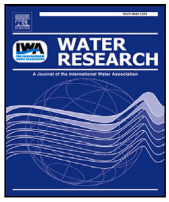
Water Research 264 (2024) 122201

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres



Online state estimation in water distribution systems via Extended Kalman Filtering

Matthew Bartos *, Meghna Thomas, Min-Gyu Kim, Matthew Frankel, Lina Sela

Fariborz Maseeh Dept. of Civil, Architectural, and Environmental Engineering, The University of Texas at Austin, 301 E Dean Keeton St, Austin, 78712, TX, USA

ARTICLE INFO

Dataset link: https://github.com/future-water/wds-state-estimation

Keywords:
State estimation
Hydraulic modeling
Extended Kalman Filter
Water distribution systems

ABSTRACT

Operators of water distribution systems (WDSs) need continuous and timely information on pressures and flows to ensure smooth operation and respond quickly to unexpected events. While hydraulic models provide reasonable estimates of pressures and flows in WDSs, updating model predictions with real-time sensor data provides clearer insights into true system behavior and enables more effective real-time response. Despite the growing prevalence of distributed sensing within WDSs, standard hydraulic modeling software like EPANET do not support synchronous data assimilation. This study presents a new method for state estimation in WDSs that combines a fully physically-based model of WDS hydraulics with an Extended Kalman Filter (EKF) to estimate system flows and heads based on sparse sensor measurements. To perform state estimation via EKF, a state-space model of the hydraulic system is first formulated based on the 1-D Saint-Venant equations of conservation of mass and momentum. Results demonstrate that the proposed model closely matches steady-state extended-period models simulated using EPANET. Next, through a holdout analysis it is found that fusing sensor data with EKF produces flow and head estimates that closely match ground truth flows and heads at unmonitored locations, indicating that state estimation successfully infers internal hydraulic states from sparse sensor measurements. These findings pave the way towards real-time operational models of WDSs that will enable online detection and mitigation of hazards like pipe leaks, main bursts, and hydraulic transients.

1. Introduction

Water distribution systems (WDSs) are critical infrastructure systems that are responsible for the provision of safe and reliable drinking water to communities. These systems span large geographic distances and must be regularly monitored and maintained to avert hazards like pipe leaks, main bursts, and contaminant intrusion events. To contend with these issues, public water utilities employ a diverse set of technologies to ensure safe and reliable WDS operation. Among these technologies, remote monitoring has been identified as a key mechanism towards improved decision-making, and is being increasingly adopted by water utilities (Daniel et al., 2023; Arnell et al., 2023). For example, Supervisory Control and Data Acquisition (SCADA) systems provide hydraulic measurements at remote points of interest such as pumping stations, storage tanks, and water treatment facilities; advanced metering infrastructure (AMI) automatically collects and transmits water consumption information at points of use; and distributed sensor networks (DSN) collect pressure and flow measurements within WDSs (Xing and Sela, 2019; Shafiee et al., 2018). However, because sensing is typically sparse, sensor data alone is not enough to provide a complete and timely assessment of the current state of the system. Therefore, automated data processing and assimilation techniques are essential for providing critical performance information to water utilities to support operations and management decisions. Among the various data assimilation techniques, this paper focuses on *state estimation*, which integrates increasingly available sensing data into hydraulic models to gain comprehensive system observability.

State estimation (SE) can be broadly defined as the process of determining the *unknown state variables* of a system based on the *mathematical model* describing the dynamics of the system and available *measurements*. The formulation of the SE problem requires the definition of the state variables, inputs, measurements, and the process model that links between them, all depending on the specific application (Simon, 2006; Tshehla et al., 2017). The processes governing WDSs include system hydraulics, such as demand- and pressure-driven, steady-state and extended-period models (Todini and Rossman, 2013; Boulos et al., 2006), and transient hydraulics (Xing and Sela, 2020). In this work, we focus only on physically-based hydraulic models, which are dominant for modeling WDSs dynamics, and do not consider model-free methods, such as graph signal processing and neural networks (Zhou et al., 2023a, 2022; Xing and Sela, 2022).

* Corresponding author.
E-mail address: mdbartos@utexas.edu (M. Bartos).

https://doi.org/10.1016/j.watres.2024.122201
Received 12 March 2024; Received in revised form 14 July 2024; Accepted 31 July 2024
Available online 5 August 2024
0043-1354/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

In the context of WDSs, the SE problem has been extensively studied over the past four decades. However, there is lack of consistency between the various definitions of states, inputs, measurements, as well as solution techniques. Predominant examples of applications of SE in WDS include estimating heads and flows (Mankad et al., 2022; Wang et al., 2022), leaks (Blocher et al., 2020; Pecci et al., 2020), consumer demands (Do et al., 2017; Ruiz et al., 2022; Chu et al., 2021), and joint parameter estimation, including pipe roughness coefficients (Waldron et al., 2020; Zhou et al., 2018; Zhang et al., 2018b). Beyond the differing applications of SE in WDSs, several solution techniques have been proposed, including least squares methods (Arsene and Gabrys, 2014; Zhang et al., 2018a), inverse optimization techniques (Pecci et al., 2020; Mankad et al., 2022), evolutionary algorithms (Preis et al., 2011), and methods based on control and dynamic system theory (Delgado-Aguiñaga and Begovich, 2017; Torres et al., 2020; Delgado-Aguiñaga and Besançon, 2018).

While in the water systems community there is a discordance related to the definition of states and solution approaches, in the control community, the SE problem is firmly grounded in mathematical theory, offering various forms of clearly established fundamental tools for analyzing and solving a broad class of estimation problems (Simon, 2006). In control theory, the state variables of a *dynamical system*, (i.e., a system that evolves in time), provide an internal description of the system that fully characterizes the system and its response to any given set of inputs at any time. State equations are a standard representation of a dynamical system, in which the time derivative of each state variable is expressed in terms of the state variables and inputs. Once a dynamical system is defined using state equations, different state estimation methods, such as Kalman Filtering (KF) and Extended Kalman Filtering (EKF), are available for analyzing and solving SE problems (Simon, 2006).

The main challenge associated with applying control-based SE methods to WDSs is due to the characteristics of extended-period hydraulics, which are predominantly used for modeling WDSs. These models primarily consist of algebraic equations, representing the linear mass balance at nodes, and nonlinear head-loss (or head-gain) at pipes and valves (or pumps), with only a few dynamic equations, accounting for the change in tank water levels (Todini and Rossman, 2013). Hence, classical KF and EKF state estimation techniques are not directly applicable to WDSs. Acknowledging this limitation, KF applications in WDSs have been focused on demand estimation (Kang and Lansey, 2009; Ruiz et al., 2022) or using transient hydraulic models (i.e., water hammer equations) (Delgado-Aguiñaga and Besançon, 2018; Torres et al., 2020). However, demand estimation provides limited information about system states, since flows and heads are the main states of interest to support decision-making regarding operational flow distribution, pressure management, and water loss control. While transient hydraulics models offer a more realistic representation of system dynamics, they introduce additional challenges. Utilizing transient models for WDSs requires more extensive data to parameterize the models and a fine numerical grid in time and space, imposing significant computational demands and hindering real-time application for even medium-size systems (Riaño-Briceño et al., 2022).

This work advances a new method for state estimation in WDSs that *combines a fully physically-based state-space model of the hydraulic system with EKF to estimate system flows and heads based on sparse sensor measurements*. To address the above modeling limitation, we formulate WDS hydraulics considering dynamic simulation based on the Saint-Venant equations that model conservation of mass and momentum, rather than conservation of mass and energy (Boulos et al., 2006). The 1-D Saint-Venant equations are discretized and solved using an implicit numerical scheme with a Preissmann slot used to capture pressurized WDS behavior. Using a series of recurrence relations, the equations are reduced to a state-space representation of a *dynamical system* with heads and flow rates as the state variables, enabling the application of EKF to assimilate sensor data into the model. We focus on daily operating horizons with sub-hourly time resolutions (on the order of 30 min), which are typical for modeling and operation of water distribution systems. This process is implemented using the *PipeDream* solver (Bartos and Kerkez, 2021), a Python-based simulation engine originally developed for stormwater system modeling that was adapted in this study to accommodate WDS hydraulics. This study makes the following major contributions: (i) we present a new physically-based hydraulic solver for WDSs based on the Saint-Venant equations for unsteady flow. We demonstrate that the proposed solver closely matches extended-period simulation results and is capable of achieving speeds that are comparable to EPANET (Rossman et al., 2020), an industry standard software for hydraulic modeling; (ii) we present the first application of SE in networked WDSs by combining the new hydraulic solver with an efficient sensor data assimilation approach based on EKF; and (iii) we show that EKF enables accurate estimation of hydraulic heads in ungaged locations from sparse and noisy sensor data.

2. Methods

This section describes the derivation, implementation, and evaluation of a real-time state estimation scheme for WDSs using a physically-based hydraulic solver. First, we present a new hydraulic model based on the Saint-Venant equations with a Preissmann slot representation of pressurized flow. Second, we verify the integrity of the hydraulic model by comparing model outputs for three benchmark WDS models against the EPANET solver. Third, we derive and apply an EKF scheme to assimilate sensor data into the hydraulic model. Finally, we evaluate the degree to which the EKF improves estimates of pressure heads at ungauged locations in WDSs using a holdout assessment.

2.1. Hydraulic solver

In this section, we describe a numerical scheme for modeling the flow dynamics in pressurized WDS using the Saint-Venant equations. We first show how the Saint-Venant equations may be adapted to represent pressurized flow using a Preissmann slot formulation. We then describe a method for applying the Saint-Venant equations to a networked pipe system, using the SUPERLINK scheme (Ji, 1998). Finally, we show how the dynamical equations can be expressed in state-space form, thereby facilitating the use of EKF to assimilate real-time sensor data.

Unsteady flow within a one-dimensional conduit or open channel is described by the Saint-Venant equations. This pair of partial differential equations consists of a continuity equation (Eq. (1)) and a momentum equation (Eq. (2)):

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = q_{in} \tag{1}$$

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x}(Qu) + gA\left(\frac{\partial h}{\partial x} - S_0 + S_f + S_L\right) = 0 \tag{2}$$

Where Q is the flow rate, A is the cross-sectional area of flow, u is the average velocity, h is the pressure hydraulic head above the bottom of the conduit, x is distance, t is time, q_{in} is the exogenous flow input per unit length, g is the acceleration due to gravity, and S_o, S_f and S_L represent the conduit bottom slope, friction head loss slope, and local head loss slope (due to contractions and expansions), respectively. The continuity equation specifies that the change in mass of water within a control volume equal the mass flow rate in minus the mass flow rate out. The momentum equation corresponds to Newton's second law of motion applied to a fluid volume, and requires that the time rate of change of momentum in the volume equal the sum of forces acting on that volume (including hydrostatic, gravitational, frictional, and local contraction and expansion forces).

The Saint-Venant equations are most commonly applied to model open channel flow; however, they may also represent flow in pressurized pipe systems through the inclusion of a Priessmann slot. Initially proposed by Preissmann (1961), the Preissmann slot enables

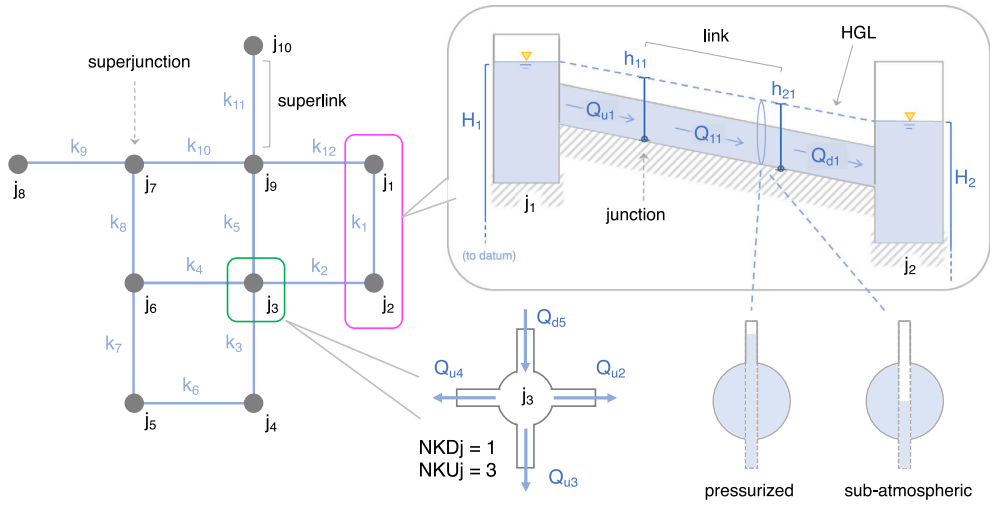


Fig. 1. Model structure with computational elements. (Left): Network model with superjunctions (indexed by j) and superlinks (indexed by k) shown. (Top right): Superlink with internal links and junctions shown. (Bottom right): Visualization of Priessmann slot. (Bottom center): Superjunction with inflow from one superlink and outflow to three superlinks.

pressurized flow to be represented using the Saint-Venant equations through the addition of a fictitious narrow groove at the crown of the pipe (Preissmann, 1961). The Preissmann slot is particularly useful for situations where pipes alternate between full and partially-full flow. In this situation, transitioning between pressurized (parabolic) and open-channel (hyperbolic) equations leads to numerical instability, which the Preissmann slot addresses by using only the open-channel form of the equations (Vasconcelos and Wright, 2007). Sub-atmospheric pressures may be represented through the use of a "negative" Preissmann slot that extends below the pipe crown (Kerger et al., 2011). To match the model assumptions of the traditional hydraulic solvers, such as EPANET, we use the positive and negative Priessmann slot formulations for pressurized and sub-atmospheric pressures, respectively (see Fig. 1, lower-right).

To solve the Saint-Venant equations across a networked WDS, we adapt the SUPERLINK scheme initially proposed by Ji (Ji, 1998), and modified by Bartos and Kerkez (Bartos and Kerkez, 2021). In this formulation, the pipe network is divided into four types of computational elements, including links, junctions, superlinks, and superjunctions. *Links* (indexed by ik) represent linear sections of conduit or pipe. *Junctions* (indexed by Ik) connect links together, and may possess volume or may simply represent computation points. *Superlinks* (indexed by k) consist of linear chains of alternating junctions and links connected end-to-end without branching. *Superjunctions* (indexed by j) connect superlinks together, and may represent storage tanks, invert discontinuities, or simply computational points where multiple superlinks meet.

Under this scheme, the Saint-Venant equations are discretized and solved within each superlink using an implicit staggered-grid formulation, wherein the continuity equation and momentum equation are applied to separate control volumes. Specifically, the continuity equation is applied around each internal junction:

$$\left(\frac{B_{ik}\Delta x_{ik}}{2}+\frac{B_{i-1k}\Delta x_{i-1k}}{2}+A_{s,Ik}\right)\cdot\frac{h_{Ik}^{t+\Delta t}-h_{Ik}^{t}}{\Delta t}+Q_{ik}^{t+\Delta t}-Q_{i-1k}^{t+\Delta t}=Q_{in,Ik}^{t+\Delta t} \quad (3)$$

Where B is the Preissmann slot width, A_s is the junction surface area, Δx is the longitudinal length, and Q_{in} is the exogenous flow input. The momentum equation is applied around each internal link:

$$\begin{aligned}&(Q_{ik}^{t+\Delta t}-Q_{ik}^{t})\frac{\Delta x_{ik}}{\Delta t}+u_{I+1k}Q_{I+1k}^{t+\Delta t}-u_{Ik}Q_{Ik}^{t+\Delta t}\\&+gA_{ik}(h_{I+1k}^{t+\Delta t}-h_{Ik}^{t+\Delta t})+gA_{ik}(S_{f,ik}+S_{L,ik}-S_{o,ik})\Delta x_{ik}=0\end{aligned} \quad (4)$$

The system of equations is closed by applying a mass balance to each superjunction:

$$\sum_{l=1}^{NKDj}Q_{dk_l}^{t+\Delta t}-\sum_{m=1}^{NKUj}Q_{uk_m}^{t+\Delta t}+Q_{in,j}^{t+\Delta t}=\frac{A_{sj}(H_j^{t+\Delta t}-H_j^{t})}{\Delta t} \quad (5)$$

Where $Q_{in,j}$ is the external inflow into superjunction j, A_{sj} is the surface area of superjunction j, Q_{dk} is the flow rate at the furthest downstream element of superlink k, Q_{uk} is the flow rate at the furthest upstream element of superlink k, and H_j is the total head at superjunction j. NKD_j is the number of superlinks with their downstream ends attached to superjunction j, while NKU_j is the number of superlinks with their upstream ends attached to superjunction j.

In summary, the SUPERLINK numerical scheme applies (i) the continuity equation around each junction; (ii) the momentum equation around each link; (iii) the continuity equation around each superjunction; and (iv) the momentum equation at the boundary between superjunctions and superlinks. Taken together, this implicit system of equations describes the full unsteady equations of flow for a pipe network, and may be expressed as a large time-varying linear system at each time step given initial states. Considering the single superlink and adjacent superjunctions shown in Fig. 1 (upper-right), and assuming no inflows/outflows from adjacent superlinks and no local head losses, the system of equations may be written as a matrix equation:

$$\left[\begin{array}{cccc|ccc}\frac{1}{\Delta t}\frac{d\forall_1}{dH_1} & & & & 1 & & \\ & \frac{1}{\Delta t}\frac{d\forall_{11}}{dh_{11}} & & & -1 & 1 & \\ & & \frac{1}{\Delta t}\frac{d\forall_{21}}{dh_{21}} & & & -1 & 1 \\ & & & \frac{1}{\Delta t}\frac{d\forall_2}{dH_2} & & & -1 \\ \hline -gA_{u1} & gA_{u1} & & & b_{u1} & c_{u1} & \\ & -gA_{11} & gA_{11} & & a_{11} & b_{11} & c_{11} \\ & & -gA_{d1} & gA_{d1} & & a_{d1} & b_{d1}\end{array}\right]\left[\begin{array}{c}H_1^{t+\Delta t}\\ h_{11}^{t+\Delta t}\\ h_{21}^{t+\Delta t}\\ H_2^{t+\Delta t}\\ \hline Q_{u1}^{t+\Delta t}\\ Q_{11}^{t+\Delta t}\\ Q_{d1}^{t+\Delta t}\end{array}\right]$$

$$=\left[\begin{array}{c}\frac{1}{\Delta t}\frac{d\forall_1}{dH_1}H_1^{t}+Q_{in,1}^{t+\Delta t}\\ \frac{1}{\Delta t}\frac{d\forall_{11}}{dh_{11}}h_{11}^{t}+Q_{in,11}^{t+\Delta t}\\ \frac{1}{\Delta t}\frac{d\forall_{21}}{dh_{21}}h_{21}^{t}+Q_{in,21}^{t+\Delta t}\\ \frac{1}{\Delta t}\frac{d\forall_2}{dH_2}H_2^{t}+Q_{in,2}^{t+\Delta t}\\ \hline \frac{\Delta x_{u1}}{\Delta t}Q_{u1}^{t}+gA_{u1}(S_{o,u1}-S_{f,u1}-z_1)\\ \frac{\Delta x_{11}}{\Delta t}Q_{11}^{t}+gA_{11}(S_{o,11}-S_{f,11})\\ \frac{\Delta x_{d1}}{\Delta t}Q_{d1}^{t}+gA_{d1}(S_{o,d1}-S_{f,d1}-z_2)\end{array}\right] \quad (6)$$

Where $\frac{d\forall_{Ik}}{dh_{Ik}}$ refers to the change in volume within control volume Ik per unit change in hydraulic head (i.e., the first term in Eq. (3)); z_j is the bottom elevation of superjunction j; and a_{ik}, b_{ik} and c_{ik} are coefficients related to the inertial component of the momentum balance (e.g. $a_{ik} = -\max(u_{Ik}, 0)$, $c_{ik} = -\max(-u_{I+1k}, 0)$). It should be noted that several parameters like the friction slope S_f depend nonlinearly on the states at time $t + \Delta t$, and thus this system of equations is solved iteratively at each time step using Newton's method. After convergence, the above linear system may be written compactly as follows:

$$K_t \mathbf{s}_{t+\Delta t} = \mathbf{b}_t \tag{7}$$

Where K_t is the matrix representing system dynamics, $\mathbf{s}_{t+\Delta t}$ is a vector containing all the states of the system, including the total heads at each superjunction (H_j), the pressure heads above the bottom of each junction (h_{Ik}), the flows within each link (Q_{ik}), the flows at the upstream and downstream ends of each superlink (Q_{uk} and Q_{dk}, respectively). Using a series of recurrence relations, this large system of equations can be reduced to a smaller linear system that depends only on the pressure heads at each superjunction (see derivation in Bartos and Kerkez (2021)):

$$E_t \mathbf{x}_{t+\Delta t} = F_t \mathbf{x}_t + G_t \mathbf{u}_t \tag{8}$$

Where $\mathbf{x}_t$ is the vector of superjunction pressure heads ($H_j \ \forall \ j$) at time t, $\mathbf{u}_t$ is the exogenous flow input (or demand) at each superjunction at time t, and E_t, F_t, and G_t are matrices describing the dynamics of the system at each timestep. Unlike a descriptor system, the matrix E_t is full-rank, and thus invertible for systems that contain some nonzero volume of water. Pre-multiplying by the inverse of E_t puts the system in standard state-space form:

$$\mathbf{x}_{t+\Delta t} = A_t \mathbf{x}_t + B_t \mathbf{u}_t \tag{9}$$

In the context of WDSs, A_t describes the interconnection between nodes and coupling of states, and B_t is the input matrix describing the interconnection between exogenous flow inputs or demands. Given initial state $\mathbf{x}_0$ and inputs $\mathbf{u}_t$, at each time t matrices A_t and B_t are updated and new state variables $\mathbf{x}_{t+\Delta t}$ are computed. With the system in state-space form, an EKF may now be applied to assimilate pressure head data into the hydraulic model.

2.2. Modifications to hydraulic solver

We extend the previous model in three primary ways to enable representation of pressurized water networks: (i) a new hydraulic geometry that represents pressurized full-pipe flow is implemented, (ii) new friction losses formulations that are consistent with headlosses models in pressurized flow are implemented, and (iii) valve and pump control rules are implemented to achieve feature parity with common hydraulic solvers, such as EPANET. The state-space WDS model is implemented as an extension to the PipeDream hydraulic solver developed in Bartos and Kerkez (2021).

2.2.1. Force main hydraulic geometry

First, we define a new hydraulic geometry type that is consistent with the pressurized pipe model. For this hydraulic geometry, the pipe is assumed to always be flowing full with a circular cross section. The cross-sectional area is thus assumed to remain constant:

$$A_{ik} = \frac{\pi}{4} D^2 \tag{10}$$

Where D is the pipe diameter. The Preissman slot width B_{ik} of each element is set to correspond to an acoustic wave speed corresponding to the speed of sound in water ($c = 1496$ [m/s]):

$$B_{ik} = \frac{g A_{ik}}{c^2} \tag{11}$$

2.2.2. Friction formulation

Second, we modify the existing friction force formulation used by PipeDream to represent the Hazen–Williams and Darcy–Weisbach formulations (SI units):

$$S_{f,hw} = 10.667 \cdot C^{-1.852} D^{-4.871} \tag{12}$$

$$S_{f,dw} = 0.0252 \cdot f(\epsilon, D, Q) D^{-5} \tag{13}$$

Where $S_{f,hw}$ is the friction slope under the Hazen–Williams formulation, $S_{f,dw}$ is the friction slope under the Darcy–Weisbach formulation, C is Hazen–Williams factor, ϵ is pipe relative roughness. The friction factor f is computed using the Hagen–Poiseuille formulation for laminar flow, and the Swamee and Jain approximation for turbulent flow (Rossman et al., 2020).

2.2.3. Pump implementation

Pumps are implemented using the following pump curve function (Rossman et al., 2020):

$$H_{e,p}^{t+\Delta t} = a_p - b_p (Q_p^{t+\Delta t})^{c_p} \tag{14}$$

Where $Q_p^{t+\Delta t}$ is the flow rate through the pump. The parameters a_p, b_p, and c_p define the shape of the pump curve and are empirically determined. $H_{e,p}^{t+\Delta t}$ is the effective pump head, and is defined as follows:

$$H_{e,p} = \begin{cases} H_{max,p}, & H_{jdp} - H_{jup} > H_{max,p} \\ H_{jdp} - H_{jup}, & H_{min,p} < H_{jdp} - H_{jup} < H_{max,p} \\ H_{min,p}, & H_{jdp} - H_{jup} < H_{min,p} \end{cases} \tag{15}$$

Where H_{jup} is the head at the superjunction upstream of the pump, H_{jdp} is the head at the superjunction downstream of the pump, and $H_{min,p}$ and $H_{max,p}$ are the minimum and maximum allowable head differentials across the pump, respectively.

2.2.4. Tank implementation

Storage tanks are represented as superjunctions in PipeDream, with the tank storage area represented by superjunction surface area A_{sj}. Tank rules related to the minimum and maximum tank water levels are implemented using controllable orifices that connect the inlet/outlet of storage tanks to an upstream/downstream superjunction. To model tank rule behavior, the orifice is specified to close when the tank water level reaches its minimum or maximum level.

2.3. Extended Kalman filter

Kalman filtering is an optimal state estimation technique that estimates the internal states of a linear dynamical system by combining observations of the system state together with estimates from a process model. This technique can be adapted to nonlinear systems like the Saint-Venant equations through the use of the EKF (Simon, 2006). In this section, we describe a method for applying the EKF to the WDS state-space model developed in the preceding section.

To start, we may express the internal states and observed outputs of the water distribution system in terms of a hidden state equation (Eq. (16)) along with an observation equation (Eq. (17)):

$$\mathbf{x}_{t+\Delta t} = A_t \mathbf{x}_t + B_t \mathbf{u}_t + \mathbf{v}_t \tag{16}$$

$$\mathbf{y}_t = C_t \mathbf{x}_t + \mathbf{w}_t \tag{17}$$

Where $\mathbf{y}_t$ is a vector of observations at time t, C_t is the observation matrix, $\mathbf{v}_t$ is a random vector corresponding to the process (i.e., hydraulic model) noise, and $\mathbf{w}_t$ is a random vector corresponding to the measurement noise.

The state equation describes how the internal system states (e.g., pressure heads) evolve over time, while the observation equation describes our observations of the system states (e.g., pressure sensor measurements). The observation matrix C_t represents a linear function that maps internal states to observed outputs. In this case, the

observation matrix is a subset of the rows of the identity matrix that when multiplied by the state vector returns the pressure heads at the subset of gaged junctions. Both the state and observation equations are stochastic. The state equation is subject to *process noise*, which reflects the uncertainty in the exogenous input (e.g., nodal demands), and is represented by the random vector $\mathbf{v}_t$. The observation equation is subject to *measurement noise*, which reflects uncertainty in the measurement (e.g., sensor noise), and is represented by the random vector $\mathbf{w}_t$.

The conventional KF is the minimum mean-squared error estimator of system states if (i) the process model is perfectly accurate; (ii) both the process noise $\mathbf{v}_t$ and the measurement noise $\mathbf{w}_t$ are zero-mean and uncorrelated, with known covariance matrices V_t and W_t, respectively; and (iii) all errors are Gaussian. For a nonlinear system like the Saint-Venant equations, these assumptions do not hold because nonlinear transformations of a Gaussian random variable are not Gaussian in general. However, EKF may be applied to produce a near-optimal estimate of the internal system states. EKF works by approximating the true nonlinear system at each time step using its Jacobian. For our implicit discretization of the Saint-Venant equations, the Jacobian is precisely the A_t matrix described in Eq. (9).

Using EKF, our best estimate $\hat{\mathbf{x}}_{t+\Delta t}$ of the internal system state is given by feeding back the difference between the observed and modeled output multiplied by a gain factor $L_{t+\Delta t}$:

$$\hat{\mathbf{x}}_{t+\Delta t} = A_t\hat{\mathbf{x}}_t + B_t\mathbf{u}_t + L_{t+\Delta t}[\mathbf{y}_{t+\Delta t} - C_t(A_t\hat{\mathbf{x}}_t + B_t\mathbf{u}_t)] \tag{18}$$

The optimal Kalman gain L_t at time step t is given by:

$$L_t = P_tC_t^T(C_tP_tC_t^T + W_t)^{-1} \tag{19}$$

And the estimation error covariance matrix $P_{t+\Delta t}$ is given by:

$$P_{t+\Delta t} = A_t(P_t - P_tC_t^T(C_tP_tC_t^T + W_t)^{-1}C_tP_t)A_t^T + V_t \tag{20}$$

Thus, given an initial estimate of the system state $\hat{\mathbf{x}}_0 = \mathbb{E}[\mathbf{x}_0]$ and an initial estimate of the error covariance $P_0 = \mathbb{E}[(\mathbf{x}_0 - \hat{\mathbf{x}}_0)(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T]$, EKF estimates the system state and error covariance at each time step using the following recursion:

1. The estimation error covariance $P_{t+\Delta t}$ is computed (Eq. (20)).
2. The error covariance is used to compute the Kalman Gain $L_{t+\Delta t}$ (Eq. (19)).
3. The posterior estimate of the state vector $\hat{\mathbf{x}}_{t+\Delta t}$ is computed using the Kalman gain (Eq. (18)).

This process is repeated at each time step until the simulation is terminated.

2.4. Implementation and validation

The hydraulic solver is implemented in Python as an extension to the PipeDream software package. The *numba* just-in-time compiler is used to accelerate numerical code. To evaluate the applicability of the Saint-Venant equations for WDSs, we compare the output of the PipeDream solver against the output of the EPANET solver, an industry standard public domain software for hydraulic modeling (Rossman et al., 2020). The hydraulic model is constructed by extracting all the WDS characteristics, such as pipe diameters, length, loss coefficients nodal elevations and demands from the EPANET standard input file (`.inp`). The following conversion is used between the computational elements. First, nodes and pipes in EPANET are represented as superjunctions and superlinks in PipeDream, respectively. Nodal demands in EPANET are represented as negative nodal inflows in PipeDream, with the base demand at each superjunction scaled by the demand pattern specified in EPANET. Reservoirs are represented as fixed head boundary conditions at superjunctions, and initial tank levels are assumed to be known. Demand patterns are sampled at user-specified timesteps using the same method as EPANET, such that both models are executed with the same inputs. The duration of the extended period simulation, simulation time step, number of convergence iterations, head tolerance, and friction model can be specified by the user. Python codes for executing the simulations are made publicly available on a GitHub repository (Bartos et al., 2024).

Table 1
Comparison between PipeDream and EPANET simulation results.

Network	Head MAD (m)	Flow MAD (10^{-4} m^3/s)
NET2	0.26	0.15
CA1	0.04	2.30
PA1	0.24	0.69

3. Application and results

In this section, we assess the accuracy and running times for PipeDream compared to EPANET and evaluate the proposed state estimation procedure. We demonstrate the results using three benchmark networks extensively used in the WDS research community (Ormsbee et al., 2022):

- **Net2** is based on the distribution system in the city of Cherry Hill, NJ. NET2 consists of 1 tank, 40 pipes, and 35 junctions.
- **CA1** is based on the WDS for the city of Fairfield, CA, and contains 1 tank, 126 pipes, and 111 junctions.
- **PA1** is based on the North Penn Water Authority WDS, and contains 2 tanks, 399 pipes, and 337 junctions.

3.1. Hydraulic solver accuracy & speed

First, the applicability of the PipeDream model to WDSs is assessed by comparing model results against EPANET—an industry standard model for WDS simulation. The three test networks are simulated using both EPANET and PipeDream. Next, the superjunction heads and superlink flows predicted by PipeDream are compared against the nodal heads and link flows predicted by EPANET. To compare results between the two models, we compute the mean absolute deviation (MAD) between EPANET and PipeDream results over the duration of the simulation for each network. In addition to testing the accuracy of the new solver, we also test its computational speed with respect to (i) the solver time step and (ii) the number of convergence iterations at each timestep.

Solver accuracy. PipeDream closely matches the output of EPANET for all networks considered. Table 1 shows the MAD in head and flow between EPANET and PipeDream results over the 24-h simulation duration for all three networks using a 30 min simulation time step with 40 convergence iterations per time step. Fig. 2 shows maps of the MAD in nodal heads (top) and link flows (bottom) for the three networks. Differences in nodal heads between EPANET and PipeDream are extremely small with maximum MAD across all nodes in all networks being 0.26 [m], occurring in NET2. Differences in link flows between EPANET and PipeDream are greater, but are still within acceptable tolerances for practical applications, with maximum difference of $2.30 \cdot 10^{-4}$ $[m^3/s]$ for CA1. Some discrepancies are expected, given that EPANET and PipeDream use different governing equations (conservation of energy vs. conservation of momentum), feature different control volumes (with PipeDream accounting for compressibility via the Priessman slot and nodal surface areas), and use different convergence criteria. However, for perspective, these head and flow differences are less than the precision level of common measurement instruments (Walski, 2021; AWWA, 2018). Overall, the results indicate that PipeDream produces hydraulic simulation results that are highly consistent with EPANET for networks of varying sizes and configurations.

Solver speed. While a trade-off between simulation run time and accuracy is observed, we find that using a time step on the order of

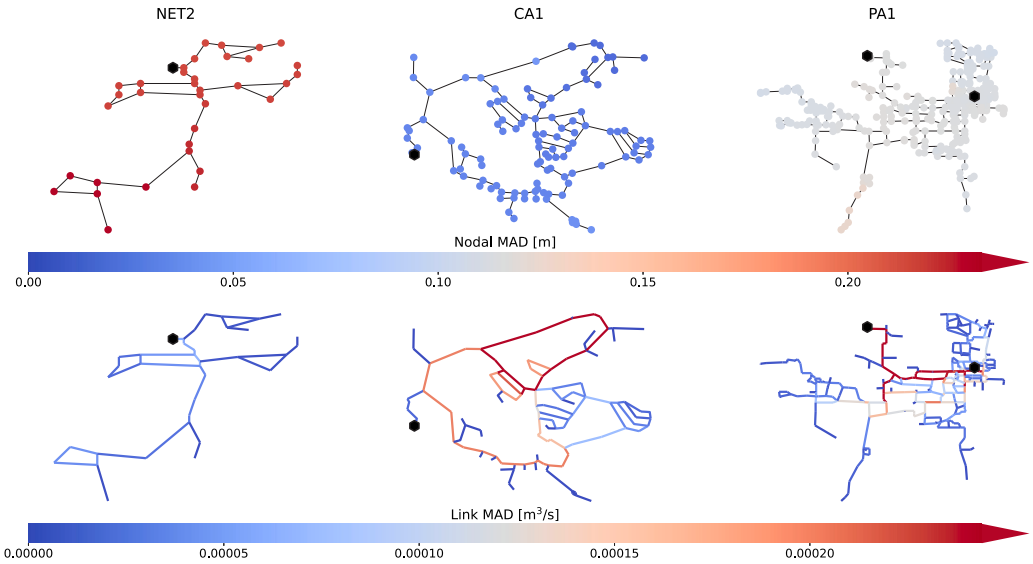


Fig. 2. Comparison between PipeDream and EPANET results: (top) heads and (bottom) flow rates.

Table 2
Comparison between PipeDream and EPANET run times.

Converge iterations	Simulation run time (multiples of EPANET)					MAD in node heads (m)				
	Time step (s)					Time step (s)				
	60	300	600	900	1800	60	300	600	900	1800
40	19.27	6.17	3.72	2.95	1.95	0.02	0.03	0.06	0.09	1.18
20	19.35	6.01	3.66	2.87	1.93	0.02	0.03	0.06	0.09	1.18
10	19.53	5.62	3.50	2.55	1.54	0.02	0.03	0.06	0.09	1.18
5	18.82	4.81	2.56	1.81	0.99	0.02	0.03	0.06	0.09	1.18
1	8.00	1.55	0.83	0.55	0.32	0.05	0.07	0.07	0.09	0.16

30 min and between 5–10 convergence iterations generally ensures both accurate results and fast run times (between 1–2 times that of EPANET). Table 2 (left panel) shows the mean simulation run times of the PipeDream solver across all networks as multiples of EPANET run time by time step and number of convergence iterations, using 10 trials per network. For smaller time steps (on the order of 60 s), PipeDream is roughly 20 times slower than EPANET. However, for time steps on the order of 30 min (1800 s), which is typical for hydraulic simulations, PipeDream achieves speeds on par with EPANET, with run times between 0.32 and 1.95 times that of EPANET depending on the number of convergence iterations chosen. PipeDream is capable of achieving speeds that are comparable to EPANET while also incurring minimal numerical error. Table 2 (right panel) shows the mean MAD in node heads across all networks by time step and number of convergence iterations. The smallest errors are obtained using a small time step; however, for all realizations except those using a single convergence iteration, the number of convergence iterations does not affect the nodal heads. Overall, using a time step of 30 min and 5 convergence iterations ensures a mean run time equal to that of EPANET, with a mean difference of 1.18 [m] in nodal heads over all networks considered. The results of this assessment demonstrate that PipeDream is a suitable numerical model for water distribution networks.

3.2. State estimation

A holdout assessment is performed to evaluate the performance of EKF at estimating hydraulic heads at ungauged locations. This assessment represents an important use-case for WDSs, given that operators desire real-time information about hydraulic states at the system scale but typically only have access to a small number of measurement sites. To conduct our holdout assessment, synthetic pressure sensor data are first generated from a 'ground-truth' WDS model and then assimilated into a 'de-calibrated' WDS model at a subset of nodes using EKF. Pressure heads are then examined at the holdout sites to determine whether EKF improves the interpolation of hydraulic states at ungauged locations compared to a model-only approach.

The PipeDream model is first simulated under a calibrated parameter set to produce 'ground truth' time series of pressure heads at each node for each network. These 'ground truth' time series represent the true physical behavior of the underlying WDSs. Synthetic sensor data are then generated by adding random Gaussian noise to these ground truth time series. In order to simulate realistic conditions in which model outputs do not exactly align with real-world system behavior, modifications are then made to "de-calibrate" the ground truth hydraulic model, thereby producing an imperfect (de-calibrated) hydraulic model. Random noise is added to node demands, node elevations, and pipe friction coefficients such that the model results differ from the ground truth results. These three changes represent common sources of uncertainty in WDS models, and thus the de-calibrated model reflects operator uncertainty about the real-world behavior of the WDS.

Next, EKF is applied to assimilate sensor data into the de-calibrated hydraulic models and produce posterior estimates of the hydraulic states at ungauged locations. To quantify the performance of the EKF at interpolating hydraulic states, both the EKF estimate and the de-calibrated model are compared against the ground truth time series of pressure at each ungauged node. The performance of the estimates are characterized in terms of the mean absolute deviation (MAD) between the estimate and the ground truth pressure heads. For this assessment, the standard deviation of the process noise, $\mathbf{v}_t$, and measurement noise, $\mathbf{w}_t$, are selected as 0.5 [m] (the same standard deviation used to add noise to the artificial sensor data). We also evaluate the improvement in estimation accuracy under differing numbers of sensors. Because the results of EKF are sensitive to the placement of sensors, a method for selecting sensor locations within the WDSs is needed. For this purpose, we use a greedy approximation algorithm, which has previously been used for placement of pressure and water quality sensors in WDSs (Sela and Amin, 2018; Krause et al., 2008). The greedy approach consists of selecting at each iteration the sensor location that results in the lowest MAD, until no more sensors are available or no additional improvement is achieved. For the three networks in this study, we assess the incremental improvement in accuracy offered by each additional sensor, starting with zero sensors (corresponding to the model-only approach) and terminating at seven sensors.

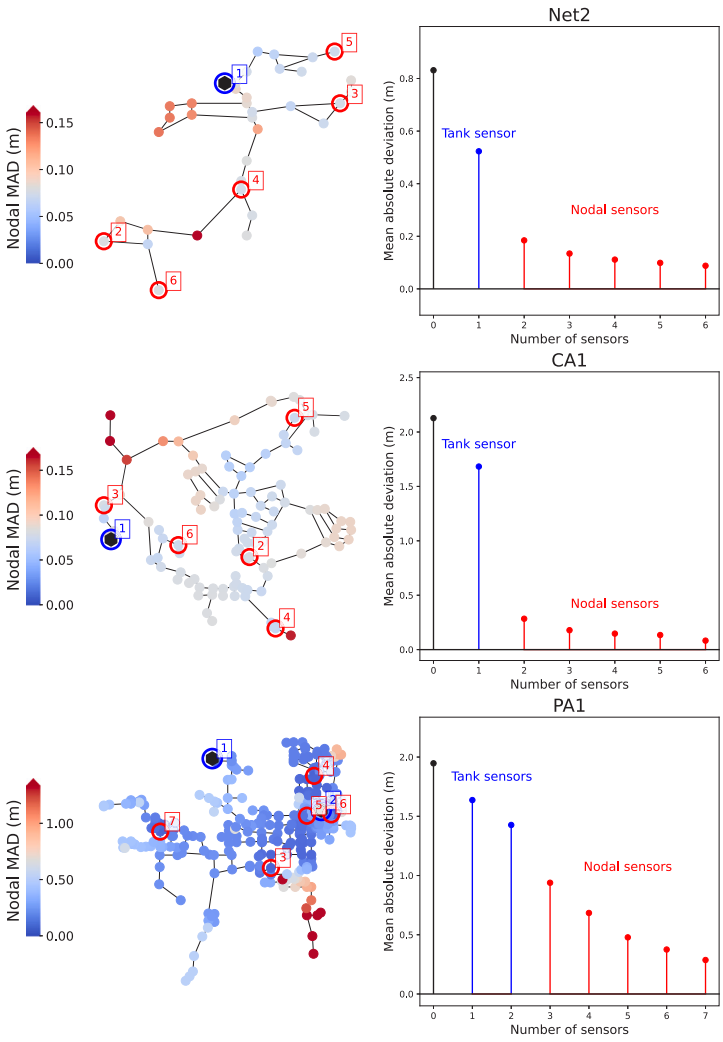


Fig. 3. EKF results: (left) sensor locations and MAD of nodal heads where numbers indicate the order in which sensors are added, and (right) mean MAD across all nodes in the networks with each added sensor. The red labels indicate the order in which sensors are placed in the networks. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The results of the holdout assessment show that EKF significantly improves estimates of pressure heads at ungaged locations in WDSs, even when only a small number of sensors are used. Fig. 3 (right) shows the mean estimation error over all junctions and time steps (in terms of pressure head MAD) for increasing numbers of sensors. Similarly, Fig. 3 (left) shows the MAD across all nodes for all three networks along with sensor locations. Here, the sensor number refers to order in which sensors are added to the system. Water level sensors at the tanks are added first followed by pressure sensors at the nodes based on the greedy site selection algorithm. We observe that overall, EKF yields a close estimate of hydraulic states with the greatest mean deviation ranging from 1.6 [m] with one sensor and 0.45 [m] with 7 sensors for PA1. In addition, as expected, the MAD curves exhibit a monotonic decrease and diminishing return, indicating that more sensors is better, but the benefits of including more sensors decreases as additional sensors are added. Indeed, for Net2 and CA1, accurate pressure estimates are obtained with a small number of sensors. For Net2, the MAD is decreased from roughly 0.8 [m] with no sensors to roughly 0.2 [m] with only one tank sensor and one nodal sensor. Similarly, for CA1, the MAD is decreased from roughly 2.1 [m] with no sensors to less than 0.4 [m] with one tank sensor and one nodal sensor. These results suggest that state estimation is effective in WDSs even when the number of deployed sensors is small.

To further examine the potential of EKF to estimate hydraulic states at ungauged locations in WDSs, we assess time series of estimated pressure heads at select holdout locations (Fig. 4) along with the spatial distribution of estimation errors as more sensors are added (Fig. 5). Fig. 4 (left) shows four selected holdout locations in PA1 (Nodes 29, 311, 366, and 641), along with time series of pressure heads at these nodes. These time series show pressure head estimates from the 'de-calibrated' model (blue), the true heads at holdout locations (dotted black), and the EKF estimates under increasing number of sensors: from 1 sensor (light red) to 7 sensors (dark red). We observe that EKF accurately estimates the hydraulic heads at the ungauged locations even in circumstances when the hydraulic model both underestimates and overestimates the true state. We also observe that EKF estimation accuracy improves and noise decreases as additional sensors are added. Fig. 5 shows the spatial distribution of pressure head MAD as more sensors are added. These results demonstrate that the estimation error is smallest close to the sensor locations and improves as sensor coverage increases. These results suggest that sensor placement is critical for improving the estimates and should be carefully designed.

3.3. Sensitivity analysis

Lastly, we explore how the process and measurement uncertainty impact the state estimates. The process covariance, V_t, represents the uncertainty associated with the hydraulic model, while the measurement covariance, W_t, represents the uncertainty associated with the

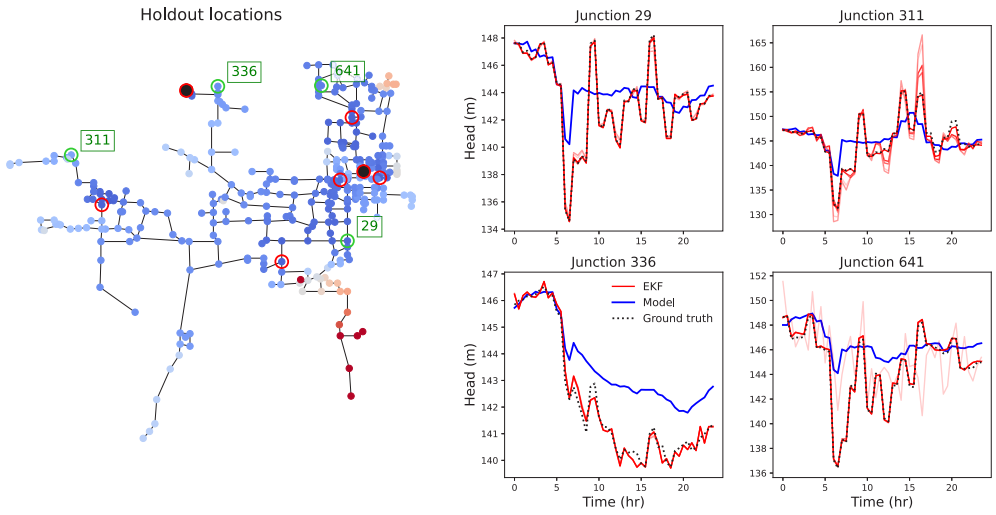


Fig. 4. State estimation performance in PA1: model-based estimation (blue), true heads at holdout locations (Nodes 29, 311, 366, and 641) (dotted black), and the EKF estimates with increasing number of sensors: 1 sensor (light red) – 7 sensors (dark red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

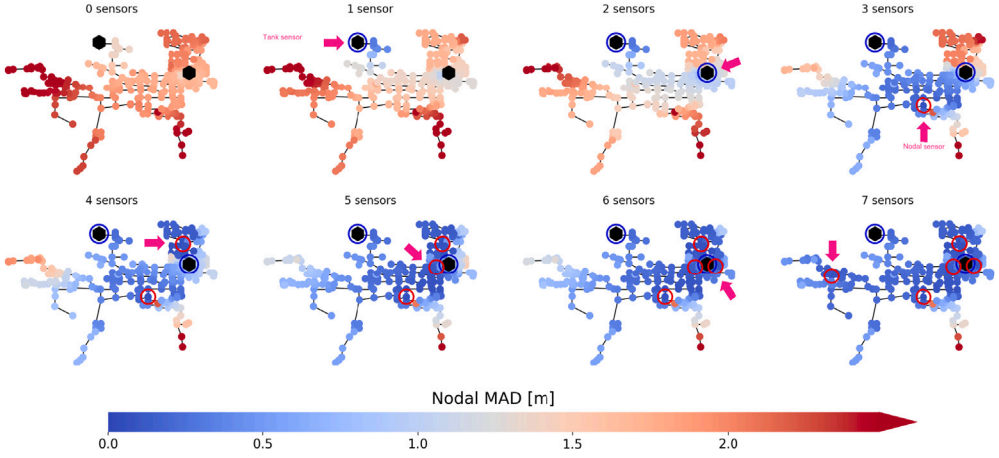


Fig. 5. State estimation performance in PA1: spatial distribution of MAD with added sensors.

measurements. Therefore, the ratio between the process and measurement noise quantifies the degree to which EKF trusts the hydraulic model versus the measurements. A higher process noise suggests that the filter relies more on observations and the estimates are likely to be closer to the measurements. On the other hand, a higher measurement noise implies that the filter will assign more emphasis on the model and the estimates are likely to align more closely with the hydraulic model. To test this assumption, we vary the process noise while keeping measurement noise constant and explore the effect on the EKF estimates. Fig. 6 shows the response in the estimates at the four holdout locations for small (top row), medium (middle row), and large (bottom row) process to measurement noise ratios (corresponding to process-to-measurement noise variance ratios of 2.5×10^{-8}, 2.5×10^{-6}, and 10, respectively). As expected, when the process noise is small compared to the measurement noise (top row), the EKF estimates (dotted black lines) closely align with the hydraulic model predictions (blue lines). As process noise increases, EKF estimates assign more emphasis to the measurements, thus approaching the ground truth (bottom row). Fig. 7 demonstrates the overall MAD in pressure heads between EKF and the hydraulic model as a function of the process to measurement noise ratio. For the WDSs explored here, we observe the expected results wherein smaller ratios of process to measurement variance result in closer alignment to the hydraulic model. Additionally, the dependence is not linear and at some point further changes in the ratios do no impact the estimates. These results demonstrate that the proposed state estimation approach can be readily adapted to account for individual operator knowledge regarding the relative trustworthiness of sensor and model data. For situations where operators place greater confidence in the model, for instance, the ratio of process to measurement covariance can be lowered to account for operator preference and place greater weight on model estimates.

4. Conclusions

This research presents a new method for state estimation in water distribution networks that will enable operators to more reliably track pressures and flow rates within large, distributed WDSs. We first demonstrate that the Saint-Venant equations with a negative Preissman slot scheme can accurately and efficiently model pressurized pipe system dynamics. Using this new process model, we subsequently show that EKF significantly improves the estimation of pressure heads at ungaged locations in WDSs, even with a relatively small number of sensors. The state estimation approach proposed in this study provides operators with an accurate real-time method to estimate heads and flows, even when sensor observations are sparse and hydraulic models are uncertain. These results provide a foundation for several avenues of further exploration. First, the new formulation of system hydraulics, which conforms to state-space representation of dynamical systems,

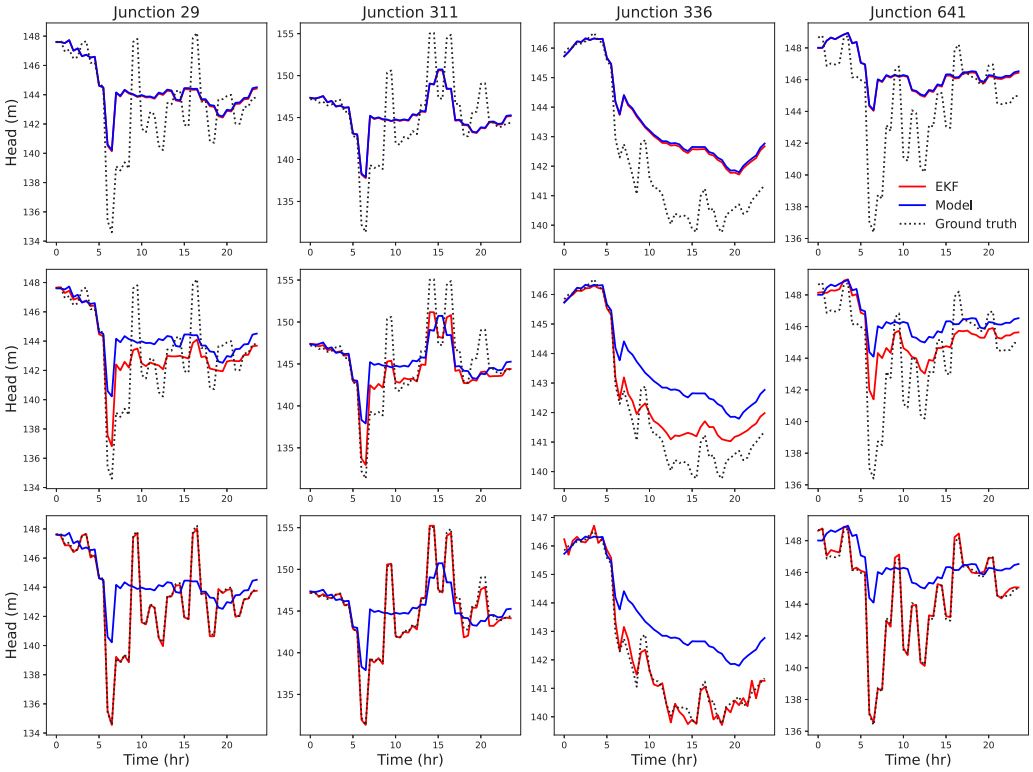


Fig. 6. Sensitivity of EKF estimates to process noise: (top) low, (middle) medium, and (bottom) high process noise at holdout locations (Nodes 29, 311, 366, and 641) in PA1. Hydraulic model (blue), true heads (dotted black), and EKF estimates (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

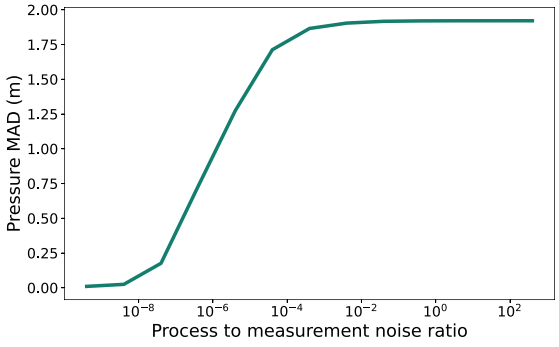


Fig. 7. Sensitivity of EKF estimates to process and measurement noise: mean absolute difference in pressure heads between hydraulic model and EKF estimates.

lends itself to rich mathematical theory and fundamental tools for solving a broad class of estimation and control problems (Simon, 2006). Second, this new formulation can also be extended to capture both steady and transient hydraulics in the same modeling framework, and can thus be used to explore estimation and control of WDSs at different temporal scales. Third, the state estimation approach can be extended to include other water management problems that are critical to water utilities, such as leak detection and demand estimation (Chu et al., 2021; Zhou et al., 2023a; Rajabi et al., 2023).

Future work is needed to increase the usability of the proposed approach for state estimation in WDSs, specifically: (1) *Hydraulic devices and controls.* The current approach has been tested with models of benchmark WDSs widely used in the research community with systems including reservoirs, storage tanks, and pumps. To facilitate applicability to a wider range of systems, PipeDream should be extended to handle common hydraulic devices that are characteristic of real-world WDSs, such as pressure reducing and flow control valves, and tank level controls. (2) *Scalability.* The proposed method demonstrates results comparable with EPANET with respect to computational times and accuracy for medium-size WDSs models having hundreds of nodes and links. Realistic WDSs typically comprise tens of thousands of nodes and links, and hence the accuracy and computational complexity of the proposed SE approach should be tested and evaluated for networks of larger sizes. (3) *Computational enhancements.* Scaling to larger networks will be facilitated by improving the computational efficiency of the solver and state estimation algorithm. For large networks, our solver may be further optimized by incorporating parallelization of computations across superlinks, reducing linear chains of links and nodes into superlinks, and utilization of specially-structured (banded) matrices and linear algebra routines. The effect of these optimizations on model scalability should be investigated rigorously in future work. Moreover, future work should explore numerical optimizations to the classic EKF algorithm that improve computational speed and numerical stability for larger systems, including the square root covariance filter, the square root information filter, the Chandrasekhar square root filter, and reduced-rank variants of the Kalman Filter (Verhaegen and Van Dooren, 1986; Farrell and Ioannou, 2001). (4) *Joint state-parameter estimation.* A major advantage of our proposed approach is that it can be readily integrated into joint state-parameter estimation schemes to infer uncertain states and static parameters simultaneously. Future work should explore the potential to further improve interpolation and forecasting of hydraulic states in WDSs through joint state-parameter estimation schemes like expectation–maximization (EM) (Dempster et al., 1977). (5) *Data sparsity and sensor placement.* In this work, data was synthetically generated for all locations and used to place sensors in the WDSs. In realistic settings, only limited number of sensors, and thus, limited information will be available. Hence, there is a need to create optimization methods for sensor placement that can account for data (in)adequacy and additional performance objectives. (6) *Water quality*

estimation. As water quality is critical to maintaining public health and continuous water quality monitoring within WDSs is scarce, it is important to extend the modeling to include water quality parameters, such as disinfectant residuals, and explore state estimation suitable for utilizing grab and intermittent sampling. While further research is warranted, the methodology presented in this paper is the first to provide a fast, scalable, and 'batteries-included' approach to state estimation in WDSs that can be readily applied to real-world networks currently modeled by hydraulic software like the EPANET. To improve accessibility and promote the use of data assimilation within the field of water resources, Python code for reproducing the analysis presented in this work are made publicly available in a GitHub repository (Bartos et al., 2024). We expect that this work will contribute to the development of new online models informed by real-time sensor data that will support decision-making for water utilities.

CRediT authorship contribution statement

Matthew Bartos: Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Conceptualization. **Meghna Thomas:** Writing – review & editing, Visualization, Validation, Software, Methodology. **Min-Gyu Kim:** Software, Methodology. **Matthew Frankel:** Software, Investigation, Data curation. **Lina Sela:** Writing – original draft, Supervision, Funding Acquisition, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Code and data for this study are available at: https://github.com/future-water/wds-state-estimation.

Acknowledgments

This work was supported in part by the National Science Foundation under Grants 1943428 and 2220516.

References

Arnell, M., Miltell, M., Olsson, G., 2023. Making waves: A vision for digital water utilities. Water Res. X 19, 100170. http://dx.doi.org/10.1016/j.wroa.2023.100170, URL https://www.sciencedirect.com/science/article/pii/S2589914723000063.

Arsene, C.T., Gabrys, B., 2014. Mixed simulation-state estimation of water distribution systems based on a least squares loop flows state estimator. Appl. Math. Model. 38 (2), 599–619.

AWWA, 2018. Addendum to water meters—Selection, installation, testing, and maintenance. In: Manual of Water Supply Practices M6, fifth ed., American Water Works Association, Denver.

Bartos, M., Kerkez, B., 2021. Pipedream: An interactive digital twin model for natural and urban drainage systems. Environ. Model. Softw. 144, 105120.

Bartos, M., Thomas, M., Frankel, M., Kim, M.-G., Sela, L., 2024. State estimation in water distribution systems code. URL https://github.com/future-water/wds-state-estimation.

Blocher, C., Pecci, F., Stoianov, I., 2020. Localizing leakage hotspots in water distribution networks via the regularization of an inverse problem. J. Hydraul. Eng. 146 (4), 04020025.

Boulos, P.F., Lansey, K.E., Karney, B.W., 2006. Comprehensive Water Distribution Systems Analysis Handbook for Engineers and Planners. MWH Soft, Inc..

Chu, S., Zhang, T., Yu, T., Wang, Q.J., Shao, Y., 2021. A noise adaptive approach for nodal water demand estimation in water distribution systems. Water Res. 192, 116837. http://dx.doi.org/10.1016/j.watres.2021.116837, URL https://www.sciencedirect.com/science/article/pii/S004313542100035X.

Daniel, I., Ajami, N.K., Castelletti, A., Savic, D., Stewart, R.A., Cominola, A., 2023. A survey of water utilities' digital transformation: drivers, impacts, and enabling technologies. npj Clean Water 6 (1), 51.

Delgado-Aguiñaga, J.A., Begovich, O., 2017. Water leak diagnosis in pressurized pipelines: A real case study. In: Verde, C., Torres, L. (Eds.), Modeling and Monitoring of Pipelines and Networks: Advanced Tools for Automatic Monitoring and Supervision of Pipelines. Springer International Publishing, Cham, pp. 235–262.

Delgado-Aguiñaga, J., Besançon, G., 2018. EKF-based leak diagnosis schemes for pipeline networks. IFAC-PapersOnLine 51 (24), 723–729, 10th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS 2018.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B Stat. Methodol. 39 (1), 1–22. http://dx.doi.org/10.1111/j.2517-6161.1977.tb01600.x.

Do, N.C., Simpson, A.R., Deuerlein, J.W., Piller, O., 2017. Particle filter based model for online estimation of demand multipliers in water distribution systems under uncertainty. J. Water Resour. Plan. Manag. 143 (11), 04017065.

Farrell, B.F., Ioannou, P.J., 2001. State estimation using a reduced-order Kalman filter. J. Atmos. Sci. 58 (23), 3666–3680. http://dx.doi.org/10.1175/1520-0469(2001)058<3666:seuaro>2.0.co;2.

Ji, Z., 1998. General hydrodynamic model for sewer/channel network systems. J. Hydraul. Eng. 124 (3), 307–315. http://dx.doi.org/10.1061/(asce)0733-9429(1998)124:3(307).

Kang, D., Lansey, K., 2009. Real-time demand estimation and confidence limit analysis for water distribution systems. J. Hydraul. Eng. 135 (10), 825–837.

Kerger, F., Archambeau, P., Erpicum, S., Dewals, B., Pirotton, M., 2011. An exact Riemann solver and a Godunov scheme for simulating highly transient mixed flows. J. Comput. Appl. Math. 235 (8), 2030–2040. http://dx.doi.org/10.1016/j.cam.2010.09.026.

Krause, A., Leskovec, J., Guestrin, C., VanBriesen, J., Faloutsos, C., 2008. Efficient sensor placement optimization for securing large water distribution networks. J. Water Resour. Plan. Manag. 134 (6), 516–526.

Mankad, J., Natarajan, B., Srinivasan, B., 2022. Integrated approach for optimal sensor placement and state estimation: A case study on water distribution networks. ISA Trans. 123, 272–285.

Ormsbee, L., Hoagland, S., Hernandez, E., Hall, A., Ostfeld, A., 2022. Hydraulic model database for applied water distribution systems research. J. Water Resour. Plan. Manag. 148 (8), 04022037.

Pecci, F., Parpas, P., Stoianov, I., 2020. Sequential convex optimization for detecting and locating blockages in water distribution networks. J. Water Resour. Plan. Manag. 146 (8), 04020057.

Preis, A., Whittle, A.J., Ostfeld, A., Perelman, L., 2011. Efficient hydraulic state estimation technique using reduced models of urban water networks. J. Water Resour. Plan. Manag. 137 (4), 343–351.

Preissmann, A., 1961. Propagation des intumescences dans les canaux et rivières. In: 1st Congress of the French Association for Computation. Grenoble, France.

Rajabi, M.M., Komeilian, P., Wan, X., Farmani, R., 2023. Leak detection and localization in water distribution networks using conditional deep convolutional generative adversarial networks. Water Res. 238, 120012.

Riaño-Briceño, G., Hodges, B.R., Sela, L., 2022. PTSNet: A parallel transient simulator for water transport networks based on vectorization and distributed computing. Environ. Model. Softw. 158, 105554. http://dx.doi.org/10.1016/j.envsoft.2022.105554, URL https://www.sciencedirect.com/science/article/pii/S1364815222002547.

Rossman, L., Woo, H., Tryby, M., Shang, F., Janke, R., Haxton, T., 2020. EPANET 2.2 User Manual. Tech. Rep., (EPA/600/R-20/133), U.S. Environmental Protection Agency, Washington, D.C..

Ruiz, E., Díaz, S., González, J., 2022. Potential performance of hydraulic state estimation in water distribution networks. Water Resour. Manag. 1–18.

Sela, L., Amin, S., 2018. Robust sensor placement for pipeline monitoring: Mixed integer and greedy optimization. Adv. Eng. Inform. 36, 55–63.

Shafiee, M.E., Barker, Z., Rasekh, A., 2018. Enhancing water system models by integrating big data. Sustainable Cities Soc. 37, 485–491.

Simon, D., 2006. Optimal State Estimation: Kalman, H-Infinity, and Nonlinear Approaches. John Wiley & Sons.

Todini, E., Rossman, L.A., 2013. Unified framework for deriving simultaneous equation algorithms for water distribution networks. J. Hydraul. Eng. 139 (5), 511–526. http://dx.doi.org/10.1061/(ASCE)HY.1943-7900.0000703.

Torres, L., Jiménez-Cabas, J., González, O., Molina, L., López-Estrada, F.-R., 2020. Kalman filters for leak diagnosis in pipelines: Brief history and future research. J. Mar. Sci. Eng. 8 (3).

Tshehla, K.S., Hamam, Y., Abu-Mahfouz, A.M., 2017. State estimation in water distribution network: A review. In: 2017 IEEE 15th International Conference on Industrial Informatics. INDIN, pp. 1247–1252.

Vasconcelos, J.G., Wright, S.J., 2007. Comparison between the two-component pressure approach and current transient flow solvers. J. Hydraul. Res. 45 (2), 178–187. http://dx.doi.org/10.1080/00221686.2007.9521758.

Verhaegen, M., Van Dooren, P., 1986. Numerical aspects of different Kalman filter implementations. IEEE Trans. Autom. Control 31 (10), 907–917. http://dx.doi.org/10.1109/tac.1986.1104128.

Waldron, A., Pecci, F., Stoianov, I., 2020. Regularization of an inverse problem for parameter estimation in water distribution networks. J. Water Resour. Plan. Manag. 146 (9), 04020076.

Walski, T., 2021. Discussion of "regularization of an inverse problem for parameter estimation in water distribution systems" by Alexander Waldron, Filippo Pecci, and Ivan Stoianov. J. Water Resour. Plan. Manag. 147 (12), 07021020.

Wang, S., Taha, A.F., Gatsis, N., Sela, L., Giacomoni, M.H., 2022. Probabilistic state estimation in water networks. IEEE Trans. Control Syst. Technol. 30 (2), 507–519. http://dx.doi.org/10.1109/TCST.2021.3066102.

Xing, L., Sela, L., 2019. Unsteady pressure patterns discovery from high-frequency sensing in water distribution systems. Water Res. 158, 291–300.

Xing, L., Sela, L., 2020. Transient simulations in water distribution networks: TSNet Python package. Adv. Eng. Softw. 149, 102884. http://dx.doi.org/10.1016/j.advengsoft.2020.102884, URL https://www.sciencedirect.com/science/article/pii/S0965997819309342.

Xing, L., Sela, L., 2022. Graph neural networks for state estimation in water distribution systems: Application of supervised and semisupervised learning. J. Water Resour. Plan. Manag. 148 (5), 04022018.

Zhang, H., Wang, K., Zhou, X., Wang, W., 2018a. Using DFP algorithm for nodal demand estimation of water distribution networks. KSCE J. Civ. Eng. 22, 2747–2754.

Zhang, Q., Zheng, F., Duan, H.-F., Jia, Y., Zhang, T., Guo, X., 2018b. Efficient numerical approach for simultaneous calibration of pipe roughness coefficients and nodal demands for water distribution systems. J. Water Resour. Plan. Manag. 144 (10), 04018063.

Zhou, X., Liu, S., Xu, W., Xin, K., Wu, Y., Meng, F., 2022. Bridging hydraulics and graph signal processing: A new perspective to estimate water distribution network pressures. Water Res. 217, 118416.

Zhou, X., Xu, W., Xin, K., Yan, H., Tao, T., 2018. Self-adaptive calibration of real-time demand and roughness of water distribution systems. Water Resour. Res. 54 (8), 5536–5550.

Zhou, X., Zhang, J., Guo, S., Liu, S., Xin, K., 2023a. A convenient and stable graph-based pressure estimation methodology for water distribution networks: Development and field validation. Water Res. 233, 119747.