Delft University of Technology

A Multi-step Data Assimilation Framework to Investigate the Effect of Measurement Uncertainty in the Reduction of Water Distribution Network Model Errors

Fayaz, Ibrahim Miflal; Castro-Gama, Mario; Alfonso, Leonardo

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

Check for updates

A Multi-step Data Assimilation Framework to Investigate the Effect of Measurement Uncertainty in the Reduction of Water Distribution Network Model Errors

Ibrahim Miflal Fayaz [1] · Mario Castro-Gama[1,2,3] · Leonardo Alfonso[1]

Abstract

Water distribution network (WDN) models are a common decision support tool for understanding the behavior and performance of WDNs, aiding in the planning and management of WDN systems. The increasing availability of real-time data has recently promoted the exploration of Data Assimilation (DA) techniques to improve these models. However, flow, pressure and demand data are uncertain, particularly due to sensor characteristics such as precision and noise. An open question is to what extent DA can still improve hydraulic models when the data used to this end is uncertain. This paper proposes a three-step Ensemble Kalman Filter based DA approach for WDNs (3-EnKF-WDN), building on previous approaches, and advancing in two main fronts: the use of extended period simulation, and the use of pressure-dependent demand (PDD) analysis. Different scenarios considering uncertain sensor data, with varied precision and noise, are applied to two networks of different sizes, representative of real-world WDNs. The computational demand of the 3-EnKF-WDN method is also assessed. Results show that increasing sensor's precision and decreasing the noise in state measurements reduce model error, as expected. However, we also found that model errors: 1) are reduced more effectively by using 3-EnKF-WDN than by increasing sensors' precision; 2) are not reduced if certain noise thresholds are surpassed; 3) can be reduced without assimilating demand data if the WDNs are fully monitored with head sensors in all the nodes and flow sensors in all the links.

1 Introduction

WDNs are critical infrastructures that deliver potable water to consumers. Proper design and operation of these networks is essential to guarantee a reliable and safe water supply, guaranteeing public health and economic growth Adedeji et al. (2018). Hydraulic models are used to simulate and analyze WDNs, and are important tools to assist decisions. Model results

✉ Ibrahim Miflal Fayaz
  ibrahim.miflal@gmail.com

Extended author information available on the last page of the article

can be used to design and operate WDNs, ensuring reliability and efficiency. Traditionally, these models have been built and operated using historical data from sensors. However, the digital transformation of the water sector is promoting Advanced Metering Infrastructure (AMI), such as smart water meters and Supervisory Control and Data Acquisition (SCADA) systems that open the door to real-time modelling of WDNs, which is becoming increasingly important (Grievson et al. 2022). The use of real-time sensor data in models can help capturing the complexity and variability of real-world systems, leading to improved and timely decision-making (Rossman 1993; Antonowicz et al. 2018). Although there are different ways to use real-time data in modelling, the integration of technology and digitalization has given rise to new approaches to updating model states. One of them is Data Assimilation (DA), which has the potential of improving model accuracy in real-time by utilizing long-term measurement data (Hill et al. 2014). DA synthesizes prior knowledge of model states with available measurements to provide an optimized estimate of current model states and reduce uncertainties. However, these measurements can be unstable and contain larger errors. The ability to address measurement errors using calibration methods and efficiently utilizing a large amount of data is challenging Zhou et al. (2018).

The use of Kalman Filters (KF) for WDNs was first introduced by Todini (1999) for calibrating pipe roughness coefficients in WDNs with a simple linear structure. As KF can only be used for linear systems, the Extended Kalman Filter (EKF) was applied by Shang et al. (2006, 2008) to estimate nodal demands in a small hypothetical network by approximating nonlinear systems with tangent linear operators. These studies showed good results with KF and EKF in cases of limited nonlinearity and uncertainty, but their efficacy may be limited in highly looped networks (Van Den Bossche 2013) or the presence of large measurement errors Shang et al. (2006, 2008).

The effectiveness of the EnKF was proven in updating water demands and water demand model parameters for a Water Demand Forecasting Model under the assumption of known pipe roughness values and no leakage in the system (Okeya et al. 2014). They explored the possibility of burst detection using Kalman filtering of flow observations and forecasts from the hydraulic model, and an extension of this study by Okeya et al. (2014) showed that the applied methodology was effective in detecting bursts in real-time and estimating the leak flow. Ruzza (2017) carried out a similar leak detection study in WDNs using KF, EnKF, Ensemble Smoothing, and Normal-Score EnKF to identify nodal leakages. Ensemble-based methods are also effective in providing stable calibration results to ensure the long-term accuracy of models as demonstrated by Zhou et al. (2018, 2022).

EnKF avoids model linearization by simulating model states using an ensemble of parameters derived from Monte Carlo perturbations. Particle Filter (PF), which extends the use of the ensemble to non-Gaussian models and increases the ensemble size, was successfully used by Do et al. (2017a, b) to estimate nodal demand patterns in WDN models using measurements with specific errors. A recent study by Bragalli et al. (2016) tested the use of EnKF in WDNs using an innovative 3-step EnKF for a small WDN which showed promising results for the capabilities of a multi-step DA in WDNs.

EnKF is an ideal and optimal method for applying DA for WDN as EnKF is stable with large nonlinear systems and a low probability of divergence from the true value. The computational demand of EnKF is also lower than PF (Simon 2006; Gillijns et al. 2006; Van Den Bossche 2013).

Despite the previous research efforts, the application of DA techniques in WDNs is still limited. In particular, the extent to which model errors can be reduced under measurement uncertainty is still unknown. Additionally, extended-period simulations have not been carried

out for a multi-step DA algorithm. In previous studies, performed by Bragalli et al. (2016) and Okeya et al. (2014), Demand Driven Analysis (DDA) was used.

In this paper, a three-step Ensemble Kalman Filter-based DA for WDNs (3-EnKF-WDN) approach is presented. The approach is innovative as the hydraulic modelling involves extended period simulation and Pressure-Dependant Demand (PDD). The objective is to understand to which extent model errors can be reduced under measurement uncertainty, in particular due to sensor precision and noise, when incrementally assimilating the system states of pressure (step 1), flow (step 2) and demand (step 3). We also propose a new evaluation metric, Combined Total Variance Ratio, to quantify the overall effectiveness of this DA process. Additional analyses include the effect of the number of ensembles in the EnKFs, and the computational demand of 3-EnKF-WDN.

The remainder of the paper presents the methodology section, outlining the approach used in this study. Two case studies are used to demonstrate the application of the proposed DA method. Afterwards, the results are presented and discussed. Conclusions and findings are drawn in the last section.

2 Methodology

The methodology consists of three parts. The first part details the implementation of the improved DA algorithm, which starts with an initialization, and moves incrementally by assimilating pressure, flow and demand data. The second part presents the new evaluation metric, Combined Total Variance Ratio, to quantify the overall effectiveness of the DA process. Finally, the third part includes an experimental setup to evaluate the effect of the measurement uncertainties on the effectiveness of the DA process, the effect of different numbers of ensembles in the EnKF and the computational demand of the DA.

2.1 Three-step Ensemble Kalman Filter-based DA for WDNs (3-EnKF-WDN)

The structure of the 3-EnKF-WDN algorithm is shown in Fig. 1, and further detailed in the sections below.

The multi-step EnKF for WDNs involves initializing the ensemble of state estimates and updating the ensembles with measurements of head, flow, and demands. This process is repeated at each time step of the simulation to estimate the hydraulic state of the network over time. The q_j after the state symbol refers to the "known" demand which is used to initialize the 3-step DA.

2.1.1 Initialization

Before proceeding with the 3-steps, it is necessary to generate the initial ensemble of states describing our prior knowledge, using the following procedure:

1.(a) Generate an ensemble of demands (q) with a mean μ_{q_j} (base demand of each node) and variance $\sigma^2_{q_j}$
  (b) Using EPANET 2.2 modelling system and the WNTR Python library (Klise et al. 2017b, a) we compute matrices of pressure (H_{q_j}) and flowrate (Q_{q_j}) initialized in the network with the ensembles of demands (q_{meas}) and their averages $H_{|q_j}$, $Q_{|q_j}$, with |, denoting the average of the respective state being calculated
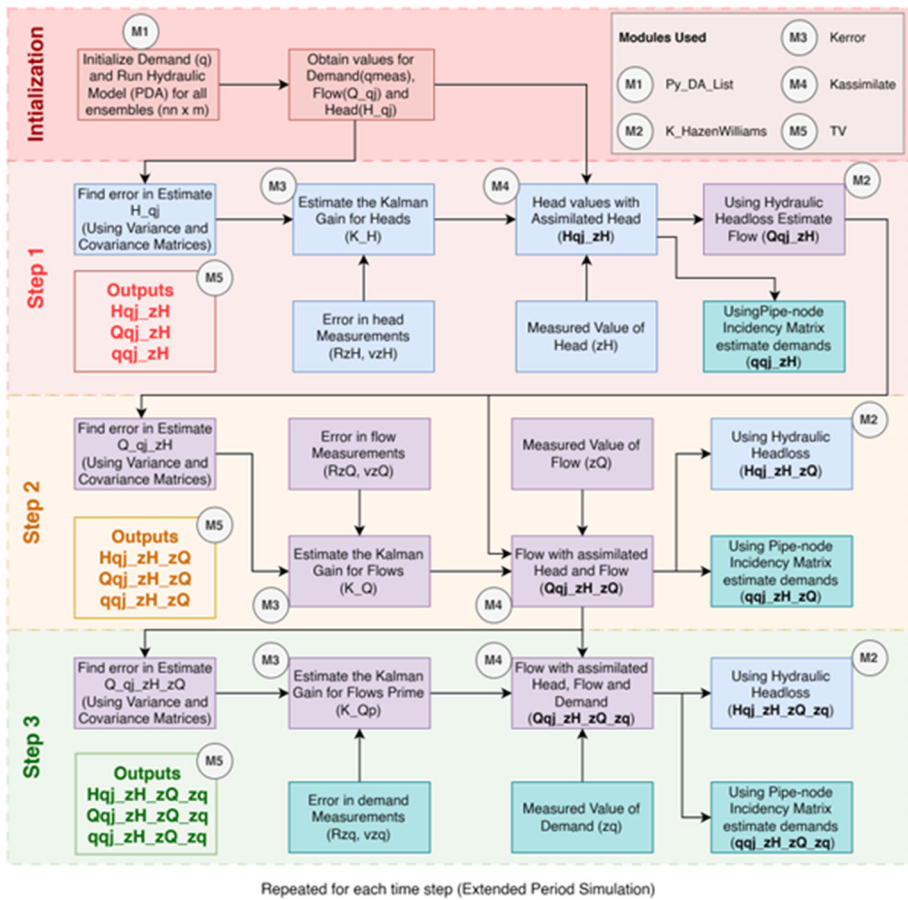
Fig. 1 Step-by-step implementation of the 3-Step DA Algorithm

(c) The number of ensembles "m" must be large enough for the estimated co-variance matrices to be inverted

Once initialised, data assimilation is carried out for up to 3-steps depending on the available type of measurements, as follows.

2.1.2 Step One - Assimilation of Pressure Head

Update the ensemble of state estimates with head measurements by calculating the Kalman gain, assimilating these measurements and estimating the flow and demand, as follows:

1.(a) Calculate the ensemble mean μ_H and ensemble prior variance of Head P_H, using Eqs. 1 and 2.

$$P_H = \frac{1}{m-1} \sum_{j=1}^{m} \left[\left(H_{|q_j} - \mu_H \right) \left(H_{|q_j} - \mu_H \right)^T \right] \tag{1}$$

$$\mu_H = \frac{1}{m} \sum_{j=1}^{m} H_{|q_j} \tag{2}$$

(b) Calculate the Kalman Gain K_H for the head using the error in the estimate and the errors in the measurement of the head (Eq. 3)

$$K_H = P_H M_H^T (M_H P_H M_H^T + R_{Z_H})^{-1} \tag{3}$$

where R_{z_H} is the precision of head sensors and v_{z_H} is the noise in head sensors.

(c) Assimilate the measurements of Head (Z_H) and update the Head values ($H_{q_j z_H}$), using Eq. 4:

$$H_{|q_j z_H} = H_{|q_j} + K_H (z_H - M_H H_{|q_j} - v z_H) \tag{4}$$

(d) Estimate Flow ($Q_{q_j z_H}$) using hydraulic head losses

(e) Estimate Demand ($q_{q_j z_H}$) using the Pipe-Node Incidence Matrix (A_{21}) as defined by Todini and Pilati (1988), Eq. 5.

$$q_{|q_j z_H} = A_{21} Q_{|q_j z_H} \tag{5}$$

2.1.3 Step Two - Assimilation of Flow

Update the ensemble of state estimates with measurements of flow by assimilating the measurements, estimating the head and demand, and calculating the Kalman gain.

1.(a) Calculate the ensemble mean μ_Q and ensemble prior variance of the Flow P_Q. Where;

$$P_Q = \frac{1}{m-1} \sum_{j=1}^{m} \left[(Q_{|q_j z_H} - \mu_Q)(Q_{|q_j z_H} - \mu_Q)^T \right] \tag{6}$$

$$\mu_Q = \frac{1}{m} \sum_{j=1}^{m} Q_{|q_j z_H} \tag{7}$$

(b) Calculating the Kalman Gain K_F for flow using the error in the estimate and the errors in the measurement of flow (Precision of flow sensors; R_{zQ}, noise in flow sensors; v_{zQ})

$$K_Q = P_Q M_Q^T (M_Q P_Q M_Q^T + R_{Z_Q})^{-1} \tag{8}$$

(c) Assimilate the measurements of Flow (Z_Q) and update the Flow values ($Q_{q_j z_H z_Q}$)

$$Q_{|q_j z_H z_Q} = Q_{|q_j z_H} + K_Q (z_Q - M_Q Q_{|q_j z_H} - v z_Q) \tag{9}$$

(d) Estimation of Demand ($q_{q_j z_H z_Q}$) using Pipe-Node Incidence Matrix($A21$)

$$q_{|q_j z_H z_Q} = A_{21} Q_{|q_j z_H z_Q} \tag{10}$$

(e) Estimation of Head ($H_{q_j z_H z_Q}$) using hydraulic head losses and Pipe-Node Incidence Matrices (A_{11}, A_{12} and A_{21}) as detailed in Bragalli et al. (2016)

2.1.4 Step Three - Assimilation of Demand

Update the ensemble of state estimates with measurements of demands by assimilating the measurements, estimating the flow and head, and calculating the Kalman gain.

1.(a) Calculate the ensemble mean μ'_Q and ensemble prior variance of the $Q_{q_j z_H z_Q}$

$$P'_Q = \frac{1}{m-1} \sum_{j=1}^{m} \left[\left(Q_{|q_j z_H z_Q} - \mu'_Q \right) \left(Q_{|q_j z_H z_Q} - \mu'_Q \right)^T \right] \tag{11}$$

$$\mu_Q = \frac{1}{m} \sum_{j=1}^{m} Q_{|q_j z_H z_Q} \tag{12}$$

(b) Calculate the Kalman Gain K'_Q for flow prime using the error in the estimate and the errors in the measurement of demands (Precision of demand sensors; R_{z_q}, noise in demand sensors; v_{z_q}).

$$K'_Q = P'_Q A_{21} M_q^T \left(M_q A_{21} P'_Q M_q^T + R_{Z_q} \right)^{-1} \tag{13}$$

(c) Assimilate the measurements of demands(z_q) and update flow values ($Q_{q_j z_H z_Q z_q}$)

$$Q_{|q_j z_H z_Q z_q} = Q_{|q_j z_H z_Q} + K'_Q (z_q - M_q A_{21} Q_{|q_j z_H z_Q} - v z_q) \tag{14}$$

(d) Estimate Demand ($q_{q_j z_H z_Q z_q}$) using Pipe-Node Incidence Matrix(A_{21})

$$q_{|q_j z_H z_Q z_q} = A_{21} Q_{|q_j z_H z_Q z_q} \tag{15}$$

(e) Estimate Head ($H_{q_j z_H z_Q z_q}$) using hydraulic head losses and Pipe-Node Incidence Matrices (A_{11}, A_{12} and A_{21}) as detailed in Bragalli et al. (2016)

2.2 Evaluation Metric

The effectiveness of the DA can be estimated using the Total Variance (TV), Eq. 16, as suggested by Bragalli et al. (2016).

$$TV\{\overline{\otimes}\} = \frac{1}{S} \sum_{i=1}^{S} \left((\overline{\otimes}_i - \otimes_i^{true})^2 + \frac{1}{S} \sum_{i=1}^{m} \left[\frac{1}{m(m-1)} \sum_{j=1}^{m} \left(\otimes_i^j - \overline{\otimes}_i \right)^2 \right] \right) \tag{16}$$

where TV is the Total Variance, $\otimes$ is the state variable (either H, Q or q), $\overline{\otimes}$ is the ensemble mean, S is the number of state variables (i.e., number of nodes or pipes), m is the number of ensembles, i is the iterator for the state variable and j is the iterator for the ensembles.

However, for extended period simulation, we use the daily average TV value, obtained by dividing TV by the number of time steps used for the DA.

$$TVR\{\overline{\otimes}\} = \frac{TV\{\overline{\otimes}\}}{TV \otimes} \tag{17}$$

where $TVR\{\overline{\otimes}\}$ is the Total Variance Ratio of the system state, $TV\{\overline{\otimes}\}$ is the posterior system state assimilation (either 1 step, 2 steps assimilated), and $TV \otimes$ is the prior system state $\otimes$ without the assimilation of measurements from the current step.

To quantify the overall effectiveness of the implemented DA method, the TV values for each system state are normalized to obtain a Total Variance Ratio (TVR), which are averaged to obtain a Combined Total Variance Ratio ($CTVR$), which indicates the overall effectiveness of all system states (head, demand and flow) of all assimilation steps.

$$CTVR = \frac{1}{N} \left[\frac{1}{t} \sum_{i=1}^{N} TVR\overline{\otimes}_H + TVR\overline{\otimes}_Q + TVR\overline{\otimes}_q \right] \qquad (18)$$

where $CTVR$ is the Combined Total Variance Ratio, N is the number of system states assimilated and being combined, $TVR\overline{\otimes}_k$ is the Total Variance Ratio for System State, and k is either head (H), flow(Q) or demand (q).

2.3 Evaluating the effect of measurement uncertainty

Measurements are always affected by a degree of uncertainty. In the case of WDNs, measurement uncertainty depends on the sensors used for measuring the system's states. The precision and noise of these sensors are important in determining how well the sensors can capture the true states of the system.

Therefore, it is important to identify the limit of applicability of the proposed 3-step DA algorithm under uncertain observations To this end, we propose a number of experiments to test the effect of uncertainty due to sensor precision and uncertainty due to sensor noise, applied to the measurements of head, flow and demand. On the one hand, to investigate the effect of the uncertainty due to noise, six different levels of noise were applied to each state measurement. The selected noise values were varied using a normal distribution with a 5% standard deviation. In total 600 simulations were carried out for each sensor type. On the other hand, the effect of the uncertainty due to sensor precision was investigated applying six different precision values for each state of sensor, and for all the possible combinations of sensor precision values.

3 Case Studies

Two networks of different sizes which are representative of real-world WDNs are taken for this study.

The first case study is the Modena network which is the same WDN used by Bragalli et al. (2016); Han et al. (2020); Bhave and Gupta (2006) and in many other similar studies. The network consists of 317 pipes, 268 nodes and 4 reservoirs with a fixed head between 72.0 m and 74.5 m. The network has a total length of 71. 8 km of pipes with diameters between 100 mm and 400 mm. Although the network of Modena is small, the topology and distribution of the network make it suitable for the proposed research as the network is comparable to real world small WDNs as seen in Fig. 2.

The second case study is the Five Reservoir network ($FiveRes$), which is much larger than Modena. The network consists of 1278 pipes, 935 nodes and 5 reservoirs Zheng and Zecchin (2014). The layout of the network is given in Fig. 3. The network has a total length of 253.7 km of pipes with a diameter of 600 mm. The $FiveRes$ network provides a suitable comparison of how the DA algorithm can handle larger and more complex WDN models.

The monitoring network in Modena is more distributed compared to the $FiveRes$ Network as seen in Fig. 4. The number of sensors is also much less in $FiveRes$ compared to the size
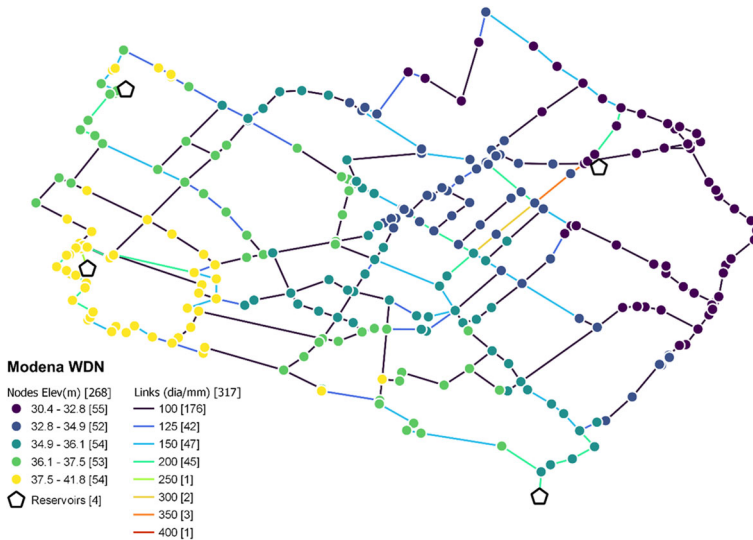
Fig. 2 Layout of the Modena network

of the network as seen in Table 1. Hence, it may not provide a good representation of the hydraulic states within the WDN for $FiveRes$. As such the experiments for measurement uncertainty were repeated for the $FiveRes$ network with sensors located at all the nodes and links.

4 Results and Discussion

The methodology in Fig. 1 was applied to the networks of Modena and FiveRes, modifying precision and noise of the measurements and evaluating their effect on the models' error.

4.1 Uncertainty Due to Noise

In the case of Modena, as seen in Fig. 5 the DA method is most sensitive to noise in the flow measurements, and any noise beyond one litre per second results in the DA algorithm being ineffective, as $CTVR$ exceeds one. The threshold of noise for the head is between 0.1 and 0.2 meters of noise. Noise in the measurement of demand, on the other hand, is resilient to an increase in noise up to 0.5 litres per second. Therefore, in the case of Modena, both flow and head sensors must be calibrated regularly to ensure that their accuracy remains within the effective threshold for DA. However, demand sensors require less maintenance and calibration as they can be effective to a higher threshold of noise compared to the other state measurement sensors.

The results for the $FiveRes$ network show that increasing noise in the measurement of the systems states results in an increase in CTVR. However, analyzing the results for $FiveRes$ in Fig. 6, we observed both head and demand exhibit significant variation in results when a suboptimal monitoring network in the $FiveRes$ WDN is used compared to the fully monitored network as shown in Table 1.
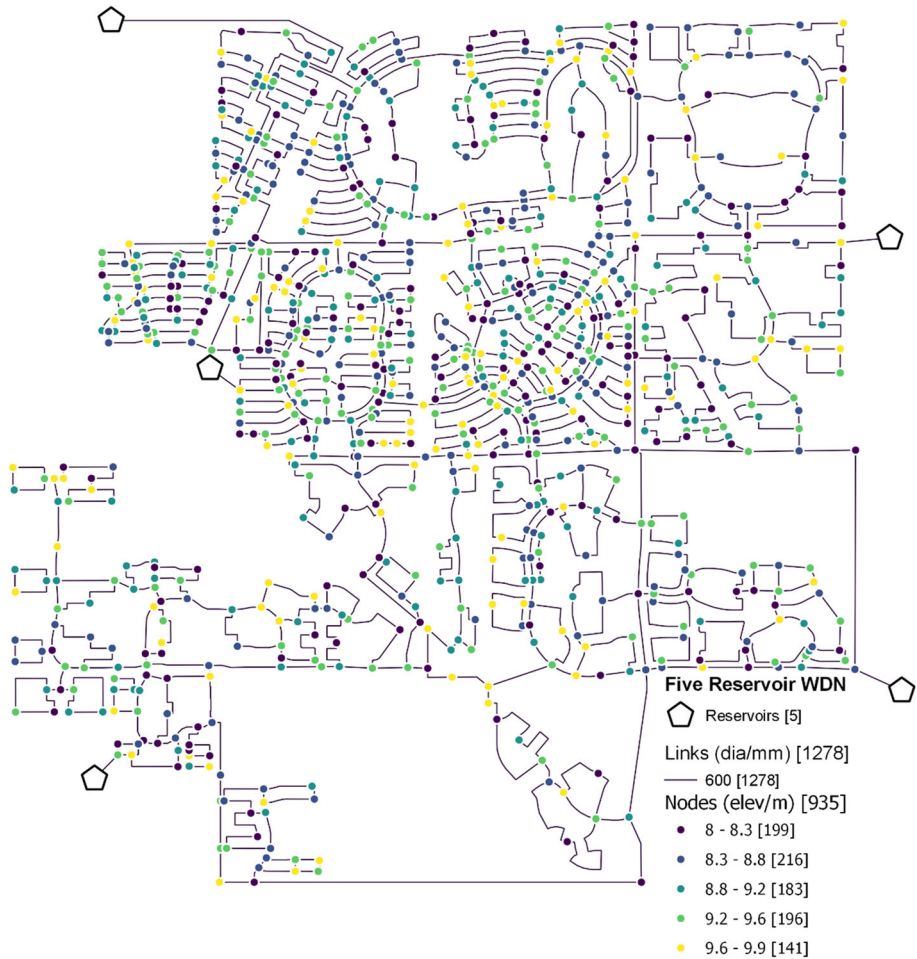
Fig. 3 Layout of the *Five Res* network

1. Head Sensors:

   (a) The DA remains effective for the entire range of heads tested based on the minimum CTVR

   (b) CTVR exceeds one even at the lowest noise levels, as indicated by the maximum CTVR

   (c) When the network is fully monitored, noise is not acceptable in head sensors for successful DA.

2. Demand Sensors:

   (a) Demand sensors become completely ineffective beyond a noise level of 0.8 litres per second.

   (b) Demand sensors show the maximum CTVR exceeding the threshold of one even at very low noise levels.

Five Reservoirs, Monitoring Network of Head Sensors   Five Reservoirs, Monitoring Network of Head Sensors
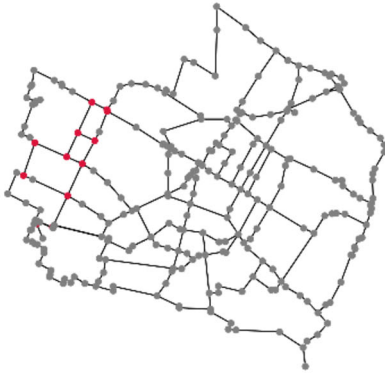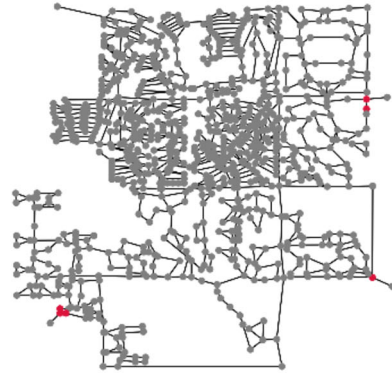
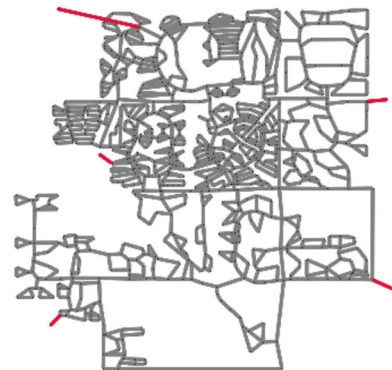Five Reservoirs, Monitoring Network of Demand Sensors   Five Reservoirs, Monitoring Network of Demand Sensors
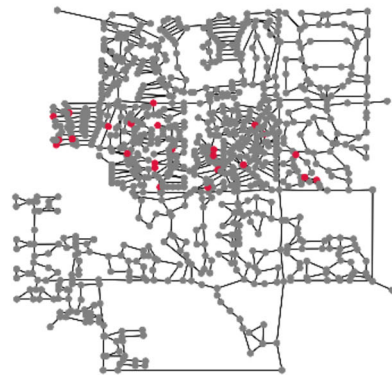
Five Reservoirs, Monitoring Network of Flow Sensors   Five Reservoirs, Monitoring Network of Flow Sensors

Fig. 4 Monitoring Networks for Head, Demand and Flow Sensors in Modena (Left) and *FiveRes* WDN (Right)

(c) When the network is fully monitored, the DA becomes less dependent on demand sensors and the noise in demand sensors does not have a significant impact on the effectiveness of the DA

Table 1 Configuration of sensors in the tested case studies

| WDN	Head Sensors (sH))	Demand Sensors (sq)	Flow Sensors (sQ)
Modena	12	11	4
FiveRes	6	25	5
FiveRes (Fully monitored)	935	935	1278

3. Flow Sensors:

   (a) The DA remains effective until approximately 2 lps based on the minimum CTVR.

   (b) The maximum CTVR shows that the DA is not effective at all ranges similar to head and demand.

   (c) When the network is fully monitored, noise is not acceptable in flow sensors for successful DA. The results consistently exceed the CTVR threshold of one for all noise levels beyond zero.

In general, it is observed that an increase in the noise in state measurements results in reduced effectiveness of the 3-EnKF-WDN, as the $CTVR$ increases when noise increases. Recall that $CTVR > 1$ indicates that the prior state yields less error compared to the assimilated states, and therefore 3-EnKF-WDN is ineffective.

4.2 Uncertainty Due to Sensor Precision

Figure 7 shows how the average TV of flow varies with varying precision of sensors for Modena. It can be seen from the sub-plots that, in general, an increase in the precision value of flow sensors results in an increase in the average TV of flow. Figure 8 shows a close-up of one of the sub-plots from Fig. 7, when the precision of demand sensors (R_{z_q}) is fixed at 0.1 litres/second and the precision of head sensors (R_{z_H}) is fixed at 0.1 meters. It shows that the average total variance of flow states increases with the increased precision value of flow sensors.

These findings suggest that an additional step of DA has a greater impact on decreasing the average TV of flow (i.e., on reducing the model error), than increasing the precision of flow sensors. In practical cases, this implies that operators may opt for less precise sensors, but a variety of sensors, to perform a multi-step DA and achieve better results. As seen in Fig. 8, carrying out a multi-step DA seems to be much more effective in reducing model error than assimilating just one system state. Also, the precision of the sensor does not improve the
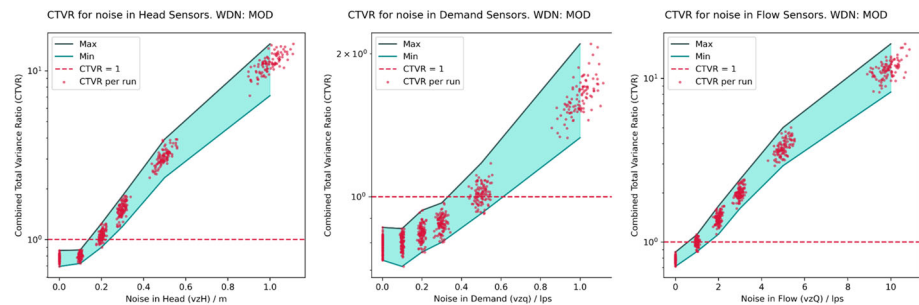
Fig. 5 CTVR against Noise in state measurements. Head (Left), Demand (Middle), Flow (Right), - for Modena Network
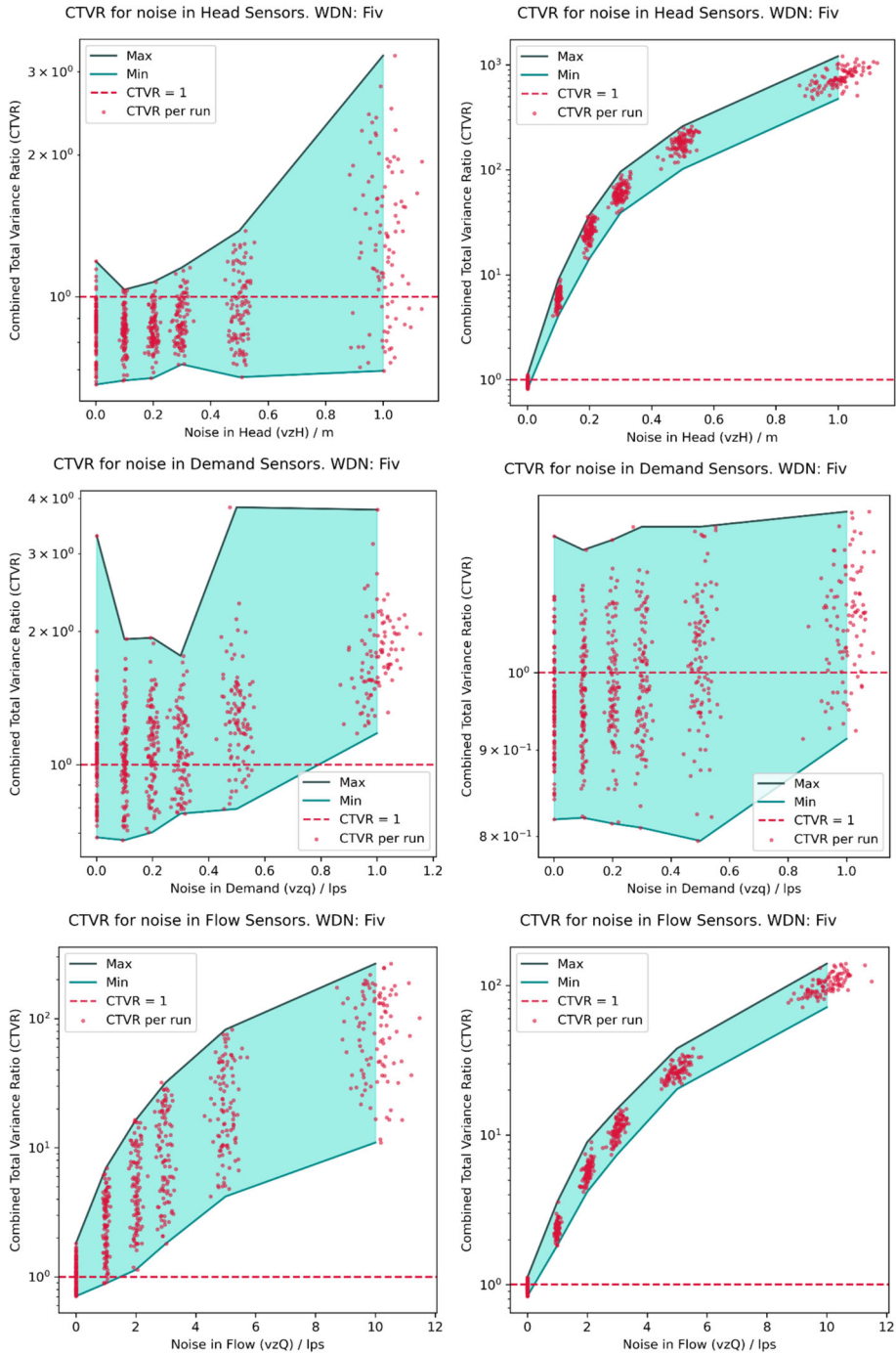
Fig. 6 CTVR against Noise in state measurements. Head (Top), Demand (Middle), Flow (Bottom), - for *FiveRes* (left) and Fully Monitored *FiveRes* (Right)

Fig. 7 Average Total Variance of Flow against precision of Sensors for Modena Network

results more considerably than the improvement obtained by an additional DA step. Similar results were obtained for both *FiveRes* and Modena WDNs with slight variations which may be due to factors such as the network topology, and hydraulic state of the WDN, among others.
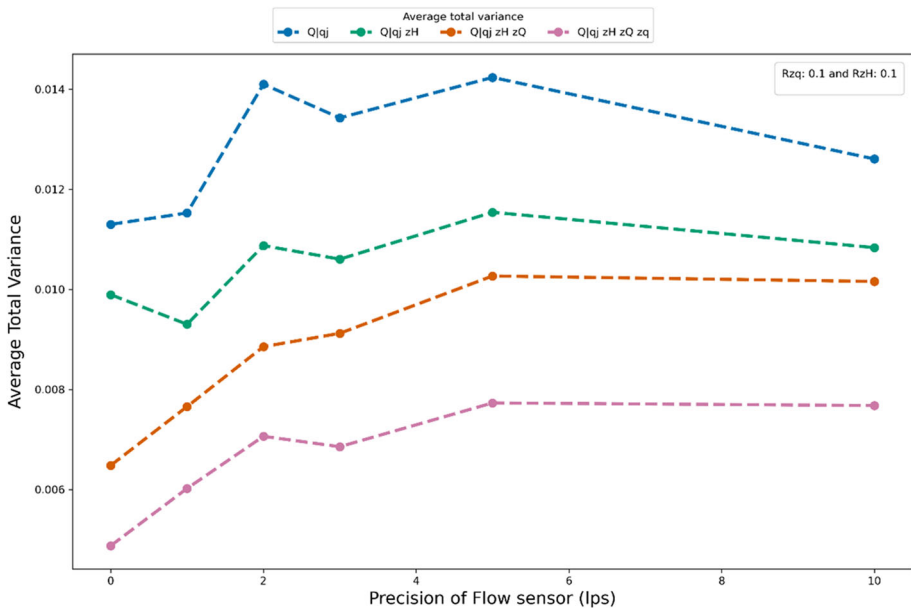
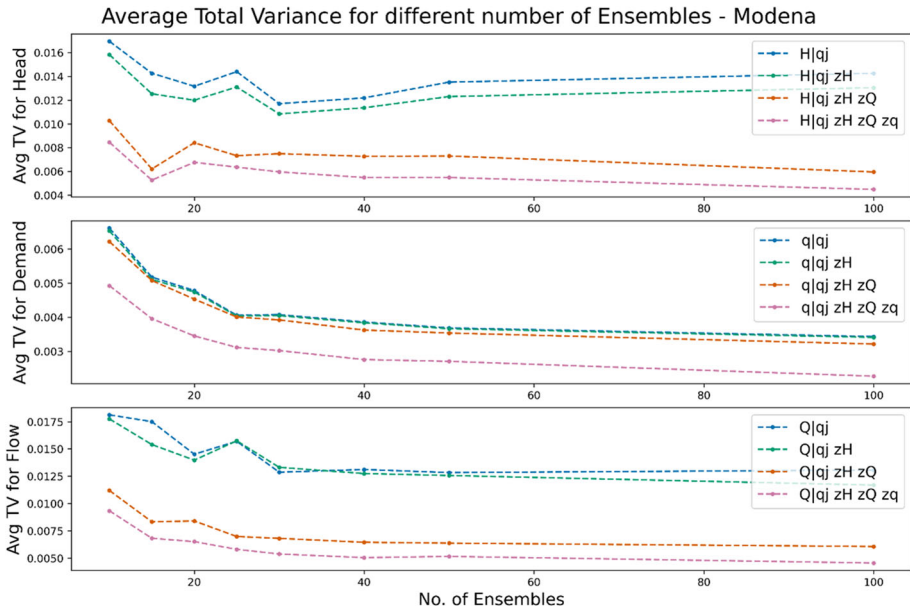**Fig. 8** Average TV of Flow when Rzq: 0.1, RzH: 0.1 against precision of Flow Sensors for Modena WDN

Fig. 9 Average TV for different numbers of ensembles for *Modena* WDN

4.3 Discussion on Number of Ensembles and Computational Demand

As each member of the ensembles in the EnKF is an independent realization of the model, we discuss how the number of ensembles affect the performance of the proposed 3-step EnKF. Generally, the higher the number of ensembles, the better the estimation of uncertainty, and computational demand is greater (Mulder 2014). Hence the implemented 3-step EnKF was simulated with ensembles between 5 members up to 100 members.

Figure 9 show that the increase in the number of ensembles improves the results of the DA as the average *TV* decreases in all cases with the increase in the number of ensembles used by the EnKF. It can also be seen that the consecutive steps of assimilation result in a reduction
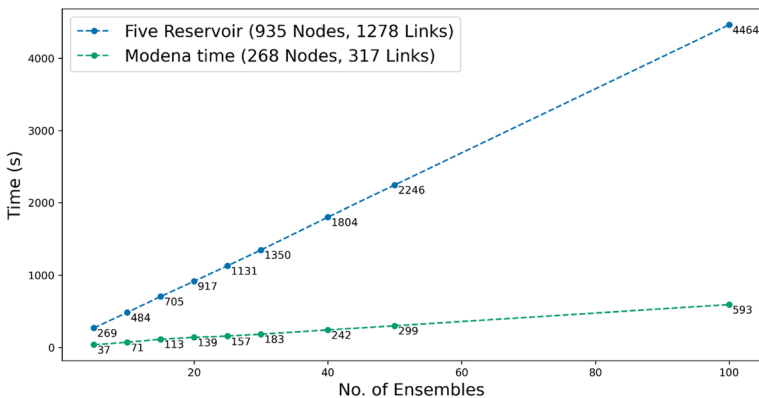

Fig. 10 Simulation time against the number of ensembles for Modena and *FiveRes* Network

**Table 2** Computer systems used for testing the proposed DA algorithm

Computer System	System 1 (Laptop)	System 2 (Desktop)	System 3 (HPC - Snellius)
Processor	Intel i5-8250	Ryzen 9-5900	AMD EPYC 7H12 (2x)
(Clock Speed)	- 4 Cores	- 12 Cores	- 128 Cores
	(1.60 - 3.40 GHz)	(3.0 - 4.7 GHz)	(2.6 - 3.3 GHz)
Memory (RAM)	1 x 8GB (2400 MHz)	2 x 32GB (3466 MHz)	16 x 16GB (3200 MHz)
Storage Type	SSD - NVMe	SSD - NVMe	SSD - NVMe
Operating System	Windows	Windows	Linux

in the model error as well. The asymptotic behaviour also indicates that few ensembles yield high average TV, but that it reduced rapidly as more ensembles are added. However, the rate of reduction of TV starts to be marginal after 30 to 50 ensembles, indicating that more ensembles are not necessary. This behaviour was seen for both Modena and $FiveRes$ networks.

The simulation time was compared for the different number of ensembles using various configurations of computer systems. It is observed that an increase in the number of ensembles from 5 to 100 results in an increase in simulation time from 37 seconds to 593 seconds for Modena and 269 seconds to 4464 seconds for $FiveRes$. With the increase in the size of the network from Modena (268 nodes, 317 links) to $FiveRes$ (935 Nodes, 1278 Links) the increase in simulation time is exponential. This can be seen by the increase in the gradient of the graph in Fig. 10. Although the increase in the size of the network is ≈ 3.5 times, the increase in simulation time is by ≈ 7.5 times.

In addition, 3-EnKF-WDN was tested on three different computer systems with different computational resources. The specifications of these systems are given in Table 2.

From the three different computer systems tested, the processors and their respective clock speeds show the most significant effect on the computational time. The current implementation of the algorithm runs serially without any parallel components, as such the computation time depends on the single-core clock speeds of the processors. Hence the results seen from Fig. 11 are representative of the base and boosted clock speeds of the processors used in the systems in Table 2 If the ensembles are generated in parallel, it will bring about a significant
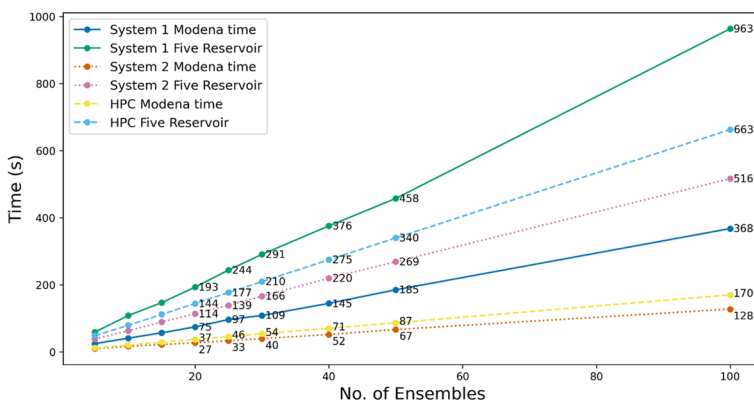
Fig. 11 Simulation time against the number of ensembles for Modena and $FiveRes$ WDNs using different computer systems

improvement in the computation time of the 3-EnKF-WDN. This will allow for the use of the 3-EnKF-WDN for larger WDNs, for running it for more time-steps without a significant computational burden.

5 Conclusions and Recommendations

In this paper, 3-EnKF-WDN, a 3-step DA method that assimilates pressure, flow and demand data, running a hydraulic model in extended-period simulation and under PDA was presented, along with a new evaluation metric called Combined Total Variance Ratio. The method was applied to two networks to evaluate its effectiveness in reducing the error in the hydraulic model under uncertain measurements.

The study demonstrated the importance of considering the effect of measurement uncertainty when using the 3-step DA algorithm. Two sources of uncertainty in the measurements were explored, namely precision and noise. It was found that the precision of sensors and the noise in measurements affect the efficacy of the 3-step DA.

When noise is added to the measurements, 3-EnKF-WDN becomes generally ineffective, within a small range of variation. The effect of the noise is significant in extensively monitored WDN. The findings also confirm the importance of maintaining the sensors with noise as small as possible. This could be achieved by carrying out regular maintenance and calibration of sensors. In practical applications, it is recommended to carry out simulations like the experiments with noise-in-state measurements used in this study to determine the respective thresholds of noise up to which the 3-step DA is still effective for the respective WDN.

It was also found that having high-precision sensors measuring one variable brings less reduction in model error than having less precise sensors measuring more variables.

The study also demonstrated that 30 to 50 ensembles are enough for the 3-EnKF-WDN to perform well, on the two studied networks, and that increasing ensembles beyond this number only introduces unnecessary computational burden.

It was also found that sensor data of demand do not improve the model error when applying 3-EnKF-WDN when the WDN is fully monitored (i.e., with head sensors in all the nodes and flow sensors in all the links). This is similar to the results obtained by Bragalli et al. (2016) where the $TVR(q)$ of demand was found to be the least sensitive to reduction in the $TVRs$ for the multi-objective optimization carried out in their study.

The proposed method has the potential to be applied to diverse WDN problems such as leak detection, anomaly detection, demand estimation, and water quality evaluation. This can be achieved by adapting the multi-step DA algorithm for the required purpose.

Some limitations of the study include the heavy computational time required. Parallelization of the algorithm using a method that can run hydraulic simulations in parallel is a solution to be explored in future research. In addition, the effect of the order and synchronicity of the assimilated data needs to be established. Other explorations to be made include the effect of the standard deviation or variation of the ensembles of demands and the effect of measurement uncertainty on the Kalman Gain.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

Adedeji KB, Hamam Y, Abe BT et al (2018) Pressure management strategies for water loss reduction in large-scale water piping networks: A review. In: Gourbesville P, Cunge J, Caignaert G (eds) Advances in hydroinformatics. Springer Singapore, pp 465–480. https://doi.org/10.1007/978-981-10-7218-5_33, http://link.springer.com/10.1007/978-981-10-7218-5_33, series Title: Springer Water

Antonowicz A, Brodziak R, Bylka J et al (2018) Use of EPANET solver to manage water distribution in smart city. E3S Web of Conferences 30:01016. https://doi.org/10.1051/e3sconf/20183001016, https://www.e3s-conferences.org/10.1051/e3sconf/20183001016

Bhave PR, Gupta R (2006) Analysis of water distribution networks. Alpha Science

Bragalli C, Fortini M, Todini E (2016) Enhancing knowledge in water distribution networks via data assimilation. Water Resour Manag 30(11):3689–3706. https://doi.org/10.1007/s11269-016-1372-0

Do N, Simpson A, Deuerlein J et al (2017) Demand estimation in water distribution systems: Solving under-determined problems using genetic algorithms. Procedia Eng 186:193–201. https://doi.org/10.1016/j.proeng.2017.03.227

Do N, Simpson AR, Deuerlein JW et al (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. https://doi.org/10.1061/(asce)wr.1943-5452.0000841

Gillijns S, Mendoza O, Chandrasekar J et al (2006) What is the ensemble kalman filter and how well does it work? In: 2006 American control conference. IEEE. https://doi.org/10.1109/acc.2006.1657419

Grievson O, Holloway T, Johnson B (2022) A Strategic Digital Transformation for the Water Industry. IWA Publishing. https://doi.org/10.2166/9781789063400 https://iwaponline.com/ebooks/book/860/A-Strategic-Digital-Transformation-for-the-Water

Han Z, Ma D, Hou B et al (2020) Seismic resilience enhancement of urban water distribution system using restoration priority of pipeline damages. Sustainability 12(3):914. https://doi.org/10.3390/su12030914 https://www.mdpi.com/2071-1050/12/3/914

Hill D, Kerkez B, Rasekh A et al (2014) Sensing and cyberinfrastructure for smarter water management: The promise and challenge of ubiquity. J Water Resour Plan Manag 140(7):01814002. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000449 https://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0000449

Klise K, Hart D, Moriarty D et al (2017a) Water network tool for resilience (wntr) user manual. Tech Report EPA/600/R-17/264, U.S. Environmental Protection Agency, Cincinnati, Ohio

Klise KA, Bynum M, Moriarty D et al (2017) A software framework for assessing the resilience of drinking water systems to disasters with an example earthquake case study. Environ Model Softw 95:420–431. https://doi.org/10.1016/j.envsoft.2017.06.022

Mulder D (2014) Applying data-assimilation and calibration in the field of urban drainage. PhD thesis, TU Delft. https://repository.tudelft.nl/islandora/object/uuid:05e6c69d-fb3e-4c25-8422-78849de8978d/datastream/OBJ/download

Okeya I, Kapelan Z, Hutton C et al (2014) Online burst detection in a water distribution system using the kalman filter and hydraulic modelling. Procedia Eng 89:418–427. https://doi.org/10.1016/j.proeng.2014.11.207

Rossman L (1993) EPANET Users Manual. Tech. Rep. No. EPA-600/R-94/057, Environmental Protection Agency, Risk Reduction Engineering Laboratory, Cincinnati, Ohio

Ruzza V (2017) Data assimilation techniques for leakage detection in water distribution systems. PhD thesis, Universit'a degli Studi di Padova. https://www.research.unipd.it/retrieve/e14fb26f-a9ce-3de1-e053-1705fe0ac030/Valentina_Ruzza_tesi.pdf

Shang F, Uber J, van Bloemen-Waanders B et al (2006) Real time water demand estimation in wds. Proc 8th Annual water distribution systems analysis symposium, Cincinnati, Ohio, USA, August 27–30, http://www.cs.sandia.gov/~bartv/papers/realtime_water.pdf

Shang F, Uber JG, van Bloemen Waanders BG et al (2008) Real time water demand estimation in water distribution system. In: Water distribution systems analysis symposium 2006. American Society of Civil Engineers, pp 1–14. https://doi.org/10.1061/40941(247)95, http://ascelibrary.org/doi/abs/10.1061/40941%28247%2995

Simon D (2006) Optimal state estimation: Kalman, H [infinity] and nonlinear approaches. Wiley-Interscience. https://isharifi.ir/teaching/2019/IoT/[Dan_Simon]_Optimal_State_Estimation_Kalman,_H_In(BookFi).pdf, OCLC: ocm64084871

Todini E (1999) Using phase-state modelling for inferring forecasting uncertainty in nonlinear stochastic decision schemes. J Hydroinformatics 1(2):75–82. https://doi.org/10.2166/hydro.1999.0007

Todini E, Pilati S (1988) A gradient algorithm for the analysis of pipe networks., Research Studies Press Ltd., Letchworth, Hertfordshire, UK, chap Computer Applications in Water Supply: vol. 1- Systems Analysis and Simulation, pp 1-20

Van Den Bossche W (2013) Data assimilation toolbox for MATLAB. Master's thesis, KU Leuven

Zheng F, Zecchin A (2014) An efficient decomposition and dual-stage multi-objective optimization method for water distribution systems with multiple supply sources. Environ Model Softw 55:143–155. https://doi.org/10.1016/j.envsoft.2014.01.028

Zhou X, Xu W, Xin K et al (2018) Self-adaptive calibration of real-time demand and roughness of water distribution systems. Water Resour Res 54(8):5536–5550. https://doi.org/10.1029/2017WR022147 https://onlinelibrary.wiley.com/doi/abs/10.1029/2017WR022147

Zhou X, Guo S, Xin K et al (2022) Maintaining the long-term accuracy of water distribution models with data assimilation methods: A comparative study. Water Res 226:119268. https://doi.org/10.1016/j.watres.2022.119268 https://linkinghub.elsevier.com/retrieve/pii/S0043135422012131

Authors and Affiliations

Ibrahim Miflal Fayaz [1] · **Mario Castro-Gama[1,2,3]** · **Leonardo Alfonso[1]**

Mario Castro-Gama
mario.castrogama@vitens.nl

Leonardo Alfonso
l.alfonso@un-ihe.org

[1] Hydroinformatics and Socio-Technical Innovation, IHE Delft, Westvest 7, Delft 2611 AX, South Holland, The Netherlands

[2] Water Expertise Center, VITENS N.V., Oude Veerweg 1, Zwolle 8019 BE, Overijssel, The Netherlands

[3] CiTG, TU Delft, Stevinweg 1, Delft 2628 CN, South Holland, The Netherlands