Revisiting the Water Quality Sensor Placement Problem: Optimizing Network Observability and State Estimation Metrics

Ahmad F. Taha[†,c], Shen Wang[†], Yi Guo[‡], Tyler H. Summers[‡], Nikolaos Gatsis[†], Marcio H. Giacomoni[§], and Ahmed A. Abokifa[*]

Abstract—**Real-time water quality (WQ) sensors in water distribution networks (WDN) have the potential to enable network-wide observability of water quality indicators, contamination event detection, and closed-loop feedback control of WQ dynamics. To that end, prior research has investigated a wide range of methods that guide the geographic placement of WQ sensors. These methods assign a metric for fixed sensor placement (SP) followed by *metric-optimization* to obtain optimal SP. These metrics include minimizing intrusion detection time, minimizing the expected population and amount of contaminated water affected by an intrusion event. In contrast to the literature, the objective of this paper is to provide a computational method that considers the overlooked metric of state estimation and network-wide observability of the WQ dynamics. This metric finds the optimal WQ sensor placement that minimizes the state estimation error via the Kalman filter for noisy WQ dynamics—a metric that quantifies WDN observability. To that end, the state-space dynamics of WQ states for an entire WDN are given and the observability-driven sensor placement algorithm is presented. The algorithm takes into account the time-varying nature of WQ dynamics due to changes in the hydraulic profile—a collection of hydraulic states including heads (pressures) at nodes and flow rates in links which are caused by a demand profile over a certain period of time. Thorough case studies are given, highlighting key findings, observations, and recommendations for WDN operators. Github codes are included for reproducibility.**

I. INTRODUCTION AND LITERATURE REVIEW

In dynamic infrastructure sciences, the sensor placement (SP) problem is concerned with the time-varying selection or one-time placement of sensors, while optimizing desired objective functions. This problem exists widely in dynamic networks such as transportation systems, electric power systems, and water distribution networks (WDN). The optimal placement of water quality (WQ) sensors is a crucial issue in WDN due to the dangers brought by accidental or intentional contamination, the expensiveness of sensors and their installation cost, and their potential in performing real-time feedback control of water quality—control that requires high-frequency WQ sensor data.

WQ sensor placement in WDN serves different purposes. The high-level one is minimizing the potential public health impacts of a contamination incident given a limited number of sensors. To quantify this high-level objective, the WQ literature considers various mathematical objectives and metrics. Specifically, the SP problem has been studied in [1], [2], [3], [4], [5], [6], [7] considering different contamination risks, optimization objectives, optimization formulations, uncertainty, the solution methodology and its computational feasibility, and the use of mobile sensors. Rathi and Gupta [8] classify methodologies from over forty studies into two categories as single- and multi-objective SP problem. Two other comprehensive surveys focusing on optimization strategies are also conducted in [9], [10].

As we mentioned, the most common objective of sensor placement in WDN is to minimize the potential public health caused by contamination incident, and it can be formulated as maximizing the coverage of water with a minimum number of sensors. Lee and Deininger introduce the concept of "Demand Coverage" and solve the problem using a mixed integer programming (MIP) method [11]. Kumar et al. [12] and Kansal et al. [13] propose heuristic methods to find optimal sensor location one by one by selecting one optimal location first and then selecting the next location by modifying the coverage matrix. To consider nodes with lower water quality, Woo et al. modify the objective by placing weights for each term and normalizing the concentrations [14]. Alzahrani et al. [15] and Afshar and Marino [16] use genetic algorithm (GA) and ant colony optimization (ACO) respectively to find optimal placement strategy to maximize the demand coverage. Ghimire et al. [17] and Rathi and Gupta [18] also suggested heuristic methods to solve the problem.

We briefly summarize the more recent literature on this problem followed by identifying the key research gap. Recently, He *et al.* [19] propose a multi-objective SP method to explicitly account for contamination probability variations. Hooshmand *et al.* [20] address SP problem with the identification criterion assuming that a limited sensor budget is available, followed by minimizing the number of vulnerable nodes using mixed integer programming (MIP). A combined management strategy for monitoring WDN is proposed in [21] based on the application of water network partitioning and the installation of WQ sensors. Winter *et al.* [22] investigate optimal sensor placements by intro-

ducing two greedy algorithms in which the imperfection of sensors and multiple objectives are taken into account. Giudicianni *et al.* [23] present a method that relies on a priori clustering of the WDN and on the installation of WQ sensors at the most central nodes of each cluster—selected according to different topological centrality metrics. Hu *et al.* [24] propose a customized genetic algorithm to solve multi-objective SP in WDN. Based on graph spectral techniques that take advantage on spectrum properties of the adjacency matrix of WDN graph, a sensor placement strategy is discussed in Di Nardo *et al.* [25]. Different objective functions leads to different placement strategies, and Tinelli *et al.* [26] discuss the impact of objective function selection on optimal placement of sensors. Zhang *et al.* [27] investigate the global resilience considering all likely sensor failures that have been rarely explored.

The research community thoroughly investigated water quality sensor placement strategies considering various socio-technical objectives (as briefly discussed above). The objective of this paper is *not* to develop a computational method to solve such SP problems with the aforementioned metrics/objectives. The objective herein is to find optimal SP of water quality sensors considering an overlooked, yet significant metric: the state observability and estimation metric jointly with Kalman filtering.* In short, this metric maps sensor placements given a fixed hydraulic profile to a scalar value to be minimized. This value quantifies the observability of unmeasured WQ states (i.e., concentrations of chlorine) in the entire water network. The observability quantification metric is depicted as a state estimation error measuring the difference between the actual WQ states and their estimates. Accordingly, this proposed metric finds the optimal WQ sensor placement that minimizes the state estimation error via the vintage Kalman filter for noisy WQ dynamics and measurement models.

To the best of our knowledge, this is the first attempt to find the optimal sensor placement jointly with optimizing Kalman filter performance for WQ dynamics. The most related research is the *ensemble Kalman filter*-based techniques by Rajakumar *et al.* [28], where the authors explore the impact of sensor placement on the final state estimation performance. However, the study *(i)* does not provide sensor placement strategy, *(ii)* mainly focuses on estimating water quality states and reaction parameters, and *(iii)* a dynamic model for WQ is not present to guide optimal SP. To that end, the objective of this study is to provide a control- and network-theoretic method that determines the optimal geographic placements of water quality sensors while optimizing the Kalman filter performance. The specific paper contributions are:

- The state-space, control-theoretic dynamics depicting the evolution of WQ states, i.e., concentrations of chlorine, are shown. Specifically, we are modeling and

tracking chlorine concentrations as a surrogate for contamination—this has been showcased in various studies depicting rapid depletion of chlorine upon the introduction of contaminants [29]. The dynamics of chlorine concentrations are represented as a time-varying state-space model. This model is then utilized to formulate the water quality sensor placement (WQSP) problem that optimizes the Kalman filter state estimation performance. This formulation *(i)* takes into account and builds a mapping between a WDN observability metric and the performance of Kalman filter, and *(ii)* is a set function optimization (an optimization problem that takes sets as variables) that is difficult to solve for large networks.

- To account for the time-varying nature of the dynamic WQ model (due to the changes in the hydraulic profiles that are caused by demand profiles), an algorithm that computes a sensor placement for the most common hydraulic profiles is presented. Furthermore, scalability of this algorithm is investigated. The algorithm is based on an important theoretical feature for set function optimization called submodularity. This feature has been identified in recent control-theoretic studies for sensor placement strategies [30], [31]. In particular, the developed approach is based on a greedy algorithm which returns a suboptimal placement strategy with guarantees on the distance to optimality. An efficient implementation of the algorithm is also presented. Compared to [30], [31], the proposed algorithm takes into account the time-varying nature of WQ dynamics.

- Thorough case studies on three water distribution networks under different conditions are presented. The case studies consider varying scales of water networks, significant demand variability, different number of allocated sensors, and their impact on the state estimation performance and WQSP solution. Important observations and recommendations for water system operators are given. Github codes are included for reproducibility.

The rest of the paper is organized as follows. Section II introduces network-oriented water quality dynamic model by presenting the models of each component in detail. An abstract, linear, state-space format for the water quality model is given first considering the first-order reaction model with known reaction rate coefficients. WQ observability and its metric (observability Gramian) are introduced in Section III, and then WQSP problem is formulated and solved by taking advantage of submodularity property of set function optimization in Section IV. A scalable implementation of the problem is showcased. Section V presents case studies to support the computational algorithms. Appendix A outlines components of the scalable implementation of the presented computational methods. The notation for this paper is introduced next.

Paper's Notation Italicized, boldface upper and lower case characters represent matrices and column vectors: a is a scalar, $\boldsymbol{a}$ is a vector, and $\boldsymbol{A}$ is a matrix. Matrix $\boldsymbol{I}_n$ denotes a identity square matrix of dimension n-by-

---

*In dynamic systems, the Kalman filter is a widely used algorithm that computes unmeasured state estimates of a system given a dynamic model and data from sensor measurements subject to noise.

n, whereas $\mathbf{0}_{m \times n}$ denotes a zero matrix with size m-by-n. The notations $\mathbb{R}$ and $\mathbb{R}_{++}$ denote the set of real and positive real numbers. The notations $\mathbb{R}^n$ and $\mathbb{R}^{m \times n}$ denote a column vector with n elements and an m-by-n matrix in $\mathbb{R}$. For any two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with same number of columns, the notation $\{\boldsymbol{A}, \boldsymbol{B}\}$ denotes $[\boldsymbol{A}^\top \ \boldsymbol{B}^\top]^\top$. For a random variable $\boldsymbol{x} \in \mathbb{R}^n$, $\mathbb{E}(\boldsymbol{x})$ is its expected value, and its covariance is denoted by $\mathbb{C}(\boldsymbol{x}) = \mathbb{E}\left((\boldsymbol{x} - \mathbb{E}(\boldsymbol{x}))(\boldsymbol{x} - \mathbb{E}(\boldsymbol{x}))^\top\right)$.

II. State-Space Water Quality Dynamic Model

We model WDN by a directed graph $\mathcal{G} = (\mathcal{W}, \mathcal{L})$. Set $\mathcal{W}$ defines the nodes and is partitioned as $\mathcal{W} = \mathcal{J} \bigcup \mathcal{T} \bigcup \mathcal{R}$ where $\mathcal{J}$, $\mathcal{T}$, and $\mathcal{R}$ are collection of junctions, tanks, and reservoirs. For the i-th node, set $\mathcal{N}_i$ collects its neighboring nodes (any two nodes connected by a link) and is partitioned as $\mathcal{N}_i = \mathcal{N}_i^{\text{in}} \bigcup \mathcal{N}_i^{\text{out}}$, where $\mathcal{N}_i^{\text{in}}$ and $\mathcal{N}_i^{\text{out}}$ are collection of inflow and outflow nodes. Let $\mathcal{L} \subseteq \mathcal{W} \times \mathcal{W}$ be the set of links, and define the partition $\mathcal{L} = \mathcal{P} \bigcup \mathcal{M} \bigcup \mathcal{V}$, where $\mathcal{P}$, $\mathcal{M}$, and $\mathcal{V}$ represent the collection of pipes, pumps, and valves. In this paper, we the use Lax-Wendroff scheme [32] to space-discretize pipes and each pipe with length L is split into s_L segments. The number of junctions, reservoirs, tanks, pipes, pumps and valves is denoted as n_J, n_R, n_{TK}, n_P, n_M, and n_V. Hence the number of nodes and links are $n_N = n_J + n_R + n_{TK}$ and $n_L = n_P \cdot s_L + n_M + n_V$.

The principal component of the presented state-space, control-theoretic water quality modeling is a state-vector defining the concentrations of the disinfectant (chlorine) in the network. Concentrations at nodes such as junctions, reservoirs, and tanks are collected in vector $\boldsymbol{c}_N \triangleq \{\boldsymbol{c}_J, \boldsymbol{c}_R, \boldsymbol{c}_T\}$; concentrations at links such as pipes and pumps are collected in $\boldsymbol{c}_L \triangleq \{\boldsymbol{c}_P, \boldsymbol{c}_M, \boldsymbol{c}_V\}$. We define WQ state $\boldsymbol{x}(t) \triangleq \boldsymbol{x}$ at time t as $\boldsymbol{x}(t) = \{\boldsymbol{c}_N, \boldsymbol{c}_L\} = \{\boldsymbol{c}_J, \boldsymbol{c}_R, \boldsymbol{c}_T, \boldsymbol{c}_P, \boldsymbol{c}_M, \boldsymbol{c}_V\} \in \mathbb{R}^{n_x}, n_x = n_N + n_L$.

We also make two assumptions: (i) the mixing of the solute is complete and instantaneous at junctions and in tanks with a continuously stirred tank reactors (CSTR) model [33], and (ii) the first-order reaction for single-species that describes disinfectant decay both in the bulk flow and at the pipe wall are assumed herein. The assumptions are widely used in the literature [34], [35], [36].

A. Conservation of mass

The water quality model represents the movement of all chemical and/or microbial species (contaminant, disinfectants, DBPs, metals, etc.) within a WDN as they traverse various components of the network. Specifically, we are considering the single-species interaction and dynamics of chlorine. This movement or time-evolution is based on three principles: (i) mass balance in pipes, which is represented by chlorine transport in differential pipe lengths by advection in addition to its decay/growth due to reactions; (ii) mass balance at junctions, which is represented by complete and instantaneous mixing of all inflows, that is the concentration of chlorine in links
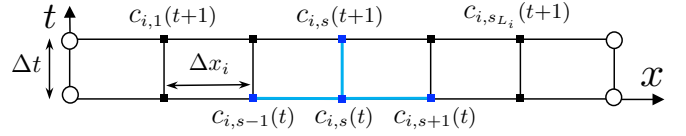
Fig. 1. Time-space discretization of Pipe i based on the L-W scheme.

flowing into this junction; and (iii) mass balance in tanks, which is represented by a continuously stirred tank reactors (CSTRs) [33] model with complete and instantaneous mixing and growth/decay reactions. The modeling of each component is introduced next.

1) *Chlorine transport and reaction in pipes:* The water quality modeling for pipes involves modeling the chlorine transport and reaction by 1-D advection-reaction (A-R) equation. For any Pipe i, the 1-D A-R model is given by a PDE:

$$\partial_t c_P = -v_i(t) \partial_x c_P + r_i c_P, \tag{1}$$

where $v_i(t)$ is flow velocity, r_i is the first-order reaction rate and certain constant, which is related with the bulk and wall reaction rate and mass transfer coefficient between the bulk flow and the pipe wall [35], [34]. Here, the Lax-Wendroff (L-W) scheme [32] shown in Fig. 1 is used to approximate the solution of the PDE (1) in space and time. This model has been used and accepted in the literature [34], [37], [38]. Pipe i with length L_i is split into s_{L_i} segments, and the discretized form for segment s is given by

$$c_{i,s}(t + \Delta t) = \underline{\alpha} c_{i,s-1}(t) + (\alpha + r_i) c_{i,s}(t) + \bar{\alpha} c_{i,s+1}(t), \tag{2}$$

where L-W coefficients for previous, current, and next segment are $\underline{\alpha} = 0.5\beta(1 + \beta)$, $\alpha = 1 - \beta^2$, and $\bar{\alpha} = -0.5\beta(1 - \beta)$. Note that $\beta \in (0, 1]$ for Pipe i at time t is a constant related with stability condition of L-W scheme, and can be decided by $v_i(t) \Delta t (\Delta x_i)^{-1}$, where Δt and Δx_i are the time step and the space-discretization step in Fig. 1. Hence, to stabilize L-W scheme, the water quality time step $\Delta t \leq \min(\Delta x_i / v_i(t))$, for all $i \in \mathcal{P}$. The L-W scheme coefficients $\underline{\alpha}$, α, and $\bar{\alpha}$ are a function of time but vary much slower than $\boldsymbol{x}(t)$, and they only change when $v_i(t)$ changes after the Δt and Δx_i are fixed. That is, they only update each hydraulic time step. Equation (2) can be lumped in a matrix-vector form for all segments s for all Pipes $i \in \mathcal{P}$ as:

$$\boldsymbol{c}_P(t + \Delta t) = \boldsymbol{A}_P(t) \boldsymbol{c}_P(t) + \boldsymbol{A}_N(t) \boldsymbol{c}_N(t), \tag{3}$$

where matrices $\boldsymbol{A}_P$ and $\boldsymbol{A}_N$ map the scalar equation (2) into the vector form (3). The Github codes of this paper [39] shows how these matrices are computed.

2) *Chlorine mass balance at junctions:* Mass conservation of the disinfectant (i.e., chlorine) for Junction i at time t can be described by

$$\sum_{k=1}^{|\mathcal{N}_i^{\text{in}}|} q_{ki}(t) c_{ki}(t) = d_i(t) c_i(t) + \sum_{j=1}^{|\mathcal{N}_i^{\text{out}}|} q_{ij}(t) c_{ij}(t),$$

where $\{ki : k \in \mathcal{N}_i^{\text{in}}\}$ and $\{ij : j \in \mathcal{N}_i^{\text{out}}\}$ represent the sets of links with inflows and outflows of Junction i; d_i is its

demand; $q_{ki}(t)$ and $q_{ij}(t)$ are the flow rate in Links ki and ij; $c_{ki}(t)$ and $c_{ij}(t)$ are the corresponding concentrations. Specifically, when links are pipes, $c_{ki}(t)$ and $c_{ij}(t)$ should be the last and first segment of Pipes ki and ij. The matrix form when considering all junctions is given as

$$c_{\mathrm{J}}(t + \Delta t) = \boldsymbol{A}_{\mathrm{J}}(t)\boldsymbol{c}_{\mathrm{J}}(t) + \boldsymbol{A}_{\mathrm{L}}(t)\boldsymbol{c}_{\mathrm{L}}(t). \qquad (4)$$

3) Chlorine mass balance at tanks: Akin to dealing with junctions, we can express the mass balance equations for each tank, the details are similar and omitted for brevity and ease of exposition. With that in mind, the provided Github codes present all of the necessary details that are required to arrive at the high-fidelity state-space description. We directly give the matrix form of all tanks as

$$c_{\mathrm{T}}(t + \Delta t) = \boldsymbol{A}_{\mathrm{T}}(t)\boldsymbol{c}_{\mathrm{T}}(t) + \boldsymbol{A}'_{\mathrm{P}}(t)\boldsymbol{c}_{\mathrm{P}}(t), \qquad (5)$$

where $\boldsymbol{A}_{\mathrm{T}}$ is in terms of tank volumes $\boldsymbol{V}_{\mathrm{T}}$, time step Δt, and flow rates flowing in or out of tanks.

4) Chlorine mass balance at reservoirs: Without loss of generality, it is assumed that the chlorine sources are only located at reservoirs, and the concentration at a reservoir is constant. That is,

$$c_{\mathrm{R}}(t + \Delta t) = \boldsymbol{c}_{\mathrm{R}}(t). \qquad (6)$$

5) Chlorine transport in pumps and valves: We consider that the *lengths* of pumps to be null, i.e., the distance between its upstream node and downstream node is zero, and hence they neither store any water nor are discretized into different segments. Therefore, the concentrations at pumps or valves equal the concentration of upstream nodes (a reservoir) they are connecting. That is

$$c_j(t + \Delta t) = c_i(t + \Delta t) = c_i(t) = c_j(t), i \in \mathcal{K}, j \in \mathcal{M},$$

and the corresponding matrix form for pumps is

$$\boldsymbol{c}_{\mathrm{M}}(t + \Delta t) = \boldsymbol{c}_{\mathrm{M}}(t). \qquad (7)$$

As for valves installed on pipes, it is simply treated as a segment of that pipe. In this case, the concentration in valves equals the segment concentrations in pipes. We next show how these matrix forms can yield state-space formulation of water quality modeling.

B. Water quality modeling in state-space form

The briefly summarized water quality model of each component from the previous section can be written as a state-space Linear Difference Equation (LDE) as in (8) where $\boldsymbol{I}$ is an identity matrix of appropriate dimension. For the ease of exposition, we consider that $\Delta t = 1\,\mathrm{sec}$ and the time-index t is replaced with another time-index k. The state-space form of the water quality model is presented as a linear time-variant (LTV) system:

$$\boldsymbol{x}(k+1) = \boldsymbol{A}(k)\boldsymbol{x}(k) + \boldsymbol{w}(k), \ \ \boldsymbol{y}(k) = \boldsymbol{C}\boldsymbol{x}(k) + \boldsymbol{v}(k), \ (9)$$

where $\boldsymbol{x}(k) \in \mathbb{R}^{n_x}$ is the state vector defined above; $\boldsymbol{y}(k) \in \mathbb{R}^{n_y}$ represents a vector of data from WQ sensors; $\boldsymbol{w}(k) \in \mathbb{R}^{n_x}$ and $\boldsymbol{v}(k) \in \mathbb{R}^{n_y}$ are the process and

$$\boldsymbol{x}(t+\Delta t) = \underbrace{\begin{bmatrix} \boldsymbol{A}_{\mathrm{J}} & & & & \boldsymbol{A}_{\mathrm{L}} & \\ & \boldsymbol{I} & & & & \\ & & \boldsymbol{A}_{\mathrm{T}} & \boldsymbol{A}'_{\mathrm{P}} & & \\ & \boldsymbol{A}_{\mathrm{N}} & & \boldsymbol{A}_{\mathrm{P}} & & \\ & & & & & \boldsymbol{I} \end{bmatrix}}_{\boldsymbol{A}(t)} \underbrace{\begin{bmatrix} \boldsymbol{c}_{\mathrm{J}} \\ \boldsymbol{c}_{\mathrm{R}} \\ \boldsymbol{c}_{\mathrm{T}} \\ \hline \boldsymbol{c}_{\mathrm{P}} \\ \boldsymbol{c}_{\mathrm{M}} \end{bmatrix}}_{\boldsymbol{x}(t)} \begin{matrix} \\ c_{\mathrm{N}} \\ \\ c_{\mathrm{L}} \\ \end{matrix}. \qquad (8)$$

measurement noise; $\boldsymbol{C} \in \mathbb{R}^{n_y \times n_x}$ is a matrix depicting the location of the placed sensors where $n_y \ll n_x$. We note the following. First, although the argument of the state-space matrices $\boldsymbol{A}(k)$ is in terms of k, this is somewhat of an abuse for the notation seeing that $\boldsymbol{A}(k)$ encodes the hydraulic profile (heads and flow rates) that does not change with the same frequency as the water quality states $\boldsymbol{x}(k)$. Hence, the state-space model (9) is time varying as system matrix $\boldsymbol{A}(k)$ changes for different hydraulic simulation, but remains the same $\boldsymbol{A}$ in a single simulation. Second, and without loss of generality, the input vector from booster stations is implicitly embedded within the state-space matrix $\boldsymbol{A}$. Third, for all $k \geq 0$, it is assumed that initial condition, process noise $\boldsymbol{w}(k)$ and the measurement noise $\boldsymbol{v}(k)$ are uncorrelated and the noise variance from each sensor is σ^2. Finally, we like to point out that extensive details for the above state-space model can be studied from our recent work on model predictive control of water quality dynamics [40].

III. Observability Metrics for WQ Dynamics

The objective of this section is two-fold. First, to introduce water system engineers and researchers to control-theoretic approaches for ensuring or optimizing the observability of the water quality dynamics. Herein, observability is defined as the ability to estimate water quality model states $\boldsymbol{x}(k)$ from available measurements $\boldsymbol{y}(k)$ via a state estimation routine. This provides situational awareness for the operator given data from few water quality sensors. Second, to define a simple observability metric that maps the number and location of fixed sensors to a scalar metric acting as a proxy for the state estimation.

A. Metrics for observability and its interpretations

In dynamic systems theory, observability is a measure of how the system state vector $\boldsymbol{x}(k) \in \mathbb{R}^{n_x}$ can be inferred from knowledge of its output $\boldsymbol{y}(k) \in \mathbb{R}^{n_y}$ over either finite- or infinite-time horizons. In particular, given sensor data $\boldsymbol{y}(0), \boldsymbol{y}(1), \ldots, \boldsymbol{y}(k_f - 1)$ for finite $k_f = k_{final}$ time-steps, observability is concerned with reconstructing or estimating the initial unknown state vector $\boldsymbol{x}(0)$ from the k_f measurements, and subsequently computing $\boldsymbol{x}(1), \ldots, \boldsymbol{x}(k_f)$ assuming noiseless system. Accordingly, a linear dynamic system (such as the water quality model (9)) is observable if and only if the observability matrix for k_f time-steps

$$\mathcal{O}(k_f) = \{\boldsymbol{C}, \boldsymbol{C}\boldsymbol{A}, \ldots, \boldsymbol{C}\boldsymbol{A}^{k_f - 1}\} \in \mathbb{R}^{k_f n_y \times n_x}$$

is full column rank [41], i.e., $\text{rank}(\mathcal{O}(k_f)) = n_x$ assuming that $k_f n_y > n_x$. In this section, and for brevity, we assume that the hydraulic variables are not changing during each hydraulic simulation period and hence $\boldsymbol{A}(k) = \boldsymbol{A}$. With that in mind, the proposed sensor placement formulations considers changing hydraulic simulations.

For the infinite-time horizon case with $k_f = \infty$ (that is, data has been collected over a long period of time), a system is observable if and only if the observability matrix $\mathcal{O}(k_f = n_x) \in \mathbb{R}^{n_x n_y \times n_x}$ is full column rank [41]. However, observability is a binary metric—it cannot indicate *how observable* a dynamic system is. Due to the complexity and dimension of the water quality model (9), this dynamic model is *not* observable, i.e., it fails the aforementioned rank condition for various water networks and hydraulic simulation profiles. Specifically, it is virtually impossible to accurately reconstruct all chlorine concentrations (states $\boldsymbol{x}(k)$) unless water quality sensors are ubiquitously available and widespread in the network, i.e., installed at each junction.

To that end, a more elaborate, non-binary quantitative metric for observability is needed for the water quality model and the sensor placement problem. One metric is based on the *observability Gramian* [41] defined as the k_f sum of matrices

$$\boldsymbol{W}(k_f) = \sum_{\tau=0}^{k_f} \left(\boldsymbol{A}^\top\right)^\tau \boldsymbol{C}^\top \boldsymbol{C} \boldsymbol{A}^\tau.$$

The system is observable at time-step k_f if matrix $\boldsymbol{W}(k_f)$ is nonsingular and is unobservable if $\boldsymbol{W}(k_f)$ is singular. Similarly, this definition extends for the infinite-horizon case with $k_f = \infty$. However, $\boldsymbol{W}$ is still a matrix and the aforementioned observability-singularity discussion is still binary. As a result, various non-binary metrics have been explored in the literature [42], [43]. This includes: the minimum eigenvalue $\lambda_{\min}(\boldsymbol{W})$, the log determinant $\log \det(\boldsymbol{W})$, the trace$(\boldsymbol{W})$, and the sums or products of the first m eigenvalues $\lambda_1, \ldots, \lambda_m$ of $\boldsymbol{W}$. These metrics differ in their practical application, interpretation, and theoretical properties; the reader is referred to [43] for a thorough discussion. In this paper, we utilize the $\log \det(\boldsymbol{W})$ metric due to various reasons outlined in the ensuing sections, but the formulations presented in the paper can be extend to other metrics.

B. Metrics for water quality observability matrix

In this section, we provide a discussion on the utilized metric for observability for the sensor placement problem. To do so, we consider the time-invariant state-space matrices for a single hydraulic simulation $k \in [0, k_f]$ which is also a single instant of hydraulic simulation and demand profile. That is, to ease the ensuing exposition we assume that the state-space matrix $\boldsymbol{A}(k) = \boldsymbol{A}$ is fixed rather than being time-varying (the actual methods consider time-varying demand pattern). The objective of this section is to formulate a water quality observability metric that maps collection of water quality data $\boldsymbol{y}(k)$ from a specific

number of sensors n_y to a scalar observability measure under the noise from water quality dynamics and measurement models.

First, consider the augmented measurement vector $\bar{\boldsymbol{y}}(k_f) \triangleq \{\boldsymbol{y}(0), \ldots, \boldsymbol{y}(k_f)\}$ for $k_f + 1$ time-steps. Given (9), this yields:

$$\bar{\boldsymbol{y}}(k_f) = \underbrace{\begin{bmatrix} \boldsymbol{C} & & & \\ \boldsymbol{C}\boldsymbol{A} & \boldsymbol{C} & & \\ \vdots & \vdots & \boldsymbol{C} & \\ \boldsymbol{C}\boldsymbol{A}^{k_f} & \boldsymbol{C}\boldsymbol{A}^{k_f-1}\cdots & & \boldsymbol{C} \end{bmatrix}}_{\mathcal{O}(k_f)} \underbrace{\begin{bmatrix} \boldsymbol{x}(0) \\ \boldsymbol{w}(0) \\ \vdots \\ \boldsymbol{w}(k-1) \end{bmatrix}}_{\boldsymbol{z}(k_f)} - \underbrace{\begin{bmatrix} \boldsymbol{v}(0) \\ \boldsymbol{v} \\ \vdots \\ \boldsymbol{v}(k) \end{bmatrix}}_{\bar{\boldsymbol{v}}(k_f)}, \quad (10)$$

where $\boldsymbol{z}(k_f)$ lumps initial unknown state $\boldsymbol{x}_0 = \boldsymbol{x}(0)$ and process noise $\boldsymbol{w}(k_f)$, and $\boldsymbol{v}(k_f)$ collects all measurement noise. Note that the left-hand side of (10) is known, whereas vectors $\boldsymbol{z}(k_f)$ and $\boldsymbol{v}(\cdot)$ are unknown vectors. To that end, the problem of estimating $\boldsymbol{z}(k_f) \triangleq \boldsymbol{z} \in \mathbb{R}^{n_z}$, where $n_z = (k+1)n_x$, is important to gain network-wide observability of water quality state $\boldsymbol{x}$ which will guide the real-time estimation. As a probabilistic surrogate to estimating $\boldsymbol{z}$, we utilize the minimum mean square estimate (MMSE) defined as $\mathbb{E}(\boldsymbol{z} - \hat{\boldsymbol{z}})$, and its corresponding posterior error covariance matrix $\Sigma_{\boldsymbol{z}}$. These two quantities provide estimates of means and variances of the unknown variable $\boldsymbol{z}(k_f)$. Interestingly, these can be written in terms of the sensor noise variance σ^2, the collected sensor data $\bar{\boldsymbol{y}}(k_f)$, the observability-like matrix $\mathcal{O}(k_f)$ in (10), and the expectation and covariance of the unknown variable $\boldsymbol{z}(k_f)$ given by

$$\mathbb{E}(\boldsymbol{z}(k_f)), \ \mathbb{C}(\boldsymbol{z}(k_f)) = \mathbb{E}\left((\boldsymbol{z} - \mathbb{E}(\boldsymbol{z}))(\boldsymbol{z} - \mathbb{E}(\boldsymbol{z}))^\top\right)$$

Given these developments, and to guide the sensor placement problem formulation, a metric is needed to map the covariance matrix $\Sigma_{\boldsymbol{z}}$ to a *scalar* value. In particular, the metric $\log \det(\Sigma_{\boldsymbol{z}})$, which maps an n_z-by-n_z matrix Σ_z to a scalar value, can be used to achieve that. Fortunately, $\log \det(\Sigma_{\boldsymbol{z}})$ has a closed form expression given by:

$$\log \det(\Sigma_{\boldsymbol{z}}) = 2n_z \log(\sigma) - \log \det\left(\sigma^2 \mathbb{C}^{-1}(\boldsymbol{z}) + \boldsymbol{W}_o\right) \quad (11)$$

where $\boldsymbol{W}_o = \mathcal{O}^\top(k_f)\mathcal{O}(k_f)$. The reader is referred to [30] for the derivation of (11). We note the following: *(i)* the closed-form expression of $\log \det(\Sigma_{\boldsymbol{z}})$ in (11) assumes a *fixed* sensor placement while associating a scalar measure of water quality observability given a collection of sensor data and the system's parameters. This closed form expression is rather too complex to be incorporated within a sensor placement formulation and does not allow for near real-time state estimation. The next section discusses simple solutions to these issues. *(ii)* We use the $\log \det(\cdot)$ metric here as it is endowed with desirable theoretical properties (namely super/sub-modularity) that makes it amenable to large-scale networks, it exhibits a closed-form expression as in (11), and has been used in various sensor placement studies in the literature. With that in mind, other metrics can be used including the trace operator.

C. Relationship with the Kalman filter

The above discussions yield a metric that can be used for quantifying observability of the water quality model (9), in addition to probabilistically estimating the unknown, initial state vector $\boldsymbol{x}(0)$. A relevant problem is the real-time state estimation via the Kalman filter, which essentially reconstructs or estimates in real-time states $\boldsymbol{x}(k)$ from output measurements $\boldsymbol{y}(k)$. This is in contrast with the batch state estimation as in (10). While the initial state estimation problem discussed in the previous section provides a starting point for reconstructing $\boldsymbol{x}$, the Kalman filter presents a more general approach to the estimation problem. In fact, ignoring the process noise $\boldsymbol{w}$ and setting variances of sensor data to $\sigma^2 = 1$, the Kalman filter becomes equivalent to a real-time version of the above probabilistic estimator. Most importantly, the metric $\log \det(\cdot)$ degenerates to:

$$\log \det (\Sigma_{\boldsymbol{z}}) = -\log \det(\boldsymbol{I}_{n_x} + \boldsymbol{W}(k_f)) \tag{12}$$
$$= -\log \det \left(\boldsymbol{I}_{n_x} + \sum_{\tau=0}^{k_f} \left(\boldsymbol{A}^{\top}\right)^{\tau} \boldsymbol{C}^{\top} \boldsymbol{C} \boldsymbol{A}^{\tau} \right)$$

where $\boldsymbol{I}_{n_x}$ is an identity matrix of size n_x. This is shown in the recent control theoretic literature [44], [31]. In short, this is a simple metric that maps the number of installed or placed sensors (i.e., number of rows of matrix $\boldsymbol{C}$) a metric that defines the quality of the state estimates. When no sensor is installed or $\boldsymbol{C}$ is a zero matrix, the observability Gramian $\boldsymbol{W}(k_f)$ is also a zero matrix, and intuitively the $\log \det(\cdot)$ metric defined above the maximum error of 0. When the network is fully sensed — that is $n_y = n_x$, $\boldsymbol{C} = \boldsymbol{I}_{n_x}$, and all states are measured — then $\boldsymbol{W}(k_f) = \boldsymbol{I}_{n_x} + \boldsymbol{A} + \ldots + \boldsymbol{A}^{k_f}$ and the smallest error is achieved.

Building on that, the control theoretic literature thoroughly investigated bounds for the estimation error and the corresponding metric with respect to the number of sensors; see [30, Theorem 2]. The objective of this paper is to build on these developments and investigate how such metric relates with the performance of the Kalman filter. The next section formulates the water quality sensor problem using the introduced metric.

IV. WATER QUALITY SENSOR PLACEMENT FORMULATION

The objective of the presented water quality sensor placement (WQSP) formulation is to minimize the error covariance of the Kalman filter while using at most r water quality sensors. In WDN, water quality sensors are installed at nodes, that is, at most r sensors are selected from the set $\mathcal{W} = \mathcal{J} \bigcup \mathcal{T} \bigcup \mathcal{R}$ where the cardinality of set $|\mathcal{W}| = n_N$, i.e., the set $\mathcal{W}$ contains n_N possible locations at various junctions, tanks, and reservoirs. This forms a sensor set $\mathcal{S} \subset \mathcal{W}$ where $|\mathcal{S}| = n_{\mathcal{S}} \leq r$. The specific geographic placement and locations of these $n_{\mathcal{S}}$ sensors are encoded in matrix $\boldsymbol{C}$ of (9) through binary indicators. In short, the presented WQSP seeks to find the optimal set

$\mathcal{S}_r^*$ that optimizes the state estimation performance with at most r WQ sensors.

The metric discussed in the previous section assumes that the state-space matrix $\boldsymbol{A}$ (encoding network and hydraulic simulation parameters) is time-varying due to varying demand and flow/head profiles. In short, the metric (12) yields a time-varying value and hence different state estimation performance for each hydraulic simulation reflected with a different $\boldsymbol{A}(k)$ matrix. As a result, considering a varying hydraulic simulation profile within the sensor placement problem is important, i.e., the sensor placement solution needs to be aware of the most probable demand and hydraulic scenarios.

Consequently, we define $\boldsymbol{D}_i \in \mathbb{R}^{n_J \times T_h k_f}, \forall i \in \{1, \ldots, n_d\}$ for all n_J junctions during T_h distinct hydraulic simulations, each lasting k_f sec. The notation $\boldsymbol{D}_{i,k}$ defines the kth column vector of matrix $\boldsymbol{D}$. Parameter n_d reflects the number of potential demand patterns; concrete examples are given in case study section. Demand profiles $\boldsymbol{D}_i \in \mathcal{D}$ essentially define the most common varying demand profiles experienced by the system operator from historical data. Each demand profile results in a different hydraulic profile and hence a different state-space matrix[†] $\boldsymbol{A}(\boldsymbol{D}_{i,k}) = \boldsymbol{A}(k)$. Given these definitions and for an a priori defined $\boldsymbol{D}_{i,k} \in \mathcal{D}$, one useful instant of the WQSP problem can be abstractly formulated as:

$$\text{WQSP:} \quad \begin{aligned} &\text{minimize} \quad f(\mathcal{S}; \boldsymbol{A}(\boldsymbol{D}_{i,k})) \\ &\text{subject to} \quad \mathcal{S} \subset \mathcal{W}, \ |\mathcal{S}| = n_{\mathcal{S}} \leq r. \end{aligned} \tag{13}$$

The design variable in the optimization problem WQSP is the location of the installed sensors reflected via set $\mathcal{S}$ defined earlier. The objective function $f(\cdot; \cdot) : \mathbb{R}^{n_{\mathcal{S}}} \times \mathbb{R}^{n_x \times n_x} \rightarrow \mathbb{R}$ maps the optimal sensor placement candidate $\mathcal{S}$ and given hydraulic demand profile $\boldsymbol{D}_{i,k}$ and its corresponding matrix $\boldsymbol{A}(\boldsymbol{D}_{i,k})$ to the state estimation, Kalman filter performance. We note that when the objective function has a set as the variable (i.e., $\mathcal{S}$ in $f(\cdot; \cdot)$), the objective function is often referred to as a *set function.* We use these terms interchangeably. In this paper, the set (objective) function takes the form of (12) which indeed takes explicitly the sensor placement set $\mathcal{S}$ through matrix $\boldsymbol{C}$ as well as the a priori known hydraulic profiles and the corresponding state-space matrices $\boldsymbol{A}(\boldsymbol{D}_{i,k})$. The constraint set of WQSP represents the number of utilized sensors and their location in the network.

For small-scale water networks, one may solve the set function optimization (13) via brute force, but this is impossible for large-scale networks—such problems are known to be an NP-hard one, i.e., a computational problem that is suspected to have no polynomial-time algorithm to optimally solve. To address this computational challenge, we resort to a widely-used approach in combinatorial optimization: exploiting special property of the set function $f(\mathcal{S}; \boldsymbol{A}(\boldsymbol{D}_{i,k}))$ via sub/super-modularity defined

---

[†]We defined $\boldsymbol{A}(k)$ earlier due to the change in the hydraulic and demand profiles. The notation $\boldsymbol{A}(\boldsymbol{D}_{i,k})$ is equivalent to $\boldsymbol{A}(k)$ but offers more clarity.

Algorithm 1: Greedy algorithm to solve WQSP problem.

Input: Number of sensors r, all demand profiles $\mathcal{D}$, water network parameters, $k = i = 1, \tilde{\mathcal{S}} = \emptyset$
Output: Optimal sensor set $\mathcal{S}^\star$

1 Compute:
$\quad A(D_{i,k}) = A, \forall i, k \in \{1, \ldots, n_d\}, \{1, \ldots, T_h k_f\}$

2 for $k \leq T_h k_f$ **do**

3   // **For each single hydraulic simulation interval** k

4 $i = 1, \bar{\mathcal{S}} = \emptyset$

5   **for** $i \leq n_d$ **do**

6   // **For each demand profile**

7 $j = 1, \mathcal{S}_j = \emptyset$

8   **while** $j \leq r$ **do**

9 $e_j \leftarrow$
$\quad\quad \text{argmax}_{e \in \mathcal{W} \setminus \mathcal{S}} [f(\mathcal{S}; A) - f(\mathcal{S} \cup \{e\}; A)]$

10 $\mathcal{S}_j \leftarrow \mathcal{S}_j \cup \{e_j\}$

11 $j \leftarrow j + 1$

12   **end while**

13 $\bar{\mathcal{S}} \leftarrow \bar{\mathcal{S}} \bigcup \mathcal{S}_j, i \leftarrow i + 1$

14   **end for**

15 $\mathcal{S}^{(k)} \leftarrow \arg\max_{\mathcal{S} \in \bar{\mathcal{S}}} f(\mathcal{S}; A)$

16 $\tilde{\mathcal{S}} \leftarrow \tilde{\mathcal{S}} \bigcup \mathcal{S}^{(k)}$

17 $k \leftarrow k + k_f$

18 end for

19 $\mathcal{S}^* \leftarrow \arg\max_{\mathcal{S} \in \tilde{\mathcal{S}}} T(\mathcal{S})$ // **Greedy-optimal sensor placement**

Fig. 2. (a) Three-node network, (b) Net1, and (c) Net3.
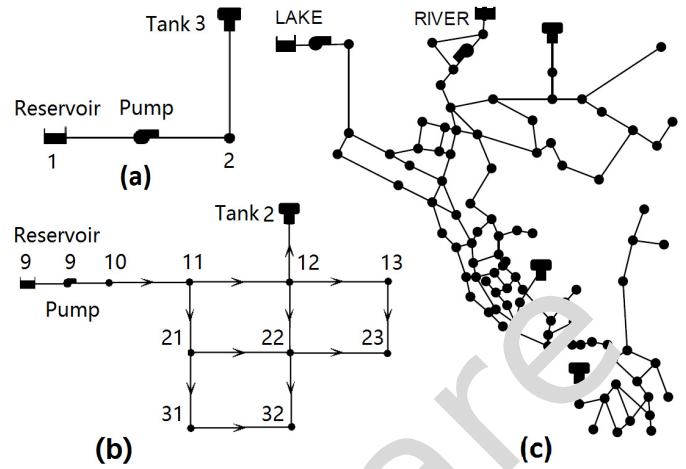
as follows.

A set function $f(\cdot)$ is submodular if and only if $f(\mathcal{A} \cup \{a\}) - f(\mathcal{A}) \geq f(\mathcal{B} \cup \{a\}) - f(\mathcal{B})$ for any subsets $\mathcal{A} \subset \mathcal{B} \subseteq \mathcal{V}$ and $\{a\} \in \mathcal{V} \setminus \mathcal{B}$. A set function $f(\cdot)$ is supermodular if $-f(\cdot)$ is submodular. Intuitively, submodularity is a diminishing returns property where adding an element to a smaller set gives a larger gain than adding one to a larger set [45].

The computational framework of submodularity of set function optimization allows one to use greedy algorithms [46] with desirable performance while being computationally tractable. Although greedy algorithms are known to return suboptimal solutions, they are also known to return excellent performance when the set function is especially sub/super-modular. Interestingly, the set function in WQSP given in (12) is indeed supermodular [30, Theorem 2]. Given this property, a vintage greedy algorithm—applied to solve the NP-hard problem WSQP—can return a solution $\mathcal{S}$ with objective function value $f(\mathcal{S})$ *at least* 63% of the optimal solution $f(\mathcal{S}^*)$ [30]. Empirically, a large body of work [30], [31], [47] shows that the solution provided by some greedy algorithms can be near-optimal, rather than being 63% optimal.

We apply a greedy algorithm to solve the WQSP for various hydraulic profiles. The details of this algorithm are given in Algorithm 1. The notation $\mathcal{S}_j$ denotes the sensor set with j placed sensors. The notation $\mathcal{S}^{(k)}$ denotes the sensor set at iteration k. The sets $\tilde{\mathcal{S}}$ and $\bar{\mathcal{S}}$ are supersets that include various sets $\mathcal{S}$. Variable $e \in \mathcal{S}$ defines an element (i.e., a junction) in the set $\mathcal{S}$. The inputs for the algorithm are the number of sensors r, all demand

profiles $D_{i,k} \in \mathcal{D}$, and WDN parameters. The output of the algorithm is greedy-optimal sensor set $\mathcal{S}^*$. The first step of the algorithm is to compute all state-space matrices $A(D_{i,k})$ for various demand profiles $D_{i,k}$. Then, given a fixed hydraulic simulation interval k, a fixed demand profile i, and fixed number of sensors j, Step 9 computes the optimal element in the set $\mathcal{W} \setminus \mathcal{S}_j$ that yields the best improvement in the set function optimization reflecting the Kalman filter performance—a core component of the greedy algorithm and supermodular optimization. At each iteration inside the while loop, the algorithm finds the optimal element e_j (i.e., the sensor through a junction ID) that results in the best improvement in the state estimation performance metric.

Then, the n_d sets $\mathcal{S}_j$ (that include the optimal sensor sets for all n_d demand profiles) are stored in a master set $\bar{\mathcal{S}}$. This is then followed by finding the optimal sensor sets from $\bar{\mathcal{S}}$ for all T_h hydraulic simulations; these are all included in another master set $\tilde{\mathcal{S}}$. Finally, the algorithm terminates by computing the final optimal sensor locations $\mathcal{S}^*$ via picking the combination that maximizes the occupation time $T(\mathcal{S})$ for all $\mathcal{S} \in \tilde{\mathcal{S}}$, i.e., a metric that defines the frequency of a specific sensor activation. Finally, we note that this algorithm returns the *greedy-optimal* solution. This solution is not necessarily the optimal solution as discussed above with the 63% optimality guarantees. Thorough case studies are given in the ensuing section.

V. Case Studies

We present three simulation examples (three-node network, Net1, and Net3 network [33], [48]) to illustrate the applicability of our approach. The three-node network is designed to illustrate the details of proposed method and help readers understand the physical meaning of results intuitively. Then, we test Net1 with looped network topology considering the impacts on final WQSP from choosing *(i)* the length of a single hydraulic simulation t, *(ii)* L-W scheme time-step Δt (or equally dynamic number of segments), *(iii)* different base demands, and *(iv)* different patterns. Net3 network is used to test scalability
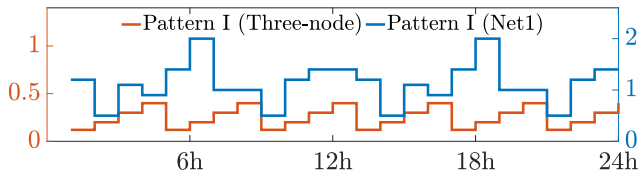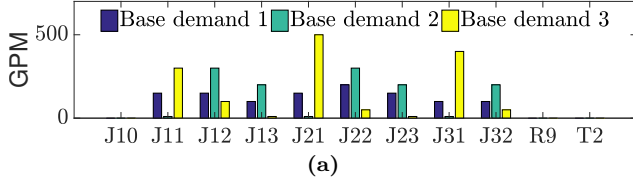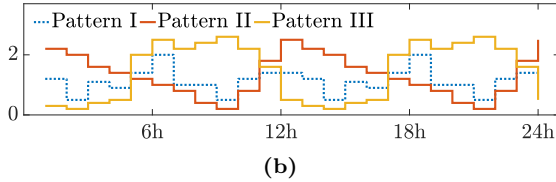
Fig. 3. Pattern for Three-node and Net1 networks.

Fig. 4. Different base demands (a) and demand patterns (b) for nodes in Net1.

Fig. 5. Sensor placement results for the three-node network (a) and Net1 (b) in 24 hours with $\kappa_f = 300\,\sec$, $\Delta t = 5\,\sec$, Pattern I, Base demand 1.

of proposed algorithm and verify our findings further. Considering that the LDE model (9) produces accurate state evolution, we eliminate the process noise and set the sensor noise standard deviation to $\sigma = 0.1$.

The simulations are performed via EPANET Matlab Toolkit [49] on Windows 10 Enterprise with an Intel(R) Xeon(R) CPU E5-1620 v3 @3.50 GHz. All code, parameters, tested networks, and results are available on Github [39] which includes an efficient and scalable implementation of Algorithm 1. The details of this implementation are included in Appendix A.

A. Three-node network

The three-node network shown in Fig. 2a includes one junction, one pipe, one pump, one tank, and one reservoir. A chlorine source ($c_1 = 0.8\,\mathrm{mg/L}$) is installed at Reservoir 1. The initial chlorine concentrations at or in the other components are 0 mg/L. Only Junction 2 consumes water, and its base demand is $d_{\mathrm{base}} = 2000$ GPM. The corresponding pattern Pattern I (viewed as a row vector) for Junction 2 in 24 hours is presented in Fig. 3. Hence, only one demand profile for a day is computed as $D = d_{\mathrm{base}} \times$ Pattern I. The pipe is split into fixed as $s_{L_{23}} = 150$ segments, and the single hydraulic simulation interval is set to $k_f = 300\,\sec$ and $T_h = 24$ hydraulic simulations are considered. To help the readers understand intuitively about the water quality modeling in state-space form and the observability (Gramian), an illustrative code including step by step comments for this small three-node network is available in our Github [39] for the convenience of readers.

For the three-node network, there are three possible sensor locations (R1, J2, and T3); therefore, r is set to 1 or 2 in Algorithm 1. The final sensor placement results are presented as Fig. 5a. When $r = 1$, J2 is the
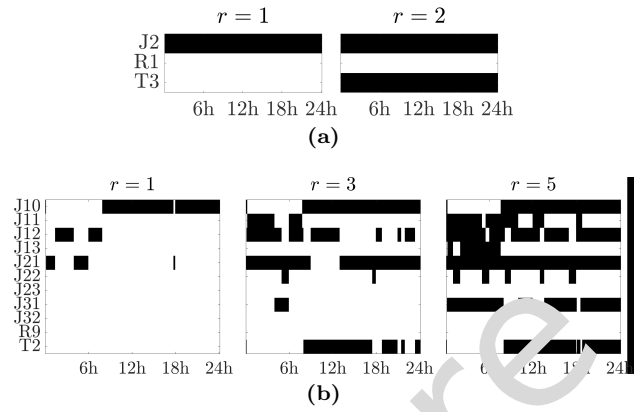
best location or the center of the network, and when $r = 2$, locations J2 and T3 are selected. To qualify the centrality or importance of a specific location during 24 hours, occupation time $T(\mathcal{S}) = \frac{\text{Selected time}}{\text{Total time}}$ is defined as a percentage of the selected time by Algorithm 1 in a day. This measure indicates the importance of the selected sensor locations. If the sensor location does not change during 24 hours, the occupation time would be 100%; see Tab. I. With that in mind, this 100% figure of sensor occupation time rarely happens for any junction in larger networks—its occurrence in the three-node network is due to its simple topology. We show more interesting results with varying occupation time in the next sections.

B. Looped Net1 network

Net1 network [33], [48] shown in Fig. 2b is composed of 9 junctions, 1 reservoir, 1 tank, 12 pipes, and 1 pump. Beyond optimal sensor placements, here we investigate the impact of the length of a single hydraulic simulation length k_f, L-W scheme time-step Δt, and the demand profile on the final sensor placement result. This network is more complex than the three-node network because its flow direction changes and flow rates or velocities vary dramatically every hour. To balance the performance of L-W scheme and computational burden, s_{L_i} for each pipe is set to an integer which is the ceiling of $\frac{L_i}{v_i(t)\Delta t}$, and dynamic number of segments setting makes $\Delta t = 5\,\sec$. Furthermore, If the parameter $\Delta t = 10\,\sec$ is needed, and this can be achieved conveniently by reducing the s_{L_i} for each pipe by half.

1) Base case scenario and its result: The base case is considered with the following settings: $\Delta t = 5\,\sec$, single hydraulic simulation $k_f = 300\,\sec$, and demand profile for a single interval $D_k =$ Base demand 1 $\times$ Pattern I shown in Fig. 4. There are 11 possible sensor locations (see Fig. 2b), and the number of sensor locations r is chosen as $[1, 3, 5]$ in (13). Similarly, we consider 24 hours in Algorithm 1. The final result is presented in Fig. 5b, and the sensor placement results in terms of occupation time T are presented in Tab. I.

TABLE I. Sensor placement results with detailed occupation time (Base case of Net1: $\Delta t = 5$ sec, $k_f = 300$ sec , Pattern I, Base demand 1).

Network	r	Result (selected positions are in blue)
Three-node	1	$T_{\text{J2}} = 100\%$
	2	$T_{\text{J2,T3}} = 100\%$
Net1 (Base case)	1	$T_{\text{J10}} = 66.4\%$, $T_{\text{J12}} = 18.6\%$, $T_{\text{J21}} = 14.8\%$
	3	$T_{\text{J10}} = 68.5\%$, $T_{\text{J12}} = 56.7\%$, $T_{\text{J21}} = 83.4\%$ $T_{\text{T2}} = 53.6\%$
	5	$T_{\text{J10}} = 69.5\%$, $T_{\text{J12}} = 87.8\%$, $T_{\text{J21}} = 100\%$ $T_{\text{T2}} = 65.7\%$, $T_{\text{J31}} = 82.1\%$, $T_{\text{J11}} = 49.4\%$

TABLE II. Sensor placement results considering the impacts of L-W scheme time-step Δt and the length of the single observation time t (Case A: $\Delta t = 10$ sec, $k_f = 300$ sec; Case B: $\Delta t = 5$ sec, $k_f = 60$ sec).

Network	r	Result (selected positions are in blue)
Net1 (Case A)	1	$T_{\text{J10}} = 71.6\%$, $T_{\text{J12}} = 12.4\%$, $T_{\text{J21}} = 14.5\%$
	3	$T_{\text{J10}} = 71.6\%$, $T_{\text{J12}} = 48.4\%$, $T_{\text{J21}} = 61.2\%$ $T_{\text{T2}} = 68.8\%$
	5	$T_{\text{J10}} = 74.7\%$, $T_{\text{J12}} = 82.7\%$, $T_{\text{J21}} = 100\%$ $T_{\text{T2}} = 69.2\%$, $T_{\text{J31}} = 73.0\%$, $T_{\text{J11}} = 62.6\%$
Net1 (Case B)	1	$T_{\text{J10}} = 86.5\%$, $T_{\text{J12}} = 8.3\%$, $T_{\text{J21}} = 5.1\%$
	3	$T_{\text{J10}} = 100\%$, $T_{\text{J12}} = 40.4\%$, $T_{\text{J21}} = 53.6\%$ $T_{\text{T2}} = 74.2\%$
	5	$T_{\text{J10}} = 100\%$, $T_{\text{J12}} = 85.7\%$, $T_{\text{J21}} = 100\%$ $T_{\text{T2}} = 78.1\%$, $T_{\text{J31}} = 39.2\%$, $T_{\text{J11}} = 36.5\%$

From Fig. 5b and Tab. I, when $r = 1$, J10 in Fig. 5b is the best sensor location most of the time ($T_{\text{J10}} = 66.4\%$) and the best location switches to J12 or J21 occasionally ($T_{\text{J12}} = 18.6\%$, $T_{\text{J21}} = 14.8\%$). Hence, the solution of WQSP is $\mathcal{S}^*_{r=1} = \{\text{J10}\}$ (marked as blue in Tab. I). Similarly, the locations with the largest r occupation time are selected as the final results when $r = 3$ and 5. These greedy-optimal placements are given by $\mathcal{S}^*_{r=3} = \{\text{J10}, \text{J12}, \text{J21}\}$ and $\mathcal{S}^*_{r=5} = \mathcal{S}^*_{r=3} \bigcup \{\text{T2}, \text{J31}\}$. This showcases supermodularity of the set function optimization, seeing that $\mathcal{S}^*_{r=3} \subset \mathcal{S}^*_{r=5}$.

2) The impacts of L-W scheme time-step and the length of single observation time: Here, we study the impact of L-W scheme time-step and the length of single observation time parameter on the final WQSP results—in comparison with the base case from the previous section. At first, only Δt is increased from 5 (from the base case) to 10 sec (Case A). Accordingly, the number of segments of all pipes is reduced by 50%, while still maintaining the accuracy of LDE state-space model compared to the EPANET water quality simulation. We also define Case B by reducing k_f from 300 sec (base case) to 60 sec. The results for this experiment are shown in Tab. II. We observe the following: *(i)* the final results are exactly the same as the ones under the base case for $r = 1, 5$, and the differences are materialized only in the slightly changed occupation time; *(ii)* the results under $r = 3$ are different from the base case as the solution changes from $\mathcal{S} = \{\text{J10}, \text{J12}, \text{J21}\}$ (base case) to $\mathcal{S} = \{\text{J10}, \text{T2}, \text{J21}\}$ (Cases A and B). This is due to the fact that the base case did not produce a clear winner in terms of the sensor placement—the occupation times ($T_{\text{J12}} = 56.7\%$, $T_{\text{T2}} = 53.6\%$) are similar.
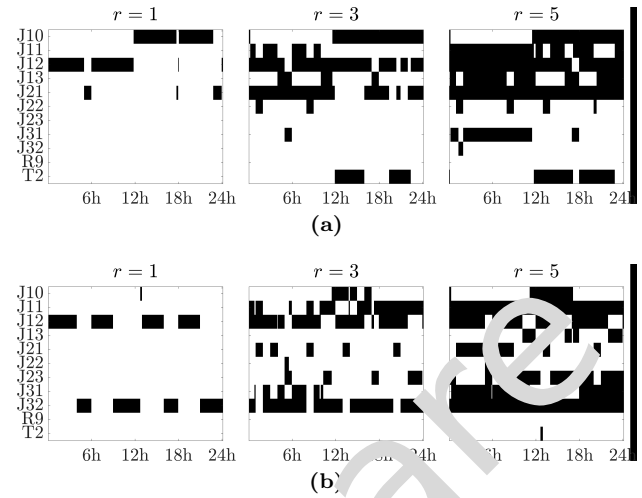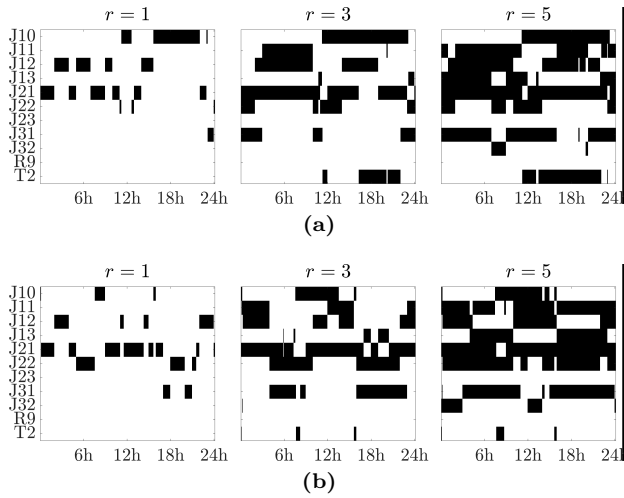

Fig. 6. Sensor placement results for Net1 with $k_f = 300$ sec, $\Delta t = 5$ sec, Pattern I under (a) base demand 2, (b) Base demand 3.

We note that even if the sensor placement strategy is changed when $r = 3$, the final performances of these three cases are comparable, and the relative error of Kalman filter performance in (13) reached between Base case and Case A (Case B) is 17.2% (7.9%) even though the difference in Δt is two times and the difference in k_f is 5 times, which is acceptable. Hence, one could make this preliminary conclusion: the impacts of L-W scheme time-step Δt and the length of hydraulic simulation k_f on the final sensor placement results are negligible assuming that the number of pipe segments (the partial differential equation space discretization parameter) is large enough to ensure the accuracy of LDE model.

3) The impact of various demand patterns: In this section, the impact of demand profiles on the final sensor placement result is explored. Note that the demand in 24 hours at a node is decided by its base demand and the corresponding patterns simultaneously. Furthermore, other demand patterns could reflect other days of the weeks such as a weekend, rather than assuming a week-long demand curve.

First, the Pattern I is fixed as the stair-shape in Fig. 3 or the dotted line in Fig. 4b, and base demands 1, 2, and 3 in Fig. 4a are used. That is, we have $n_d = 3$ different demand profiles which is an input for Algorithm 1. Note that these base demands are generated for illustrative purposes. Base demand 1 is designed to assign nearly identical base demand at each node. Base demand 2 assigns more base demands to the nodes on the right half part of the network in Fig. 2b, such as $\{\text{J12}, \text{J13}, \text{J22}, \text{J23}, \text{J32}\}$. Base demand 3 assigns larger base demands to the nodes on the left half part of the topology in Fig. 2b, such as $\{\text{J11}, \text{J21}, \text{J31}\}$.

The final sensor placement strategies under the three base demands 1, 2, and 3 are shown as Fig. 5b, Fig. 6a, and Fig. 6b, and the corresponding detailed occupation time are not shown for brevity. It can be observed that the greedy-optimal location switches from $\mathcal{S}^*_{r=1} = \{\text{J10}\}$ (under base demand 1) to $\mathcal{S}^*_{r=1} = \{\text{J12}\}$ (under base de-

(a)

(b)

Fig. 7. Sensor placement results for Net1 with $k_f = 300$ sec, $\Delta t = 5$ sec, and Base demand 1 under (a) Pattern II, (b) Pattern III.
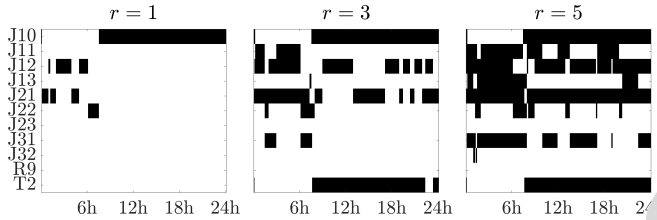
Fig. 8. Final sensor placement results for Net1 with $k_f = 300$ sec, $\Delta t = 5$ sec consider five different demand profiles (fusion of Fig. 5b, Fig. 6, and Fig. 7).

mand 3) along with changing base demand; when $r = 3$, it switches from $\mathcal{S}^*_{r=3} = \{J10, J12, J21\}$ (under base demand 1) to $\mathcal{S}^*_{r=3} = \{J11, J12, J32\}$ (under base demand 3); when $r = 5$, it switches from $\mathcal{S}^*_{r=5} = \{J10, J12, J21, J31, T2\}$ (under base demand 1) to $\mathcal{S}^*_{r=5} = \{J11, J12, J23, J31, J32\}$ (under base demand 2). This showcases changing base demands or different demand profiles indeed have an impact on the sensor placement, but Algorithm 1 still returns the best placement according to the chosen metrics.

Second, to test the impact of patterns, Patterns I, II, and III in Fig. 4b are used when base demand 1 is fixed (see Fig. 4a). We have another $n_d = 3$ different group of demand profiles. Again, these patterns are only used for illustrative purposes to test the algorithm's performance. It can be seen that Pattern I is relatively flatter compared with the other patterns, while Patterns II and III vary dramatically and are complementary to each other. The final sensor placement strategies under Patterns I, II, and III are shown as Fig. 5b, Fig. 7a, and Fig. 7b that can also be viewed as three corresponding matrices with only zeros and ones element (not selected or selected). It can be observed that the greedy-optimal location switches from $\mathcal{S}^*_{r=1} = \{J10\}$ to $\mathcal{S}^*_{r=1} = \{J21\}$ and from $\mathcal{S}^*_{r=3} = \{J10, J12, J21\}$ to $\mathcal{S}^*_{r=3} = \{J21, J22, J31\}$. With the above comparisons, we claim that both base demands and patterns would have impacts on the final sensor placement solution in Net1. In order to quantify
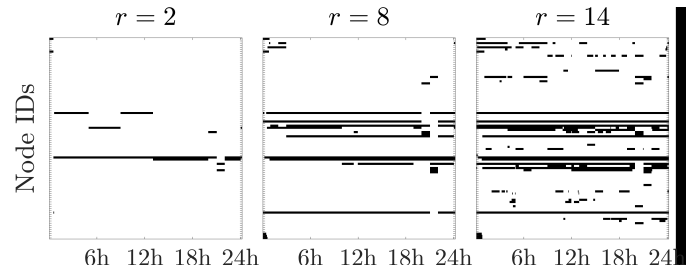
Fig. 9. Sensor placement results for Net3 with $r = 2, 8, 14$ (95 node IDs are not shown for brevity).

the similarity between two sensor placement strategies S^*_1 and S^*_2 (viewed as matrices with only zeros and ones), we define a similarity metric as $-\sum \sum \oplus(S^*_1, S^*_2)$, where $\oplus$ stands for element-wise logical operator xor. Note that this similarity metric is always a negative value, and when two matrices are the same, the largest similarity value 0 is reached. With applying this similarity metric, Fig. 7 is closer or more similar to Fig. 5b than Fig. 6, That is, the pattern tends to cause less impacts than the base demand in Net1 case. This conclusion may extend to the other networks, and it is always safe to claim that varying demand profiles at each node has significant impact on the sensor placement strategy.

We consider all discussed $n_d = 5$ demand profiles $\boldsymbol{D}_i \in \mathbb{R}^{15 \times k_f}$ where $i = 1, \ldots, n_d$, and run Algorithm 1, the final sensor placement results considering Patterns I with Base demand 1,2, and 3, and Patterns II and III are shown as Fig. 8, which is the fusion of Fig. 5b, Fig. 6, and Fig. 7. The final solution $\mathcal{S}^*_{r=1} = \{J10\}$, $\mathcal{S}^*_{r=3} = \mathcal{S}^*_{r=1} \bigcup \{J21, T2\}$, and $\mathcal{S}^*_{r=5} = \mathcal{S}^*_{r=3} \bigcup \{J12, J31\}$, thereby showcasing the greedy-optimal solution for Algorithm 1 that exploits supermodularity of the set function optimization.

C. Net3 network

In this section, the conclusions drawn from looped Net1 network in previous section are further corroborated via the Net3 water network shown in Fig. 2c with 90 junctions, 2 reservoirs, 3 tanks, 117 pipes, and 2 pumps. The base demands of all junctions are assumed as fixed, and a relative flatten pattern (varies slightly) are tested. The results selecting $r = 2, 8, 14$ from 95 node locations are shown as Fig. 9, the detailed locations are presented in Tab. III, and set $\mathcal{S}^*_{r=2} \subset \mathcal{S}^*_{r=8} \subset \mathcal{S}^*_{r=14}$ indicates the supermodularity property of the solution for this Net3 network. This showcases this property for even a larger network, further reaffirming the performance of the greedy algorithm. Besides that, the motivations behind testing Net3 network from a practical point of view come in two aspects, that are *(i)* whether it is effective or not via adding extra sensors to reduce the Kalman filter estimation error? and *(ii)* is the strategy from Algorithm 1 better than random strategies?

TABLE III. Sensor placement results for Net 3.

| Network	r	Result
Net3	2	$\mathcal{S}^*_{r=2} = \{\text{J203, J204}\}$
	8	$\mathcal{S}^*_{r=2} \bigcup \{\text{J261, J163, J169, J173, J184, J206}\}$
	14	$\mathcal{S}^*_{r=8} \bigcup \{\text{J208, J177, J179, J209, J20, J121}\}$

D. Estimation performance and comparing with random SP

This section computationally investigates two important issues closely related to the two motivations aforementioned: First, the performance of the state estimation and Kalman filter as the number of utilized sensors r in the water network varies. The second issue is whether a uniform (i.e., placing a sensor every other junction) or random sensor placement strategy yields a comparable performance—in terms of the state estimation metric—when comparison with the greedy-optimal presented in Algorithm 1. Both issues are investigated for the larger network Net3.

First, the relationship between performance of the Kalman filter $f(\mathcal{S}_r)$ (12) and the number of sensors is shown as Fig. 10a. Interestingly, Kalman filter performance $f(\mathcal{S}_r)$ decreases roughly linearly as the number of sensors r increases from 1 to 14 for three different hydraulic simulations. Specifically, Fig. 10a showcases the performance of the greedy-optimal solution when r is fixed in Algorithm 1 with fixed hydraulic profiles ($T_h = 0^{\text{th}}, 10^{\text{th}}, 20^{\text{th}}$ hour) i.e., the three figures in Fig. 10a show similar trend for three different hydraulic profiles. The best performance or lower bounds under the corresponding cases are reached when all sensor locations are placed ($r = 95$). This indicates that one would not expect a large improvement of Kalman filter performance via increasing the number of sensors even the locations of added sensors are all greedy-optimal.

Furthermore, the time-varying Kalman filter performance $f(\mathcal{S}^*_{r=14})$ for 24 hours is depicted via the blue line in Fig. 10b. the performance value can easily reach 10^5 level for this relatively large size network due to *(i)* the large dimension of $\boldsymbol{z}$ ($n_z = 3.066 \times 10^6$), *(ii)* covariance matrix $\mathbb{C}$ with tiny diagonal element (i.e., 5×10^{-3}), and *(iii)* the typical value of k_f is 200 in Net 3 resulting in $\boldsymbol{W}_o$ with huge value element in (11). Moreover, the trend of the blue line is decided by the hydraulic profile such as the flow rates for 24 hours, the plot of flow rates are not shown for brevity.

To address the second issue, we showcase the performance of a random sensor placement with a fixed number of sensors $r = 14$. Specifically, ten random sensor placements are generated for every hydraulic simulation. To quantify the performance of the proposed optimal placement, we define the relative performance of a random placement strategy $\hat{\mathcal{S}}$ as $\Delta f(\hat{\mathcal{S}}_{r=14}) = f(\hat{\mathcal{S}}_{r=14}) - f(\mathcal{S}^*_{r=14})$. A smaller value of $\Delta f(\hat{\mathcal{S}}_{r=14})$ implies a better optimal placement. The red lines in Fig. 10b are the relative performance of ten different randomizations—all of them
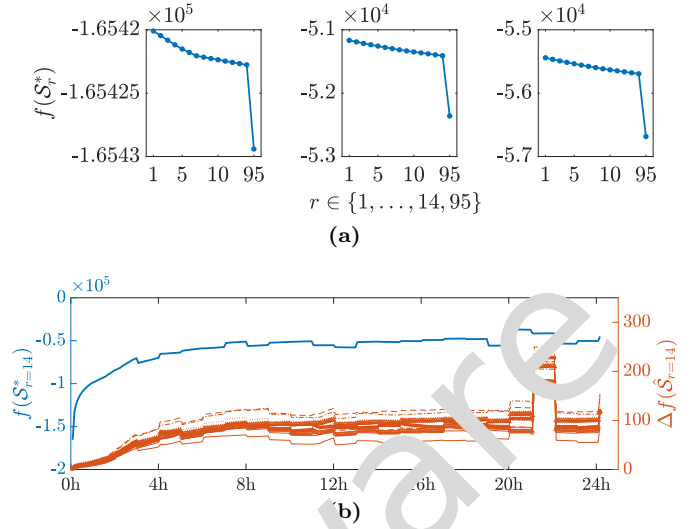
Fig. 10. Kalman filter performance $f(\mathcal{S}^*_r)$ with $r = \{1, \ldots, 14, 95\}$ when $T_h = 0^{\text{th}}, 10^{\text{th}}, 20^{\text{th}}$ hour (a), performance $f(\mathcal{S}^*_{r=14})$ for 24 hours (blue line in (b)), and the relative performance of ten randomized sensor placements $\Delta f(\hat{\mathcal{S}}_{r=14})$ (red lines in (b)).

are greater than zero showcasing a much better state estimation performance through the greedy algorithm. Even though the differences of performance are only 100-200 on average, the actual Kalman filter performance is orders of magnitude better due to fact that the log det function is used to quantify the state estimation accuracy. That is, the $\mathcal{S}^*_{r=14}$ obtained from Algorithm 1 performs significantly better than any random strategy $\hat{\mathcal{S}}^*_{r=14}$.

VI. CONCLUSIONS, PAPER LIMITATIONS, AND FUTURE DIRECTIONS

The paper presents a new computational method that results in sensor placements of WQ sensing devices in water networks. The method exclusively focuses on the WDN observability in regards to the WQ dynamics. After thoroughly testing three networks, we summarize the findings. First, the impacts of choosing L-W scheme time-step Δt (or the number of segments s_L) and the length of a single hydraulic simulation k_f on the sensor placement strategy is minor and can be neglected. Second, our proposed method can be applied to practical networks to determine sensor placement strategy because in practice historical data for demand patterns are available, thereby furnishing the sensor placement with the most common demand patterns. Hence, there is a possibility that the optimal sensor placement in terms of occupation time obtained would relatively be *time-invariant*. Third, the algorithm verifies the supermodular nature of the advocated set function optimization as corroborated via different test cases on three different networks. Fourth, and even if demand patterns change significantly, the algorithm can still be used to obtain a sensor placement that optimizes the sensor occupation time.

The paper *does not* advocate for *only* looking at state estimation metrics for water quality sensor placement. As

mentioned in Section I, a plethora of social and engineering objectives are typically considered in the literature to solve the WQSP. To that end, it is imperative that the proposed approach in this paper be studied in light of the other metrics and objectives discussed in the literature (such as minimizing the expected population and amount of contaminated water affected by an intrusion event). Consequently, an investigation of balancing engineering and state estimation objectives with more social-driven ones is needed. Hence, the objective of this paper is to provide a tool for the system operator that quantifies network observability vis-a-vis water quality sensor placements. The water system operator can also balance the objective of obtaining network-wide observability with these other metrics. Future work will focus on this limitation of the present work, in addition to considering multi-species dynamics that are nonlinear in the process model, which necessitate alternate approaches to quantify observability of nonlinear dynamic networks. This will also allow examining the potential reaction between contaminants and chlorine residuals that the sensors are monitoring.

DATA AVAILABILITY STATEMENT

Some or all data, models, or code used during the study were provided by a third party. Specifically, we provide a Github link that includes all the models, the data, and the results from the case study [39].

ACKNOWLEDGMENT

APPENDIX

This section presents a brief discussion on an efficient implementation of Algorithm 1 in light of the large-scale nature of the problem. This nature is due to the size of water networks, but mainly due to the space discretization of the PDE. This results in a large state-space dimension for the LDE model (9). Considering Net3 for example with a single hydraulic simulation (i.e., $t = 5$ minutes or equally $k_f = 300$ time-steps when $\Delta t = 1$ second to reach a decent performance in L-W scheme), the number of segments of different pipes is set as $s_L = 1000$. Given these figures, the corresponding dimension of the state-space model (8) and (9) is $n_x = 117,099$. That is, $\boldsymbol{A} \in \mathbb{R}^{117,099 \times 117,099}$, $\boldsymbol{C} \in \mathbb{R}^{95 \times 117,099}$ in (9), $\mathcal{O}(k_f) \in \mathbb{R}^{95 k_f \times 117,099 k_f}$ in (10), and $\boldsymbol{W}_o \in \mathbb{R}^{117,099 k_f \times 117,099 k_f}$ in (11) with $k_f = 300$. Next, we discuss the balance between the accuracy of our LDE and the computational burden.

From the above example, the problem dimension is determined via number of the water quality simulation time-steps k_f, the number of pipes n_P, and the number of segments of each single pipe s_L. Note that $k_f = t/\Delta t$, where $\Delta t \leq \min(\frac{L_i}{s_{L_i} v_i(t)})$ for all $i \in \mathcal{P}$. Ideally, parameter s_{L_i} should be as large as possible to ensure

accuracy of LDE model. After fixing the length of the interval t, parameter k_f (Δt) should be as small (large) as possible to reduce the dimension of $\mathcal{O}$ in (10). This subsequently reduces the computational burden of finding the log det of a large-scale matrix in (11)—and hence improves the computational tractability of Algorithm 1. Evidently, there exists a conflict between increasing of the accuracy of the LDE model (via increasing s_L for each pipe) and reducing computational burden (via decreasing k_f, increasing Δt or, equally, decreasing s_L) in a single hydraulic simulation. We next show simple approaches which can alleviate this conflict, significantly reduce the computational burden while maintaining the accuracy, and yield a scale algorithmic implementation.

- First, the state-space matrices $\boldsymbol{A}$ and $\boldsymbol{C}$, which are the major components of the dynamic water quality model and all subsequent matrices and Algorithm 1, are extremely sparse. In fact, more than 99.9% of the entries of these matrices are zeros. Thus, matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ can be expressed as the sparse matrix form thereby reducing the computational burden and the needed memory by many orders of magnitude. We use this in our Github codes.

- Second, to reduce the number of time-steps k_f or increase Δt, the dynamic number of segments of each pipe should be adopted due to the fact that the Δt is related with pipe length L_i and its velocity $v_i(t)$ which depends on the hydraulic simulation and demand profile. That is, for the short (long) Pipe i with large (small) velocity for interval t, the number of segments s_{L_i} can be chosen as a relatively small (large) value that is still enough to ensure the accuracy of L-W scheme. With varying number of segments, the dimension of $\boldsymbol{A}$ (i.e., n_x) varies in different hydraulic simulations. The interested reader is referred to our Github [39] for the details of this implementation. For example, if the velocities during interval t_1 are two times less than the ones during interval t_2, then the size of n_x can be reduced by half in t_2.

- Third, the WQSP problem parameters (13) in each single hydraulic simulation are independent on each other. This is due to $\boldsymbol{A}$ and $\boldsymbol{C}$ in (10) are time-invariant in a single interval. After obtaining initial conditions and matrices $\boldsymbol{A}$, $\boldsymbol{C}$ offline, Algorithm 1 can be implemented through parallel computing. That is, multi-intervals can be calculated simultaneously on a multi-core computer. Moreover, the bottleneck of Algorithm 1 is located in calculating the log det of large-scale matrix (11). To that end, we adopt the LU and Cholesky decompositions [50] to accelerate the evaluation of the log det objective function for different sensor placements.

- Fourth, it is clear that reducing the length of a single hydraulic simulation t would be another effective trick, since it would reduce the dimension of $\mathcal{O}(k)$ and $\boldsymbol{W}_O(k)$ significantly according to (11).

Finally, it is worth mentioning that Algorithm 1 is indeed offline seeing that it solves a *placement* problem of

water quality sensors: sensors that cannot have varying geographic locations. Powerful computational resources, at the disposal of water network operators, can hence be used to run the algorithm.

We present the tested computational time for running a horizon of 24 hours for the tested three networks. For the three-node network, it takes 45.7 sec to terminate Algorithm 1; for Net1 with larger discretization time-step $\Delta t = 10$ sec, the computational time is 129.6 sec, and the computational time becomes 188.1 sec when the time-step is reduced by half, that is, $\Delta t = 5$ sec. When we change a new demand profile for the Net1 network, the computational time can reach 555.6 sec; for the Net 3 network it takes almost four hours for the entire 24 hour simulation horizon. The reasons of the varying computational time are: (i) special demand profiles may result in time-consuming LU or Cholesky decompositions, (ii) the discretization time-step Δt and length of single observation time k_f have an impact on the size of $\boldsymbol{W}_o$ and result in different computational time, (iii) and the tested computer have different numbers of cores and capacities of memory.

REFERENCES

[1] A. Krause, J. Leskovec, C. Guestrin, J. VanBriesen, and C. Faloutsos, "Efficient sensor placement optimization for securing large water distribution networks," *Journal of Water Resources Planning and Management*, vol. 134, no. 6, pp. 516–526, 2008.

[2] A. Ostfeld, J. G. Uber, E. Salomons, J. W. Berry, W. E. Hart, C. A. Phillips, J.-P. Watson, G. Dorini, P. Jonkergouw, Z. Kapelan *et al.*, "The battle of the water sensor networks (bwsn): A design challenge for engineers and algorithms," *Journal of Water Resources Planning and Management* vol. 34, no. 6, pp. 556–568, 2008.

[3] A. Preis and A. Ostfeld, "Multiobjective contaminant sensor network design for water distribution systems," *Journal of Water Resources Planning and Management*, vol. 12, no. 4, pp. 366–377, 2008.

[4] S. Schal, A. Lothes, L. S. Bryson, and L. Ormsbee, "Water quality sensor placement guidance using va-spot," in *World Environmental and Water Resources Congress 2013: Showcasing the Future*, 2013, pp. 1022–1132.

[5] D. G. Eliades, M. Kyriakou, and M. M. Polycarpou, "Sensor placement in water distribution systems using the S-PLACE Toolkit," *Procedia Engineering*, vol. 70, no. 2010, pp. 602–611, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.proeng.2014.02.066

[6] Y. Shastri and U. Diwekar, "Sensor placement in water networks: A Stochastic programming approach," *Journal of Water Resources Planning and Management*, vol. 132, no. 3, pp. 192–203, 2006.

[7] M. M. Aral, J. Guan, and M. L. Maslia, "Optimal design of sensor placement in water distribution networks," *Journal of Water Resources Planning and Management*, vol. 136, no. 1, pp. 5–18, 2010.

[8] S. Rathi and R. Gupta, "Sensor placement methods for contamination detection in water distribution networks: A review," *Procedia Engineering*, vol. 89, pp. 181–188, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.proeng.2014.11.175

[9] W. E. Hart and R. Murray, "Review of sensor placement strategies for contamination warning systems in drinking water distribution systems," *Journal of Water Resources Planning and Management*, vol. 136, no. 6, pp. 611–619, 2010.

[10] C. Hu, M. Li, D. Zeng, and S. Guo, "A survey on sensor placement for contamination detection in water distribution systems," *Wireless Networks*, vol. 24, no. 2, pp. 647–661, 2018.

[11] B. H. Lee and R. A. Deininger, "Optimal locations of monitoring stations in water distribution system," *Journal of Environmental Engineering*, vol. 118, no. 1, pp. 4–16, 1992.

[12] A. Kumar, M. Kansal, and G. Arora, "Identification of monitoring stations in water distribution system," *Journal of Environmental Engineering*, vol. 123, no. 8, pp. 746–752, 1997.

[13] M. Kansal, T. Dorji, S. K. Chandniha, and A. Tyagi, "Identification of optimal monitoring locations to detect accidental contaminations," in *World Environmental and Water Resources Congress 2012: Crossing Boundaries*, 2012, pp. 758–776.

[14] H.-M. Woo, J.-H. Yoon, and D.-Y. Choi, "Optimal monitoring sites based on water quality and quantity in water distribution systems," in *Bridging the Gap: Meeting the World's Water and Environmental Resources Challenges*, 2001, pp. 1–9.

[15] M. A. Al-Zahrani and K. Moied, "Optimizing water quality monitoring stations using genetic algorithms," *Arabic Journal for Science and Engineering*, vol. 28, no. 1, pp. 57–75, 2003.

[16] A. Afshar and M. A. Mariño, "Multi-objective coverage-based aco model for quality monitoring in large water networks," *Water resources management* vol. 26, no. 8, pp. 2159–2176, 2012.

[17] S. R. Ghimire and B. D. Barkdoll, *A Heuristic Method for the Battle of the Water Network Sensors. Demand Based Approach*, pp. 1–10. [Online]. Available: https://ascelibrary.org/doi/abs/10.1061/40941%28247%29102

[18] S. Rathi and R. Gupta, "Locations of sampling stations for water quality monitoring in water distribution networks." *Journal of environmental science & engineering*, vol. 56, no. 2, pp. 169–178, 2014.

[19] G. He, T. Zhang, F. Zheng, and Q. Zhang, "An efficient multi-objective optimization method for water quality sensor placement within water distribution systems considering contamination probability variations," *Water Research* vol. 143, pp. 165–175, 2018. [Online]. Available: https://doi.org/10.1016/j.watres.2018.06.041

[20] F. Hooshmand, F. Amerehi, and S. A. MirHassani, "Logic-based benders decomposition algorithm for contamination detection problem in water networks," *Computers and Operations Research*, vol. 115, p. 104840, 2020. [Online]. Available: https://doi.org/10.1016/j.cor.2019.104840

[21] C. Ciaponi, E. Creaco, A. Di Nardo, M. Di Natale, C. Giudicianni, D. Musmarra, and G. F. Santonastaso, "Reducing impacts of contamination in water distribution networks: A combined strategy based on network partitioning and installation ofwater quality sensors," *Water (Switzerland)*, vol. 11, no. 6, pp. 1–16, 2019.

[22] C. de Winter, V. R. Palleti, D. Worm, and R. Kooij, "Optimal placement of imperfect water quality sensors in water distribution networks," *Computers and Chemical Engineering*, vol. 121, pp. 200–211, 2019. [Online]. Available: https://doi.org/10.1016/j.compchemeng.2018.10.021

[23] C. Giudicianni, M. Herrera, A. Di Nardo, R. Greco, E. Creaco, and A. Scala, "Topological Placement of Quality Sensors in Water-Distribution Networks without the Recourse to Hydraulic Modeling," *Journal of Water Resources Planning and Management*, vol. 146, no. 6, pp. 1–12, 2020.

[24] C. Hu, L. Dai, X. Yan, W. Gong, X. Liu, and L. Wang, "Modified NSGA-III for sensor placement in water distribution system," *Information Sciences*, vol. 509, pp. 488–500, 2020. [Online]. Available: https://doi.org/10.1016/j.ins.2018.06.055

[25] A. Di Nardo, C. Giudicianni, R. Greco, M. Herrera, G. F. Santonastaso, and A. Scala, "Sensor Placement in Water Distribution Networks based on Spectral Algorithms," vol. 3, pp. 593–584, 2018.

[26] S. Tinelli, E. Creaco, and C. Ciaponi, "Impact of objective function selection on optimal placement of sensors in water distribution networks," *Ital. J. Eng. Geol. Environ*, pp. 173–178, 2018.

[27] Q. Zhang, F. Zheng, Z. Kapelan, D. Savic, G. He, and Y. Ma, "Assessing the global resilience of water quality sensor placement strategies within water distribution systems," *Water Research*, vol. 172, p. 115527, 2020.

[28] A. G. Rajakumar, M. S. Mohan Kumar, B. Amrutur, and Z. Kapelan, "Real-Time Water Quality Modeling with Ensemble Kalman Filter for State and Parameter Estimation in Water Distribution Networks," *Journal of Water Resources Planning and Management*, vol. 145, no. 11, pp. 1–12, 2019.

[29] Y. J. Yang, J. A. Goodrich, R. M. Clark, and S. Y. Li, "Modeling and testing of reactive contaminant transport in drinking water pipes: Chlorine response and implications for online contaminant detection," *Water research*, vol. 42, no. 6-7, pp. 1397–1412, 2008.

[30] V. Tzoumas, A. Jadbabaie, and G. J. Pappas, "Sensor placement for optimal Kalman filtering: Fundamental limits, submodularity, and algorithms," *Proceedings of the American Control Conference*, vol. 2016-July, pp. 191–196, 2016.

[31] H. Zhang, R. Ayoub, and S. Sundaram, "Sensor selection for Kalman filtering of linear dynamical systems: Complexity, limitations and greedy algorithms," *Automatica*, vol. 78, pp. 202–210, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.automatica.2016.12.025

[32] P. D. Lax and B. Wendroff, "Difference schemes for hyperbolic equations with high order of accuracy," *Communications on pure and applied mathematics*, vol. 17, no. 3, pp. 381–398, 1964.

[33] L. A. Rossman *et al.*, "EPANET 2: users manual," 2000.

[34] L. A. Rossman and P. F. Boulos, "Numerical methods for modeling water quality in distribution systems: A comparison," *Journal of Water Resources planning and management*, vol. 122, no. 2, pp. 137–146, 1996.

[35] H. Basha and L. Malaeb, "Eulerian–lagrangian method for constituent transport in water distribution networks," *Journal of Hydraulic Engineering*, vol. 133, no. 10, pp. 1155–1166, 2007.

[36] F. Shang, J. G. Uber, L. A. Rossman *et al.*, "Epanet multi-species extension user's manual," *Risk Reduction Engineering Laboratory, US Environmental Protection Agency, Cincinnati, Ohio*, 2008.

[37] A. Morais, "Fast and robust solution methods for the water quality equations," 2012.

[38] P. Fabrie, G. Gancel, I. Mortazavi, and O. Piller, "Quality modeling of water distribution systems using sensitivity equations," *Journal of Hydraulic Engineering*, vol. 136, no. 1, pp. 34–44, 2010.

[39] S. Wang, "Shenwang9202/water-quality-modeling-and-sensor-placement," Apr 2020. [Online]. Available: https://github.com/ShenWang9202/Water-Quality-Modeling-and-Sensor-Placement

[40] S. Wang, A. F. Taha, and A. A. Abokifa, "How effective is model predictive control in real-time water quality regulation? state-space modeling and scalable control," *Water Resources Research*, 2020, , In review. Link: https://arxiv.org/abs/2009.01298.

[41] J. P. Hespanha, *Linear systems theory.* Princeton university press, 2018.

[42] T. H. Summers and J. Lygeros, "Optimal Sensor and Actuator Placement in Complex Dynamical Networks," jun 2013. [Online]. Available: http://arxiv.org/abs/1306.2491.

[43] T. H. Summers, F. L. Cortesi, and J. Lygeros, "On Submodularity and Controllability in Complex Dynamical Networks," *IEEE Transactions on Control of Network Systems*, 2016.

[44] S. T. Jawaid and S. L. Smith, "Submodularity and greedy algorithms in sensor scheduling for linear dynamical systems," *Automatica*, 2015.

[45] L. Lovász, "Submodular functions and convexity," in *Mathematical programming the state of the art.* Springer, 1983, pp. 235–257.

[46] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms.* MIT press, 2009.

[47] F. L. Cortesi, T. H. Summers, and J. Lygeros, "Submodularity of energy related controllability metrics," in *Proceedings of the IEEE Conference on Decision and Control*, vol. 2015-Febru, no. February, mar 2014, pp. 2883–2888. [Online]. Available: http://arxiv.org/abs/1403.6351http://dx.doi.org/10.1109/CDC.2014.7039832

[48] F. Shang, J. G. Uber, and M. M. Polycarpou, "Particle backtracking algorithm for water distribution system analysis," *Journal of environmental engineering*, vol. 128, no. 5, pp. 441–450, 2002.

[49] D. G. Eliades, M. Kyriakou, S. Vrachimis, and M. M. Polycarpou, "Epanet-matlab toolkit: An open-source software for interfacing epanet with matlab," in *Proc. 14th International Conference on Computing and Control for the Water Industry (CCWI)*, The Netherlands, Nov 2016, p. 8.

[50] L. N. Trefethen and D. Bau III, *Numerical linear algebra.* Siam, 1997, vol. 50.