



Bayesian Optimization of Booster Disinfection Scheduling in Water Distribution Networks[☆]

Mohammadreza Moeini^a, Lina Sela^b, Ahmad F. Taha^c, Ahmed A. Abokifa^{d,*}

^a Ph.D. Student; Department of Civil, Materials, and Environmental Engineering; The University of Illinois Chicago; Chicago, IL 60607, USA

^b Associate Professor; Department of Civil, Architectural, and Environmental Engineering; The University of Texas at Austin; Austin TX 78712, USA

^c Associate Professor; Department of Civil and Environmental Engineering; Vanderbilt University; Nashville, TN 37235, USA

^d Assistant Professor; Department of Civil, Materials, and Environmental Engineering; The University of Illinois Chicago; Chicago, IL 60607, USA

ARTICLE INFO

Keywords:

Water distribution networks
chlorine residual
booster chlorination systems
Bayesian optimization

ABSTRACT

Chlorine remains the most widely used disinfectant in drinking water treatment and distribution systems worldwide. To maintain a minimum residual throughout the distribution network, chlorine dosage needs to be regulated by optimizing the locations of chlorine boosters and their scheduling (i.e., chlorine injection rates). Such optimization can be computationally expensive since it requires numerous evaluations of water quality (WQ) simulation models. In recent years, Bayesian optimization (BO) has garnered considerable attention due to its efficiency in optimizing black-box functions in a wide range of applications. This study presents the first attempt to implement BO for the optimization of WQ in water distribution networks. The developed python-based framework couples BO with EPANET-MSX to optimize the scheduling of chlorine sources, while ensuring the delivery of water that satisfies water quality standards. Using Gaussian process regression to build the BO surrogate model, a comprehensive analysis was conducted to evaluate the performance of different BO methods. To that end, systematic testing of different acquisition functions, including the probability of improvement, expected improvement, upper confidence bound, and entropy search, in conjunction with different covariance kernels, including Matérn, squared-exponential, gamma-exponential, and rational quadratic, was conducted. Additionally, a thorough sensitivity analysis was performed to understand the influence of different BO parameters, including the number of initial points, covariance kernel length scale, and the level of exploration vs exploitation. The results revealed substantial variability in the performance of different BO methods and showed that the choice of the acquisition function has a more profound influence on the performance of BO than the covariance kernel.

1. Introduction

Deterioration of the water quality (WQ) in drinking water distribution networks (WDNs) is a significant challenge facing the supply of clean and safe water worldwide (Liu et al., 2017; Makris et al., 2014). WQ deterioration occurs through various physical, biological, and chemical processes that take place in the bulk flow and/or at the walls of the WDN pipes (Abokifa et al., 2016a; Biswas et al., 1993; Lu et al., 1995). Chlorine-based disinfectants are widely applied during water treatment to eliminate microbiological contaminants (Deborde and von Gunten, 2008), and a sufficient residual is typically maintained throughout the distribution system to prevent microbial

recontamination (LeChevallier, 1999). However, elevated chlorine doses have been associated with the excessive formation of harmful disinfection byproducts (DBPs) (Li et al., 2019), as well as taste and odor complaints from consumers (Fisher et al., 2011). This is particularly an issue for large WDNs, where chlorine concentrations may become completely depleted by the time the water reaches the far ends of the system. Alternatively, injecting chlorine in smaller doses at multiple distributed locations in the WDN (i.e., booster stations) has been shown to produce a more even distribution of chlorine concentrations while reducing the overall cost and amount of injected disinfectant (Abokifa et al., 2019; Boccelli et al., 1998; Maheshwari et al., 2018; Ohar and Ostfeld, 2014; Ostfeld and Salomons, 2006; Prasad et al., 2004; Tryby

[☆] Submitted to: Water Research May 19, 2023.

* To whom correspondence should be addressed: Dr. Ahmed Abokifa; Phone +1(312)-413-4636

E-mail address: abokifa@uic.edu (A.A. Abokifa).

<https://doi.org/10.1016/j.watres.2023.120117>

Received 12 October 2022; Received in revised form 19 May 2023; Accepted 22 May 2023

Available online 23 May 2023

0043-1354/© 2023 Elsevier Ltd. All rights reserved.

et al., 1999).

Optimizing the design of booster chlorination systems has been attempted by several studies that aimed to optimize the placement (i.e., locations) and/or the scheduling (i.e., injection rates) of chlorine dosing stations. In these studies, a wide range of optimization methods, including both linear and nonlinear optimization techniques were implemented (Islam et al., 2013; Mala-Jetmarova et al., 2017). To enable the formulation of the booster optimization problem as a linear programming problem, many studies relied on first-order decay kinetics to describe chlorine decay in the WDN. For instance, Boccelli et al. (1998) formulated a linear programming (LP) approach to minimize the total chlorine dose by applying the principle of linear superposition to disinfectant concentrations resulting from multiple injections over time, which was then solved using the simplex algorithm. Tryby et al., (2002) expressed the booster optimization problem as a mixed-integer linear programming (MILP) model by using both binary location variables and continuous injection rates. Propato and Uber (2004) formulated a linear least-squares (LLS) problem to determine the optimal injection rates that minimize space-time variations in residual concentrations. Lansey et al. (2007) developed a two-stage approach for optimizing booster locations followed by minimizing the injected chlorine mass using a combination of LP and a genetic algorithm (GA) or enumeration. Goyal and Patel (2017) applied an LP-based approach in Excel coupled with EPANET to optimize booster locations and scheduling. More recent works have focused on formulating WQ modeling within a linear state-space representation in which the relationship between network inputs (booster injection) and outputs (chlorine concentration at junctions) is explicitly described (Wang et al., 2022, 2021). Such control-oriented modeling enabled the implementation of scalable Model Predictive Control (MPC) algorithms for WQ control in WDNs (Wang et al., 2021).

Aside from first-order chlorine decay kinetics, a few studies also considered nonlinear reaction kinetics for the booster optimization problem. For instance, Munavalli and Kumar (2003) formulated the optimal scheduling of chlorine sources as a nonlinear optimization problem by implementing non-first-order bulk and wall reactions. The model was then solved using GA. Several studies have also formulated the WQ control problem as a multi-objective optimization problem. Prasad et al. (2004) formulated a multi-objective optimization model to minimize the total chlorine dose while maximizing volumetric demand that satisfies WQ standards, which was then solved using a multi-objective GA (MOGA). Behzadian et al. (2012) applied non-dominated sorting genetic algorithm II (NSGA-II) to optimize booster locations and scheduling using a two-phase multi-objective optimization process to optimize both chlorine residuals and DBP levels.

In addition to optimizing chlorine injection locations and scheduling, hydraulic controls have also been considered as decision variables in WQ optimization formulations. Ostfeld and Salomons, (2006) formulated a model that couples the design and operation of booster chlorination stations with the scheduling of pumping units. The model was optimized using GA with the objective of either minimizing the total costs of pumping and booster chlorination or maximizing the injected chlorine dose. Gibbs et al. (2010) applied GA to optimize both booster disinfection dosing together with daily pump scheduling for a real-life WDN. They showed that considering the hydraulics as well as the dosing regime in the optimization process can help maintain disinfectant residuals at the extremities of the network while achieving significant energy cost savings. Similarly, Kang and Lansey (2010) implemented GA to simultaneously optimize valve operation and booster scheduling, while Nono et al. (2018) used NSGA-II to integrate booster chlorination scheduling within network operational interventions to reduce the water age.

The majority of previous studies focused on optimizing the design of booster chlorination systems using either linear programming or evolutionary optimization techniques (e.g., GA). Although linear approaches are faster and simpler (Prasad et al., 2004), their application is generally limited to first-order decay reactions. On the other hand,

evolutionary optimization methods are significantly more computationally expensive since they typically involve conducting numerous evaluations of the objective function(s). The latter usually includes running a numerical model (e.g., EPANET/EPANET-MSX), in which the partial differential equation(s) governing the transport and decay of chlorine, and potentially other species, in the distribution system are numerically solved (Rossman and Boulos, 1996; Rossman et al., 1994; Shang et al., 2008). The high computational cost involved in evaluating such numerical models prohibits the real-time implementation of WQ control algorithms.

Bayesian Optimization (BO) has been recently gaining significant popularity due to its high efficiency in the derivative-free optimization of computationally-expensive objective functions (Frazier, 2018; Gelbart et al., 2014; Snoek et al., 2012; Shahriari et al., 2015; Wu et al., 2017). Instead of directly optimizing the objective function, BO builds a probabilistic model of the objective function (known as the surrogate model) that is sequentially updated by sampling the underlying numerical model. The sequential sampling process aims to balance exploration and exploitation, which is done by using an explicit acquisition function that guides the search toward the most promising solutions with potentially optimal values of the objective function and/or high uncertainty.

This study presents the first attempt at implementing BO for the optimization of water quality in drinking water distribution systems. The key contributions of this work are (i) developing a simulation-optimization framework that couples EPANET-MSX with a Bayesian Optimization algorithm, (ii) applying BO to optimize the scheduling of disinfectant booster stations within a case study featuring a real-life, mid-size WDN, (iii) conducting a systematic analysis of the performance of different BO methods (i.e., different covariance kernels and acquisition functions), as well as the role of different BO parameters to understand the capabilities and limitations of applying BO for water quality optimization under different scenarios.

The rest of this paper is organized as follows: Section 2 provides a description of the methods implemented in this study, including the formulation of the booster optimization model and the underlying theory of BO. Section 3 describes the case study WDN, showcases the results of the systematic analyses of different BO methods and parameters, and discusses some of the key limitations of BO. Section 4 offers a set of concluding remarks and highlights the key takeaways of the present study.

2. Methodology

2.1. Optimization framework

Fig. 1 depicts an overview of the general framework developed in this study for implementing BO to optimize WQ in WDNs. The Python-based framework couples BO, as implemented within the pyGPGO package (Jimenez, 2020), with the functions library toolkit of EPANET-MSX. The framework is used to conduct a thorough sensitivity analysis of BO methods and parameters. Below, details on each of the key components of the framework and how they are linked to one another are provided.

2.2. Optimization Model Formulation

The objective of the booster scheduling optimization is to minimize the cost of chlorine injection while satisfying various water quality standards at all demand nodes. In this study, the total cost of chlorine injection is calculated as the summation of both the capital cost of the booster system design (BCD) and the operational cost of booster chlorine injection (BCI), which can be described by (Abokifa et al., 2019; Ohar and Ostfeld, 2014; Ostfeld and Salomons, 2006):

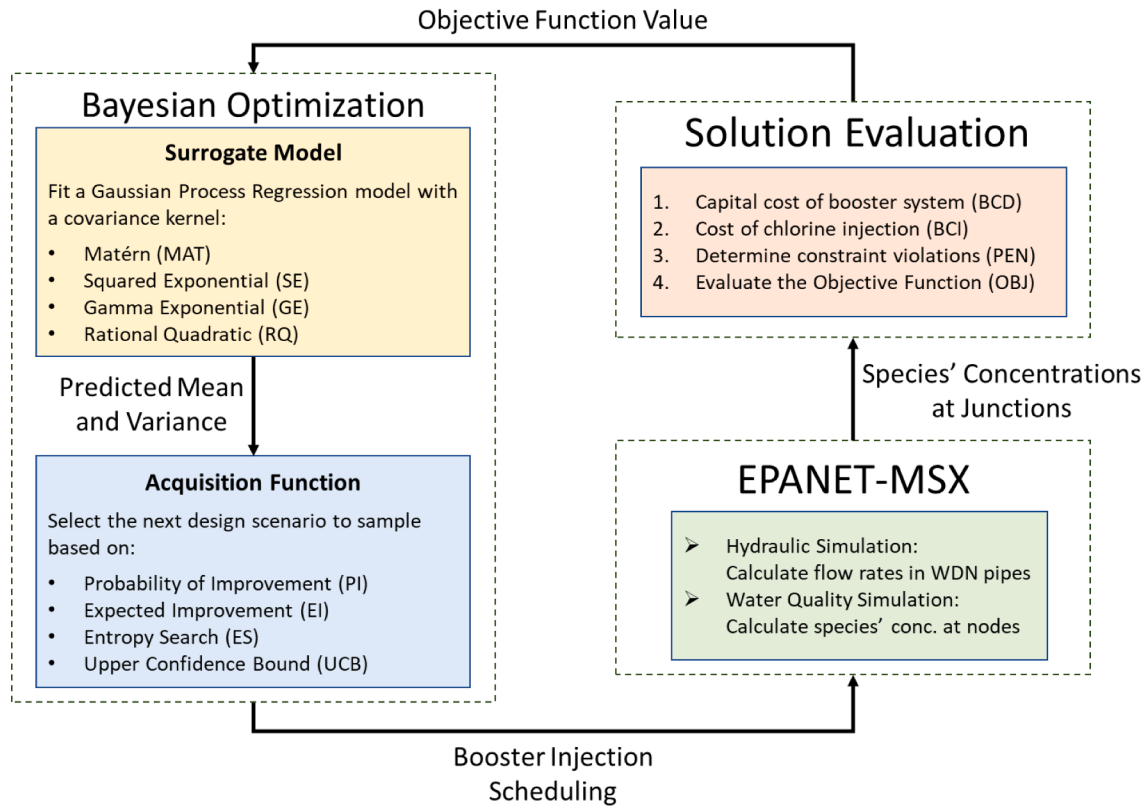


Fig. 1. Schematic of the booster scheduling optimization framework.

$$BCI = \lambda \sum_{b=1}^{n_b} \sum_{i=1}^{n_i} Cl_{b,i} \times \Delta t_i \quad (1)$$

$$BCD = DRV(AI, BLD) \left[\sum_{b=1}^{n_b} \alpha (C_b^{max})^\beta + \gamma V_b \right] \quad (2)$$

where $Cl_{b,i}$ is the chlorine mass injection rate of booster “b” during injection event “i” (kg/min); n_b is the number of chlorine boosters in the WDN; n_i is the number of injection events in one day (events/day); Δt_i is the length of the injection event “i” in minutes (min/event); λ is the chlorine injection cost per unit mass of chlorine (\$/kg); DRV is the daily return value coefficient (day⁻¹) as a function of the annual interest AI (%) and booster design lifetime BLD (years); C_b^{max} is the maximum injection rate booster station b can produce (mg/min); V_b is the total injected mass of chlorine by booster station b (mg); and α [\$(mg/min)^\beta\$], β [-], and γ [\$/mg] are empirical booster chlorination capital cost coefficients.

In order to ensure the satisfaction of water quality standards, the concentrations of various species, such as chlorine and DBPs, throughout the WDN must be constrained within a range specified by minimum and maximum thresholds. To incorporate these constraints into the objective function formulation, a penalty function (PEN) is constructed to account for violations of the upper and lower concentration bounds of the species in the EPANET-MSX simulation:

$$PEN = \sum_{s=1}^{n_s} \left(\delta_{min,s} \times \sum_{j=1}^{n_j} \max \left(1 - \frac{C_{t,j,s}}{C_s^{min}}, 0 \right) + \delta_{max,s} \times \sum_{t=1}^{n_t} \max \left(\frac{C_{t,j,s}}{C_s^{max}} - 1, 0 \right) \right) \quad (3)$$

where, $\delta_{min,s}$ and $\delta_{max,s}$ are the penalty cost coefficients for violating the minimum and maximum constraints of species “s”, respectively; n_s is the

number of species simulated by EPANET-MSX; C_s^{max} and C_s^{min} are the minimum and maximum constraints imposed on the concentration of species “s” throughout the WDN; and $C_{t,j,s}$ is the concentration of species “s” at junction “j” during time-step “t” (mg/L). The latter is obtained through running a WQ simulation in EPANET-MSX for a given booster design.

Finally, the objective function to be minimized is the summation of the aforementioned cost and penalty terms:

$$OBJ = BCI + BCD + PEN \quad (4)$$

The decision variables of the optimization model are the chlorine injection rates at each booster station during each injection event ($Cl_{b,i}$). Thus, the number of decision variables is equal to the number of injection events times the number of chlorine boosters ($n_i \times n_b$).

2.3. Bayesian Optimization

Bayesian optimization consists of two main components, (i) the surrogate function, which is a probabilistic surrogate model that is trained to predict the mean and variance of the objective function at any point within the solution space, and (ii) the acquisition function, which is a mathematical technique that uses the predicted mean and variance generated by the surrogate model to guide the selection of the next point to sample within the solution space.

2.3.1. Surrogate Function

In this study, Gaussian process regression (GPR) is implemented to build a probabilistic surrogate model for the booster optimization objective function. GPR builds a surrogate model $f(x)$ of the objective function by fitting a Gaussian distribution over the mean (μ) and covariance (k) of the evaluated values of the objective function at various points (x) (Seeger, 2004). The resulting probabilistic model can be described as (Candelieri et al., 2018):

$$f(x) \sim GP(\mu(x); k(x, x')) \quad (5)$$

The mean and covariance of this distribution are updated throughout the optimization process based on the new evaluations of the objective function to produce the posterior by means of Bayes' rule.

2.3.2. Covariance Kernel

GPR implements a covariance kernel function $k(x, x')$ to capture the correlation between the predicted values of f based on the proximity of the values of the predictor (x). In this study, four of the more commonly used covariance kernels for GPR were systematically tested (Archetti and Candelieri, 2019; Scikit-learn, 2023).

Squared-Exponential (SE). The squared-exponential kernel, also known as radial basis function (RBF), can be described by:

$$k(x_i, x_j) = \exp\left(-\frac{d(x_i, x_j)^2}{2l^2}\right) \quad (6)$$

where $d(x_i, x_j)$ is the Euclidean distance, and l is the length scale of the kernel.

Gamma-Exponential (GE). The gamma-exponential kernel is a slightly modified version of the squared-exponential kernel:

$$k(x_i, x_j) = \exp\left(-\frac{d(x_i, x_j)^{2\gamma}}{2l^2}\right) \quad (7)$$

where γ is an adjustable hyperparameter that controls the smoothness of the function.

Matérn (MAT). The Matérn kernel is a generalization of the RBF kernel, and can be described by:

$$k(x_i, x_j) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}}{l} d(x_i, x_j)\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}}{l} d(x_i, x_j)\right) \quad (8)$$

where $\Gamma(\cdot)$ is the gamma function, $K_\nu(\cdot)$ is a modified Bessel function, and ν is an adjustable hyperparameter that controls the smoothness of the function. The smaller the value of ν , the less smooth the approximated function is. As ν increases, the kernel approaches the RBF kernel.

Rational Quadratic (RQ). The rational-quadratic kernel can be seen as a scale mixture (an infinite sum) of RBF kernels with different characteristic length scales:

$$k(x_i, x_j) = \left(1 + \frac{d(x_i, x_j)^2}{2\alpha l^2}\right)^{-\alpha} \quad (9)$$

where α is the scale mixture parameter.

Kernel length scale. In this study, the length scale (l) in each of the tested covariance kernels was selected by minimizing the cross-validation error of fitting the GPR model to the objective function. To that end, 10 sets of 250 randomly generated evaluations of the objective function were generated, and the data was used to train the GPR model. To select the best value for the length scale parameter for each kernel function, an optimization problem with the objective of minimizing the cross-validation mean squared error (CV-RMSE) was formulated. The latter was then solved via a univariate bounded optimization routine using the `minimize_scalar` function in the `scipy` library.

2.3.3. Acquisition function

The acquisition function offers a simple and computationally efficient approach for finding the best points to sample in order to maximize the value of the surrogate function. The acquisition function guides the

sampling of the surrogate function by balancing the trade-off between exploitation and exploration. Exploration seeks to sample from the areas in the solution domain that have not been sampled in previous iterations (i.e., areas in the domain with higher uncertainty). Exploitation seeks to sample from the areas in the solution domain in which the optimum value of the surrogate function is expected to be located based on previous iterations. The acquisition function allows the explicit control of both exploitation and exploration to guide the optimization process to find the optimal solution in the least number of iterations (Candelieri et al., 2018). In order to select the next point to sample during each iteration, an efficient optimization algorithm, namely the Limited-Memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS-B) algorithm, was used to locate the optimal value for the acquisition function.

In this study, four of the more frequently used acquisition functions in BO literature (Archetti and Candelieri, 2019), namely probability of improvement (PI), expected improvement (EI), upper confidence bound (UCB), and entropy search (ES), are tested for the optimization of booster chlorination in WDNs.

Probability of improvement (PI). The PI acquisition function evaluates the likelihood of improvement of the surrogate function $f(x)$ by means of a normal distribution:

$$PI(x) = P(f(x) \geq f(x^+)) = \Phi(Z) \quad (10)$$

where $f(x^+)$ is the best previously observed objective function value, and Z is the standardized value of $f(x^+)$:

$$Z = \frac{\mu(x) - f(x^+)}{\sigma(x)} \quad (11)$$

where $\mu(x)$ and $\sigma(x)$ are the mean and standard deviation of $f(x)$ at candidate point x , and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

Expected improvement (EI). The EI acquisition function evaluates the expected improvement of the surrogate function $f(x)$ as follows:

$$EI(x) = \sigma(x)(Z \Phi(Z) + \varphi(Z)) \quad (12)$$

where $\varphi(\cdot)$ is the probability density function of the standard normal distribution.

Upper confidence bound (UCB). UCB uses the confidence bound concept to balance the exploitation and exploration trade-off by considering a hyperparameter (β):

$$UCB(x) = \mu(x) + \beta\sigma(x) \quad (13)$$

Using a higher value for β results in favoring the predicted variance over the mean, which results in favoring exploration over exploitation.

Entropy Search (ES). The ES acquisition function uses differential entropy to reduce the uncertainty in finding the optimal value of the acquisition function by maximizing the information obtained using new data (Wang and Jegelka, 2017). In other words, ES aims to find the data point(s) that would cause the largest possible differential entropy. Equations describing the ES approach can be found elsewhere (Hennig and Schuler, 2012; Wang and Jegelka, 2017).

3. Results and Discussion

3.1. Case study

The C-Town benchmark WDN (Figure S-1) is implemented herein to demonstrate the BO framework. The network comprises one reservoir (source), seven tanks, 388 nodes, 432 pipes, 11 pumps grouped into five pumping stations, and one valve (Creaco et al., 2016; Sousa et al., 2016).

The C-Town WDN has been used in previous literature as a case study for optimizing chlorine booster scheduling and location (Abokifa et al., 2019) as well as other WDN optimization applications (Ostfeld et al., 2012). Chlorine is assumed to be injected at four optimized locations (i.e., $n_b = 4$), namely the reservoir (R1), and junctions J201, J385, and J420 (Figure S-1). The optimized locations of the chlorine boosters were obtained from (Abokifa et al., 2019), and the proposed BO framework is used herein to optimize the scheduling of the four chlorine boosters. Each of the four booster stations is assumed to have six different injection events, each lasting 4 hours, within a periodic 24 hours-cycle (i.e., $n_i = 6$). Thus, the total number of decision variables is equal to 24. A simulation duration of 72 hours was implemented with a timestep of 15 minutes. Chlorine concentrations during the last 24 hours were considered in the calculation of constraint violations to ensure that cyclical conditions have been established (Ohar and Ostfeld, 2014). Junctions that receive no chlorine due to being present on pipes with no flow were excluded from the analysis to avoid biasing the optimization results (20 junctions). The empirical booster chlorination cost parameters were taken from (Ohar and Ostfeld, 2014): $\lambda = (\$2/\text{kg})$, $\alpha = 2.21$, $\beta = 0.13$, and $\gamma = 0$.

To compute the constraint violation penalty term in the objective function (PEN) for a given booster system design, the concentrations of all species throughout the WDN must be simulated. In this study, this is achieved by conducting an EPANET-MSX simulation every time the objective function is evaluated. The main inputs to the WQ simulations are (1) the layout of the WDN, that is, the locations, connectivity, and sizing of each of the network components (i.e., pipes, nodes, pumps, tanks, boosters, etc.), (2) water demand information, and (3) chlorine injection rates at boosters ($Cl_{b,i}$), which are the decision variables in the optimization problem. The outputs of the WQ simulations are the species' concentrations at the network junctions at each timestep ($C_{i,j,s}$).

For the case study WDN, a multiple-species WQ model that accounts for the simultaneous decay of chlorine as well as the formation of trihalomethanes (THMs) is adopted (Ohar and Ostfeld, 2014):

$$\frac{\partial C_A}{\partial t} = -kC_A C_B; \quad \frac{\partial C_B}{\partial t} = -Y_B k C_A C_B; \quad \frac{\partial C_P}{\partial t} = Y_P k C_A C_B \quad (14)$$

where, C_A , C_B , and C_P are the concentrations of the chlorine disinfectant (A), a fictitious reactant (B), and THMs (P), respectively [mg/L]; k is the reaction rate constant [$\text{L}(\text{mg} \cdot \text{hr})^{-1}$]; Y_B and Y_P are the ratios between the stoichiometric coefficients of the fictitious reactant (B) and THMs (P) to that of chlorine (A) in [mg/mg]. While the present study considers the

continuous formation of THMs throughout the WDN, more recent work on THM formation showed that the rate of THM formation can be more simply expressed as proportional to chlorine consumed by the reaction after a certain initialization period (Sathasivan et al., 2020). Values of the reaction model parameters were taken consistent with those reported by Abokifa et al., (2016a) as follows: $k = 0.1164 \text{ L}(\text{mg} \cdot \text{hr})^{-1}$, $Y_B = 1$, and $Y_P = 0.05$. The minimum and maximum bounds on the chlorine concentration throughout the WDN were set to 0.2 mg/L and 4 mg/L , respectively, and the maximum bound on THMs was set to 0.08 mg/L . By default, the constraint violation penalty cost coefficients for all species were set to the same value ($\delta_{\min, Cl} = \delta_{\max, Cl} = \delta_{\max, THMs} = 0.1$).

3.2. Evaluation of Bayesian Optimization Methods

3.2.1. BO methods comparison

First, a systematic testing of different BO methods (i.e., different combinations of acquisition functions and covariance kernels) was performed. For each BO method, ten different optimization runs were conducted, where each optimization run started from a different seed of randomly generated initial points. Fig. 2 shows box-and-whisker plots representing the ranges of the final values of the objective function achieved by each of the tested BO methods. The figure shows significant variability in the performance of different BO methods, where the best-performing method (UCB acquisition function with MAT kernel) produced a median objective function value that is less than half of that produced by the worst-performing BO method (ES acquisition with SE kernel). This variability shows the importance of the systematic testing conducted herein to ensure that the best BO method is identified for WQ optimization. The results also imply that similar systematic analyses must be conducted before applying BO to other WDN optimization applications. It is also worth noting that the variability between the performance of different acquisition functions was generally found to be greater than that between different covariance kernels, which is further discussed in the following sections.

3.2.2. Choice of acquisition function and covariance kernel

Overall, the performance of BO was found to be more sensitive to the choice of the acquisition function than that of the covariance kernel. This can be seen in Fig. 3, which depicts the mean and standard deviation of the best objective values obtained by each of the four (a) acquisition functions and (b) covariance kernels, respectively, over the ten optimization runs. As can be seen from the figure, significant

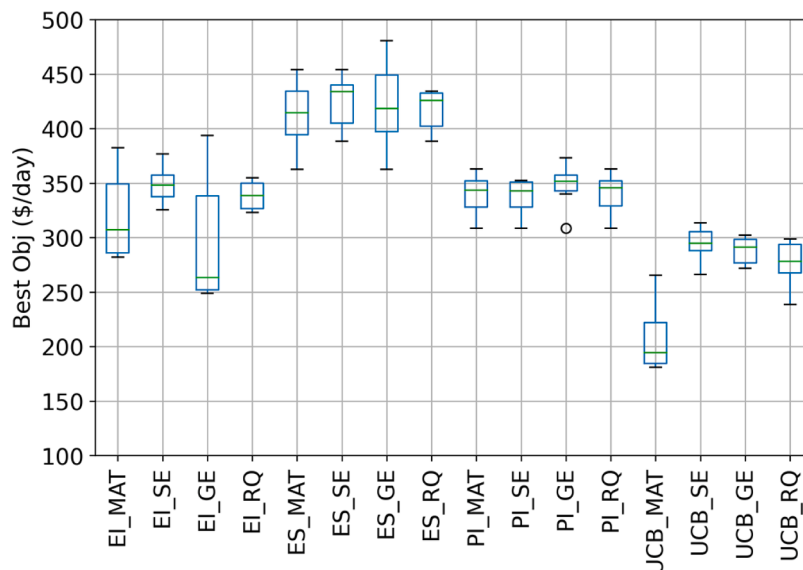


Fig. 2. Ranges of best objective function values achieved by each of the tested BO methods.

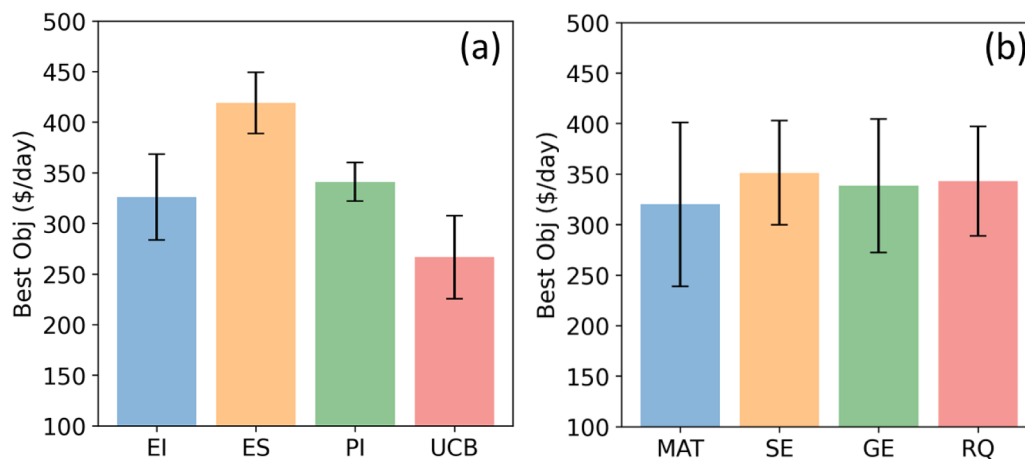


Fig. 3. Mean and standard deviation of the best objective function values achieved by each of the tested (a) acquisition functions and (b) covariance kernels.

variability between the performance of different acquisition functions is observed. On average, UCB acquisition was found to produce the best performance among all four tested acquisition functions. This is followed by both EI and PI, who appeared to produce similar performances. Finally, ES showed the worst performance among all four acquisition functions, with an average final objective function value that's more than 1.7X that obtained by UCB. On the other hand, no significant variability was observed between the average performance of different covariance kernels, with the MAT and SE kernels showing the best and worst average performances, respectively. These results imply that the choice of the acquisition function is more influential than that of the covariance kernel, which can be attributed to the fact that the acquisition function dictates the pathway of the optimization process by controlling the successive selection of the solution candidates to sample.

3.2.3. Convergence profile and number of iterations

Fig. 4 shows the average convergence profiles of different (a) acquisition functions, and (b) covariance kernels over the ten optimization runs. The figures show that the convergence speed of BO is mainly controlled by the choice of the acquisition function rather than the covariance kernel, further affirming the dominant role of the acquisition function in controlling the optimization process. For instance, UCB acquisition typically reached the best solution in less than 250 iterations (Fig. 4-a). On the other hand, minimal differences were observed between the convergence profiles of different covariance kernels (Fig. 4-b). It's also worth noting that little to no enhancement was generally

observed in the value of the objective function after 500 iterations for all of the tested BO methods. Additionally, stopping the optimization when no enhancement is achieved for 100 consecutive iterations could help avoid unnecessary computational costs.

3.2.4. Best solution and constraint violations

Further investigation of the best solutions achieved by the best-performing BO method (UCB-MAT) revealed that BO was indeed able to obtain high-quality solutions that minimize the cost of chlorine injection in the WDN while maintaining the concentrations of chlorine and THMs within the specified constraints throughout the distribution system. The latter can be seen in Fig. 5, which shows the distribution of the hourly chlorine and THMs concentrations at all nodes in the WDN for the best solution in the randomly generated initial population (Fig. 5-a, b), and for the final solution obtained by the best performing BO method (Fig. 5-c, d) for one of the ten optimization runs. As can be seen from the figure, the frequency of constraint violations was significantly reduced after applying BO. The percentage of hourly concentrations ($C_{t,j,s}$) in violation of the maximum chlorine and THM constraints dropped from 6.9% and 39.6% before BO (Fig. 5-a) to 0.4% and 26.8% after applying BO (Fig. 5-c). The decrease in maximum constraint violations can be attributed to the 30% decrease in the total injected chlorine dose achieved by BO. Surprisingly, despite this decrease in the total applied chlorine dose, the percentage of hourly concentrations in violation of the minimum chlorine constraint remained the same (13.6%) after applying BO. This reflects BO's ability to find the best distribution for chlorine

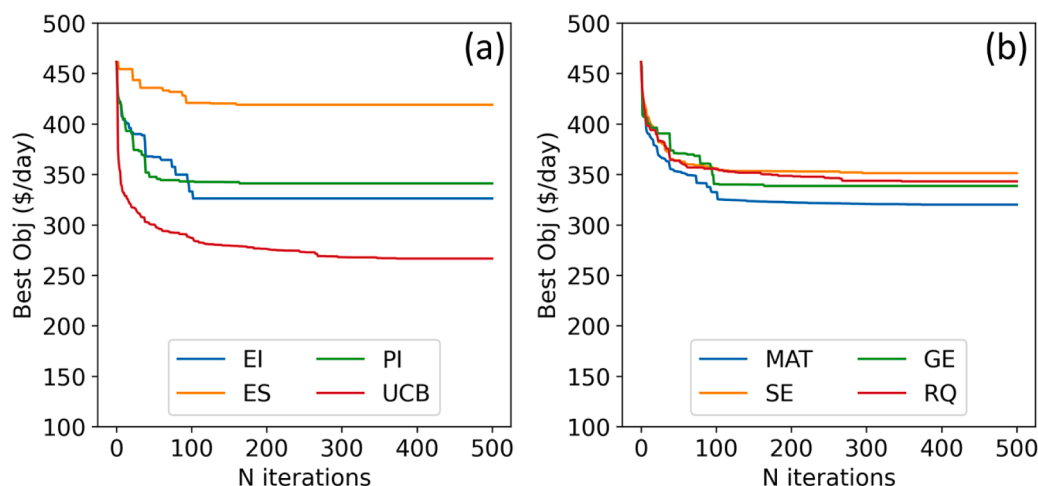


Fig. 4. Average convergence profile of each of the tested (a) acquisition functions and (b) covariance kernels.

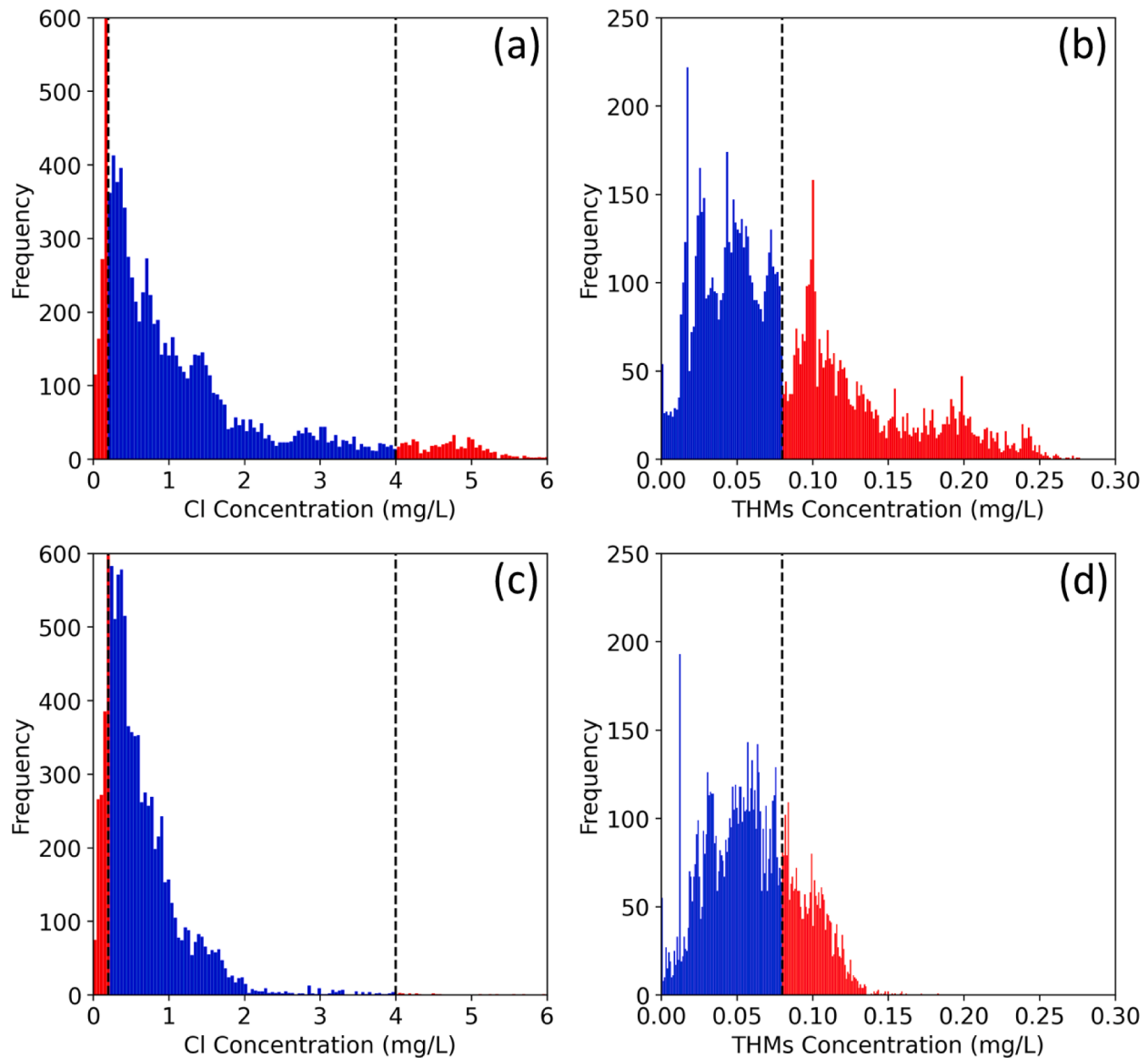


Fig. 5. Distribution of the hourly concentrations of chlorine and THMs at all junctions in the distribution system for (a), (b) the best solution out within the randomly generated initial population, and (c), (d) the final solution obtained by the best performing BO method (UCB-MAT).

injection that minimizes DBP formation while maintaining chlorine residuals in the system. In addition to the notable decrease in constraint violation frequency, the extent of constraint violations, that is, the differences between the hourly concentrations ($C_{t,j,s}$) and the min/max concentration thresholds (C_s^{min} , C_s^{max}), diminished after applying BO. This is evident by the decrease in the value of the penalty function (Eq. 3) from 329.3 \$/day before applying BO to 107.4 \$/day after applying BO. Taken together, the results showed that BO was able to reduce the total objective function by more than 50% from 434 \$/day to 200 \$/day, while reducing both the cost of chlorine injection and constraint violations for chlorine and THMs throughout the network.

In many cases, the importance of maintaining a minimum chlorine residual throughout the WDN, and/or minimizing the formation of DBPs, is greater than that of preventing occasional exceedance of the maximum chlorine concentration threshold. To simulate these scenarios, the BO optimization was repeated twice. In the first scenario, the constraint violation penalty cost coefficient for the minimum chlorine constraint was set to five times that of the maximum chlorine constraint ($\delta_{min,cl} = 0.5$ and $\delta_{max,cl} = 0.1$). As expected, the best solution obtained using these settings resulted in decreasing the percentage of hourly

concentrations violating the minimum chlorine constraint from 13.6% to 3.9%. Nevertheless, the cost of chlorine injection (BCI) increased by 45% as a result of adopting a more stringent penalty cost for violating the minimum concentration constraint. Furthermore, the frequency of maximum THMs violations increased from 26.8% to 48.9%. This can be attributed to the increase in the total chlorine dose injected in the system, which is needed to maintain a minimum residual throughout the WDN at all times.

In the second scenario, the maximum THMs constraint penalty cost coefficient was set to ($\delta_{max,THMs} = 0.3$), while those for chlorine constraints were kept unchanged ($\delta_{min,cl} = \delta_{max,cl} = 0.1$). This scenario resulted in decreasing the frequency of maximum THMs violations from 26.8% to 11.6%. However, adopting a more stringent penalty cost for violating the maximum THMs constraint also resulted in significantly increasing the total penalty for violating the minimum chlorine constraint from 13.6% to 24.1%, which can be attributed to the decrease in the total chlorine dose injected in the system (BCI decreased by 15%). Taken together, these results highlight the important trade-off between maintaining a sufficient residual while limiting the formation of DBPs within the WDN.

3.3. Influence of Bayesian Optimization parameters

3.3.1. Covariance kernel length scale

The performance of BO was found to be greatly influenced by the choice of the length-scale parameter of the covariance kernel function (l). This parameter dictates the strength of the relationship between the predicted values by the surrogate GPR model at different points along the function domain. Fig. 6 depicts the cross-validation root mean squared error (CV-RMSE) between the true value of the objective function and the mean value predicted by the surrogate GP model (y-axis) for different values of the length scale parameter (x-axis) using different covariance kernels, namely (a) Matérn (MAT), (b) squared exponential (SE), (c) gamma exponential (GE), and (d) rational quadratic (RQ). The shaded area represents the range of CV-RMSE values for ten different randomly generated sets of 250 solutions. For each set, the GPR model was trained using different values of l in the range of $[1, 1 \times 10^{10}]$, and the five-fold cross validation RMSE was computed for each value. As can be observed from Fig. 6, the accuracy of the GPR model in predicting the values of the objective function appears

to be greatly dependent on the choice of the length-scale parameter of the covariance kernel. This can be seen from the orders-of-magnitude variability in the CV-RMSE values at different length scales for different kernel functions. More importantly, optimal values of the length-scale parameter for the different covariance kernels spanned a wide range $[1 \times 10^4, 1 \times 10^6]$. Taken together, these results highlight the importance of carefully tuning the length scale parameter to ensure that the best GPR model performance is achieved for each covariance kernel.

3.3.2. Number of initial points

Another important parameter that was found to significantly influence the performance of BO was the number of initial points used to fit the GPR surrogate model before starting the iterations. Typically, the larger and more diverse the initial population is, the better the performance of the GPR model in predicting the mean and standard deviation of the objective function will be. However, the computational cost of fitting the GPR model increases as the number of initial points increases. Fig. 7 shows the CV-RMSE scores of fitting the GPR model using a

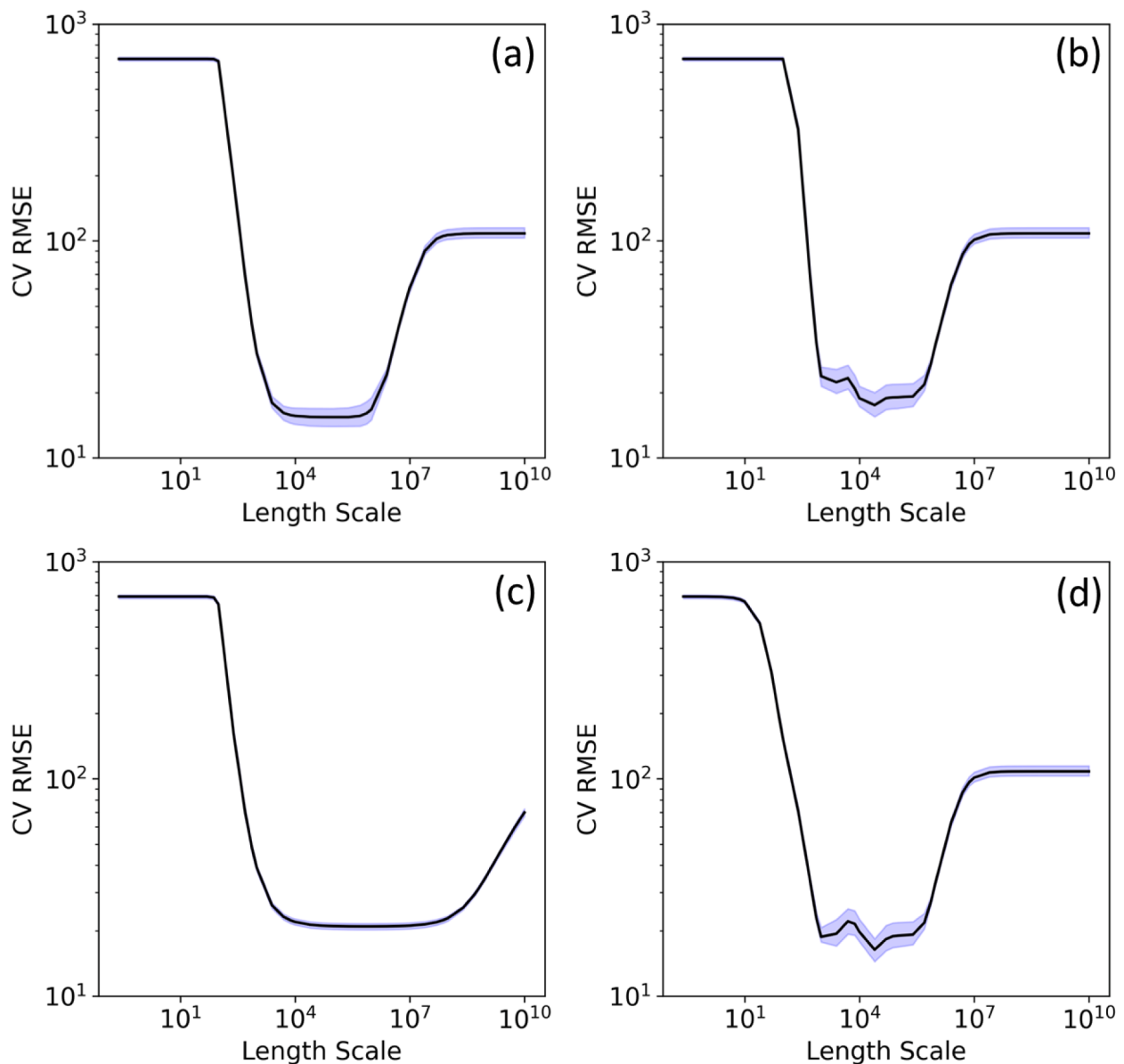


Fig. 6. Cross-validation RMSE of the GPR model vs. the length-scale of each of the tested covariance kernels: a) Matérn, (b) squared-exponential, (c) gamma-exponential, and (d) rational-quadratic.

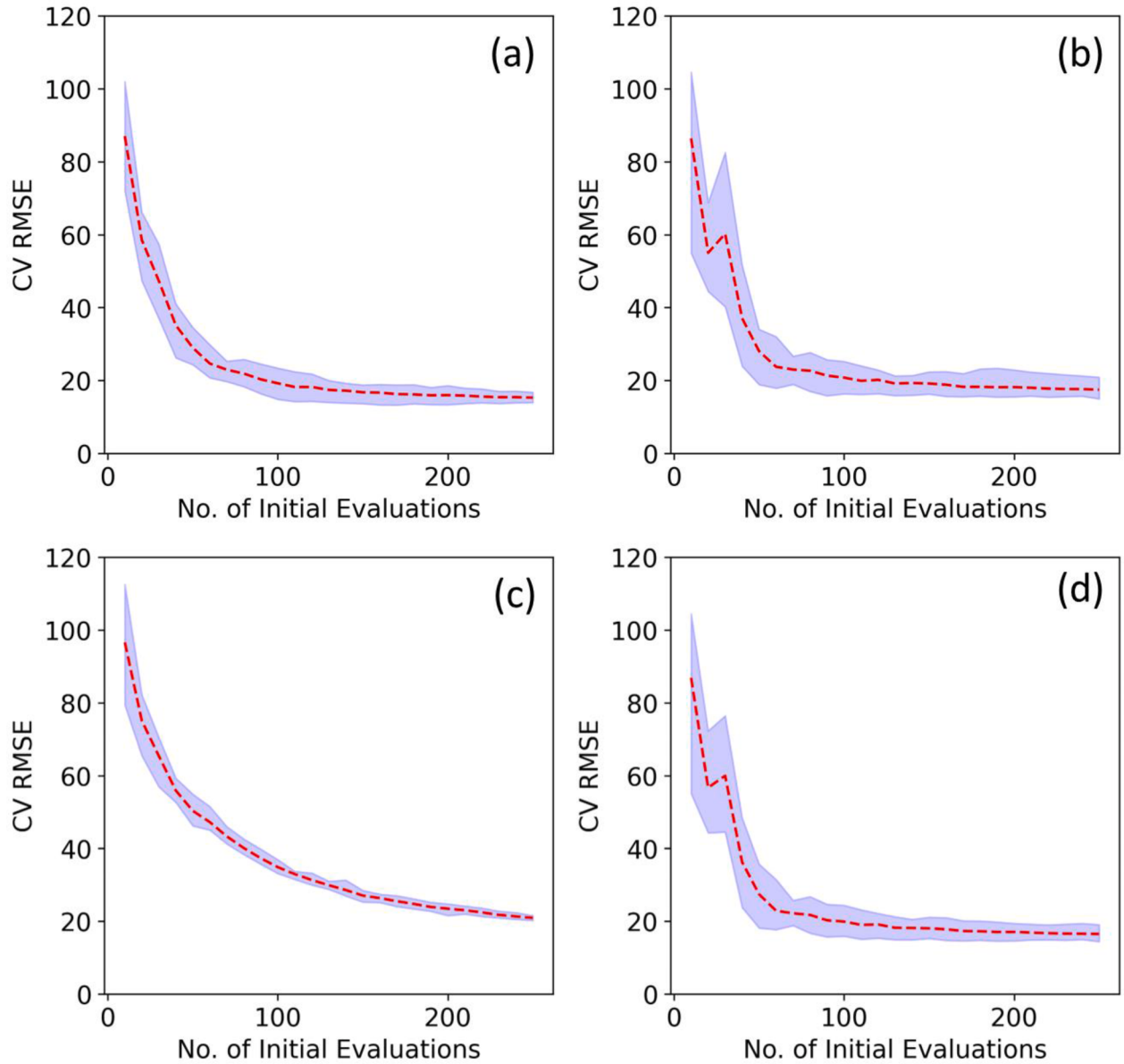


Fig. 7. Cross-validation RMSE of the GPR model vs. the number of initial evaluations used to fit the GPR model using each of the tested covariance kernels: a) Matérn, (b) squared-exponential, (c) gamma-exponential, and (d) rational-quadratic.

different number of initial points for different covariance kernels. The shaded area represents the range of CV-RMSE values for ten different randomly generated sets of 250 solutions. The figure shows that the performance of the GPR model improves as the number of initial points increases. Nevertheless, the additional enhancement in the GPR model's performance decreases as the number of initial points increases, with most of the improvement realized in the first 100 initial points. Overall, little enhancement is observed beyond 100 initial points for all of the tested kernels with the exception of GE. Thus, an initial population size of 100 points was found to achieve a sufficient trade-off between GPR model accuracy and computational cost. It is also worth noting that the GPR model showed remarkable accuracy in learning the highly non-linear objective function. This can be seen from Figure S-2, where the five-fold cross-validation r^2 scores were greater than 0.9 for all four covariance kernels when only 150 points were used to train the GPR model.

3.3.3. Exploration vs exploitation level

Another key advantage of BO is that the choice of the acquisition

function enables the tuning of the trade-off between exploration and exploitation to achieve the best performance. To further investigate this property, we examined the performance of the UCB acquisition function, which showed the best performance among all tested acquisition functions, at different exploration/exploitation levels. In UCB, this level can be directly adjusted via the (β) parameter. Increasing the value of β favors the uncertainty component of the acquisition function (i.e., exploration) over the mean value of the function (i.e., exploitation). A range of different β values [0.5-2.5] was sampled, and the results showed that the final value of the objective function achieved by different choices of the beta parameter after 500 iterations is consistent. Yet, the convergence speed seemed to increase with increasing the value of β . This can be seen from Figure S-3, which shows the convergence of UCB during the first 200 iterations using different β values. The results indicate that increasing the exploration level over that of the exploitation generally improves the convergence speed of BO.

3.4. Computational Cost

To understand the computational cost of conducting BO, the time taken by the different BO processes under different scenarios, including the number of decision variables and the size of the WDN, was analyzed. All simulations were executed on a desktop computer with Intel(R) Xeon (R) W-2133 CPU @ 3.60GHz processor and 32 GB RAM.

3.4.1. BO initialization

First, running one water quality simulation of the C-TOWN network using EPANET-MSX takes approximately 4.3 seconds, which constitutes the cost of evaluating the objective function. Therefore, the computational cost of conducting the initial 100 simulations was 7.2 minutes. The cost of fitting the Gaussian process model to the 100 initial points was approximately 0.1 seconds.

3.4.2. Acquisition function optimization

After the initialization stage, three processes take place with each new iteration: (1) the optimization of the acquisition function to select the new point to sample, (2) evaluating the objective function at the new point, and (3) updating the Gaussian process model. Out of the three processes, we found that the optimization of the acquisition functions takes up the most time. This cost was found to vary from one acquisition function to another. On average, optimization of the UCB function takes approximately 13.2 seconds during each iteration, which is higher than what the other three acquisition functions require (approximately 8.7 seconds on average). However, considering how fast UCB converges to the optimal solution compared to the other acquisition functions (Fig. 4), the overall time needed by UCB to achieve convergence remains comparable to the other acquisition functions. More importantly, the optimization of the acquisition function can be parallelized over multiple threads. In this study, a 6-8 fold reduction in the computational cost of each iteration was achieved when all 12 threads of the computer were used to optimize the acquisition function using the multi-start L-BFGS-B optimization routine compared to using only one core. Further enhancements in the computational cost of BO can be achieved by utilizing high-performance computing (HPC) resources.

3.4.3. Updating the surrogate model

The cost of updating the GPR surrogate model was found to increase as the optimization process progressed. This can be attributed to the continuous increase in the number of points to which the GPR model is being fitted. In this study, the cost of fitting the GPR model increased from ~0.1 seconds during the first iteration (101 points) to ~0.6 seconds after 500 iterations (600 points). Although this increase remains insignificant compared to the cost of optimizing the acquisition function, this may not be the case for systems that require a large number of iterations to converge. This can be seen from Figure S-4, which shows the computational cost of fitting the GPR model (y-axis) to a different number of points (x-axis), compared to that of a random forest (RF) regression model. While the RF model displays linear scaling in the computational cost, the cost of fitting the GPR model increases exponentially with the number of points. The latter can be attributed to the cost of inverting the covariance matrix via LU decomposition during each iteration, which scales with the cube of the number of points (n^3). Thus, for optimization applications requiring a significant number of iterations (i.e., more than 5000), other surrogate models (e.g., RF) may provide significant computational cost reductions over the GPR model.

3.4.4. Number of decision variables

To understand how the computational cost of BO scales with the number of decision variables, another optimization scenario was conducted on the case study network. In this scenario, only one chlorine source was placed at the reservoir (R1) instead of four boosters, thus reducing the number of decision variables to six instead of 24. Expectedly, the cost of conducting the EPANET-MSX simulation for the one-

source scenario was similar to that of the four-source scenario (i.e., ~4.3 second). Yet, the one-source scenario was found to require a smaller number of initial points to achieve a satisfactory performance by the GPR model, where only 50 initial points were needed to achieve an average cross-validation $r^2 > 0.95$. Thus, the cost of BO initialization for the one-source scenario would be half that for the four-source scenario. Furthermore, the average cost of optimizing the UCB acquisition function was found to drop 3-folds from 13.2 seconds for the four-source scenario to 4.6 seconds for the one-source scenario.

Taken together, the results of this study indicate that (i) for small WDNs, the overall computational cost of BO is limited by the number of decision variables in the optimization problem, since the cost of evaluating the objective function is minimal compared to that of optimizing the acquisition function. On the other hand, for large WDNs, the cost of evaluating the objective function would be significant compared to that of optimizing the acquisition function. For instance, running an EPANET-MSX simulation of the fairly large BWSN-2 network (Ostfeld et al., 2008), containing 12,523 nodes and 14,822 pipes, was found to take approximately 82 seconds, which is greater than the cost of optimizing the acquisition function.

3.5. Future research needs

Although BO showed satisfactory performance in the optimization of booster chlorination scheduling, several research questions still require further investigation.

3.5.1. Comparison against evolutionary algorithms

The results of this study indicate that, unlike evolutionary optimization algorithms, BO is capable of converging to high-quality solutions within a small number of iterations. For instance, in a previous study (Abokifa et al., 2019), it was found that a population size of 100 and a maximum number of stall generations of 150 are necessary for optimizing booster chlorination in the C-TOWN WDN. Thus, GA would require at least 15,000 evaluations of the objective function. Nevertheless, it is worth noting that evolutionary algorithms can benefit from parallel processing to distribute the computational cost of evaluating the objective function over several threads, while for BO, the objective function is sampled sequentially during each iteration. However, BO still benefits from parallelizing the optimization process of the acquisition function, which was found to significantly reduce the computational cost of BO in the present study, which makes BO a very efficient optimization method overall. Taken together, the results of this study highlight the need for conducting further research into how the performance of BO compares to evolutionary optimization algorithms (e.g., GA or PSO).

3.5.2. Water quality simulation model

In this study, the multiple-species model EPANET-MSX was adopted for conducting the water quality simulations needed for the evaluation of the objective function. The presented framework allows for setting constraints on the concentrations of any number of species in the WDN. Since chlorine is a highly reactive oxidant, it undergoes side reactions with various species. Therefore, chlorine concentrations in the system can influence the concentrations of different species of interest from a water quality perspective, including DBPs, microbiological species (e.g., bacteria/biofilms), and lead concentrations. Future studies can leverage the presented framework to study the optimization of chlorine dosage that maintains the concentrations of such species within regulatory standards. For instance, Maheshwari et al., (2020) proposed an approach for the optimization of disinfectant dosage for simultaneous control of lead and disinfection-byproducts. Furthermore, previous studies highlighted the importance of considering dispersive solute transport in water quality simulations used for booster optimization (Abokifa et al., 2019), especially in the dead-end branches that experience frequent stagnations and excessive residence times (Abokifa et al.,

2016b). Other studies indicated that 2-D models can provide a more accurate representation of the radial transport of chlorine in WDN pipes than 1-D models (Ozdemir and Metin Ger, 1999).

4. Conclusions

This study presents the first attempt to apply Bayesian Optimization (BO) to the optimization of water quality in drinking water distribution networks (WDNs). A systematic analysis of the performance of different BO methods in the optimization of booster chlorination scheduling was conducted. To that end, various combinations of gaussian process regression (GPR) covariance kernels and BO acquisition functions were systematically tested on a case study featuring a mid-size real-life WDN. The results revealed that the performance of BO was mainly dependent on the choice of the acquisition function, whereas little variability was observed between the performance of different covariance kernels. The results also highlighted the importance of carefully selecting the length scale of the covariance kernel and the number of initial evaluations used to fit the GPR model to ensure the best performance is achieved. Optimization of the acquisition function was found to be the most computationally demanding component of BO, with the computational cost increasing with the number of decision variables.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

We report financial support by the National Science Foundation as described in the acknowledgments section.

Data availability

Data will be made available on request.

Acknowledgments

Support by the National Science Foundation under Grants 2015603, 2015671, and 2015658, is gratefully acknowledged.

References

- Abokifa, A.A., Yang, Y.J., Lo, C.S., Biswas, P., 2016a. Investigating the role of biofilms in trihalomethane formation in water distribution systems with a multicomponent model. *Water Res* 104, 208–219.
- Abokifa, A.A., Yang, Y.J., Lo, C.S., Biswas, P., 2016b. Water quality modeling in the dead end sections of drinking water distribution networks. *Water Res* 89, 107–117.
- Abokifa, A.A., Maheshwari, A., Gudi, R.D., Biswas, P., 2019. Influence of Dead-End Sections of Drinking Water Distribution Networks on Optimization of Booster Chlorination Systems. *J. Water Resour. Plan. Manag.* 145 (12), 04019053.
- Archetti, F., Candelieri, A., 2019. *Bayesian Optimization and Data Science*. Springer.
- Behzadian, K., Alimohammadnejad, M., Ardeshir, A., Jaliliani, F., Vasheghani, H., 2012. A novel approach for water quality management in water distribution systems by multi-objective booster chlorination. *Int. J. Civ. Eng.* 10 (1), 51–60.
- Biswas, P., Lu, C., Clark, R.M., 1993. A model for chlorine concentration decay in pipes. *Water Res* 27 (12), 1715–1724.
- Boccelli, D., Tryby, M.E., Uber, J.G., Rossman, L.A., Zierolf, M.L., Polycarpou, M., 1998. Optimal Scheduling of Booster Disinfection in Water Distribution Systems. *J. Water Resour. Plan. Manag.* 124 (2), 99–111.
- Candelieri, A., Perego, R., Archetti, F., 2018. Bayesian optimization of pump operations in water distribution systems. *J. Glob. Optim.* 71 (1), 213–235.
- Creaco, E., Alvisi, S., Franchini, M., 2016. Multistep Approach for Optimizing Design and Operation of the C-Town Pipe Network Model. *J. Water Resour. Plan. Manag.* 142 (5).
- Deborde, M., von Gunten, U., 2008. Reactions of chlorine with inorganic and organic compounds during water treatment-Kinetics and mechanisms: A critical review. *Water Res* 42 (1–2), 13–51.
- Fisher, I., Kastl, G., Sathasivan, A., 2011. Evaluation of suitable chlorine bulk-decay models for water distribution systems. *Water Res* 45 (16), 4896–4908.
- Frazier, P.I., 2018. A Tutorial on Bayesian Optimization. *arXiv Preprint arXiv:1807.02811*.
- Gelbart, M.A., Snoek, J., Adams, R.P., 2014. Bayesian optimization with unknown constraints. In: *Uncertain. Artif. Intell. - Proc. 30th Conf. UAI 2014*, pp. 250–259.
- Gibbs, M.S., Dandy, G.C., Maier, H.R., 2010. Calibration and Optimization of the Pumping and Disinfection of a Real Water Supply System. *J. Water Resour. Plan. Manag.* 136 (4), 493–501.
- Goyal, R.V., Patel, H.M., Goyal, R.V., Patel, H.M., 2017. Optimal location and scheduling of booster chlorination stations for drinking water distribution system. *J. Appl. Water Eng. Res.* 5 (1), 51–60.
- Hennig, P., Schuler, C.J., 2012. Entropy search for information-efficient global optimization. *J. Mach. Learn. Res.* 13, 1809–1837.
- Islam, N., Sadiq, R., Rodriguez, M.J., 2013. Optimizing booster chlorination in water distribution networks: A water quality index approach. *Environ. Monit. Assess.* 185 (10), 8035–8050.
- Jimenez, J., 2020. *pyGPGO Documentation*.
- Kang, D., Lansey, K., 2010. Real-Time Optimal Valve Operation and Booster Disinfection for Water Quality in Water Distribution Systems. *J. Water Resour. Plan. Manag.* 136 (4), 463–473.
- Lansey, K., Pasha, F., Pool, S., Elshorbagy, W., Uber, J., 2007. Locating Satellite Booster Disinfectant Stations. *J. Water Resour. Plan. Manag.* 133 (4), 372–376.
- LeChevallier, M.W., 1999. The case for maintaining a disinfectant residual. *J. Am. Water Works Assoc.* 91 (1), 86–94.
- Li, R.A., McDonald, J.A., Sathasivan, A., Khan, S.J., 2019. Disinfectant residual stability leading to disinfectant decay and by-product formation in drinking water distribution systems: A systematic review. *Water Res* 153, 335–348.
- Liu, G., Zhang, Y., Knibbe, W.J., Feng, C., Liu, W., Medema, G., van der Meer, W., 2017. Potential impacts of changing supply-water quality on drinking water distribution: A review. *Water Res* 116, 135–148.
- Lu, C., Biswas, P., Clark, R.M., 1995. Simultaneous transport of substrates, disinfectants and microorganisms in water pipes. *Water Res* 29 (3), 881–894.
- Maheshwari, A., Abokifa, A.A., Gudi, R.D., Biswas, P., 2018. Coordinated Decentralization-Based Optimization of Disinfectant Dosing in Large-Scale Water Distribution Networks. *J. Water Resour. Plan. Manag.* 144 (10), 04018066.
- Maheshwari, A., Abokifa, A., Gudi, R.D., Biswas, P., 2020. Optimization of disinfectant dosage for simultaneous control of lead and disinfection-byproducts in water distribution networks. *J. Environ. Manage.* 276 (July), 111186.
- Makris, K.C., Andra, S.S., Botsaris, G., 2014. Pipe scales and biofilms in drinking-water distribution systems: Undermining finished water quality. *Crit. Rev. Environ. Sci. Technol.*
- Mala-Jetmarova, H., Sultanova, N., Savic, D., 2017. Lost in optimisation of water distribution systems? A literature review of system operation. *Environ. Model. Softw.* 93, 209–254.
- Munavalli, G.R., Kumar, M.S.M., 2003. Optimal Scheduling of Multiple Chlorine Sources in Water Distribution Systems. *J. Water Resour. Plan. Manag.* 129 (6), 493–505.
- Nono, D., Basupi, I., Odile, P.T., Parida, B.P., 2018. Integrating booster chlorination and operational interventions in water distribution systems. *J. Hydroinformatics* 20 (5), 1025–1041.
- Ohar, Z., Ostfeld, A., 2014. Optimal design and operation of booster chlorination stations layout in water distribution systems. *Water Res* 58, 209–220.
- Ostfeld, A., Salomons, E., 2006. Conjunctive optimal scheduling of pumping and booster chlorine injections in water distribution systems. *Eng. Optim.* 38 (3), 337–352.
- Ostfeld, A., Uber, J.G., Salomons, E., Berry, J.W., Hart, W.E., Al, E., 2008. The Battle of the Water Sensor Networks (BWSN): A Design Challenge for Engineers and Algorithms. *J. Water Resour. Plan. Manag.* 134 (6), 556–568.
- Ostfeld, A., Salomons, E., Ormsbee, L., Uber, J.G.J.G., Bros, C.M.C.M., Kalungi, P., Burd, R., Zazula-Coetzee, B., Belrain, T., Kang, D., Lansey, K., Shen, H., McBean, E., Yi Wu, Z., Walski, T., Alvisi, S., Franchini, M., Johnson, J.P.J.P., Ghimire, S.R.S.R., Barkdoll, B.D.B.D., Koppel, T., Vassiljev, A., Hoon Kim, J., Chung, G., Guen Yoo, D., Diao, K., Zhou, Y., Li, J., Liu, Z., Chang, K., Gao, J., Qu, S., Yuan, Y., Devi Prasad, T., Laucelli, D., Vamvakieridou Lyroudia, L.S., Kapelan, Z., Savic, D., Berardi, L., Barbaro, G., Giustolisi, O., Asadzadeh, M., Tolson, B.A.B.A., McKillop, R., Wu, Z.Y., Walski, T., Alvisi, S., Franchini, M., Johnson, J.P.J.P., Ghimire, S.R.S.R., Barkdoll, B.D.B.D., Koppel, T., Vassiljev, A., Kim, J.H., Chung, G., Yoo, D.G., Diao, K., Zhou, Y., Li, J., Liu, Z., Chang, K., Gao, J., Qu, S., Yuan, Y., Prasad, T.D., Laucelli, D., Lyroudia Vamvakieridou, L.S., Kapelan, Z., Savic, D., Berardi, L., Barbaro, G., Giustolisi, O., Asadzadeh, M., Tolson, B.A.B.A., McKillop, R., 2012. Battle of the water calibration networks. *J. Water Resour. Plan. Manag.* 138 (5), 523–532.
- Ozdemir, O.N., Metin Ger, A., 1999. Unsteady 2-D chlorine transport in water supply pipes. *Water Res* 33 (17), 3637–3645.
- Prasad, T.D., Walters, G.A., Savic, D.A., 2004. Booster disinfection of water supply networks: Multiobjective approach. *J. water Resour. Plan. Manag.* 130 (5), 367–376.
- Propato, M., Uber, J.G., 2004. Linear Least-Squares Formulation for Operation of Booster Disinfection Systems. *J. Water Resour. Plan. Manag.* 130 (1), 53–62.
- Rossman, L.A., Boulous, P.F., 1996. Numerical Methods for Modeling Water Quality in Distribution Systems: A Comparison. *J. Water Resour. Plan. Manag.* 122 (2), 137–146.
- Rossman, L.A., Clark, R.M., Grayman, W.M., 1994. Modeling Chlorine Residuals in Drinking-Water Distribution Systems. *J. Environ. Eng.* 120 (4), 803–820.
- Sathasivan, A., Kastl, G., Korotta-Gamage, S., Gunasekera, V., 2020. Trihalomethane species model for drinking water supply systems. *Water Res* 184.
- Scikit-learn, 2023. In: 1.7. Gaussian Processes — scikit-learn [WWW Document]. URL: https://scikit-learn.org/stable/modules/gaussian_process.html#kernels-for-gaussian-processes (accessed 5.18.23).
- Seeger, M., 2004. Gaussian processes for machine learning. *Int. J. Neural Syst.* 14 (2), 69–106.
- Shahriari, B., Swersky, K., Wang, Z., Adams, R.P., De Freitas, N., 2015. Taking the human out of the loop: A review of Bayesian optimization. *Proc. IEEE* 104 (1), 148–175.
- Shang, F., Uber, J.G., Rossman, L.A., 2008. Modeling reaction and transport of multiple species in water distribution systems. *Environ. Sci. Technol.* 42 (3), 808–814.

- Snoek, J., Larochelle, H., Ryan, P.A., 2012. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems* 25.
- Sousa, J., Muranho, J., Sá Marques, A., Gomes, R., 2016. Optimal Management of Water Distribution Networks with Simulated Annealing: The C-Town Problem. *J. Water Resour. Plan. Manag.* 142 (5), 1–9.
- Tryby, M.E., Boccelli, D.L., Koehling, M.T., Uber, J.G., Summers, R.S., Rossman, L.A., 1999. Booster chlorination for managing disinfectant residuals. *J. Am. Water Works Assoc.* 91 (1), 95–108.
- Tryby, M., Boccelli, D., Uber, J., Rossman, L., 2002. Facility Location Model for Booster Disinfection of Water Supply Networks. *J. Water Resour. Plan. Manag.* 128 (5), 322–333.
- Wang, Z., Jegelka, S., 2017. Max-value entropy search for efficient Bayesian optimization. *34th Int. Conf. Mach. Learn. ICML 2017* 7, 5530–5543.
- 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 Resour. Res.* 57 (5), e2020WR027771.
- Wang, S., Taha, A.F., Chakrabarty, A., Sela, L., Abokifa, A.A., 2022. Model Order Reduction for Water Quality Dynamics. *Water Resour. Res.* 58 (4), e2021WR029856.
- Wu, J., Poloczek, M., Wilson, A.G., Frazier, P.I., 2017. Bayesian optimization with gradients. *Adv. Neural Inf. Process. Syst.* 2017-Decem (3), 5268–5279.