Particle Filter-Based Model for Online Estimation of Demand Multipliers in Water Distribution Systems under Uncertainty

N.C. Do, A.R. Simpson, J.W. Deuerlein, O. Piller

A particle filter - based model for online estimation of demand multipliers in water distribution systems under uncertainty

Nhu C. Do[1*], Angus R. Simpson[2], Jochen W. Deuerlein[3], Olivier Piller[4]

[1] - *School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide SA 5005, Australia. Email: nhu.do@adelaide.edu.au. * - Corresponding author*

[2] - *Professor, School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide SA 5005, Australia. Email: angus.simpson@adelaide.edu.au*

[3] - *Senior Researcher, 3S Consult GmbH, Karlsruhe, Germany. Email: deuerlein@3sconsult.de*

[4] - *Senior Research Scientist, Irstea, UR ETBX, Department of Waters, France, Bordeaux. Email: olivier.piller@irstea.fr*

Abstract

Accurate modeling of water distribution systems is fundamental for the planning and operating decisions in any water network. One important component that directly affects model accuracy is the knowledge of nodal demands. Conventional models simulate flows and pressures of a water distribution network either assuming constant demands at nodes or using a short-term sample of demand data. Due to the stochastic behavior of the water demands, this assumption usually leads to an inadequate understanding of the full range of operational states in the water system. Installation of sensor devices in a network can provide information about some components in the system. However, the requirement for a reliable water distribution model that can assist with understanding of real-time events in the entire water distribution system is still an objective for hydraulic engineers.

This paper proposes a methodology for the estimation of online (near real-time) demand multipliers. A predictor-corrector approach is developed which predicts the hydraulic behaviors of the water network based on a nonlinear demand prediction model, and corrects the prediction by integrating online observation data. The standard particle filter and an improved particle filter method, which incorporates the evolutionary scheme from genetic algorithms into the resampling process to prevent particle degeneracy, impoverishment and convergence problems, are investigated to implement the predictor-corrector approach. Uncertainties of model outputs are also quantified and evaluated in terms of confidence intervals. Two case studies are presented to demonstrate the effectiveness of the proposed particle filter model. Results show that the model can provide a reliable estimate of demand multipliers in near real-time contexts.

Keywords: Particle filters, sequential Monte Carlo method, real-time demand estimation, water distribution systems, uncertainty.

Introduction

Water distribution systems (WDS) are constructed to supply water for domestic, industrial and commercial consumers. The design, operation and management of these distribution systems is usually supported by the application of hydraulic models, which are built to replicate the behavior of real systems. These conventional models simulate flows and pressures of a WDS either under steady state conditions (constant demands and operational conditions) or under a short term extended period simulation (time-varying demands and operational conditions), for example a day or a week (USEPA 2005). The outputs from hydraulic models, therefore, usually represent the distribution system behavior during the sampling period (Preis et al. 2009). This leads to an inadequate understanding of the full range of operational states in the water system.

The installation of sensor devices as well as the Supervisory Control and Data Acquisition (SCADA) systems within the WDS can provide information on the status of some components in the system. However, the use of this additional data is currently limited to computing gross differences between the model outputs and reality (Kang & Lansey 2009). Modification of the hydraulic models to maintain the consistency between observed data and simulated data is still a challenge that needs to be dealt with. Estimation of the model states/parameters, hence, is required so that the model is able to represent the real system.

Estimation is the process of fitting the outputs from the computer model, usually the pressures and flow rates at particular locations in the water network, with the field measurements, in order to calculate unknown variables of interest. Initial estimation studies in WDSs were pioneered by Rahal et al. (1980), Walski (1983) and Bhave (1988) with the proposal of the ad hoc (trial-and-error) calibration schemes, in which an iterative process to update unknown model parameters was implemented. Due to the slow convergence rate, this method is only applicable to small water networks. Later, explicit calibration methods were introduced (Ormsbee & Wood 1986; Boulos & Wood 1990; Boulos & Ormsbee 1991). These methods solved an even-determined set of water network equations where the number of unknown parameters is grouped to be equal to the number of measurements. As the measurement errors were also neglected, these methods usually did not represent real-world practical outputs. Therefore, explicit calibration models were often used to analyse historic events in water systems (Savic et al. 2009). Subsequently, implicit methods were developed using either mathematical techniques or evolutionary optimization techniques, for example: Complex Method (Ormsbee 1989), Weighted Least Squares approaches (Lansey & Basnet 1991; Datta & Sridharan 1994), Singular value decomposition (SVD) method (Sanz & Pérez 2015) or Genetic Algorithms (GA) (Preis et al. 2009; Abe & Peter 2010; Do et al. 2016). These methods have drawn a high degree of attention from researchers. However, these models are mostly impractical due to either a requirement for a large quantity of ‘good’ observation data (Savic et al. 2009) or ignoring model uncertainties. Furthermore, few approaches have attempted to estimate model parameters and model states in conjunction with model uncertainties. Bargiela and Hainsworth (1989) found that a good approximation of pressure uncertainty bounds can be obtained by a linearization of the mathematical network model. Piller (1995) and Bush and Uber (1998) used a sampling design method to estimate the model parameters and approximate the uncertainties. Lansey et al. (2001) applied a first-order approximation method to identify pipe roughness uncertainty. Nagar and Powell (2002) applied a linear fractional transformation and semi-definite programming method to estimate the pressure heads and their confidence bounds. In addition, some probabilistic methods (Xu & Goulter 1998; Kapelan et al. 2007; Hutton et al. 2013) have also been investigated for the estimation of model parameters. Due to the complexity of the uncertainties, estimation methods associated with uncertainty quantification are still a continuing research area, especially for real-time estimation purposes.

The complexity of uncertainties in WDS modeling has been addressed in Hutton et al. (2012b), in which the uncertainty is divided into three categories: (1) structural uncertainty, (2) parameter uncertainty and (3) measurement/data uncertainty. Structural uncertainty derives from the mathematical representation of the real system, such as network skeletonization and model aggregation. Skeletonized and/or aggregated models are predominantly used instead of all-pipes models to reduce the complexity of the network being analysed as well as to increase computational speed. It has been shown that skeletonized/aggregated network models can closely resemble the behaviour of full sized systems under steady state conditions (e.g. Perelman et al. (2008) and Preis et al. (2011)). The second category, parameter uncertainty, refers to the errors of the parameters used to represent system components (e.g. pipe roughnesses, pipe diameters). According to Kang and Lansey (2009), these parameters are time invariant or vary slowly over time. Hence, this source of uncertainty can be neglected for real-time estimation problems. Finally, measurement/data uncertainty is the uncertainty from measurement devices and, more importantly, uncertainty from the inability to capture the temporal and spatial variation of consumer demands. Because of their high impact on model uncertainty during short periods of time (or in real-time), nodal demands are therefore usually selected as the time varying parameters to be estimated.

The issue of short term demand forecasting and real-time demand estimation under uncertainties can be found in some recent studies. Note that the short-term demand forecasting and demand estimation are two different problems. The former focuses on predicting future demands (e.g. Cutore et al. (2008), Hutton and Kapelan (2015) and Alvisi and Franchini (2017)). The latter focuses on estimation of the current demands, which is also the main interest of this paper. This is useful, as demand estimation can be used at regular time steps to verify the accuracy of the predicted value and update the system operations. The problem of near real-time demand estimation has been studied using different approaches. Shang et al. (2006) applied an extended Kalman filter, an iterative linear algorithm for nonlinear state estimation, to approximate water demand patterns. In that paper, water demand patterns were predicted by an ARIMA time series model and were refined using real-time observations. Similarly, Hutton et al. (2012a) introduced a particle filter method and an ensemble Kalman filter for the estimation of a single district meter area, which was assumed to follow a linear time series model. The particle filter model was implemented with and without measurement error to show its effect on the demand prediction uncertainty. An alternative for the demand estimates can be found in Kang and Lansey (2009). In their paper, two comprehensive methods for the demand estimation problem were introduced, the Kalman filter and the tracking state estimator (TSE). For the Kalman filter model, the water demand patterns were also assumed to follow a linear time series model, while the TSE model involved recursively computing the sensitivity matrix (i.e. the Jacobian matrix of the measurement vector with regards to the change in the state vector). The uncertainties of the demand estimates were suggested to be quantified by applying the first-order second moment formula. The two models were then tested on a case study (116 pipes, 90 nodes, 1 source and 1 tank) with an assumption that 19 flow measurement sites and 5 pressure measurement sites were available. It should be noted that the demand estimation problem is sensitive to the locations and types of the measurements (Do et al. 2016). Demand estimation models usually perform better with flow measurements rather than pressure/head measurements. However, due to the cost and difficulty of installing flow measurement devices compared to pressure measurement devices, flow measurement devices are usually not as commonly used as pressure measurement devices in real WDS networks.

In summary, water demands in WDS studies are usually assumed to be known and varied based on a diurnal curve. However, this assumption might lead to large approximations of WDS states in real-time due to the unpredictable variation of the water demands. Some efforts have been focused on the real-time demand estimation. By assuming that the water demand follows a linear time series prediction model, these models approximated the water demand patterns with some linear algorithms such as the Kalman filter or extended Kalman filter. Given the nonlinear stochastic nature of the water demands as well as the need for practical applicability, real-time estimation modeling of WDS still requires much research effort.

This paper presents a model framework for the online (near real-time) demand estimation of a WDS, which is named the *DMFLive* model. A predictor-corrector methodology is adopted in the *DMFLive* model to predict the hydraulic behaviors of the water network based on a nonlinear demand prediction sub-model, and to correct the prediction by using online pressure observation data. A particle filter method is applied to implement the predictor-corrector approach. The typical problems of the particle filter approach (particle degeneracy, impoverishment and particle convergence) are investigated by two different resampling schemes: systematic resampling (SR) algorithm and systematic resampling integrated with a genetic algorithm process (SRGA). Uncertainties of model outputs are quantified and evaluated in terms of confidence intervals.

The paper is structured as follows. First, an explanation of the state estimation problem and its conceptual solution is introduced. Second, the basic concepts of particle filter methods to solve the estimation problem are explained. This is followed by a detailed description of the particle filter methodology applied for water demand state estimation in WDS. Two case studies are then used to evaluate the model. Finally, conclusions and suggestions for future work are given.

State estimation problem and its conceptual solution

The problem of state estimation involves finding a target state vector x_k that evolves according to a discrete time stochastic model (Ristic et al. 2004):

$$x_k = f_{k-1}(x_{k-1}, \upsilon_{k-1}) \quad (1)$$

where k is the index of discrete time steps; f_{k-1} is a known, possibly nonlinear function of the previous state and υ is the process noise sequence. The value of x_k can be found from measurements z_k, which are related to x_k via the measurement equation:

$$z_k = h_k(x_k, w_k) \quad (2)$$

where h is a known implicit or explicit, possibly nonlinear function and w is the measurement noise sequence. The noise terms υ_k and w_k are usually assumed to be white noise and independent. From a statistical and probabilistic perspectives, the state model can be represented by a probability density function (pdf). The state estimation problem, therefore, becomes a process of recursively quantifying some degree of belief in the state x_k given the measurement series Z_k $(z_i,\ i=1,\dots,k)$ up to time k. This process can be obtained by two stages: prediction and correction/update. The prediction stage involves applying the system model to predict the prior pdf of the state:

$$p(x_k|Z_{k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|Z_{k-1})dx_{k-1} \quad (3)$$

where $p(x_k|x_{k-1})$ is the probabilistic model of the state model, or the transitional probability density function, which is defined by the system equation Eq. (1) with the known statistics of υ_{k-1}; and $p(x_{k-1}|Z_{k-1})$ is the pdf of the model at time *k-1*, which is supposed to be known.

The correction/update stage implements Bayes' rule to compute the posterior probability density of the state model when the measurement z_k becomes available:

$$p(x_k|Z_k) = \frac{p(z_k|x_k)p(x_k|Z_{k-1})}{\int p(z_k|x_k)p(x_k|Z_{k-1})dx_k} \quad (4)$$

where $p(z_k|x_k)$ is the likelihood function, defined by the measurement equation (Eq. (2)) with the known statistics of w_k.

According to Ristic et al. (2004), the recursive propagation of the posterior pdf shown in Eq. (3) and Eq. (4) is only a conceptual solution that cannot be analytically solved. The solution requires the storage of a fully non-Gaussian pdf, corresponding to an infinitive dimensional vector. Since the true solution is too complex and almost impossible to compute, an implementation of approximation techniques or suboptimal Bayesian algorithms is developed. The following section introduces an approximation technique, namely the particle filter, to solve the aforementioned state estimation problem.

Particle filters

Over the last decade, particle filters have been successfully applied to the state and parameter estimation of complex system models in various environmental engineering fields, such as hydrology (Moradkhani et al. (2005), Weerts and El Serafy (2006)), hydraulic (Hutton et al. 2012a) and geoscience (van Leeuwen (2010)). Unlike the Kalman filter (for linear problems), extended

Kalman filter (which requires a linearization of the nonlinear problems) or the unscented Kalman filter (which uses a small number of deterministically chosen samples), the particle filter can use a large number of Monte Carlo samples to estimate fully nonlinear, possibly non-Gaussian target states. The key concept of a particle filter is to approximate the posterior pdf of states, defined in Eq. (4), by an ensemble of samples (N_p), each of which contains an associated weight (w_k^i), and to compute estimates based on these samples and weights:

$$p(x_k|Z_k) \approx \sum_{i=1}^{N_P} w_k^i \, \delta(x_k - x_k^i) \tag{5}$$

$$w_k^i = w_{k-1}^i \frac{p(z_k|x_k^i)p(x_k^i|x_{k-1}^i)}{p(x_k^i|x_{k-1}^i, z_k)} \tag{6}$$

where δ is the Dirac delta function; i is the particle index; and $p(x_k^i|x_{k-1}^i, z_k)$ is the importance density function. In order to simplify the weight update of the particle, the importance density function is usually chosen as the transitional density function, $p\left(x_k^i \middle| x_{k-1}^i, z_k\right) = p\left(x_k^i \middle| x_{k-1}^i\right)$, which yields with scaling:

$$w_k^i = \frac{p(z_k|x_k^i)}{\sum_{i=1}^{N_P} p(z_k|x_k^i)} \tag{7}$$

These equations form the basis of most particle filters. However, it has been shown by Doucet et al. (2000) that the variance of the weights will increase over time if the particle filtering process is limited at executing only these equations. Since the particles drift away from the "truth" as well as obtain negligible weights (Moradkhani et al. 2005), the model will fail to estimate the real states of the system. To avoid this problem, a resampling process, which replaces samples with low importance weights by the samples with high importance weights, is added to the procedure of particle filter models. In this paper, the systematic resampling method, also called the stochastic universal resampling, introduced by Kitagawa (1996), is selected for the resampling procedure of the particle filter model. A comprehensive explanation of the systematic resampling and the full review of particle filtering methods are described in (van Leeuwen 2009). In addition, an improved resampling method which integrates the evolutionary scheme from genetic algorithms into the resampling process, is also proposed to improve the efficiency of the particle filter model.

Particle filters applied for water demand state estimation in WDS

In this study, the predictor-corrector approach implemented by a particle filter model for the estimation of water demands in real-time is proposed, namely the *DMFLive* model. The demand prediction sub-model presented by van Zyl et al. (2008) has been applied to predict the water demand multipliers (DMF) in a WDS. The hydraulic EPANET toolkit (Rossman 2000) which solves the hydraulic equations was used to compute the model equivalent of the measurement data (i.e. the nodal pressures, flow rates at measurement locations or the final tank levels at the end of each time step). These computed values then were integrated with the corresponding field measurements in order to correct/update the particle weights. Particles were, thereafter, resampled (with either SR or SRGA) and subsequently used as input for the prediction model. Simultaneously, the estimated demand multipliers were computed and selected for uncertainty quantification. The uncertainties of the demand multipliers caused by the errors from measurement devices were computed using the first-order approximation formula. The flowchart of the *DMFLive* model is shown in Figure 1.

Initialization of particles

The *DMFLive* model starts with a creation of an ensemble of the particles (N_p). The particles are the demand residuals, driven by the demand prediction model to predict the demand multipliers. In addition, each particle is assigned an initial weight equal to $1/N_p$.

Demand prediction sub-model

The initial particles (for the first iteration) or the particles after resampling (from the second iteration onwards) are transferred to the demand prediction sub-model. Demand residual information carried by the particles is used to track the states and predict the demand multipliers via the following equations (van Zyl et al. 2008):

$$\ln x_k^j = \sum_{i=1}^{m} \phi_i^j \ln x_{k-i}^j + \ln \upsilon_k^j \tag{8}$$

where x_k^j is the demand residual state at time step k of the j^{th} DMF; i is the lag counter; m is the number of autocorrelation lags (for the state estimation problem $m=1$ as referred to Eq. (1)); ϕ_i is the auto-regression coefficient for lag i and $\upsilon_k(0,\sigma_h)$ is the white noise with mean zero and standard deviation σ_h.

The j^{th} DMF is calculated as:

$$DMF_k^j = C_k^j x_k^j \tag{9}$$

where C_k^j is the value at time k of a typical diurnal demand pattern of the j^{th} DMF. The C value can be identified based on meter information of different water users (e.g. in Beal and Stewart (2014)).

Real-time hydraulic data

In practice, hydraulic data can be captured in real-time via the SCADA system or sensor devices. For the *DMFLive* model, two types of real-time hydraulic data are required. First are the tank levels, pump and valve statuses, and second are the nodal heads and pipe flow rates at measurement locations. Tank levels, pump and valve statuses are used as boundary conditions for the hydraulic simulation of the water network model while the observations at measurement locations are used to correct/update the weight of the particles.

In order to validate the performance of the proposed model as well as its practical applicability to real WDS networks, all case studies in this research are assumed to have pressure measurements only. The input data sets to evaluate the *DMFLive* model are synthetically generated based on deterministic models, where the water network parameters are fully known, as follows: (1) known demand patterns are assigned to nodal demands; (2) EPANET is run to record tank levels, pump statuses, and pressures at selected measurement locations; (3) to introduce the measurement errors, a normal distributed random error in an allowable range ($\pm\Delta^{meas}$) is added to each nodal pressure.

Simulator

The hydraulic behavior of the water distribution network at each time step is simulated using an EPANET steady state simulation. The inputs are the predicted DMFs, tank levels, and pump and valve statuses. The water network characteristics such as pipe lengths, diameters, roughness coefficients, node elevations, pump curves, etc. are assumed to be known and constant. The outputs from the EPANET hydraulic solver is the model equivalent of the observations, i.e. the simulated nodal heads and pipe flow rates at measurement locations.

Corrector

The weights of the particles are corrected/updated by associating the simulated heads and flows with the actual observations via Eq. (7) where the likelihood function is assumed to be Gaussian:

$$p\left(z_k \middle| x_k^i\right) = \frac{1}{\sqrt{2\pi|R|}} e^{\left(-\frac{1}{2}\left[z_k - h\left(x_k^i\right)\right]^T R^{-1}\left[z_k - h\left(x_k^i\right)\right]\right)} \qquad (10)$$

where $h(x_k^i)$ is the model equivalent of the observations z_k (simulated nodal heads and flow rates), and R is the covariance matrix of the observation errors, which in general is caused by errors from two main sources: forward model error and measurement device error. The forward model error,

$$\Delta^{true} = Z^{true} - h(x^{true}) \qquad (11)$$

is the difference between the true observation vector, Z^{true}, and the corresponding vector output from the hydraulic simulation model EPANET using the true state x^{true}. The true observation vector is a theoretical vector that represents observations measured by perfect measurement devices. It is linked to the actual measured values via the expression:

$$Z = Z^{true} + \Delta^{meas} \quad (12)$$

The observation error covariance matrix, therefore, can be estimated as $R = R^{true} + R^{meas}$, where R^{true} and R^{meas} denote the covariance of the forward model error and the covariance of measurement error, respectively (see Waller (2013) for a detailed explanation and calculation of the observation error covariance matrix). To produce good estimates of the model state in real case studies, the error covariance matrix must be well understood and properly calibrated. As previously mentioned in this paper, the measured data in all case studies were synthetically generated from the EPANET model based on "true" demand patterns. The forward model error, therefore, equals to zero. The covariance matrix R, as a result, is the diagonal matrix where the diagonal elements are the variances of the measurement errors, since observations are independently measured at different locations of the network by different measurement devices. The measurement errors with specified ranges are assumed to be known so that the covariance matrix R can be identified. ***Resampling***

Resampling is applied to create new ensembles of particles from the posterior pdf of the previous step. In this paper, two alternatives of resampling are tested: systematic resampling algorithm (SR) and systematic resampling integrated with the GA operators (SRGA).

The SR algorithm generates a random number u_s from the uniform density U[0, *1/Np*], and consequently creates N_p ordered numbers (Hol et al. 2006):

$$u^i = \frac{i-1}{N_P} + u_s \quad (i = 1, \dots, N_P) \quad (13)$$

New particles are then selected that satisfy Eq. (14):

$$x^i_{new} = x(F^{-1}(u^i)) \quad (14)$$

where F^{-1} denotes the generalized inverse of the cumulative probability distribution of the normalized particle weights.

To reduce the convergence problem of the particles (i.e. all the particle weights are equal to zero) when applying the model for large networks with multiple demand patterns, the SRGA method is also applied. Three GA operators of selection, crossover and mutation are responsible for modifying the predicted demands before computing the weight of a particle by Eq. (10). In the selection step, particles are compared to each other through tournament selection and the best particles are selected as parents. Parent particles are then paired and go through crossover and mutation to generate offspring solutions. While the details of GA can be found in Nicklow et al. (2010), it is important to know that new parameters need to be introduced: the probability of crossover P_c, the probability of mutation P_m and the number of generations N_{gen}.

Demand multiplier outputs and uncertainty quantification

The estimate of the state x_k is obtained by taking the mean of the particle filter sample set (Salmond & Gordon 2005):

$$\hat{x}_k \approx \frac{1}{N_p}\sum_{i=1}^{N_p} x_k^{i\,*} \tag{15}$$

where $x_k^{i\,*}$ is the state updated based on the posterior analysis of the model weights.

For particle filter models, the uncertainty of the model output can be computed by taking the variance of the samples:

$$var(x_k) \approx \frac{1}{N_p}\sum_{i=1}^{N_p}\left(x_k^{i\,*} - \hat{x}_k\right)\left(x_k^{i\,*} - \hat{x}_k\right)^T \tag{16}$$

For the demand multiplier estimation problem, it should be noted that a small change in the demand multiplier can cause a large change in nodal demands (for nodes with large base demands) and consequently result in large variations of nodal pressures, especially at nodes that are sensitive to nodal demands. Most of the demand forecasting models are required to capture both peak-demand hours and off-peak demand hours, with a demand multiplier factor that can vary from 0 to 4 (Chin et al 2000). The weight of the particles via Eq. (10) can, therefore, easily approach zero which leads to either particle degeneracy or particle non-convergence. Using a larger number of particles can prevent this problem, however, if the dimension of the state vector increases, the required number of particles increases exponentially. One way to solve these issues is to incorporate the covariance of the forecasting nodal heads/ pipe flow rates into the likelihood function:

$$p\left(z_k \middle| x_k^i\right) = \frac{1}{\sqrt{2\pi|R^*|}} e^{\left(-\frac{1}{2}\left[z_k - h\left(x_k^i\right)\right]^T (R*)^{-1}\left[z_k - h\left(x_k^i\right)\right]\right)} \quad (17)$$

where R* = R+Σ, Σ is the covariance matrix of the forecast nodal heads or pipe flow rates, computed based on the forecast demands. This covariance matrix can be estimated by running the demand forecasting model multiple times to obtain the range of forecast demand multipliers, then applying these values into the hydraulic model to compute the variance of simulated nodal heads and pipe flow rates at measurement locations.

Although the method can ensure some of the particles always contain weights to avoid particle non-convergence and degeneracy, this would increase the noise of the output model. The variance of the model output (i.e. the uncertainty of the model output) is required to be computed by a different method instead of using Eq. (16).

Another way to overcome the convergence and degeneracy issues is to integrate the GA operators into the resampling process as mentioned in the previous sections. The integrated GA approach can prevent the model from experiencing these problems by exploring the state-space region and selecting the best particles (including the replication of good solutions). However, it might lead to another problem for the particle filter, referred to as particle impoverishment. The distribution of the state model, because of the particle impoverishment, is poorly represented by only one or a few particles which significantly reduces the variance of the model state.

To ensure reliable outputs from the particle filter model, it is proposed to approximate the uncertainty of the model state by an independent method, such as the first-order approximation (FOA) method adopted from Piller (1995). This also has the advantage of significantly decreasing the computational time, as it will be shown in the case studies. The model outputs, therefore, are the estimate of the demand multipliers computed by Eq. (15) and the confidence intervals computed by FOA method. For example, the 95% confidence interval of the estimated demand multiplier (i.e. the range in which the true demand multipliers are expected to be 95% of the time) can be obtained by the following expression:

$$\begin{gathered} \|\Delta DMF_k\| \leq 1.96(W^{\frac{1}{2}}J)^{\dagger} \\ \left|\Delta DMF_k^j\right| \leq 1.96\sum_{j=1}^{m}\left|S_{ij}\right|, \text{ with } S = (W^{\frac{1}{2}}J)^{\dagger} \end{gathered} \qquad (18)$$

where J is the Jacobian matrix of flows and heads with respect to the water nodal demand at time k; W is the weight matrix where the diagonal elements are the reciprocals of the variances of measurement errors ($W=R^{-1}$); superscript $\dagger$ represents the pseudo-inverse operator. The derivation of Eq. (18) is explained in detail in Appendix A.

By considering the Jacobian (sensitivity) matrix, the uncertainty of the output model from FOA method can provide meaningful information about the sensitivity of the pressure with respect to the change in the nodal demand. This information can be used to guide where to place measurement stations. However, the method requires calculation of the sensitivity matrix, which may be time consuming when applied to large and complex networks.

Summary of assumptions and input requirements for the DMFLive model

Several assumptions are made for this study: (1) the model of the water distribution network perfectly represents the real system with known network characteristics (e.g. pipe roughness coefficients, length and diameters, etc.), and only demand multipliers are required to be estimated; (2) typical demand patterns for different homogeneous demand groups in WDS are assumed to be known. The homogeneous demand groups can be identified based on a multi-criteria demand zones clustering algorithm presented in Preis et al. (2010). There is uncertainty of the model outputs associated with demand groupings, but this is not considered here. Therefore, (3) the source of uncertainty is only from the errors from measurement devices; (4) the errors of the measurement devices are assumed to be known and to follow a Gaussian distribution; (5) the observation data for the online (near real-time) estimation model is available every 10, 15 minutes, 1 hour or larger time steps. The influence of slow transients (mass oscillations) are, therefore, ignored in this context.

The inputs required for the *DMFLive* model consist of the number of particles, the inputs for the demand prediction sub-model, inputs for the hydraulic simulation model (EPANET), input for the correction step and the parameters for the integrated GA operators (P_c, P_m *and* N_{gen}). The prediction sub-model requires the data of typical demand patterns, the auto-regression coefficient (ϕ_i) and the variance of noise of demand residuals (σ_h^2). These parameters are calibrated independently based on historical demand data for specific networks, for example $\phi_I = 0.7$ and $\sigma_h^2 = 0.13^2$ as in van Zyl et al. (2008). The EPANET model requires the known data of tank levels, pump and valve statuses. The correction step requires the observation data at measurement sites. Note that the particle filter model associated with the GA process can only be applied to networks with multiple demand patterns (e.g. the second case study in this paper). Two-point crossover operator with the probability of crossover $P_c=0.7$, bitwise mutation with the probability of $P_m = 1/N_{DM}$ (N_{DM} is the number of demand patterns in the network, $N_{DM} = 5$, corresponding with $P_m=0.2$ for the second case study) and the number of generations $N_{Gen} = 50$ were selected for the GA process.

Case study 1

The first case study used to evaluate the model is shown in Figure 2. The network has 9 nodes (8 nodes with demands), 12 pipes, one tank and one reservoir. The network characteristics can be found from the EPANET example, namely the Net1 network. Three pressure measurements (with a precision of $\Delta^{meas}=\pm 0.2$ m, consistent with a standard deviation of $\sigma^{meas} = 0.1$ for the measurement error at 95% confidence interval) are assumed to be placed at three random locations (nodes 13, 22 and 31). All nodal demands are assumed to follow a single demand pattern that varies every 15 minutes, (represented by the continuous line in Figure 2.b). The demand pattern is a random daily demand pattern (from a yearly demand pattern) for 100 households obtained from the BESS model (Thyer et al. 2011). The *DMFLive* model is required to track this demand pattern using the three pressure measurements, which are also obtained every 15 minutes.

In this case study, the default demand pattern given in the Net1 example (represented by the dashed line in Figure 2.b) was selected as the typical demand pattern. Different values of the auto-regression coefficient (ϕ) as well as variance of noise (σ_h^2) were applied for the demand prediction sub-model.

The accuracy of the demand estimates from the *DMFLive* model were evaluated in terms of the coefficient of determination (R^2) and the root mean squared error (RMSE). For a number of particles $N_p = 100$, the results of the demand estimates from the DMFLive model are presented in Table 1.

The *DMFLive* model performed very well when the auto-regression coefficient was selected in the range of $0.3 \leq \phi \leq 0.9$ and the noise variance was selected in the range of $0.25 \leq \sigma_h^2 < 0.64$. Due to the large difference between the typical demand value and the actual demand value at each time step (Figure 2.b), the selection of small values of the auto-regression coefficient and noise variance resulted in relatively poorer performance of the model (e.g. R^2 = 0.465 and RMSE =0.198 for $\phi=0.3$ and $\sigma_h^2=0.04$). The best output of the *DMFLive* model was obtained at $\phi=0.7$ and $\sigma_h^2=0.25$, with R^2 = 0.988 and RMSE =0.028, respectively.

For this best estimated demand pattern, the confidence intervals and the scattergram between actual demand multipliers and estimated demand multipliers are plotted in Figure 3.a.

In Figure 3.a, the estimated demand pattern yields a very good match with the actual demand pattern during the time period (24 hours, corresponding to 96 time steps). The actual demand pattern is entirely covered by the range of the 95% confidence intervals calculated from FOA method. This confidence interval range, which is expected to bracket the "true" demand multipliers in 95% of the cases, represents the uncertainty magnitude of the estimated demand due to the error from measurement devices.

The model has also been run with the number of particles $N_p=100$ and $N_p=20$ to provide a comparison between the FOA method (i.e. Eq. (18)) and the posterior analysis (i.e. Eq. (16)) for uncertainty quantification, as shown in Figures 3.b, 3.c and 3.d. Figures 3.a and 3.c show the uncertainty quantified by the FOA method while Figures 3.b and 3.d shown the uncertainty quantified by the variance of particles. For $N_p= 100$ particles, the 95% confidence intervals from both methods are comparable to each other, which demonstrates that the FOA method can provide reliable results compared to the variance of the particle filter samples.

A good estimate of the demand multipliers ($RMSE=0.047$) is obtained by the *DMFLive* model even when the number of particles is reduced by a factor of five ($N_p=20$), as seen in Figures 3.c, and 3.d. The uncertainty boundary calculated by the FOA method in Figure 3.c has a similar range to the case with $N_P=100$ particles and covers most of the actual values. On the other hand, the uncertainty bounds calculated by Eq. (16) in Figure 3.d are collapsed into single value at some time steps due to an insufficient number of the particles. Application of Eq. (16) for uncertainty quantification, therefore, requires an in-depth evaluation of the number of particles in the model if it is selected for the uncertainty quantification.

The range of demand multipliers predicted in time according to the evolution of the particles is presented in Figure 4.a. The predicted values range from DMF_{min} = 0.1 to *DMFmax* = 7.0, indicating that the demand prediction sub-model can predict a large range of demand multipliers, and cover the range $0 \leq DMF \leq 4$ suggested by Chin et al. (2000). Figure 4.b plots the scattergrams of the actual demand multipliers versus the predicted demand multipliers (i.e. the mean of the prediction) and actual demand multipliers versus estimated demand multipliers. The scattergram shows a constant and strong correlation between actual demand multipliers and estimated demand multipliers over time with R^2 being close to unity. Due to large difference between the typical demand pattern and the actual demand pattern, the forecasting model does not provide good prediction, resulting in weak and skewed correlation between the actual values and the predicted values. Despite this, the *DMFLive* model is still capable to provide very good estimates of the demand multipliers.

Effects of tank level update on the estimation

In extended period simulations of most hydraulic solvers (including EPANET), the nodal demands are considered to be constant during the time step. The levels of the tanks in the network at the end of the time step are consequently computed based on this assumption and are used as the initial tank level for the next step. Due to continuously unpredictable change of the water demand in practice, the actual tank level at the end of the time step is usually different to the tank level computed by the model. As a result, the estimated total volume of water used during the time step is also different from the actual volume of water used in practice. This issue can be overcome by minimizing the difference between actual tank levels at the beginning of the time step and the final estimated tank level at the end of the previous step. The demand estimation model, however, will be delayed until the information of the tank level at the beginning of the next time step becomes available. In other words, the model outputs will be the estimates of the demand multiplier at the previous time step.

In order to evaluate the effect of including tank level information at the end of every time step, an additional test is conducted. Instead of assuming that the observations are available at every 15 minutes, in this test it is assumed that the data can be obtained every hour and the model is required to estimate the demand pattern at each hour time step (while the actual demand pattern is varied every 15 minutes).

Figure 5 plots the two estimated demand patterns with and without tank level information (herein referred to as *DMF-WTLive* and *DMFLive*). Note that the *DMF-WTLive* model is the modified version of *DMFLive* model at which the final tank level information is taken into account.

It can be seen that the estimates for both cases are matched with the actual demand pattern at every hour time step. The inclusion of tank information only causes a slight difference between two estimated demand patterns at some of the time steps. The root mean squared errors between estimated demand multipliers and actual demand multipliers at every hour step indicates that the *DMFLive* model obtained slight better results than the *DMF-WTLive* model (RMSE =0.046 compared to RMSE =0.080, respectively). However, the total estimated water usages tabulated in Table 2 shows that the DMF-WTLive model is more accurate in predicting the volume of water delivered to the users.

The total estimated water usage during the 24-hour simulation period from *DMFLive* model was 5942.43 m^3/day, 46.81 m^3/day (or 0.78%) less than the actual water usage. On the other hand, total estimated water usage from *DMF-WTLive* model was 6007.31 m^3/day, only 18.07 m^3/day (or 0.30%) more than the actual value. Therefore, if the estimation can be delayed one time step, the final tank level information should be included into the model to improve the accuracy of the estimated total volume of water used.

Case study 2

In order to evaluate the performance of the proposed model in large networks that contain more than one demand pattern, the C-Town network from Ostfeld et al. (2011) is selected as the second case study. The network consists of 429 pipes, 1 reservoir, 7 tanks, 5 pump stations (with a total of 11 pumps), 4 PRV valves and 388 nodes (334 nodes with demands), which are divided into five district demand areas. Each district demand area follows a different hourly demand pattern. As the data of the demand patterns is available for seven days, the first 24 hours of these demand patterns are assumed to be the typical demand patterns for the demand prediction sub-model. The performance of the particle filter model is then evaluated by estimating the remaining 6-day hourly demand patterns.

It is assumed that there are 14 pressure measurement sites (from P1 to P14) that are randomly located at 14 places. These pressure measurements, again, are assumed to have a measurement error of $\Delta^{meas}=\pm 0.2$ m. The inputs for the real-time demand estimation model are, therefore, the pressures at these locations, the tank levels of seven tanks and the pump statuses of 11 pumps at each hour time step. The topology and measurement locations of the C-Town network are shown in Figure 6. Five different demand prediction sub-models were used to predict the five demand patterns. The parameters of the five demand prediction sub-models were assumed to have the same values of $\phi = 0.7$ for the auto-regression coefficients and $\sigma_h^2 = 0.16$ for the variances of noise.

The standard particle filter model (i.e. using systematic resampling), herein referred as the *DMFLive-I* model, provides good results only if $N_P \geq 25{,}000$ particles. The estimates of five different demand patterns for 6 days (from 25h to 168h) are shown in Figure 7. It is seen that the estimated demand patterns closely match the actual demand patterns, especially for DMF 2 (RMSE = 0.021), DMF 3 (RMSE = 0.024), DMF 1 (RMSE = 0.029) and DMF 4 (RMSE= 0.036). The estimated demand pattern DMF 5 is less accurate, with the root mean squared error of RMSE = 0.061.

Figure 7 also plots the 95% confidence intervals for calculated by the FOA formula. The intervals for the estimated DMF 1, DMF 2 and DMF 3 (in Figure 7.a, 7.b and 7.c, respectively) are narrow and they cover almost the entire set of the actual demand multiplier values. The actual values of DMF 4 are also within the confidence interval of estimated DMF 4 (Figure 7.d) for most of the time. However, due to the locations of the measurements (P7 and P9 - Figure 6), the confidence interval of estimated DMF 4 pattern is relatively large compared to the others. The effect of measurement locations on the confidence intervals of the estimates is discussed later in the paper. In Figure 7.e, approximately 37% of the actual demand values of the demand pattern DMF 5 are outside the 95% confidence intervals, which is caused by the relatively poor estimates for DMF 5. Figure 8 displays the scattergrams and coefficients of determination of the five predicted demand patterns, as well as the estimated demand patterns versus their actual values.

The predicted DMFs in this case show an average correlation to the actual DMFs with the R^2 ranging from 0.69 to 0.74, while the estimated DMFs are strongly correlated to the actual ones with all R^2 values being close to unity. The estimation for these five DMFs are also reliable during the simulation period (six days), as the spreads of the scattered dots are close to bisector lines.

Improving DMFLive model performance by SRGA and modified likelihood function

The *DMFLive-I* model can only perform well with a large number of particles ($N_P \geq 25{,}000$). Smaller numbers of particles result in weak estimates of the DMFs due to particle collapse at some steps. Since increasing the number of demand patterns requires an exponentially increasing number of particles, it is necessary to improve the efficiency of particle filter model so that it can be applied to complex systems.

Two methods have been investigated as mentioned previously in the paper: (1) incorporating the variance of the forecasting nodal heads into the likelihood function. The weights of particles in the model, referred as *DMFLive-II* model, are then calculated by the modified likelihood function (Eq. (17)); and (2) by the integration of a GA process into the systematic resampling of the model, herein referred as *DMFLive-III* model.

Table 3 presents results (in terms of the RMSE of each demand pattern) of running these models with $N_P=1000$ and $N_P=5000$ for *DMFLive-I*, *II* and with $N_P^{GA}=20$ and $N_P^{GA}=100$ for *DMFLive-III*. It may be seen that for both N_P values, the *DMFLive-I* gives very poor estimates of the DMFs. On the other hand, the *DMFLive-II* model only requires $N_P=1000$ (corresponding to $1.43*10^5$ evaluations for 143 hours) to provide fairly good results, while the *DMFLive-III* performs well when $N_P^{GA}=100$. The results of *DMFLive-II* ($N_P=5000$) and *DMFLive-III* ($N_P^{GA}=100$) give similar to the results of *DMFLive-I* running at $N_P=25{,}000$ (corresponding to total evaluations of $3.575*10^6$). This means the computation can be reduced by approximately a factor of five times.

Figure 9 shows the DMF 1 uncertainty ranges from 25 to 49 hours of the three models *DMFLive I, II* and *III* computed by FOA method and by variance of the particles Eq. (16). As can be seen from Figures 9.a and c, due to particle impoverishment, the uncertainty computed by particle variance, represented by the dashed lines, is merged into a single line at almost all of the time steps. The uncertainty in Figure 9.b computed by this method is wide due to the incorporation of the forecasting nodal heads into the likelihood function.

On the other hand, the uncertainties by FOA method, which are directly computed from the sensitivity matrix and the measurement errors, show consistent ranges in both cases. Given good estimates of the demand multipliers (as in Figures 9.b and c) these ranges can cover the actual values most of the time.

Effect of the locations of measurements on the quantification of demand uncertainty

As discussed in a number of studies such as in Piller (1995) and Do et al. (2016), the locations of the measurements have a strong impact on the results of the demand estimation models. Furthermore, the selection of measurement locations also affects the confidence intervals of the estimation outputs.

From the mathematical point of view, the uncertainty of estimated demands depends on the sensitivity of the flows/heads at measurement locations in relation to the change in the water nodal demands. This sensitivity is represented by the sensitivity matrix J (Eq. (18)), which is, in this case study, the Jacobian matrix of the heads with respect to the demand multipliers. The sensitivity of the heads with respect to the change of the demand multipliers depends on two factors: (1) the position of the nodes in the network and (2) the base demands at the nodes. In fact, the nodes close to fixed-head nodes (tanks or reservoirs) are less sensitive than the ones far from the fixed-head nodes. This is because of a change in nodal demands will result in a smaller change in the pressures of the closer nodes than the farther nodes. In a similar way, small base demands in the same pattern will result in small friction losses and consequently small changes in pressures. Therefore, nodes selected in these regions may cause large uncertainty in demand multiplier estimation. The sensitivity matrix takes into account these two factors. Small values in the sensitivity matrix values mean that the nodes are less sensitive to the demands and the estimation might have large uncertainty. Therefore, the uncertainty of the estimated DMFs can be reduced by selecting the more sensitive locations in the network.

Let us conduct an additional test to evaluate the effect of the measurement locations on the uncertainty of the estimated demand multipliers, for example the uncertainty of the estimated DMF 4. For this test, the locations of measurements P7 (with the base demand of D_7^0 = 0.50 L/s) and P9 (D_9^0 = 0.59 L/s) are relocated to P7A (D_{7A}^0 = 1.33 L/s) and P9A (D_{9A}^0 = 1.13 L/s). The *DMFLive* model was implemented with the same conditions and the other measurement locations are fixed at the same places as the original test.

Figure 10 shows the sensitivity matrixes J_0 (for the original test) and $J^A{}_0$ (for the modified test) corresponding to a set of estimated values DMFs = [0.46; 0.54; 0.65; 0.47; 0.62]

It is seen that, for this network, the heads at measurement locations are only sensitive to the change of the DMF that they belong to. For example, the variation in the DMF 4 pattern only affects the sensitivity of the heads at measurement locations P7 and P9 (for original test) and at measurement locations P7A and P9A (for the modified test). The non-zero values in the sensitivity matrices, therefore, correspond to the measurement locations. For the sensitivity of the heads, the new locations P7A ($\frac{\partial H}{\partial DMF4} = 5.31$) and P9A ($\frac{\partial H}{\partial DMF4} = 11.76$) are considerably more sensitive than the locations P7 ($\frac{\partial H}{\partial DMF4} = 2.59$) and P9 ($\frac{\partial H}{\partial DMF4} = 2.55$). As a result, the confidence intervals of the estimated DMF 4 for the modified test, as shown in Figure 11, are much narrower than the confidence intervals of the estimated DMF 4 for the original test presented in Figure 7.d. Note that in this network case study, the demand patterns are well geographically distributed. The heads at measurement locations are, therefore, affected by independent demand patterns, which results in a narrow uncertainty range for the estimate. For non-geographically distributed DMF networks, the sensitivity of the heads at measurement locations are required to be accounted and accumulated for all the related DMFs. This might cause much larger uncertainty and likewise bring difficulty for the estimation of the demand multipliers, as has been addressed in Sanz and Pérez (2014).

The relocation of the pressure measurements also improves the estimation of DMF 4, with a RMSE = 0.028 for the modified test, compared to a RMSE =0.036 of the original test. The placement of the two new measurement sites also causes a slight difference in the results of other estimated DMFs due to the change in the particle weights. However, the results of the four remaining DMFs are still very good and similar to the estimated values of the original test.

To sum up, the uncertainty of estimated demand multipliers caused by the errors of measurement devices is influenced by the measurement locations. It is suggested to choose the locations that are more sensitive to the demand multipliers (see Do et al. (2016) for an example of optimal measurement location). However, it has also been shown that the *DMFLive* model can be used to estimate the demand multipliers even when the measurement devices are located at some less sensitive places. The uncertainty of the estimated demand multipliers can be used to identify which measurement locations need to be improved. This is another advantage of the *DMFLive* model.

Conclusions and recommendations

Real-time demand estimation under uncertainties is exceptionally difficult due to the unpredictable stochastic behavior of the water demand as well as the nonlinearities of hydraulic systems. In this paper, the *DMFLive* model framework has been introduced, which can be used to estimate the demand multipliers of a WDS in near real-time. A predictor-corrector approach has been adopted and solved by a particle filter method. A nonlinear demand prediction model is applied to predict water demand multipliers at each time step, while the online pressure observations are used to correct the prediction. Output uncertainty caused by the measurement errors has also been quantified by the first-order approximation formula. The performance of the *DMFLive* model is evaluated by two WDS case studies. The results showed that the nonlinear demand prediction model combined with the particle filter method used in the paper are well suited for the near real-time demand estimation problem.

Within the first case study, the benefits of having additional information about the tank level of the next time step have been explored. If the estimation of the demand multipliers can be delayed one time step, the tank level at the beginning of the next time step can be used by the model to improve the estimation of the total volume of water used.

Within the second case study, three versions of the *DMFLive* model were developed to be used in large networks with multiple demand patterns. All versions provided good results, showing that the models are capable to be used in large networks. Finally, the effect of the measurement locations on the uncertainty of the estimated demand multipliers has been explored. Results showed that the uncertainty can be used to identify which measurement locations need to be improved. Future work involves considering adding additional uncertainties into the DMFLive model. Moreover, testing the model for non-geographically distributed demand networks is also necessary to show its capability when applied in practice.

Appendix

The problem of demand calibration involves finding the demands of the network hydraulic model to best fit the data set. Consider the nonlinear regression equation:

$$y_i^{Meas} = y_i(x) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_i) \tag{A1}$$

where x is the n_d by 1 vector of parameters to calibrate (the demand multiplier factors that depend on time); $y_i(x)$ is the scalar multivariate function of predictions from the network hydraulic model, given the parameter x; ε_i is the residual between model prediction and observation, which is assumed to be Gaussian with mean of zero and standard deviation of σ_i; y_i^{Meas} is the i^{th} measurement site in the data set.

The demand calibration can be formulated as a box-constrained Least Squares problem that minimizes the differentiable criterion at each time step:

$$f(x) = \frac{1}{2}\sum_{i=1}^{m}\left(\frac{y_i(x) - y_i^{Meas}}{\sigma_i}\right)^2 = \frac{1}{2}\sum_{i=1}^{m}\varepsilon_R^2 \qquad \text{(A2)}$$

$$s.t\ x^{min} \leq x \leq x^{max}$$

where m is the number of measurement sites, ε_R is the reduced residual, which is the residual divided by the corresponding standard deviation, $\varepsilon_R \sim N(0,1)$.

The gradient of f at x^0 is:

$$\nabla f_0 = J(x^0)^T W(y(x^0) - y^{Meas}) \qquad \text{(A3)}$$

where W is the weight matrix where the diagonal elements are the reciprocals of the variances of measurement errors; $J(x^0)^T = \partial_x y(x^0)^T$ is the transposed Jacobian matrix of the prediction function at $x = x^0$.

The Hessian approximation takes the simple form of the symmetrical, positive semi-definite matrix:

$$H_0 = J(x^0)^T W J(x^0) \qquad \text{(A4)}$$

It is essential for the Jacobian to be full rank of the size of x, so that H_0 is invertible and a definite matrix.

An approximation of function f to minimize Eq. (A2) by a quadratic function at x^0 leads to the approximation of x:

$$x = x^0 - (H_0)^{-1}\nabla f_0 \qquad \text{(A5)}$$

By replacing Eq. (A2) and Eq. (A3) into Eq. (A5), the approximation of x can be expressed as:

$$x = x^0 - (J(x^0)^T W J(x^0))^{-1} J(x^0)^T W(y(x^0) - y^{Meas})$$

Using Eq. (A1):

$$x(\varepsilon) = x^0 + (J(x^0)^T W J(x^0))^{-1} J(x^0)^T W \varepsilon \tag{A6}$$

The influence of the measurement errors with regards to the parameter estimates, therefore, can be obtained at the first-order of Eq. (A6):

$$\Delta x = (J(x^0)^T W J(x^0))^{-1} J(x^0)^T W \varepsilon = (W^{\frac{1}{2}} J(x^0))^{\dagger} W^{\frac{1}{2}} \varepsilon = (W^{\frac{1}{2}} J(x^0))^{\dagger} \varepsilon_R \tag{A7}$$

The uncertainty in term of confidence limits can be expressed as:

- For 99% confidence intervals ($|\varepsilon_i| \leq 2.58\sigma_i$):

$$\|\Delta x\| \leq 2.58 \left\| \left(J(x^0)^T W J(x^0)\right)^{-1} J(x^0)^T W^{\frac{1}{2}} \right\| = 2.58 \left\| (W^{\frac{1}{2}} J(x^0))^{\dagger} \right\|$$
$$|\Delta x_i| \leq 2.58 \sum_{j=1}^{m} |S_{ij}|, \text{ with } S = (W^{\frac{1}{2}} J)^{\dagger} \tag{A8}$$

- For 95% confidence intervals ($|\varepsilon_i| \leq 1.96\sigma_i$):

$$\|\Delta x\| \leq 1.96 \left\| (W^{\frac{1}{2}} J(x^0))^{\dagger} \right\|$$
$$|\Delta x_i| \leq 1.96 \sum_{j=1}^{m} |S_{ij}|, \text{ with } S = (W^{\frac{1}{2}} J)^{\dagger} \tag{A9}$$

References

Abe, N & Peter, B 2010, 'Epanet Calibrator-An integrated computational tool to calibrate hydraulic models', *Integrating Water Systems. Boxall & Maksimovic (eds)*.

Alvisi, S & Franchini, M 2017, 'Assessment of predictive uncertainty within the framework of water demand forecasting using the Model Conditional Processor (MCP)', *Urban Water Journal*, vol. 14, no. 1, pp. 1-10.

Bargiela, A & Hainsworth, G 1989, 'Pressure and Flow Uncertainty in Water Systems', *Journal of Water Resources Planning and Management*, vol. 115, no. 2, pp. 212-229.

Beal, C & Stewart, R 2014, 'Identifying Residential Water End Uses Underpinning Peak Day and Peak Hour Demand', *Journal of Water Resources Planning and Management*, vol. 140, no. 7, p. 04014008.

Bhave, PR 1988, 'Calibrating water distribution network models', *Journal of Environmental Engineering*, vol. 114, no. 1, pp. 120-136.

Boulos, PF & Ormsbee, LE 1991, 'Explicit network calibration for multiple loading conditions', *Civil Engineering Systems*, vol. 8, no. 3, pp. 153-160.

Boulos, PF & Wood, DJ 1990, 'Explicit calculation of pipe-network parameters', *Journal of Hydraulic Engineering*, vol. 116, no. 11, pp. 1329-1344.

Bush, CA & Uber, JG 1998, 'Sampling design methods for water distribution model calibration', *Journal of Water Resources Planning and Management*, vol. 124, no. 6, pp. 334-344.

Chin, DA, Mazumdar, A & Roy, PK 2000, *Water-resources engineering*, vol. 12, Prentice Hall Englewood Cliffs.

Cutore, P, Campisano, A, Kapelan, Z, Modica, C & Savic, D 2008, 'Probabilistic prediction of urban water consumption using the SCEM-UA algorithm', *Urban Water Journal*, vol. 5, no. 2, pp. 125-132.

Datta, R & Sridharan, K 1994, 'Parameter estimation in water-distribution systems by least squares', *Journal of Water Resources Planning and Management*, vol. 120, no. 4, pp. 405-422.

Do, NC, Simpson, AR, Deuerlein, JW & Piller, O 2016, 'Calibration of Water Demand Multipliers in Water Distribution Systems Using Genetic Algorithms', *Journal of Water Resources Planning and Management*, p. 04016044.

Doucet, A, Godsill, S & Andrieu, C 2000, 'On sequential Monte Carlo sampling methods for Bayesian filtering', *Statistics and computing*, vol. 10, no. 3, pp. 197-208.

Hol, JD, Schon, TB & Gustafsson, F 2006, 'On resampling algorithms for particle filters', *Nonlinear Statistical Signal Processing Workshop, 2006 IEEE,* IEEE, pp. 79-82.

Hutton, C, Kapelan, Z, Vamvakeridou-Lyroudia, L & Savić, D 2013, 'Application of Formal and Informal Bayesian Methods for Water Distribution Hydraulic Model Calibration', *Journal of Water Resources Planning and Management*, vol. 140, no. 11, p. 04014030.

Hutton, CJ & Kapelan, Z 2015, 'A probabilistic methodology for quantifying, diagnosing and reducing model structural and predictive errors in short term water demand forecasting', *Environmental Modelling & Software*, vol. 66, pp. 87-97.

Hutton, CJ, Kapelan, Z, Vamvakeridou-Lyroudia, L & Savic, DA 2012a, 'Real-time demand estimation in water distrubtion systems under uncertainty ', paper presented at WDSA 2012: 14th Water Distribution Systems Analysis Conference.

Hutton, CJ, Kapelan, Z, Vamvakeridou-Lyroudia, L & Savić, DA 2012b, 'Dealing with uncertainty in water distribution system models: A framework for real-time modeling and data assimilation', *Journal of Water Resources Planning and Management*, vol. 140, no. 2, pp. 169-183.

Kang, D & Lansey, K 2009, 'Real-time demand estimation and confidence limit analysis for water distribution systems', *Journal of Hydraulic Engineering*, vol. 135, no. 10, pp. 825-837.

Kapelan, ZS, Savic, DA & Walters, GA 2007, 'Calibration of water distribution hydraulic models using a Bayesian-type procedure', *Journal of Hydraulic Engineering*, vol. 133, no. 8, pp. 927-936.

Kitagawa, G 1996, 'Monte Carlo filter and smoother for non-Gaussian nonlinear state space models', *Journal of computational and graphical statistics*, vol. 5, no. 1, pp. 1-25.

Lansey, K, El-Shorbagy, W, Ahmed, I, Araujo, J & Haan, C 2001, 'Calibration assessment and data collection for water distribution networks', *Journal of Hydraulic Engineering*, vol. 127, no. 4, pp. 270-279.

Lansey, KE & Basnet, C 1991, 'Parameter estimation for water distribution networks', *Journal of Water Resources Planning and Management*, vol. 117, no. 1, pp. 126-144.

Moradkhani, H, Hsu, KL, Gupta, H & Sorooshian, S 2005, 'Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter', *Water Resources Research*, vol. 41, no. 5.

Nagar, AK & Powell, RS 2002, 'LFT/SDP approach to the uncertainty analysis for state estimation of water distribution systems', *Control Theory and Applications, IEE Proceedings-,* IET, vol. 149, pp. 137-142.

Nicklow, J, Reed, P, Savic, D, Dessalegne, T, Harrell, L, Chan-Hilton, A, Karamouz, M, Minsker, B, Ostfeld, A, Singh, A & Zechman, E 2010, 'State of the Art for Genetic Algorithms and Beyond in Water Resources Planning and Management', *Journal of Water Resources Planning and Management*, vol. 136, no. 4, pp. 412-432.

Ormsbee, LE 1989, 'Implicit network calibration', *Journal of Water Resources Planning and Management*, vol. 115, no. 2, pp. 243-257.

Ormsbee, LE & Wood, DJ 1986, 'Explicit pipe network calibration', *Journal of Water Resources Planning and Management*, vol. 112, no. 2, pp. 166-182.

Ostfeld, A, Salomons, E, Ormsbee, L, Uber, JG, Bros, CM, Kalungi, P, Burd, R, Zazula-Coetzee, B, Belrain, T & Kang, D 2011, 'Battle of the water calibration networks', *Journal of Water Resources Planning and Management*, vol. 138, no. 5, pp. 523-532.

Perelman, L, Maslia, ML, Ostfeld, A & Sautner, JB 2008, 'Using aggregation/skeletonization network models for water quality simulations in epidemiologic studies', *Journal (American Water Works Association)*, vol. 100, no. 6, pp. 122-133.

Piller, O 1995, 'Modeling the behavior of a network-Hydraulic analysis and sampling procedures for parameter estimation', *Applied Mathematics thesis from the University of Bordeaux (PRES)*, vol. PhD Thesis.

Preis, A, Whittle, A & Ostfeld, A 2009, 'Online hydraulic state prediction for water distribution systems', *World Environmental and Water Resources Congress 2009: Great Rivers, May 17, 2009 - May 21, 2009,* American Society of Civil Engineers, Kansas City, MO, United states, vol. 342, pp. 323-345.

Preis, A, Whittle, AJ, Ostfeld, A & Perelman, L 2010, 'On-line hydraulic state estimation in urban water networks using reduced models', *10th International Conference on Computing and Control for the Water Industry: Integrating Water Systems, CCWI 2009, September 1, 2009 - September 3, 2009,* CRC Press, Sheffield, United kingdom, pp. 319-324.

Preis, A, Whittle, AJ, Ostfeld, A & Perelman, L 2011, 'Efficient Hydraulic State Estimation Technique Using Reduced Models of Urban Water Networks', *Journal of Water Resources Planning and Management*, vol. 137, no. 4, pp. 343-351.

Rahal, CM, Sterling, MJH & Coulbeck, B 1980, 'Parameter tuning for simulation models of water distribution networks', *Proceedings of the Institution of Civil Engineers*, vol. 69, no. 3, pp. 751-762.

Ristic, B, Arulampalm, S & Gordon, N 2004, *Beyond the Kalman filter : particle filters for tracking applications*, Artech House Boston, Ma., London.

Rossman, LA 2000, 'EPANET 2: users manual', US Environmental Protection Agency (EPA), USA.

Salmond, D & Gordon, N 2005, 'An introduction to particle filters', *State space and unopbserved component models theory and applications*, pp. 1-19.

Sanz, G & Pérez, R 2014, 'Comparison of demand pattern calibration in water distribution networks with geographic and non-geographic parameterization', *11th International Conference on Hydroinformatics,* pp. 1-8.

Sanz, G & Pérez, R 2015, 'Sensitivity analysis for sampling design and demand calibration in water distribution networks using the singular value decomposition', *Journal of Water Resources Planning and Management*, vol. 141, no. 10, p. 04015020.

Savic, DA, Kapelan, ZS & Jonkergouw, PM 2009, 'Quo vadis water distribution model calibration?', *Urban Water Journal*, vol. 6, no. 1, pp. 3-22.

Shang, F, Uber, JG, Van Bloemen Waanders, BG, Boccelli, D & Janke, R 2006, 'Real time water demand estimation in water distribution system', *8th Annual Water Distribution Systems Analysis Symposium 2006, August 27, 2006 - August 30, 2006,* American Society of Civil Engineers, Cincinnati, OH, United states, 10.1061/40941(247)95, p. 95.

Thyer, MA, Micevski, T, Kuczera, GA & Coombes, P 2011, 'A behavioural approach to stochastic end use modelling', paper presented at Ozwater 11, Australia's National Water Conference and Exhibition.

USEPA 2005, *Water Distribution System Analysis: Field Studies, Modeling and Management. A Reference Guide for Utilities*, O USEPA Cincinnati, USA, USEPA Cincinnati, Ohio, USA.

van Leeuwen, PJ 2009, 'Particle filtering in geophysical systems', *Monthly Weather Review*, vol. 137, no. 12, pp. 4089-4114.

van Leeuwen, PJ 2010, 'Nonlinear data assimilation in geosciences: an extremely efficient particle filter', *Quarterly Journal of the Royal Meteorological Society*, vol. 136, no. 653, pp. 1991-1999.

van Zyl, JE, Piller, O & le Gat, Y 2008, 'Sizing municipal storage tanks based on reliability criteria', *Journal of Water Resources Planning and Management*, vol. 134, no. 6, pp. 548-555.

Waller, JA 2013, 'Using observations at different spatial scales in data assimilation for environmental prediction', PhD Thesis, Department of Mathematics and Statistics, University of Reading. Online at: www. reading. ac. uk/maths-and-stats/research/theses/maths-phdtheses. aspx.

Walski, TM 1983, 'Technique for calibrating network models', *Journal of Water Resources Planning and Management*, vol. 109, no. 4, pp. 360-372.

Weerts, AH & El Serafy, GY 2006, 'Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models', *Water Resources Research*, vol. 42, no. 9.

Xu, C & Goulter, IC 1998, 'Probabilistic model for water distribution reliability', *Journal of Water Resources Planning and Management*, vol. 124, no. 4, pp. 218-228.

Table 1: Coefficient of determination (R^2) and root mean squared error (RMSE) of demand estimates corresponding to different parameter values of the demand prediction model for case study 1

No	Auto-regression coefficient (ϕ)	Variance of demand residual (σ_h^2)	R^2	RMSE
1		0.04	0.465	0.198
2	0.3	0.25	0.986	0.030
3		0.64	0.983	0.033
4		0.04	0.528	0.189
5	0.5	0.25	0.986	0.030
6		0.64	0.987	0.029
7		0.04	0.982	0.033
8	0.7	0.25	**0.988**	**0.028**
9		0.64	0.986	0.031
10		0.04	0.987	0.029
11	0.9	0.25	0.986	0.031
12		0.64	0.985	0.031

*Bold – Best estimated result

Table 2: Actual and estimated total volume of water usage during calculated period

Cases	Total (m^3/day)	Difference (m^3/day)	% Difference (%)
Actual daily water usage	5989.25		
Estimated water usage with *DMFLive*	5942.43	46.81	0.78
Estimated water usage with *DMF-WTLive*	6007.31	18.07	0.30

Table 3: Performance of DMFLive model with SR (I), modified likelihood function (II) and SRGA (III)

Model type	DMFLive-I		DMFLive-II		DMFLive-III (N_{Gen}=50)	
No. Particles	N_P=1000	N_P=5000	N_P=1000	N_P=5000	N_P^{GA}=20	N_P^{GA}=100
No. Eval.	$1.43*10^5$	$7.15*10^5$	$1.43*10^5$	$7.15*10^5$	$1.08*10^5$	$5.43*10^5$
$RMSE_{DMF1}$	0.386	0.405	0.050	0.027	0.107	0.030
$RMSE_{DMF2}$	0.365	0.422	0.026	0.021	0.067	0.025
$RMSE_{DMF3}$	0.416	0.237	0.029	0.027	0.068	0.023
$RMSE_{DMF4}$	0.385	0.229	0.043	0.038	0.086	0.032
$RMSE_{DMF5}$	0.366	0.246	0.074	0.049	0.190	0.050