beta_l(k)$ for all $l \in \mathcal{D}_u(k)$. The I-O model can be used to estimate future values of the output as follows:

$$\mathbf{y}(k_N) = M(k)\mathbf{u}(k_N) \quad (7)$$

where $\mathbf{y}(k_N) = [y(k+1), y(k+2), \dots, y(k+N)]^\top$, $\mathbf{u}(k_N) = [u(k+1), u(k+2), \dots, u(k+N)]^\top$ and $M(k)$ is defined as the Impulse Response matrix for the prediction horizon N :

$$M(k) = \begin{bmatrix} \beta_{k+1}(k) & \beta_{k+2}(k) & \dots & \beta_{k+N}(k) \\ \beta_{k+1}(k+1) & \beta_{k+2}(k+1) & \dots & \beta_{k+N}(k+1) \\ \dots & \dots & \dots & \dots \\ \beta_{k+1}(k+N) & \beta_{k+2}(k+N) & \dots & \beta_{k+N}(k+N) \end{bmatrix} \quad (8)$$

An alternative method to calculate the IR matrix is by using the backtracking approach, where a water parcel is backtracked through the network to find the input source (Zierolf *et al.* 1998). Using this approach, the set of paths $\mathcal{P}(k)$ which bring water into the output node at time-step k can also be identified. Path $p \in \mathcal{P}(k)$ is defined as the set of links that the water parcel, currently arriving at the output, has traversed since it left an input node. Each identified path can be regarded as a separate simulation scenario and can be used to calculate an impulse response, using the same logic as in (8). The impulse response matrix for each path is referred to as a *path-IR matrix* $M_p(k)$. Assuming complete mixing of chemicals in the water, the superposition of the path-IR matrices results in the IR matrix $M(k)$ of (8):

$$M(k) = \sum_{p \in \mathcal{P}(k)} M_p(k) \quad (9)$$

4.4. Impulse response with flow uncertainty

Let us consider the availability of uncertain hydraulic-state estimates represented by an interval, as in (2). We propose an algorithm that uses as input the interval-flow estimates to calculate the set of delays $\mathcal{D}_p(k)$, $\forall p \in \mathcal{P}(k)$, which indicate previous inputs by which the current output $y(k)$ depends on. This is a modification of the BUBA, which we will refer to as the *uncertain backtracking impulse response* (UBIR) algorithm. The algorithm also provides the set of coefficients $B_p(k)$ associated with each delay in $\mathcal{D}_p(k)$:

$$[\mathcal{D}_p(k), B_p(k)] = f_{\text{UBIR}}(\tilde{Q}(k_N); \mathcal{M}(\mathcal{G}; \hat{\theta}, \mathbf{m})), \quad \forall p \in \mathcal{P}(k) \quad (10)$$

The elements of an uncertain path-IR matrix for a path p are then given by:

$$\tilde{M}_p(k, i) = \beta_p(k), \quad \beta_p(k) \in B_p(k), \quad i \in \mathcal{D}_p(k) \quad (11)$$

The overall uncertain IR matrix is given by the superposition of the uncertain path-IR matrices:

$$\tilde{M}(k) = \sum_{p \in \mathcal{P}(k)} \tilde{M}_p(k) \quad (12)$$

The uncertain IR matrix $\tilde{M}(k)$ is essentially a superposition of all the IR matrices $M(k)$, given the considered input time-delay uncertainty. Note that this matrix cannot be used as is for the estimation of the output.

4.5. Optimization formulation for disinfection scheduling under flow uncertainty

An optimization formulation is proposed to calculate the control input $u(k)$ for a prediction horizon $k \in k_N$ using the uncertain IR matrix $\tilde{M}(k)$. The objective is to find an input that minimizes the deviation of expected output disinfectant concentration $\hat{y}(k)$ from a reference concentration $r(k)$. For simplicity, we assume that the past inputs of the system be

zero, thus the reference concentration $r(k) = r_y(k)$; note that this assumption is not limiting because any initial disinfectant concentration decays after a period, specific to each water network, thus only the controlled injection of disinfectants affects the water-quality state. Moreover, the rapid changes of disinfectant injection $\Delta u(k) \triangleq u(k) - u(k - 1)$ should be minimized. The objectives should be achieved under the following constraints: (a) The output $y(k)$ should not violate the defined lower bound $r^l(k)$ for all possible scenarios given uncertainty bounds on water demands, and (b) The control input should reside within a predefined interval given by $u^l \leq u(k) \leq u^u$.

The optimization formulation for the calculation of the control inputs $u(k)$ for the prediction horizon $k \in k_N$ will then be:

$$u(k_N) = \begin{cases} \arg \min_U & \|\hat{Y} - R\|_2^2 + \|\Delta U\|_2^2 \\ \text{s.t.} & \hat{Y}_k = \sum_{p \in \mathcal{P}} [\tilde{M}_{kj}^p U_j], \quad \forall j, k \in \{1, \dots, N\} \\ & r^l(k) \leq Y_k, \quad \forall k \\ & \mathbf{1}_{n_u} u^l \leq U \leq \mathbf{1}_{n_u} u^u \end{cases} \quad (13)$$

where $\hat{Y} = \hat{y}(k_N)$ is calculated using (7), $R = r(k_N)$, $U = u(k_N)$, $\hat{Y}_k = \hat{y}(k)$, and $\mathbf{1}_{n_u} \in \mathbb{N}^{n_u}$ is a vector of ones. The formulation in (13) is a Quadratic Programming problem with linear constraints that can be efficiently solved by available solvers (Gurobi 2023).

The first constraint is the key to addressing the uncertain time-delay issue. Essentially it represents the physics of the system using the I–O model, i.e., the impulse response matrix. However, using the uncertain IR matrix \tilde{M} in the formulation, the physics of multiple models are considered simultaneously. The use of path-IR matrices \tilde{M}_p in the formulation is essential, in order to distinguish between models that represent a different reality, and models that represent different paths in the same reality.

5. CASE STUDIES

5.1. Illustrative example

The example network of Figure 3 is used to illustrate the proposed disinfection scheduling methodology. Chlorine is injected at the input node 1 and is measured at the output node 3. The network has two flow paths from input to output: path 1 through links {1, 2} and path 2 through links {1, 3, 4, 5, 6}. Node demands follow a diurnal pattern, similar to the observed consumption patterns of a typical DMA in a city. Note that, in this network, the flow of link 6 may change direction depending on the node demands. The network is simulated for the duration of 24 h, with a time-step corresponding to $\Delta t = 5$ min.

We compare the proposed approach with a scheduling algorithm that uses the nominal model of the network and does not consider flow uncertainty (Munavalli & Kumar 2003). The implementation of the ‘nominal’ scheduling algorithm in this work considers a single-input single-output system. Essentially, it uses the formulation of (13), with the first constraint substituted by an IR model, as in (7), that is derived using the nominal hydraulic model of the network.

Both the proposed and the nominal scheduling algorithms aim to regulate the output concentration to follow, as closely as possible, the lower bound reference signal, i.e., $r(k) = r^l$. The lower bound is equal to zero for the first 100 time-steps, in order to account for the delay between input and output and to demonstrate the performance of the proposed methodology during abrupt changes of the reference signal. Using the EPANET-MATLAB Toolkit (Eliades et al. 2016), Monte-Carlo simulations

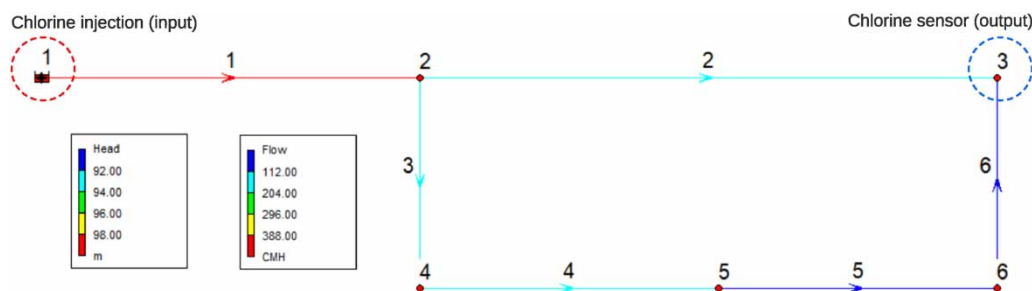


Figure 3 | Case study network for illustration of the disinfection scheduling methodology: it consists of one chlorine injection point (input) and one chlorine sensor (output). There are two flow paths from input to output. Head at nodes is in meters (m). Flow in pipes is in cubic meters per hour (CMH).

are performed in which the node demands are varied between $\pm 20\%$ from their nominal values, generating 100 different flow scenarios. The output is monitored and plotted when the control input $\bar{u}(k_N)$ calculated by the nominal scheduler is applied to all scenarios (Figure 4(a)), and when the control input $u(k_N)$ from the proposed scheduler is applied (Figure 4(b)).

The results of Figure 4 illustrate the benefits of considering time-delay uncertainty when calculating chlorine injection inputs. When the nominal model is used (Figure 4(a)), we observe the following deficiencies in a number of uncertainty scenarios compared to when the proposed scheduling approach is used (Figure 4(b)):

- $k = 100$: (a) there are lower bound violations during abrupt changes; (b) the proposed approach predicts the worst-case delay and avoids violations.
- $k = [100, 288]$: (a) there are lower bound violations because of the effect of varying time-delay on chlorine decay; (b) there are small violations when using the proposed approach due to the proposed approach not considering modeling and discretization uncertainty $\eta(k)$.
- $k = 220$: (a) There is an overshoot of chlorine concentration in some scenarios due to the effect of path 2 starting earlier in some scenarios; (b) The same effect is observed here since there is no violation of the lower bound.
- $k = 240$: (a) There is a significant drop in chlorine concentration and lower bound violation due to flow reversal occurring at link 6 in some scenarios; (b) In contrast, the proposed approach anticipates the flow reversal and keeps the chlorine concentration above the threshold in all considered scenarios.

5.2. DMA example

The scalability of the proposed approach is demonstrated using the ‘CY-DMA’ network (illustrated in Figure 5), a DMA of a real network on Cyprus (Vrachimis *et al.* 2020), with 120 pipes, 90 junctions, and 1 reservoir which serves as network water source. The disinfectant input of this network is administered at the inlet node $R - 1$, while the sensor is positioned at the far end of the network, at node 80. Node demands follow a diurnal pattern. The network has multiple flow paths from input to output, with some links changing direction during simulation. The network is simulated for the duration of 24 h, with a time-step corresponding to $\Delta t = 5$ min.

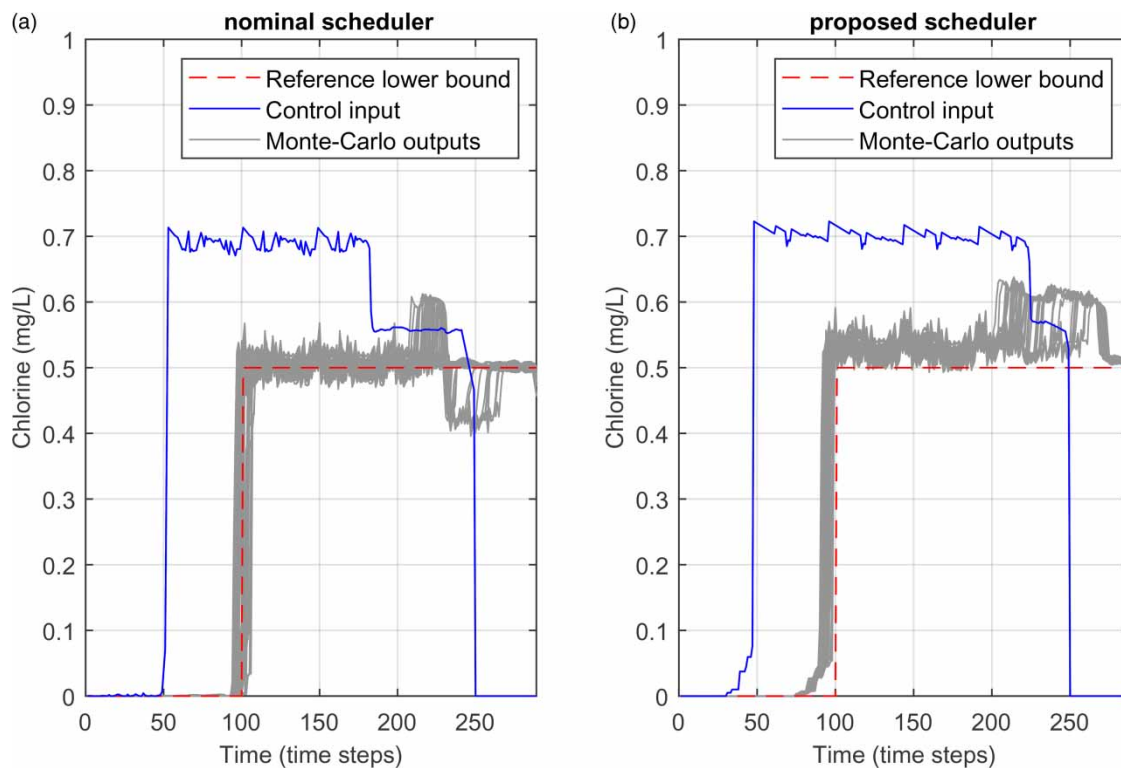


Figure 4 | Comparison between a nominal scheduler (a) and the proposed scheduler (b) when applied to 100 Monte-Carlo scenarios in which the water demands of the example network are randomized between $\pm 20\%$ of their nominal values.

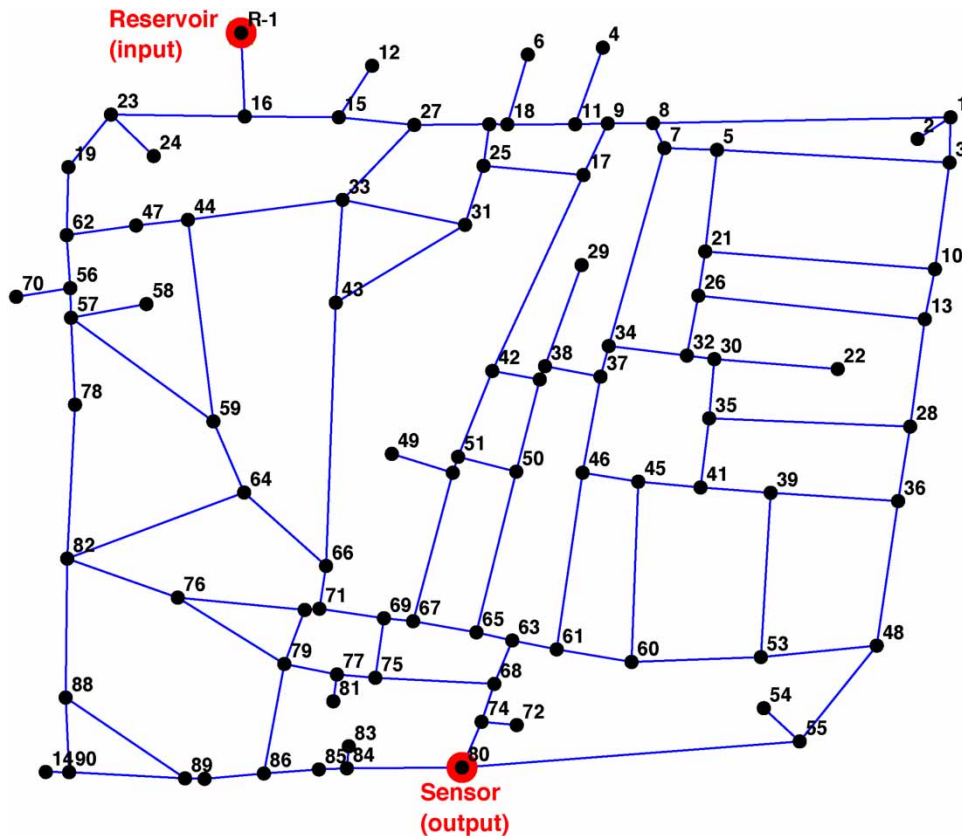


Figure 5 | Case study network 'CY-DMA' representing a district metered area with one chlorine injection point (input) and one chlorine sensor (output).

The results when the proposed approach is applied to 'CY-DMA' are shown in Figure 6. It can be observed that when the nominal scheduler is used (Figure 6(a)), there is a lower bound violation in several scenarios with varying flows, in contrast to when the proposed scheduler is employed (Figure 6(b)). On the other hand, the scheduler appears to be conservative, regulating the chlorine concentration above the target lower bound. This conservatism arises from the use of the modified uncertainty bounding backtracking algorithm (Vrachimis *et al.* 2021), which considers all flow scenarios generated by uncertain demands, not just the 100 Monte-Carlo scenarios used for testing in this work. Interestingly, this conservatism is not evident in the case study of the example network in Section 5.1, where the impact of varying time-delay on chlorine decay appears to be more prevalent than in the DMA network case study.

6. CONCLUSIONS, LIMITATIONS, AND FUTURE WORK

In this work, we have demonstrated a new disinfection scheduling methodology, which considers the time-delay uncertainty when regulating the disinfectant concentration at specific locations of a water distribution network. The results indicate that this approach is able to avoid violations of predefined bounds on disinfectant concentration and regulates the output to follow closely the reference signal, even when the modeling uncertainty on water demands is up to 20%. The use of a modified version of the uncertainty bounding backtracking algorithm proposed in Vrachimis *et al.* (2021) ensures the consideration of all different flow scenarios when calculating the scheduling input, as opposed to using a Monte-Carlo simulation approach that only samples a subset of potential scenarios.

While this work addresses the challenge associated with time-delay uncertainty when regulating disinfectant concentration, it is important to acknowledge certain limitations. Firstly, the methodology focuses on single-input single-output water-quality systems, thus limiting its applicability to networks with multiple disinfectant booster stations and multiple sensors. Additionally, the approach assumes networks without water tanks or those where disinfectant concentrations in tanks are known.

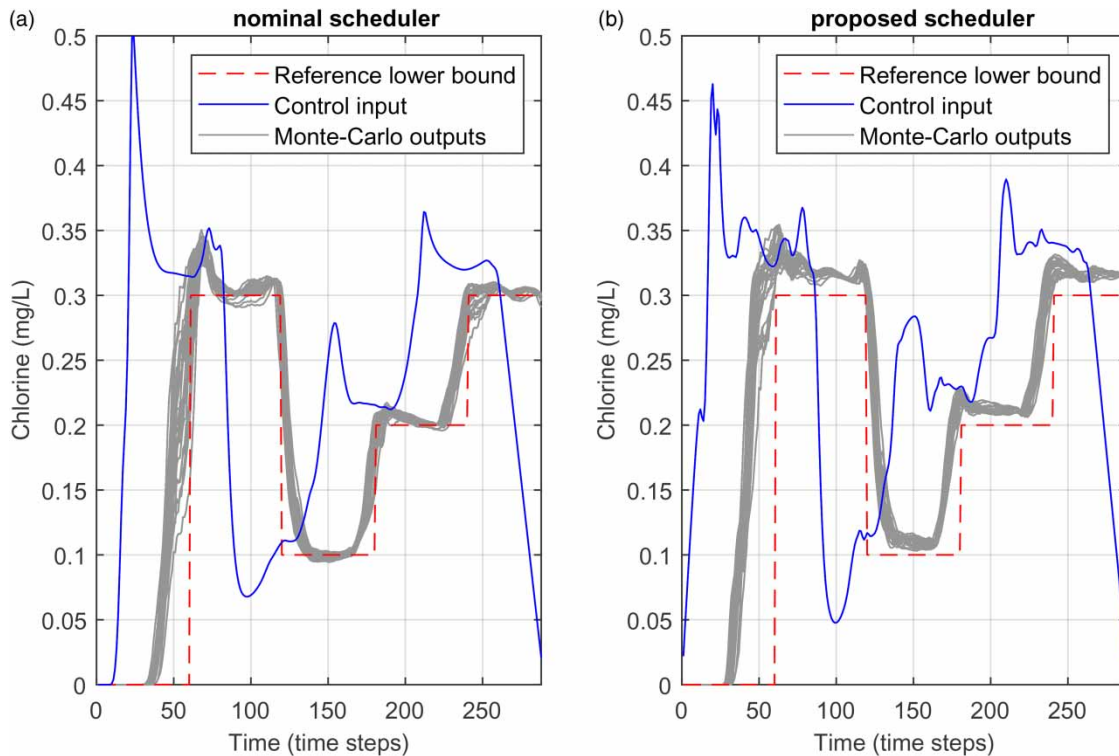


Figure 6 | Comparison between a nominal scheduler (a) and the proposed scheduler (b) when applied to 100 Monte-Carlo scenarios in which the water demands of the 'CY-DMA' network are randomized between $\pm 20\%$ of their nominal values.

Another limitation is the exclusion of uncertainty in decay rates, as our model primarily accounts for time-delay uncertainty arising from fluctuating water flows. Finally, as indicated in the case studies, the proposed scheduling algorithm may exhibit a degree of conservatism in disinfectant concentration regulation, depending on the given uncertainty of water flows.

Future work will develop this methodology to consider multiple inputs and outputs in the network and study its performance and computational efficiency when applied to large-scale water networks. Moreover, a natural extension of this work is the consideration of disinfectant measurements in a feedback loop, to improve the water-quality model and provide better estimations, thus also improving the disinfectant regulation. Finally, the effect of modeling and discretization uncertainty on the performance of the proposed methodology will be analyzed.

ACKNOWLEDGEMENTS

This work was funded by the European Research Council (ERC) under the ERC Synergy Grant 'Water-Futures' (Grant agreement No. 951424) and by the European Union Horizon 2020 programme Grant Agreement No. 739551 (KIOS CoE), and the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy.

DATA AVAILABILITY STATEMENT

The models, code and data generated during this study are available in <https://doi.org/10.5281/zenodo.10226047>.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Braun, M., Piller, O., Deuerlein, J., Mortazavi, I. & Iollo, A. 2020 Uncertainty quantification of water age in water supply systems by use of spectral propagation. *Journal of Hydroinformatics* **22**, 111–120.

- Díaz, S., Mínguez, R. & González, J. 2017 Topological observability analysis in water distribution systems. *Journal of Water Resources Planning and Management* **143**, 06017001.
- Eliades, D. G., Kyriakou, M., Vrachimis, S. & Polycarpou, M. M. 2016 EPANET-MATLAB toolkit: An open-source software for interfacing EPANET with MATLAB. In: *Proceedings of the 14th International Conference on Computing and Control for the Water Industry (CCWI)*. The Netherlands, p. 8.
- Eliades, D. G., Vrachimis, S. G., Moghaddam, A., Tzortzis, I. & Polycarpou, M. M. 2023 Contamination event diagnosis in drinking water networks: A review. *Annual Reviews in Control* **55**, 420–441.
- Elsharif, S. M., Wang, S., Taha, A. F., Sela, L., Giacomoni, M. H. & Abokifa, A. A. 2022 Control-theoretic modeling of multi-species water quality dynamics in drinking water networks: Survey, methods, and test cases. *Annual Reviews in Control*.
- European Union 2020 Directive (EU) 2020/2184 of the European Parliament and of the Council of 16 December 2020 on the quality of water intended for human consumption (recast).
- Frankel, M., Katz, L. E., Kinney, K., Werth, C. J., Zigler, C. & Sela, L. 2023 A framework for assessing uncertainty of drinking water quality in distribution networks with application to monochloramine decay. *Journal of Cleaner Production* **407**, 137056.
- Gao, Y., Brennan, M. J. & Joseph, P. F. 2006 A comparison of time delay estimators for the detection of leak noise signals in plastic water distribution pipes. *Journal of Sound and Vibration* **292**, 552–570.
- Gurobi Optimization, LLC 2023 Gurobi optimizer reference manual.
- Hart, D., Rodriguez, J. S., Burkhardt, J., Borchers, B., Laird, C., Murray, R., Klise, K. & Haxton, T. 2019 Quantifying hydraulic and water quality uncertainty to inform sampling of drinking water distribution systems. *Journal of Water Resources Planning and Management* **145**, 04018084.
- Kim, S. H. 2005 Extensive development of leak detection algorithm by impulse response method. *Journal of Hydraulic Engineering* **131**, 201–208.
- Li, C., Yang, R., Zhou, L., Zeng, S., Mavrouniotis, M., Yang, M., Yang, S. & Wu, M. 2021 Adaptive multipopulation evolutionary algorithm for contamination source identification in water distribution systems. *Journal of Water Resources Planning and Management* **147**, 04021014.
- Li, Z., Liu, H., Zhang, C. & Fu, G. 2023 Generative adversarial networks for detecting contamination events in water distribution systems using multi-parameter, multi-site water quality monitoring. *Environmental Science and Ecotechnology* **14**, 100231.
- Monteiro, L., Figueiredo, D., Covas, D. & Menaia, J. 2017 Integrating water temperature in chlorine decay modelling: A case study. *Urban Water Journal* **14**, 1097–1101.
- Moore, R. E., Kearfott, B. R. & Cloud, M. J. 2009 *Introduction to Interval Analysis*. SIAM.
- Mouly, D., Joulín, E., Rosin, C., Beaudeau, P., Zeghnoun, A., Olszewski-Ortar, A., Munoz, J. F., Welté, B., Joyeux, M., Seux, R., Montiel, A. & Rodriguez, M. J. 2010 Variations in trihalomethane levels in three French water distribution systems and the development of a predictive model. *Water Research* **44**, 5168–5179.
- Munavalli, G. R. & Kumar, M. S. M. 2003 Optimal scheduling of multiple chlorine sources in water distribution systems. *Journal of Water Resources Planning and Management* **129**, 493–504.
- Polycarpou, M. M., Uber, J. G., Wang, Z., Shang, F. & Brdys, M. 2002 Feedback control of water quality. *IEEE Control Systems* **22**, 68–87.
- Rossman, L. A. 2000 EPANET 2: Users manual.
- Sun, L., Yan, H., Xin, K. & Tao, T. 2019 Contamination source identification in water distribution networks using convolutional neural network. *Environmental Science and Pollution Research* **26**, 36786–36797.
- Vrachimis, S. G., Timotheou, S., Eliades, D. G. & Polycarpou, M. M. 2019 Iterative hydraulic interval state estimation for water distribution networks. *Journal of Water Resources Planning and Management* **145**, 04018087.
- Vrachimis, S. G., Lifshitz, R., Eliades, D. G., Polycarpou, M. M. & Ostfeld, A. 2020 Active contamination detection in water-distribution systems. *Journal of Water Resources Planning and Management* **146**, 04020014.
- Vrachimis, S. G., Eliades, D. G. & Polycarpou, M. M. 2021 Calculating chlorine concentration bounds in water distribution networks: A backtracking uncertainty bounding approach. *Water Resources Research* **57**, e2020WR028684.
- Wang, Z., Polycarpou, M. M., Uber, J. G. & Shang, F. 2006 Adaptive control of water quality in water distribution networks. *IEEE Transactions on Control Systems Technology* **14**, 149–156.
- Wang, S., Taha, A. F. & Abokifa, A. A. 2021 How effective is model predictive control in real-time water quality regulation? State-space modeling and scalable control. *Water Resources Research* **57**, e2020WR027771.
- WHO 2017 *Guidelines for Drinking-Water Quality, 4th edn, Incorporating the 1st Addendum*. WHO.
- Xie, M. & Brdys, M. 2015 Nonlinear model predictive control of water quality in drinking water distribution systems with DBPs objectives. *International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering* **9**, 354–360.
- Xie, M., Wang, P., Zhang, X.-P. & Jayaweera, D. 2018 Robust parameter estimation and output prediction for nonlinear water quality control in water distribution systems. *Journal of Water Resources Planning and Management* **144**, 04017092.
- Zierolf, M. L., Polycarpou, M. M. & Uber, J. G. 1998 Development and autocalibration of an input–output model of chlorine transport in drinking water distribution systems. *IEEE Transactions on Control Systems Technology* **6**, 543–553.

First received 25 April 2023; accepted in revised form 23 December 2023. Available online 17 January 2024