



MIT Open Access Articles

Efficient Hydraulic State Estimation Technique Using Reduced Models of Urban Water Networks

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation	Preis, Ami et al. "Efficient Hydraulic State Estimation Technique Using Reduced Models of Urban Water Networks." <i>Journal of Water Resources Planning and Management</i> 137.4 (2011): 343.
As Published	http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000113
Publisher	American Society of Civil Engineers
Version	Author's final manuscript
Citable link	http://hdl.handle.net/1721.1/69965
Terms of Use	Creative Commons Attribution-Noncommercial-Share Alike 3.0
Detailed Terms	http://creativecommons.org/licenses/by-nc-sa/3.0/

Efficient Hydraulic State Estimation Technique using Reduced Models of Urban Water Networks

Ami Preis¹; Andrew J. Whittle²; Avi Ostfeld³; and Lina Perelman⁴

1

Abstract: This paper describes and demonstrates an efficient method for online hydraulic state estimation in urban water networks. The proposed method employs an online predictor-corrector (PC) procedure for forecasting future water demands. A statistical data-driven algorithm (M5 Model-Trees algorithm) is applied to estimate future water demands and an evolutionary optimization technique (genetic algorithms) is used to correct these predictions with online monitoring data. The calibration problem is solved using a modified least-squares (LS) fit method (Huber function) in which the objective function is the minimization of the residuals between predicted and measured pressure at several system locations, with the decision variables being the hourly variations in water demands. To meet the computational efficiency requirements of real-time hydraulic state estimation for prototype urban networks that typically comprise tens of thousands of links and nodes, a reduced model is introduced using a water system-aggregation technique. The reduced model achieves a high-fidelity representation for the hydraulic performance of the complete network, but greatly simplifies the computation of the PC loop and facilitates the implementation of the online model. The proposed methodology is demonstrated on a prototypical municipal water-distribution system. DOI: 10.1061/(ASCE)WR.1943-5452.0000113. © 2011 American Society of Civil Engineers.

CE Database subject headings: Water distribution systems; Monitoring; Hydraulic models; Urban areas.

Author keywords: Water-distribution system; Online hydraulic state estimation; Predictor-corrector model; Real-time monitoring.

Introduction

Integration of near real-time hydraulic data with hydraulic computer-simulation models allows water utility engineers to operate and control their large-scale, urban water-distribution systems in real time. In conventional practice, hydraulic models are calibrated off-line (USEPA 2005) using a short-term (e.g., 1 wk) sample of flow rate and pressure measurements within the network. Thereafter, uncertain system parameters (e.g., water demands and pipe roughness) are adjusted until an acceptable match is achieved between the model outputs and physical observations. Ormsbee (1989) and Lansley and Basnet (1991) were among the first to develop formal optimization methods for determining the uncertain system elements. Datta and Sridharan (1994) and Reddy et al. (1996) used the regression approach in which parameter uncertainties were estimated as part of the calibration process. Greco and Del Giudice (1999) used a sensitivity matrix to minimize the least-squares differences between observed and predicted values, and Lingireddy and Ormsbee (1998) developed a calibration method using artificial neural networks (ANNs).

¹Postdoctoral Associate, Center for Environmental Sensing and Modeling, MIT-SMART Center, Singapore (corresponding author). E-mail: amipreis@smart.mit.edu

²Professor and Dept. Head, Dept. of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA. E-mail: ajwhittl@mit.edu

³Associate Professor, Faculty of Civil and Environmental Engineering, Technion—I.I.T., Haifa, Israel. E-mail: ostfeld@tx.technion.ac.il

⁴PhD student, Faculty of Civil and Environmental Engineering, Technion—I.I.T., Haifa, Israel. E-mail: lina@tx.technion.ac.il

Note. This manuscript was submitted on January 31, 2010; approved on July 26, 2010; published online on July 31, 2010. Discussion period open until December 1, 2011; separate discussions must be submitted for individual papers. This paper is part of the *Journal of Water Resources Planning and Management*, Vol. 137, No. 4, July 1, 2011. ©ASCE, ISSN 0733-9496/2011/4-0-0/\$25.00.

More recently, Kapelan et al. (2007) used the shuffled complex evolution metropolis (SCEM-UA) global optimization algorithm to solve a least-squares-type calibration problem in which both calibration parameter values and associated uncertainties were considered in a single optimization model run. Genetic algorithms (Holland 1975) have been used by Savic and Walters (1995), Wu and Simpson (2001), Kapelan et al. (2002), Wu et al. (2002), Walski et al. (2006), and Clark and Wu (2006).

Overall, the main limitation of all off-line calibration procedures is that they approximate the unknown parameters using a short-term sample of hydraulic data. The calibration results may represent the system hydraulics during the short period of the sampling procedure, but they are not expected to represent accurately the system conditions for the full range of operational conditions that can occur. In the case of water demands, this issue is even more critical, because water demands have dynamic/stochastic pattern variations that fluctuate with time-changing economic and demographic characteristics and may even show trends with local climatic conditions (Maidment and Miao 1986; Kenward and Howard 1999; Zhou et al. 2000). In principle, much more realistic predictions can be achieved by updating the hydraulic state estimation using continuous online hydraulic measurements, provided by a sensor network installed on the distribution system.

This paper describes and demonstrates an efficient method for online hydraulic state estimation in urban water networks. Several studies have assimilated online measurements into hydraulic state-estimation models. Andersen and Powell (2000) proposed a constrained weighted least-squares technique to investigate the effect of measurement bounds with the assumption that demands were measured or estimated from knowledge of water consumers' characteristics.

Davidson and Bouchart (2006) proposed proportional and target demand methods. These are two techniques for adjusting estimated demands in hydraulic models of water-distribution networks to produce solutions that are consistent with available Supervisory

2

Control and Data Acquisition (SCADA) data. The two techniques assume that pipe resistances and SCADA data are accurate and that the combination of SCADA data and demand estimates produce overdetermined problems. Nodal demands are regarded as stochastic variables that fluctuate about an estimated mean value. The method of weighted least squares is used to obtain solutions that satisfy all of the constraints imposed by SCADA data with adjusted nodal demands that most closely resemble the estimates. The methods are intended for use in real-time modeling but are limited to quasi-steady-state flow.

Shang et al. (2006) presented a predictor-corrector (PC) method, implemented in an extended Kalman filter to estimate water demands within distribution systems in real time. A time-series ARMA model is used to predict the water demands on the basis of the estimated demands at previous steps. The predictions are corrected using measured nodal water heads or pipe flow rates. The proposed methodology is in a preliminary stage and aimed mainly at studying the impact of spatial correlation between demand forecast errors on demand estimations. The methodology is demonstrated on EPANET example 3, having 59 demand nodes, through three simulation studies with 20 pressure, 20 flow rate, and 40 flow rate sensors. The main conclusion was that the model performances depend on sampling design, measurement uncertainty, demand forecast error, and the spatial correlations among the demand forecast errors. Although only preliminary results were presented, the study provided a modeling framework and mathematical tools for further implementations on more complex case studies.

This paper, which extends early work of the writers (Preis et al. 2009a, b), uses a PC approach that integrates a limited number of continuous observations to update predictions of the hydraulic state of a real urban water-supply network at regular time intervals. The M5 Model-Trees algorithm (Quinlan 1992) is used to forecast future water demands for a rolling planning horizon of 24 h ahead, and genetic algorithms are used to correct (i.e., calibrate) these predicted values in real time. Thereafter, at each subsequent time step, the corrected outputs of previous iterations are used as inputs for the prediction model. This a priori estimation of the calibration parameter values repeats itself at each subsequent time step whereas the forecasting model inputs correspond to the corrected outputs of previous iterations, thus improving the model performances over time and providing adequate information on the hydraulic state of the system for real-time operation and control. To meet the computational efficiency requirements of this online procedure, the urban network model is condensed to an equivalent system, with a reduced number of links and nodes through a system aggregation technique (Ulanicki et al. 1996).

System reduction and aggregation techniques were first discussed by Hamberg and Shamir (1988) who presented a simplification methodology using a stepwise combination of system elements, and through a nonlinear continuum representation of the system as a bundle. Anderson and Al-Jamal (1995) introduced a methodology for the simplification of complicated hydraulic networks using a parameter-fitting approach. The layout of the final simplified network was specified a priori. Pipe conductances and demand distribution were then determined by maximizing the fitness between the simplified and full system performances.

Ulanicki et al. (1996) presented a mathematical model for the hydraulic aggregation of water-distribution systems. The approach computes a network model equivalent to the original system with fewer components by analytical elimination of system components. The reduced nonlinear model preserves the nonlinearity of the original model and approximates the original system within a wide range of operating conditions.

Recent studies that implemented the method developed by Ulanicki et al. (1996) include Shamir and Salomons (2008) who developed a method for near-optimal online operation of an urban water-distribution system by using a reduced model of the network with optimization by a genetic algorithm. Perelman and Ostfeld (2008) have also developed a similar methodology for a conjunctive hydraulic and water quality model.

Methodology

Step 1: Off-Line Hydraulic Model Reduction

Appendix A summarizes the algorithm developed by Ulanicki et al. (1996) that is used to create an equivalent reduced system. The algorithm proceeds in a step-by-step elimination of pipes and nodes, allocating the demand at the node being eliminated to its neighboring nodes. On the basis of reducing the algebraic system of mass and energy conservation equations by eliminating variables using Gauss elimination (Hammerlin and Hoffman 1991) all reservoirs, pumps, valves, tanks, and critical nodes (e.g., nodes in which pressure is monitored and nodes that represent significant water customers) remain in the reduced network. The validity of the system's reduction is measured by the similarity of the connectivity of the simplified system with that of the original system and its hydraulic performance (e.g., similarity of pressure at nodes, water levels at tanks, and/or pumps operation) over a wide range of operating conditions.

Step 2: Predictor-Corrector Model

Step 2.1: Demand Multiplication Factors Prediction

The patterns in demand for hourly basis time steps are described by the Demand Multiplication Factors (DMFs). In any time step, the water consumption at a given node can be found by multiplying the relevant DMF by the baseline demand.

There are thousands of water consumers with highly variable and stochastic individual demand behavior to be estimated in a typical urban water system and only a relatively small number of direct measurements are available. This creates an ill-posed, highly underdetermined calibration problem that may result with no solution at all, no unique solution, or an unstable solution. This can be addressed by grouping the unknown parameters. The main advantage of grouping is that the size of the problem is reduced. By reducing the problem uncertainty, it is more feasible to predict aggregated demands than individual ones.

In this study, the consumption nodes are grouped into demand zones on the basis of a spatial analysis of the system and each group of consumption nodes is assigned its own set of DMFs. The grouping is on the basis of the assumption that water customers in a given area of the system will have the same characteristics and will not need large adjustments to achieve calibration (Walski et al. 2003).

Thereafter, the demand zones DMFs are predicted using the M5 Model-Trees algorithm (Quinlan 1992), with the inputs being the calibrated DMFs from past hours $t - 24$, $t - 25$, $t - 168$, and $t - 169$. These time cycles are based on the common approach (Shang et al. 2006, Alvisi et al. 2007, and Ghiassi et al. 2008) for the time-series forecasting of water demands that relies on direct identification of patterns existing in the archived system data.

Water demand patterns usually follow a 24-h cycle. This cycle is called the diurnal demand pattern and is used by many urban water utilities to plan the system operation 1 day ahead (i.e., to schedule pump operation and plan tank storage). Weekend demand patterns often differ from weekday patterns and usually follow a 168-h (1 wk) cycle.

Persistence on hourly level (e.g., DMF from past hour $t - 1$) was not considered in this study because the aim of this study was to develop a decision-support tool that can be used to predict the water system hydraulic states 24 h ahead at 1-h time-step intervals. At each time step, the future DMFs at $t + 24$ hrs are predicted with the following inputs: DMFs at the current time step t ; DMFs at $t - 1$; DMFs at $t - 144$ (i.e., $t - 168 + 24$); and DMFs at time step $t - 145$ (i.e., $t - 169 + 24$).

Taking into consideration persistence on hourly level (e.g. $t - 1$) may improve the overall model results, but the primary drawback to this approach is that it will limit the forecasting horizon to only 1 h ahead, which will not be enough for forecasting the system response to possible future operational scenarios that can range from several hours to several days.

The prediction tool, M5 Model-Trees algorithm, presented by Quinlan (1992) builds rule-based predictive models by using a top-down induction approach. Preis et al. (2009a) provides a comprehensive description of the model-trees technique and, therefore, only a short description of the method is presented here. A model tree is fitted to a training data set by recursively partitioning the data into homogeneous subsets on the basis of its attributes. The training data set includes combinations of inputs (i.e., DMF_{t-169} , DMF_{t-168} , DMF_{t-25} , and DMF_{t-24}) and outputs (i.e., DMF_t). The input training data set is recursively partitioned into smaller and smaller fractions ('tree leaves') until the data at the groups constitute relatively homogeneous subsets such that a linear regression equation can explain the remaining variability of each homogeneous subset (e.g., $\alpha \cdot DMF_{t-169} + \beta \cdot DMF_{t-168} + \gamma \cdot DMF_{t-25} + \delta \cdot DMF_{t-24} + \varepsilon = DMF_t$).

The overall set of linear regression models constructs the model tree with all training cases being predicted by the tree leaves. To simplify the tree structure, and thus to improve its ability to classify new instances, the tree is then pruned from the bottom up by quantifying the contribution of each attribute to the overall predicted value and removing those attributes that add little to the model. At the last stage, a smoothing process is performed to compensate for the sharp discontinuities that will inevitably occur between adjacent linear models at the leaves of the pruned tree.

The system hydraulic behavior is simulated by using the steady-state mode of EPANET, with the predicted DMFs as inputs. Possible simulation outputs are nodal pressures and pipe flow rates.

Step 2.3: Online Hydraulic Data Integration

Pressure and/or flow rate measurements (from a set of sensors) are inserted to the model at each time-step interval. This procedure is related directly to the calibration Step 2.4, at which the integration of the measurements in the model is been done by assimilating a

vector of updated hydraulic records in the model at each subsequent time-step interval (i.e., data is streamed continually to the model).

Step 2.4: DMF Correction/Calibration

A calibration problem is formulated and solved by using a genetic algorithm, which is a heuristic combinatorial search technique that imitates a set of evolutionary rules involving processes of selection, crossover, and mutation. This is accomplished by creating a random search technique that combines survival of the fittest with a randomized information exchange.

The objective of the calibration process is to match the computed and measured sensor-node data (pressure and/or flow rates), taking into consideration possible noise in the measurements. In this application, calibration is achieved by minimizing a modified least squares of differences function known as the Huber function (Huber 1973). The Huber function implementation to the hydraulic state-estimation problem is described as follows:

1. The differences (i.e., residuals) between modeled and observed pressures and flow rates at each time step, at sensor node i - are defined as $R_{i,t=k}^P$ and $R_{i,t=k}^Q$, respectively
2. The Huber function of each residual R is defined as

$$f(R_{i,t}^{P \text{ or } Q}) = \begin{cases} \frac{1}{2}(R_{i,t}^{P \text{ or } Q})^2, & |R_{i,t}^{P \text{ or } Q}| \leq h \\ h|R_{i,t}^{P \text{ or } Q}| - \frac{1}{2}h^2, & |R_{i,t}^{P \text{ or } Q}| > h \end{cases} \quad (1)$$

where h = predefined value that represent the tolerance to noise in measurements. For small residuals ($|R| \leq h$) that represent low to zero values of noise in sensor measurements, the Huber function minimizes the usual least-squares function (i.e., l_2 norm approximation); for large R ($|R| > h$) that represent high values of noise in sensor measurements, it minimizes a linear penalty function that is relatively insensitive to noise (i.e., l_1 norm approximation)

3. The overall calibration problem objective function to be minimized at each hydraulic time step t is defined as

$$\sum_{i=1}^{N_p} f(R_{i,t}^P) + \sum_{i=1}^{N_Q} f(R_{i,t}^Q) \quad (2)$$

where i = sensor nodes (i.e., monitoring station) index; N_p = total number of pressure sensors; and N_Q = total number of flow rate sensors. In this application, the value of h in each sensor node at each time-step is equal to the average of all previous time-steps' sensor-node residuals multiplied by a factor of 2.

Step 2.5: DMFs Delay

The calibrated DMFs are delayed for 24, 25, 168, and 169 h before being used as inputs in the prediction model. The DMFs delay step

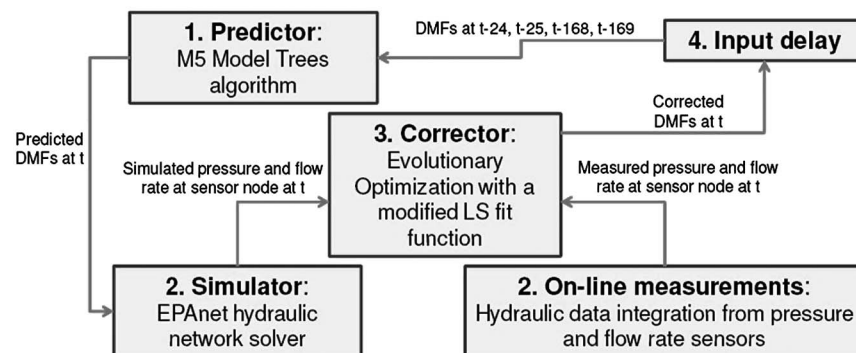


Fig. 1. PC loop for DMFs prediction at the t th time step

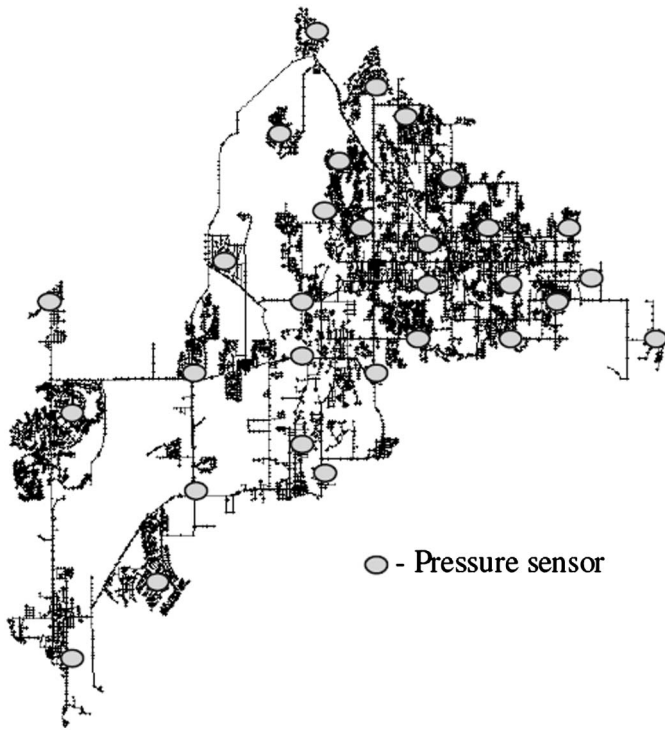


Fig. 2. Network 2 full model with the sensor node locations

is used to synchronize the input data set of the model trees according to the daily and weekly time cycles that are used in the prediction process.

Steps 2.1 to 2.5 (Fig. 1) start at $t = 169$ h, after performing an off-line calibration procedure for the first 168 h (1 week) of the collected data; the aim of this off-line calculation is to generate initial values for the input data-set of the prediction model. No a priori information is used during the first 168 h apart from constraints on the minimum and maximum values of DMF (0.0–3.0, respectively)

that are based on prior publications by Walski et al. (2003) and Jonkergouw et al. (2008).

The PC method is implemented for the reduced model of the water-distribution network to meet the computational efficiency requirements of this online procedure. The predicted water demands are effectively the same in the reduced- and full-scale models.

Results

The PC approach developed in this study was tested against the input data of Network 2 (Fig. 2) of the “Battle of the Water Sensor Networks (BWSN): A Design Challenge for Engineers and Algorithms” (Ostfeld et al. 2008). The network corresponds to an anonymous but real water-distribution system comprising 12,523 nodes, two constant head sources, two tanks, 14,822 pipes, four pumps, and five valves. The system was subject to highly variable demand patterns over a period of 934 h (~39 days). Hydraulic simulations for this system are considered valid for this entire duration. The original EPANET input file was downloaded from the University of Exeter Centre for Water Systems (ECWS) website: (www.exeter.ac.uk/cws/bwsn). The current application assumes that continuous in-line data are available from 30 pressure sensors (Fig. 2).

The nodal pressure records from these locations were generated by the EPANET model by using real input data for the system. The reservoirs and tank water levels were considered as known inputs. The model was implemented in an online manner at which the synthetic pressure measurements were streamed to the model at each time-step interval.

BWSN Network 2 Aggregation Results

The method of Ulanicki et al. (1996) was used to create four reduced hydraulic models of the full network (Fig. 3). All the reduced models include the system’s reservoirs, tanks, pumps, valves, pressure monitoring nodes, and significant consumption nodes (i.e., significant nodes). Reduced model 1 is a reduced network comprised of the significant nodes and nodes connected to pipes with diameter above 10 in. (total of 2,612 nodes and 3,748 links);

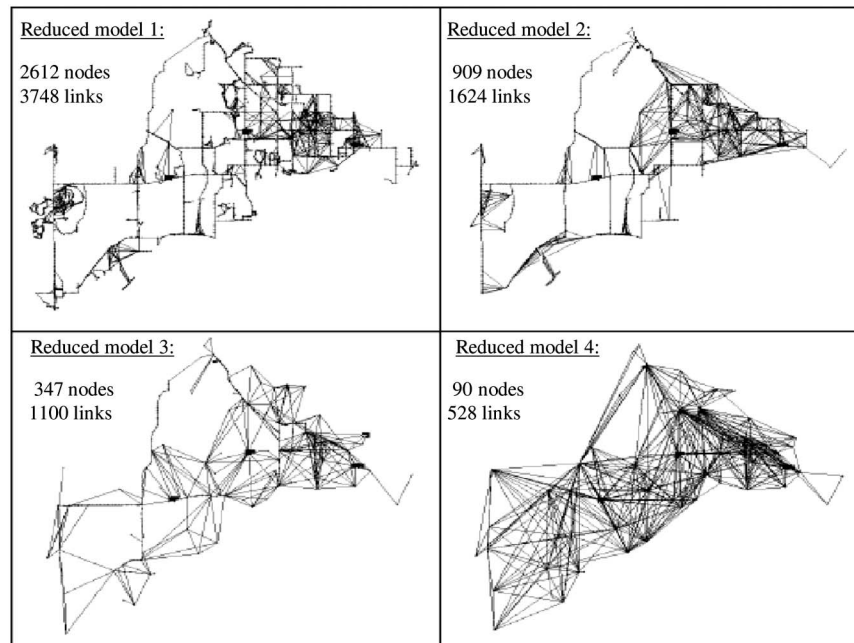


Fig. 3. Four reduced hydraulic models of Network 2

Table 1. Comparison of Sensor Nodes Pressure Data Over 168 h in the Reduced Model with That Calculated by the Full Model

		Fraction of the total sample population (%)			
		Reduced model 1: (2,612 nodes; 3,748 links)	Reduced model 2: (909 nodes; 1,624 links)	Reduced model 3: (347 nodes; 1,100 links)	Reduced model 4: (90 nodes; 528 links)
Ranges of pressure data accuracy (psi)	within 0.02 psi	99.96	99.87	99.81	93.79
	within 0.04 psi	99.98	99.95	99.93	97.39
	within 0.06 psi	100	100	100	98.53
	within 0.08 psi	—	—	—	99.08
	within 0.1 psi	—	—	—	99.41
	within 0.2 psi	—	—	—	100

Table 2. Reduction in the Hydraulic Simulation Computational Time

Reduced model option	Hydraulic simulation computational time reduction (%)
Reduced model 1: (2,612 nodes; 3,748 links)	51
Reduced model 2: (909 nodes; 1,624 links)	77
Reduced model 3: (347 nodes; 1,100 links)	89
Reduced model 4: (90 nodes; 528 links)	93

reduced model 2 is comprised of the significant nodes and nodes connected to pipes with diameter above 14 in. (total of 909 nodes and 1,624 links); reduced model 3 is comprised of the significant nodes and nodes connected to pipes with diameter above 24 in. (total of 347 nodes and 1,100 links); and reduced model 4 contains only the significant nodes (total of 90 nodes and 528 links).

The validity of the four possible reductions of the full hydraulic network is measured by the similarity of sensor nodal pressure data

over time in the reduced models with that calculated by the full model (as will be shown in Table 1). The pressure data for the validation tests was generated by using a representative sample of hourly demand pattern for 168 h (1 week) of the utility operation. The average range of operating pressures during this week was 20 psi. The results in Table 1 show that all of these reduced models resemble the original system's hydraulic performances with high accuracy.

Table 2 summarizes the reduction in the hydraulic simulation computational time as a percentage of the hydraulic simulation time of the full network that equals to 10 s for extended period simulation of 168 time steps on a DELL PC (2.66 GHz, 3.0 GB of RAM).

Reduced model 3 was chosen for the online hydraulic simulations as it has an equivalent accuracy rate in simulating pressure values as the less reduced networks (e.g., models 1 and 2) and its simulation time reduction is equivalent to the most reduced hydraulic model (e.g., reduced model 4). This aggregated network contains 347 nodes and 1,100 pipes, and thereby the computation time for the hydraulic simulation is reduced by 89%.

The analysis considers 20 demand zones [i.e., 20 groups of demand nodes (see 20 indexed squares in Fig. 4)], which were chosen on the basis of a spatial analysis of the system. It is expected that the

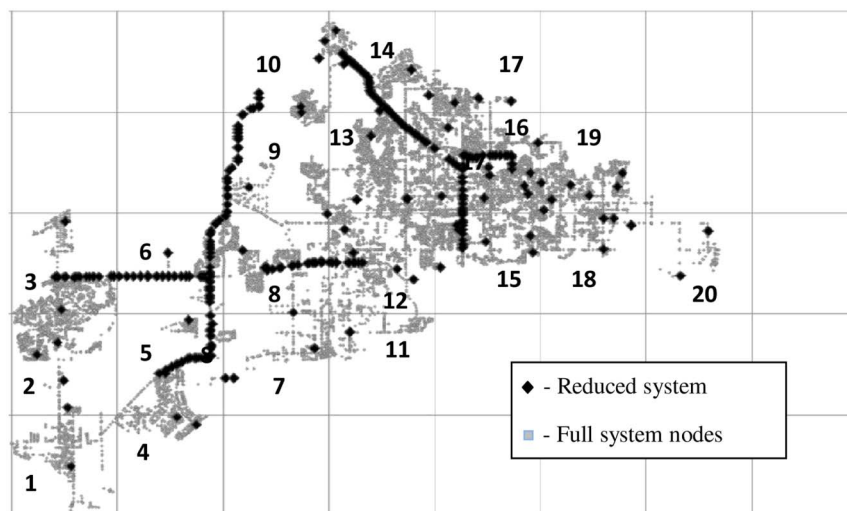
**Fig. 4.** Demand nodes groups (20 demand zones) on a plane grid of the system

Table 3. Predictive Metrics for DMFs in 20 Demand Zones for $t = 169$ h to $t = 934$ h

Demand zone	CC (Δt_1)	CC (Δt_2)	CC (Δt_3)
1	0.72	0.83	0.88
2	0.69	0.81	0.86
3	0.71	0.86	0.93
4	0.76	0.86	0.92
5	0.73	0.85	0.92
6	0.68	0.82	0.90
7	0.71	0.83	0.88
8	0.75	0.83	0.90
9	0.74	0.85	0.89
10	0.68	0.82	0.87
11	0.73	0.86	0.92
12	0.72	0.84	0.88
13	0.72	0.85	0.91
14	0.75	0.86	0.92
15	0.74	0.84	0.89
16	0.69	0.80	0.86
17	0.70	0.82	0.88
18	0.72	0.84	0.91
19	0.73	0.84	0.90
20	0.71	0.86	0.92
Average:	0.72	0.84	0.90

consumption nodes in each zone will follow the same demand pattern and each nodal base demand at each zone will be multiplied with the same DMFs.

DMFs Prediction Accuracy

The total running time of the GA calibration process (i.e., with a GA population of 120 decision variable strings and 100 GA iterations) is less than 5 min and the total running time of the data-driven prediction process is less than 5 s. The predictive ability of the model can be evaluated with several prediction metrics. In this application, the commonly used correlation coefficient (CC) was applied to evaluate the fit between predicted (p) and

actual (a) DMF values. The CC (Eq. (3)) measures the degree of correlation between predicted and actual values:

$$CC = \text{Cov}(p, a) / \sigma_p \sigma_a \quad (3)$$

where $\text{Cov}(p, a)$ = covariance between p and a ; and σ_p, σ_a = their standard deviations.

A CC equal to zero value indicates that there is no correlation between two parameters and values of 1 and -1 indicate total correlation, i.e., dependence between two parameters. A CC of 1 indicates that the maximum (or minimum) values of one parameter coincide with the maximum (or minimum) value of the other parameter; whereas a CC of -1 indicates that the maximum (or minimum) value of one parameter coincides with opposite, the minimum (or maximum) value of the other parameter.

The accuracy of the 20 zones DMF predictions starting at $t = 169$ h is summarized in Table 3. The improvement achieved in the PC model predictions through experience, is also demonstrated in the table by using three data segments of results at which the data set from $t = 169$ to $t = 934$ h was divided into three time segments ($\Delta t_1 = 169\text{--}424$ h; $\Delta t_2 = 425\text{--}679$ h; and $\Delta t_3 = 680$ to $t = 934$ h).

The relatively low CC values of Δt_1 (average of 0.72) are explained by insufficient input data for the model-trees predictor in forecasting future DMFs. For the second (Δt_2) and third (Δt_3) time periods, with the increase in training data, there is an improvement in the PC performances that is reflected in higher CCs (e.g., average $CC(t_2) = 0.84$ and average $CC(t_3) = 0.9$).

Figs. 5–7 illustrate the improvement achieved in the PC model predictions through experience, as demonstrated on one of the system demand nodes (junction 7457, which is located in zone 3). Fig. 5 shows a correlation coefficient, $CC = 0.71$ for time period Δt_1 . Figs. 6 and 7 show that with the increase in training data, there is an improvement in the PC performances that is reflected in higher CCs, $CC = 0.84$ and 0.91 , for periods Δt_2 and Δt_3 , respectively. These results highlight the importance of the prediction step in the overall process that helps to initialize the calibration procedure from a better starting point, in the search space, at each cycle of the online model, thus improving the calibration results over subsequent time-step intervals.

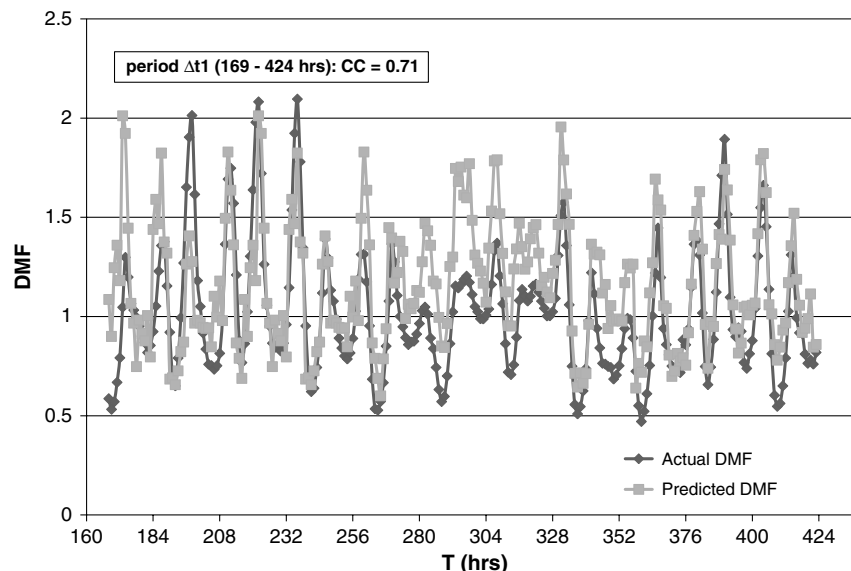


Fig. 5. Comparison between predicted and actual DMFs for time period 1

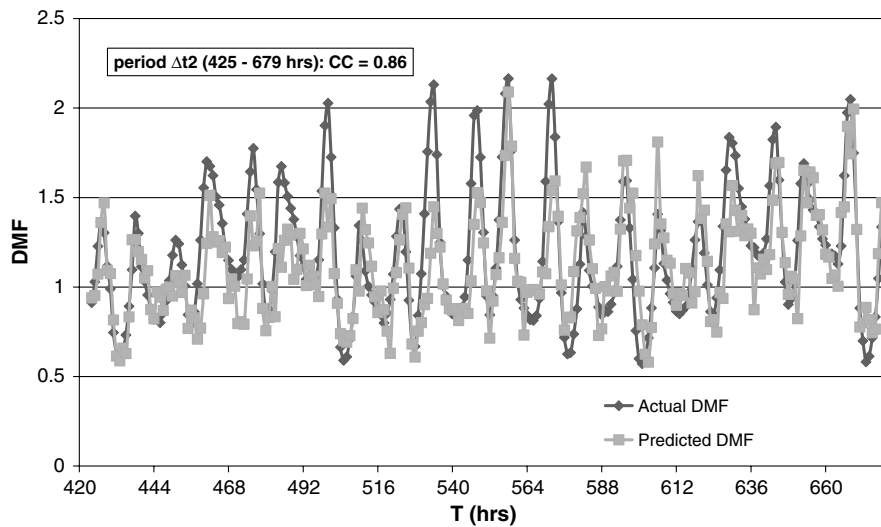


Fig. 6. Comparison between predicted and actual DMFs for time period 2

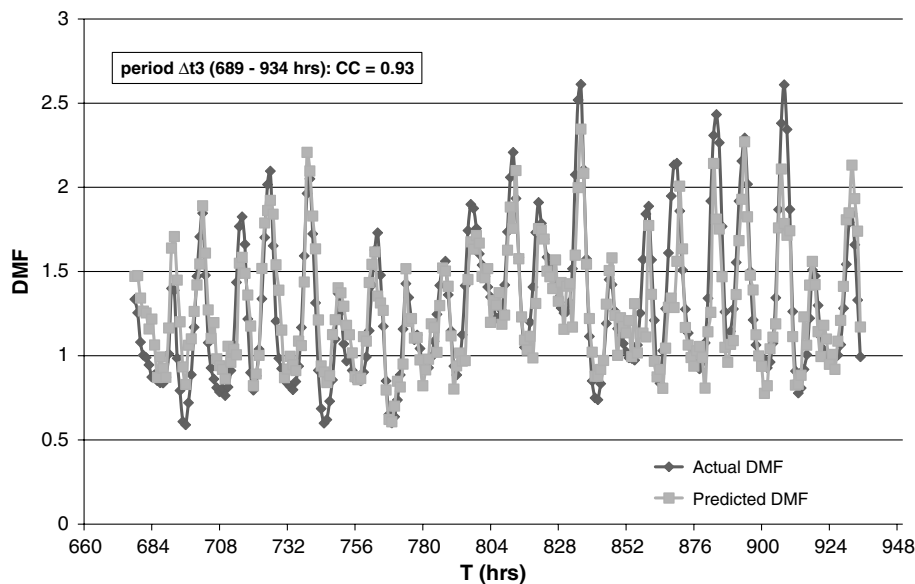


Fig. 7. Comparison between predicted and actual DMFs for time period 3

Conclusions

This study has presented and demonstrated a PC model for online, hydraulic state prediction of urban water networks. The method uses a statistical data-driven algorithm (M5 Model-Trees algorithm) to estimate future water demands, whereas near real-time field measurements are used to correct (i.e., calibrate) these predicted values online. The calibration problem is solved by using genetic algorithms with a modified least-squares fit method (Huber function) to account for noisy measurements.

The urban network model, which is comprised of over 10 thousand pipelines and nodes, was reduced by using a water system-aggregation technique thus allowing an efficient implementation of the PC procedure. As shown in Table 2, the condensed network that was chosen to resemble the full network contains only 347 nodes and 1,100 pipes instead of 12,523 nodes and 14,822 pipes—thereby the computation time for the hydraulic simulation is reduced by 89%.

The prediction step in the PC model helps initializing the evolutionary optimization search procedure from a better starting point, in the search space, at each cycle—thus improving the calibration results over time. The initial values for the calibration parameters are the predicted demand values at the previous time-steps and because the PC loop repeats itself at each subsequent time-step with the forecasting model inputs being the corrected outputs of previous iterations, the calibration model performances improve over time as shown in Table 3.

Ongoing research efforts are now focusing on the implementation of the proposed methodology for a large-scale, urban water system using physical data from a recently installed in situ sensor network. The ability to detect anomalies such as leakage and burst events in real time will be evaluated by using an analysis of residuals between predicted and observed pressures and flow rates at different locations across the system.

Additional efforts will focus on developing an automated method for demand zones selection that will incorporate graph

algorithms and heuristic search techniques to find representative clusters of consumption nodes within the complex topology of water systems. The results of this procedure will be used for the online hydraulic state estimation of the system.

Future work will also deal with testing different configurations of the PC approach at which different optimization and prediction methods will be used to replace the basic building blocks of the current procedure (e.g., GA—Levenberg Marquardt method, Cross-Entropy, Ant-Colony, or GRASP-heuristic instead of the basic GA that is currently used in the calibration process; and ANN, Gaussian Process Regression, or Support Vector Machine instead of the M5 Model-Trees algorithm that is currently used in the prediction process).

Appendix A. Hydraulic Aggregation

The Ulanicki et al. (1996) algorithm proceeds in a step-by-step elimination of pipes and nodes, allocating the demand at the node being eliminated to its neighboring nodes. All reservoirs, pumps, valves, tanks, and critical nodes (e.g., nodes in which pressure is monitored and nodes that represent significant water customers) remain in the reduced network. The validity of the system's reduction is measured by the similarity of the connectivity of the simplified system with that of the original system and its hydraulic performance (e.g., similarity of pressure at nodes, water levels at tanks, and/or pumps operation) over a wide range of operating conditions. The method of Ulanicki et al. (1996) is based on reducing the algebraic system of mass and energy conservation equations by eliminating variables by using Gauss elimination (Hammerlin and Hoffman 1991). The method involves the following stages:

Full Nonlinear Modeling of the System Hydraulics

The complete nonlinear mathematical description of the system hydraulics can be described by formulating mass conservation equation for each node (junction nodes and fixed demand nodes) of the network in which its head is unknown:

$$AQ = q \quad (4)$$

And the energy balance equation for all links of the network is

$$A^T h = \Delta h \quad (5)$$

where $A(\text{nodes, links})$ = directed incidence matrix of the network graph $G(\text{nodes, links})$ at which $A_{ij} = +1$ if the pipe p_j leaves node n_i , $A_{ij} = -1$ if the pipe p_j enters node n_i , and 0 otherwise; Q = vector of unknown flows in the links; h = vector of unknown nodal heads; q = vector of known demands at the nodes; Δh = vector of the head-losses along the links (i.e., pipes).

The relationship between the head loss and the flow in pipe, i , can be expressed with the pipes component law by using the Hazen-Williams coefficient:

$$Q_i(\Delta h_i) = g_i(D_i, CHW_i, L_i)\Delta h_i^{1/e_1} \text{sign}(\Delta h_i) \quad (6)$$

where g_i = pipe conductance (i.e., a function of the pipe diameter D_i , the Hazen-Williams head-loss coefficient CHW_i , and the pipe length L_i , with the constant $e_1 = 1.852$); $\text{sign}(h_i) = 1$ if $\Delta h_i \geq 0$; and $\text{sign}(h_i) = -1$ if $\Delta h_i < 0$

Linearization of the System's Hydraulic Model

For a given operating point defined by the nodal head h^0 and demand q^0 , the linearized approximation describes the relationships between small changes in nodal head δh and demand δq around the

chosen operating point. After linearization, Eqs. (4) and (5) take the following form:

$$G\delta h = \delta q \quad (7)$$

where $G = A[dQ(\Delta h)/d(\Delta h)]A^T$ is the symmetric Jacobian matrix whose diagonal elements are linear pipe conductances. These linear pipe conductances (i.e., \tilde{g}) can be evaluated by using Eq. (8), and $\delta\Delta h$ and δq = fluctuations in the nodal head and demand, respectively.

The elements of the Jacobian matrix are computed by using

$$\tilde{g}_{i,j} = -\frac{1}{e_1} g_{i,j} |\Delta h_{i,j}|^{(1/e_1)-1}, \quad \tilde{g}_{i,j} = \sum_j \frac{1}{e_1} g_{i,j} |\Delta h_{i,j}|^{(1/e_1)-1} \quad (8)$$

$\forall i, j$

where $\tilde{g}_{i,j}$ and $g_{i,j}$ = linear and nonlinear conductances of a pipe connecting nodes i and j , respectively; $\tilde{g}_{i,i}$ = linearized node, i , conductance (i.e., the sum of the linearized pipe conductances of the pipes that connected to the node); and $\Delta h_{i,j}$ = pipe's head loss.

The linear system of equations [Eq. (7)] describes head-demand relationship around an operating point.

Linear Model Reduction Using Gauss-Elimination Procedure

Following the linearization process, the network is then condensed by applying Hammerlin and Hoffman's Gauss-elimination process (1991) at which node, i , is removed from the network by eliminating the corresponding equation (equation, i). The demand of that node, δq_i , is redistributed among other nodes connected to node i , proportionally to the conductance of the connecting pipes. The connecting pipes of the removed node are removed as well, and new linear conductances and nodal demands are calculated for the remaining elements of the network.

Reduced Nonlinear Model Recovery from the Reduced Linear Model

At the last stage of the aggregation procedure, the reduced nonlinear model is retrieved by using the relationships formulated in Eq. (8). The aggregated model contains fewer nodes and links, forms a new network topology, and resembles the hydraulic performance of the original system with high accuracy as is shown in the "Results" section.

Acknowledgments

This work has been supported by the National Research Foundation of Singapore (NRF) and the Singapore—MIT Alliance for Research and Technology (SMART) through the Center for Environmental Modeling and Sensing.

References

- Alvisi, S., Franchini, M., and Marinelli, A. (2007). "A short-term, pattern-based model for water-demand forecasting." *J. Hydroinf.*, 9(1), 39–50.
- Andersen, J. H., and Powell, R. S. (2000). "Implicit state-estimation technique for water network monitoring." *Urban Water*, 2(2), 123–130.
- Anderson, E. J., and Al-Jamal, K. H. (1995). "Hydraulic-network simplification." *J. Water Resour. Plann. Manage.*, 121(3), 235–240.
- Clark, C., and Wu, Z. Y. (2006). "Integrated hydraulic model and genetic algorithm optimization for informed analysis of a real water system."

- 8th Annual Water Distribution Systems Analysis Symp., Cincinnati, Ohio, CD-Rom.
- Datta, R. S. N., and Sridharan, K. (1994). "Parameter estimation in water-distribution systems by least squares." *J. Water Resour. Plann. Manage.*, 120(4), 405–422.
- Davidson, J. W., and Bouchart, F. J.-C. (2006). "Adjusting nodal demands in SCADA constrained real-time water distribution network models." *J. Hydraul. Eng.*, 132(1), 102–110.
- EPANET [Computer software]. U.S. EPA, Washington, DC.
- Ghiassi, M., Zimbra, D. K., and Saidane, H. (2008). "Urban water demand forecasting with a dynamic artificial neural network model." *J. Water Resour. Plann. Manage.*, 134(2), 138–146.
- Greco, M., and Del Giudice, G. (1999). "New approach to water distribution network calibration." *J. Hydraul. Eng.*, 125(8), 849–854.
- Hamberg, D., and Shamir, U. (1988). "Schematic models for distribution systems design. I: Combination concept." *J. Water Resour. Plann. Manage.*, 114(2), 129–140.
- Hammerlin, G., and Hoffman, K. H. (1991). "Numerical mathematics." *Graduate texts in mathematics*, Springer, New York.
- Holland, J. H. (1975). *Adaptation in natural and artificial systems*, University of Michigan Press, Ann Arbor, MI.
- Huber, P. J. (1973). "Robust regression: Asymptotics, conjectures, and Monte Carlo." *Ann. Stat.*, 1(5), 799–821.
- Jonkergouw, P. M. R., Khu, S.-T., Kapelan, Z. S., and Savic, D. A. (2008). "Water quality model calibration under unknown demands." *J. Water Resour. Plann. Manage.*, 134(4), 326–336.
- Kapelan, Z. S., Savic, D. A., and Walters, G. A. (2002). "Hybrid GA for calibration of water distribution system hydraulic models." *Proc., 1st Annual Environmental & Water Resources Systems Analysis (EWRSA) Symp.*, CD-Rom.
- Kapelan, Z. S., Savic, D. A., and Walters, G. A. (2007). "Calibration of water distribution hydraulic models using a Bayesian-type procedure." *J. Hydraul. Eng.*, 133(8), 927–936.
- Kenward, T. C., and Howard, C. D. (1999). "Forecasting for urban water demand management." *Proc., 26th Annual Water Resources Planning and Management Conf.*, ASCE, Reston, VA.
- Lansley, K. E., and Basnet, C. (1991). "Parameter estimation for water distribution networks." *J. Water Resour. Plann. Manage.*, 117(1), 126–144.
- Lingireddy, S., and Ormsbee, L. (1998). "Neural networks in optimal calibration of water distribution systems." *Artificial neural networks for civil engineers: Advanced features and applications*, ASCE, Reston, VA.
- Maidment, D. R., and Miaou, S. P. (1986). "Daily water use in nine cities." *J. Water Resour. Plann. Manage.*, 110(1), 90–106.
- Ormsbee, L. E. (1989). "Implicit network calibration." *J. Water Resour. Plann. Manage.*, 115(2), 243–257.
- Ostfeld, A., et al. (2008). "The battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms." *J. Water Resour. Plann. Manage.*, 134(6), 556–568.
- Perelman, L., and Ostfeld, A. (2008). "Water distribution system aggregation for water quality analysis." *J. Water Resour. Plann. Manage.*, 134(3), 303–309.
- Preis, A., Whittle, A. J., and Ostfeld, A. (2009a). "On-line hydraulic state prediction for water distribution systems." *Proc. 11th Water Distribution Systems Analysis Symp. (WDSA09)*, World Environmental & Water Resources Congress, Kansas City, MO.
- Preis, A., Whittle, A. J., Ostfeld, A., and Perelman, L. (2009b). "On-line hydraulic state estimation in urban water networks using reduced models." *Proc. Computing & Control in the Water Industry (CCWI 2009) Integrating Water Systems*, University of Sheffield, September.
- Quinlan, J. R. (1992). "Learning with continuous classes." *Proc. 5th Australian Joint Conf. on Artificial Intelligence*, World Scientific, Singapore, 343–348.
- Reddy, P. V. N., Sridharan, K., and Rao, P. V. (1996). "WLS method for parameter estimation in water distribution networks." *J. Water Resour. Plann. Manage.*, 122(3), 157–164.
- Savic, D. A., and Walters, G. A. (1995). "Genetic algorithm techniques for calibrating network models." *Rep. No. 95/12*, Centre for Systems and Control Engineering, Univ. of Exeter, Exeter, UK, 41.
- Shamir, U., and Salomons, E. (2008). "Optimal real-time operation of urban water distribution systems using reduced models." *J. Water Resour. Plann. Manage.*, 134(2), 181–185.
- Shang, F., Uber, J., van Bloemen Waanders, B., Boccelli, D., and Janke, R. (2006). "Real time water demand estimation in water distribution system." *8th Annual Water Distribution Systems Analysis Symp.*, Cincinnati, OH, CD-Rom.
- Ulanicki, B., Zehnpfund, A., and Martinez, F. (1996). "Simplification of water distribution network models." *Proc., 2nd Int. Conf. on Hydroinformatics*, Zurich, Switzerland, 493–500.
- U.S. EPA. (2005). *Water distribution system analysis: Field studies modeling and management. A reference guide for utilities*, Cincinnati, OH.
- Walski, T. M., Chase, D. V., Savic, D. A., Grayman, W., Beckwith, S., and Koelle, E. (2003). *Advanced water distribution modeling and management*, Haestad Methods, Inc., Waterbury, CT.
- Wu, Z. Y., and Simpson, A. R. (2001). "Competent genetic algorithm optimization of water distribution systems." *J. Comput. Civ. Eng.*, 15(2), 89–101.
- Wu, Z. Y., Walski, T. M., Mankowski, R., Herrin, G., Gurrieri, R., and Tryby, M. (2002). "Calibrating water distribution models via genetic algorithms." *Proc. AWWA Information Management Technology Conf.*, Kansas City, MO.
- Zhou, S. L., McMahon, T. A., and Lewis, W. J. (2000). "Forecasting daily urban water demand: A case study of Melbourne." *J. Hydrol. (Amsterdam)*, 236(3–4), 153–164.

Queries

1. Please check that ASCE Membership Grade Marks (M.ASCE, etc.) are provided for all authors that are members.
2. There was no Andersen et al. (2001) reference in the list, so I changed this to Andersen and Powell (2000) to match the existing reference in the list. Andersen and Powell was the only one not cited at all. Is that correct or is there a reference missing?
3. Please supply a definition for ARMA.
4. What is “example 3” that you are referring to? Is it a section in the article?
5. Are the terms $t - 24$, $t - 25$, etc. saying “t minus 24 hours” in essence as we have $t + 24$ hrs (t plus 24 hours) in the next paragraph? Is the styling with an italic t OK?
6. Is the hyphen necessary with “*i*- are defined”? It isn’t in the formula, and it doesn’t appear to be a negative sign. May I remove it?
7. Were the images use in Figures 2, 3, and 4 generated by software or were the images supplied by someone? Please supply a photo credit if it is the latter case.
8. What is the reason behind citing the website where the EPANET file was downloaded? I have tagged EPANET in this sentence to be cross referenced to its listing in the references. Perhaps it might be better to create reference and place the University of Exeter information in the reference list where we can include more detail.
9. ASCE style is to use SI units. Please supply the SI equivalent for 10 in. and indicate if we should replace the existing standard units or add the SI equivalent in parentheses.
10. Please supply SI equivalent for 14 in.
11. Please supply SI equivalent for 24 in.
12. The caption for Table 1 is not quite clear - what is “...with That Calculated...” referring to? The reduced model or the pressure data. Can the caption be rephrased or shortened?
13. Please supply a definition for GA. Is it genetic algorithm? If so, may I define it a first use in the article body and replace all uses with GA for consistency?
14. Is one of the other sections supposed to be Appendix B or is this the only Appendix? If it is the only one, I would like to delete “A”.
15. Note that the numbering of the display formulas was changed to be consistent sequentially throughout the article per ASCE style.
16. Should this formula $\delta\Delta h$ be in a nearby equation? I don’t see it referenced anywhere.
17. Are there any specific grants or grant numbers tied to this project?
18. A check of online databases revealed a possible error in this reference. (Alvisi, S., Franchini, M., and Marinelli, A. [2007]) The page number 39–50 has been added. Please confirm this is correct.
19. Andersen and Powell (2000) was not cited anywhere in the article. I changed the one use of Andersen et al. (2001) to match this existing Andersen and Powell reference in the list. Is that correct or is there a reference missing? If there is a new reference for Andersen et al. (2001), please advise what should happen with Andersen and Powell (2000).
20. Please provide the sponsor/publisher of the proceedings, and their business/publishing location (not the location of the conference).
21. For the Davidson and Bouchart 2006 reference, issue 1 and page numbers 102-110 were added. Please verify that they are correct.
22. We deleted the web address for EPANET and restyled per ASCE standards.
23. A check of online databases revealed a possible error in this reference. (Ghiassi, M., Zimbra, D. K., and Saidane, H. [2008]) The page number has been changed to 138–146. Please confirm this is correct.
24. For the Huber 1973 citation, issue number 5 was added. Please confirm that it is correct.

25. For the Jonkergouw et al 2008 reference, page numbers 326–336 were added. Please confirm that they are correct.
26. Please provide the sponsor/publisher of the proceedings, and their business/publishing location (not the location of the conference).
27. There is something wrong with the Maidment and Miaou 1984 citation. The article title does not exist in the journal cited. Our databases pulled up a doi that matches the journal, volume and pages, by author Maidment et al in 1984 titled “Time Patterns of Water Use in Six Texas Cities.” But “Daily Water Use in Nine Cities” 1986 seems to be published in Water Resources Research. So it looks like two references got mixed together. Please verify the correct reference for this article (or supply both in full if appropriate). If a second reference is supplied, we will need to verify that the in-text citations are supplied/corrected as well.
28. Please verify the published proceedings title, the sponsor/publisher of the proceedings (was the University of Sheffield the host /publisher or location?), and their business/publishing location (not the location of the conference).
29. Please provide the sponsor/publisher of the proceedings (was the World Scientific the host /publisher or part of the name of the conference?), and their business/publishing location (not the location of the conference).
30. Please provide the sponsor/publisher of the proceedings, and their business/publishing location (not the location of the conference).
31. Please provide the sponsor/publisher of the proceedings, and their business/publishing location (not the location of the conference).
32. Is there a report number available for the USEPA 2005 item?
33. Please provide the sponsor/publisher of the proceedings, and their business/publishing location (not the location of the conference).
34. A check of online databases revealed a possible error in this reference. (Zhou, S. L., McMahon, T. A., and Lewis, W. J. [2000]) The issue has been changed to '3–4'. Please confirm this is correct.
35. issue