



Genetic Algorithms and Particle Filtering for Calibrating  
Water Demands and Locating Partially Closed Valves in  
Water Distribution Systems

Nhu Cuong Do

M.Sc (Civil and Environmental), B.CE

Thesis submitted in fulfilment of the requirements of the degree of

Doctor of Philosophy

Faculty of Engineering, Computer and Mathematical Sciences  
School of Civil, Environmental and Mining Engineering

August 2017



# Contents

List of Figures .....	iv
List of Tables.....	vi
Abstract .....	vii
Statement of Originality.....	ix
Acknowledgements .....	xi
List of Publications.....	xii
<b>Chapter 1 Introduction.....</b>	<b>1</b>
1.1 Research objectives.....	3
1.2 Thesis outline.....	5
<b>Chapter 2 Literature review .....</b>	<b>7</b>
2.1 Basic equations for steady state simulation of WDSs.....	8
2.2 Calibration of WDS steady state models.....	10
2.3 Selection of calibrated parameters .....	12
2.3.1 Unknown demand parameters.....	13
2.3.2 Unknown valve settings .....	17
2.4 Summary.....	19
<b>Chapter 3 Synopsis of Publications .....</b>	<b>21</b>
<b>Chapter 4 Calibration of Water Demand Multipliers in Water Distribution Systems Using Genetic Algorithm.....</b>	<b>25</b>
4.1 Introduction.....	28
4.2 GA Calibration model .....	31
4.2.1 Objective function.....	31
4.2.2 Decision variables.....	32
4.2.3 GA process and operators.....	33
4.3 Selection of measurement locations .....	34
4.3.1 Sampling design method .....	35
4.4 Case study 1 .....	37
4.4.1 Input for calibration model.....	37

4.4.2 Non-uniqueness of the solutions.....	39
4.4.3 Results of GA calibration model.....	39
4.5 Case study 2 .....	46
4.6 Case study 3 .....	48
4.7 Conclusions and recommendations .....	51
<b>Chapter 5 A particle filter-based model for online estimation of demand multipliers in water distribution systems under uncertainty.....</b>	<b>53</b>
5.1 Introduction.....	56
5.2 State estimation problem and its conceptual solution .....	60
5.3 Particle filters .....	61
5.4 Particle filters applied for water demand state estimation in WDS .....	63
5.4.1 Initialization of particles .....	63
5.4.2 Demand prediction sub-model .....	64
5.4.3 Real-time hydraulic data .....	64
5.4.4 Simulator .....	65
5.4.5 Corrector.....	65
5.4.6 Resampling.....	66
5.4.7 Demand multiplier outputs and uncertainty quantification .....	67
5.4.8 Summary of assumptions and input requirements for the DMFLive model.....	69
5.5 Case study 1 .....	70
5.5.1 Effects of tank level update on the estimation.....	74
5.6 Case study 2 .....	76
5.6.1 Improving DMFLive model performance by SRGA and modified likelihood function.....	79
5.6.2 Effect of the locations of measurements on the quantification of demand uncertainty .....	81
5.7 Conclusions and recommendations .....	84
5.8 Appendix A.....	85
<b>Chapter 6 Localization of inadvertently partially closed valves in water distribution systems.....</b>	<b>87</b>
6.1 Introduction.....	90

6.2 Modeling of partially closed valves in EPANET .....	95
6.3 Methodology .....	97
6.3.1 Observability based on sensitivity analysis.....	98
6.3.2 Application of the GA model.....	100
6.3.3 Application of the Levenberg-Marquardt algorithm .....	102
6.4 Case studies .....	103
6.4.1 Case study 1 .....	104
6.4.2 Case study 2 .....	107
6.5 Conclusions .....	115
<b>Chapter 7 Conclusions and future work .....</b>	<b>121</b>
7.1 Research contributions .....	121
7.2 Recommendations for future research.....	124
<b>Bibliography .....</b>	<b>127</b>
<b>Appendix 1 Final Published Version of Publication 1 (Chapter 4).....</b>	<b>141</b>
<b>Appendix 2 Final Published Version of Publication 2 (Chapter 5).....</b>	<b>157</b>
<b>Appendix 3 Final Published Version of WDSA2016 Paper.....</b>	<b>175</b>
<b>Appendix 4 Final Published Version of CCWI 2017 Paper.....</b>	<b>187</b>

# List of Figures

Figure 1.1 Connections between the six objectives.....	4
Figure 4.1 Example of GA chromosome and decoding for the demand calibration problem.....	33
Figure 4.2 Flowchart for the GA calibration of demand multiplier factors.....	34
Figure 4.3 Case study 1 for calibration problem .....	37
Figure 4.4 Example of non-uniqueness of solutions with 2 available measurement locations.....	39
Figure 4.5 Results of flow rates from GA calibration model for case study 1 ....	40
Figure 4.6 Results of nodal heads from GA calibration model for case study 1.	41
Figure 4.7 Results of nodal demands from GA calibration model for case study 1.....	41
Figure 4.8 Calibrated demands & estimated flows with different increment steps for case study 1.....	45
Figure 4.9 Case study 2 network for calibration problem .....	47
Figure 4.10 Case study 3 water distribution network (Net2 network) .....	48
Figure 4.11 Comparison of average calibrated demands versus actual demands of Net2 network .....	50
Figure 4.12 Comparison of average simulated flows versus actual flows of Net2 network.....	50
Figure 5.1 Process of particle filter model for real-time demand estimation in WDS.....	63
Figure 5.2 (a) Case study 1 network; (b) typical and actual demand patterns case study 1 network.....	71
Figure 5.3 Estimated demand pattern and confidence intervals: (a), (c) uncertainty quantification based on first-order approximation (FOA) method (Equation (5.18)) for $N_p = 100$ and $N_p = 20$ ; (b), (d) uncertainty quantification based on variance of the particle samples (Equation (5.16)) for $N_p = 100$ and $N_p = 20$ .....	72
Figure 5.4 (a) Prediction range of demand multipliers during simulation period; (b) Predicted demand multipliers and estimated demand multipliers.....	74
Figure 5.5 Estimated demand patterns with and without tank level updated...	75

Figure 5.6 Case study 2 network (the C-town network) .....	77
Figure 5.7 Five estimated demand patterns for case study 2 network ( $N_p = 25,000$ ) using DMFLive-I.....	78
Figure 5.8 Scattergrams and coefficients of determination for five estimated demand patterns in case study 2.....	79
Figure 5.9 DMF 1 uncertainty ranges from 25 to 49 hours computed by FOA method and posterior analysis .....	80
Figure 5.10 Sensitivity matrixes of nodal heads at measurement locations wrt demand multipliers at t=0 .....	82
Figure 5.11 Estimated DMF 4 and its confidence interval with the relocated measurement 7A and 9A .....	83
Figure 6.1 Modelling of partially closed valves in EPANET, in which simulated values are underlined and in SI unit (mH <sub>2</sub> O for nodal heads, L/s for flow rates and m/s for velocities).....	96
Figure 6.2 Process for identifying partially closed valves in WDS .....	98
Figure 6.3 (a) Case study 1 network; (b) Results from sensitivity analysis and GA model .....	104
Figure 6.4 (a) Case study 2 Network; (b) Results from sensitivity analysis ....	108
Figure 6.5 (a) Results from GA model; (b) Results from LM model at time step T=2 .....	109
Figure 6.6 Sensitivity of nodal heads and flow rates at measurement locations wrt minor loss of pipe P973 and P995 .....	112

# List of Tables

Table 4.1 Example of selection matrix S for 4 measurements in the 9 pipe network in Figure 4.3 .....	36
Table 4.2 Pipe characteristics for Case study 1.....	37
Table 4.3 Selection of the measurement locations for case study 1.....	38
Table 4.4 Comparison of mean estimated pipe flows and actual flows for case study 1.....	43
Table 4.5 Comparison of mean estimated nodal heads and actual heads for case study 1.....	43
Table 4.6 Comparison of mean calibrated nodal demands and actual demands for case study 1 .....	44
Table 4.7 Effects of measurement locations on GA calibration results .....	45
Table 4.8 Comparison of SVD model (Cheng and He 2010) and GA model for case study 2 .....	47
Table 4.9 Summary of average GA output errors for Net2 network with different increment steps.....	49
Table 5.1 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.....	71
Table 5.2 Actual and estimated total volume of water usage during calculated period.....	76
Table 5.3 Performance of DMFLive model with SR (I), modified likelihood function (II) and SRGA (III).....	80
Table 6.1 Sensitivity of flow rates and nodal heads at measurement locations wrt variation of minor losses .....	105
Table 6.2 Convergence of case study 1 network with two partially closed valves.....	106
Table 6.3 Correction of valve setting values and statuses by the LM model at time step T=2 .....	110
Table 6.4 Minor loss setting from LM method for 11 time steps from T=0 to T=10 (Case study 2) .....	113
Table 6.5 Minor loss setting from LM method for an extended period T=48 hours (Case study 2) .....	115



# Abstract

Water distribution systems 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. Conventional demand driven models simulate flows and pressures of a water distribution system requiring assumptions of known demands and known valve statuses. Due to the stochastic behavior of the water demands as well as the complexity of the piping network, these assumptions usually lead 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, calibration of the water demands and identification of valve statuses is either still not feasible or very difficult being attributable to the usual limited number of available measurement devices in most real water networks.

This dissertation addresses three main issues of water distribution modelling, which include: (1) calibration of water demands under ill-posed conditions where the number of measurements is less than the number of parameter variables, (2) estimation of water demands under uncertainty in a near real-time context, and (3) calibration and localization of unknown partially/fully closed valves in a water network. The solutions for these problems, which are the main contributions of the research, are described by three journal papers included in this dissertation.

The first journal paper presents a novel approach to calibration of the water demand multipliers under ill-posed (i.e. underdetermined) conditions by the multiple runs of a Genetic Algorithm model. The results from three case studies show that the average values of multiple runs of the Genetic Algorithm model can deliver very good estimates of the water demand multipliers, the flow rates and nodal heads at non-measured locations with a limited number of the measurements. In addition, the effects of the location and the number of measurement sites to the output of the demand calibration model are also analysed in the paper.

The second journal paper introduces a predictor-corrector approach for the online (near real-time) estimation of demand multipliers. A conventional particle filter and an improved particle filter method, which incorporates an 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. Furthermore, the paper proposes a first order approximation method to quantify the uncertainties of the model outputs caused by measurement errors. Two case studies are presented to demonstrate the effectiveness of the proposed particle filter model.

The third journal paper proposes an innovative methodology for the identification of unknown partially/fully closed valves in a water distribution network. Three sequentially applied methods are proposed in the methodology, which include: a local sensitivity analysis, an application of Genetic Algorithms and an application of the Levenberg-Marquardt algorithm. In the first method, the sensitivity of the flow rates and nodal heads at measurement locations with respect to the change in the minor losses of the valves is computed. This computation is used to identify the valves that are unable to be localized by the measurement data. The second method applies a Genetic Algorithm in an extended period simulation in order to preliminarily identify the locations of the partially/fully closed valves and their setting values, i.e. the degree of opening of the valve. Finally, the application of the Levenberg-Marquardt algorithm to a steady state simulation is implemented to correct the results from the Genetic Algorithm model.

This research has made significant contributions to the body of knowledge. Two novel methodologies have been developed for calibration of demands in water distribution systems. The impact of the location and number of measurement sites on the output of the demand calibration models has been evaluated in detail. In addition, a novel methodology for the calibration of unknown valve statuses has also been proposed. Results from realistic-size case studies have shown that the proposed methodologies are capable of solving real world problems, which enhances calibration approaches for water distribution system models.

## Statement of Originality

I, *Nhu Cuong Do*, certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

---

**Nhu Cuong Do**

August 04, 2017

This page has been intentionally left blank.

# Acknowledgements

I would like to gratefully acknowledge everyone whose support and encouragement have assisted me to complete my PhD.

First and foremost, I would like to thank my supervisors, Prof. Angus Simpson and Dr. Jochen Deuerlein who have always been very supportive and patient in discussing my research questions and reviewing my technical papers. I am indebted to Dr. Olivier Piller for his invaluable help and guidance in mathematical issues, technical areas and research direction.

I acknowledge the Australian Government for providing me the Australian Awards Scholarship to pursue my PhD and the University of Adelaide for such a great research environment. I wish to thank everyone in the School of Civil, Environmental and Mining Engineering, especially Prof. Martin Lambert, Dr. Aaron Zecchin, Dr. Angela Marchi, Dr. Giang Nguyen, Julie Ligertwood, Josie Peluso, Lisa Blinco and Hong Xuan Do for their help and advice with all matters of research and life.

Finally, I would like to thank to my parents, my sister, housemates and all friends for their love and support.

# List of Publications

The following publications were produced from the research associated with the work presented within this thesis.

## Journal papers:

[1] Do, N.C., Simpson, A.R., Deuerlein, J.W. & Piller, O. 2016, 'Calibration of Water Demand Multipliers in Water Distribution Systems Using Genetic Algorithms', *Journal of Water Resources Planning and Management*, vol. 142, no. 11, p. 04016044.

[2] Do, N.C., Simpson, A.R., Deuerlein, J.W. & Piller, O. 2017a, 'Particle filter-based model for online estimation of demand multipliers in water distribution systems under uncertainty', *Journal of Water Resources Planning and Management*, vol. 143, no.11, p. 04017065

[3] Do, N.C., Simpson, A.R., Deuerlein, J.W. & Piller, O. 2017b, 'Localization of inadvertently partially closed valves in water distribution systems ', *Journal of Water Resources Planning and Management*, (Submitted July 2017)

## Conference papers:

[1] Do, N.C., Simpson, A.R., Deuerlein, J.W. & Piller, O. 2017c 'Demand Estimation In Water Distribution Systems: Solving Underdetermined Problems Using Genetic Algorithms', *Procedia Engineering*, vol. 186, pp. 193-201, ISSN 1877-7058.

[2] Do, N.C., Simpson, A.R., Deuerlein, J.W. & Piller, O. 2017d, 'Online Demand Estimation of Geographical and Non-Geographical Distributed Demand Pattern in Water Distribution Networks', 15<sup>th</sup> Conference on Computing and Control for the Water Industry (CCWI 2017), Sheffield, UK, DOI: 10.15131/shef.data.5364133.v1.

# Chapter 1 Introduction

As an indispensable component of the urban infrastructure, a water distribution system (WDS) has to accommodate large water transfer volumes on a daily basis. From the first basic aqueducts thousands of years ago, WDSs nowadays have become a labyrinth of networks of pipes, valves, tanks, pumps and monitoring systems. The growth of cities due to rapidly increasing population as well as limited capacity of water sources have required additional linkages between water sources and demand regions. As a result, the difficulty in designing, managing and understanding the operation of the system has also increased.

The management and control of WDSs require understanding of the dynamics of hydraulics (e.g. the variations of flow rate in pipes and the pressure at customer's taps) and water quality within the distribution piping networks. These tasks can be completed by the use of hydraulic simulation models, which are built to imitate the real systems. A WDS model is an indispensable tool for providing engineering insight into the WDS, assessing water quality conditions as well as improving operational performance. In addition, the hydraulic model can also be used to simulate and evaluate extreme conditions, predict system responses to abnormal events without interrupting the actual system. Such simulation applications significantly enhance the ability of water companies to proficiently control, manage, maintain their WDSs and to provide reliable service to their customers.

A simulation model is the mathematical approximation of a real physical system. The system could be simulated with high level of confidence if perfect understanding of the system was available. However, it is not the case for most of the WDSs. In fact, a WDS is an environment of change and uncertainty. Water consumption from domestic, residential, restaurants, hospitals, parklands or industrial users is always changing and cannot be predicted. The piping system deteriorates over time as a result of aging and corrosion, resulting in changes of pipe characteristics and also increasing the frequency of

leaks or main breaks. Human interactions with the system is another factor that can also introduce uncertainty to the WDS. These factors often contribute to sometimes significant approximations of the simulation outputs to the actual physical system. As a result, a calibration analysis should be performed before the hydraulic model is used.

Calibration of a WDS model is the process of adjusting network parameters so that the output from the computer model matches the field measurements. The network parameters may include the water demands, roughness coefficients or the settings of the valves in the system, whilst the field measurements usually are the flows in pipes and pressures at some particular locations. A calibration process is essential for the following reasons: (1) improving the level of confidence for engineers and modellers in making decision as calibration shows the model's ability to mimic the present conditions of the system; (2) offering better understanding of the system. Analyses during calibration processes can evaluate which parameters are most sensitive in the model, thereby providing better knowledge of different components in the system as well as their contributions to the reaction of the system; and (3) discovering potential issues that are happening in the system. A large difference found during calibration process can be caused by an abnormal problem, which may be a pipe break, an unknown throttled or a closed valve.

The sensor technology recently applied in WDSs can assist in providing the flowrates and pressures at specific locations in real-time. In such cases, robust calibration of WDS models needs to be implemented so that the resulting model can represent important dynamic aspects of the actual system. The research implemented in this dissertation is dedicated to developing innovative methods for water demand calibration/estimation as well as innovative calibration techniques for determining the unknown valve statuses, which contribute toward enhancing the calibration approaches for WDS steady state models.



## 1.1 Research objectives

The overall aim of the research is to develop reliable and effective methodologies for calibrating WDS steady state and extended period simulation models, which includes calibration of water demand multiplier factors and calibration of valve settings. In fact, water demands are the main parameters that affect the uncertainty of the output of simulation models in the short term (e.g. hourly, daily) while a change in the valve settings can cause a huge discrepancy between the actual system and its simulation model. The presence of unknown valves/valve statuses is also one of the largest sources of uncertainty in terms of the physical properties of the system. Other properties include pipe lengths, pipe diameters and pump curves. Therefore, it is important to calibrate these parameters. Given the knowledge gaps in current research identified in Chapter 2, the aim can be divided into six objectives:

**Objective 1:** to develop a methodology for calibration of the water demand multipliers and estimating of flow rates, nodal heads under ill-posed (underdetermined) conditions where the number of measurements is less than the number of unknown demands being estimated.

**Objective 2:** to investigate the effect of measurement locations and the number of measurements on the outputs of demand calibration models.

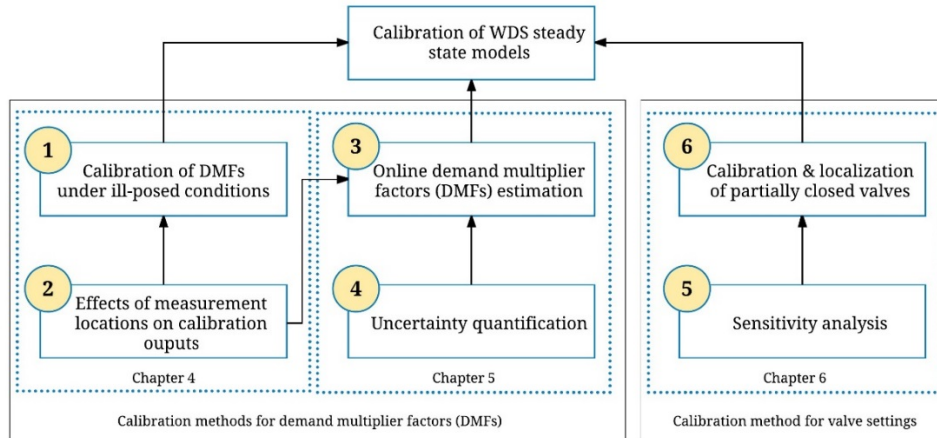
**Objective 3:** to improve estimation methodologies for the online (near real-time) estimation of the water demand multiplier factors.

**Objective 4:** to provide a methodology quantifying the uncertainty of demand estimation models caused by different sources of error.

**Objective 5:** to formulate a sensitivity method of the flow rates in pipes and pressures at nodes with respect to changes in the settings of valves to better understand the system

**Objective 6:** to develop a framework model for the calibration and localization of inadvertently partially/fully closed valves in a water network.

Figure 1.1 shows the relationships between the six objectives that contribute to the main aim of the research. The first four objectives (from Objectives 1 to 4) focus on the problem of demand calibration while the last two objectives (Objectives 5 and 6) investigate the problem of identifying partially/fully closed valves using calibration techniques.



**Figure 1.1** Connections between the six objectives

The ill-posed problem resulting from a limited number of measurements for calibration of a steady state model is investigated in Objective 1. The second objective evaluates the impacts of measurement locations and the number of measurement sites to the accuracy of the demand calibration model that was developed in Objective 1. These two objectives are presented in Chapter 4, which is also the first journal publication of the research. Objectives 3 and 4 focus on the development of an online demand estimation model of aggregated networks, in which the uncertainty of the demand outputs caused by measurement errors is computed using the proposed first-order approximation method. The effects of measurement locations to the outputs of the online model (Objective 2) are again analysed. The work implemented to achieve these objectives is shown in Chapter 5, and published in the second journal publication. Chapter 6 (i.e. the third journal publication) explores the potential issue of inadvertently partially/fully closed valves, which may occur in a WDS. This chapter consists of Objectives 5 and 6, in which Objective 5 is a necessary step for achievement of Objective 6.

## 1.2 Thesis outline

This thesis is presented as a collection of three journal publications that have been published, accepted or submitted to internationally recognized journals and is arranged in seven chapters. **Chapter 2** gives a detailed review of the relevant literature on the topic of WDS model calibration. A summary of the three publications that make up this research is presented in **Chapter 3**.

**Chapter 4** introduces the first publication (Do et al. 2016a): 'Calibration of water demand multipliers in water distribution systems using genetic algorithms', published in the *Journal of Water Resources Planning and Management*. This paper presents an approach to calibration of the demand multiplier factors (DMFs) under an ill-posed condition where the number of measurements is less than the number of parameter variables (Objective 1). The problem is solved using a Genetic Algorithm (GA). The results from three synthetic (i.e. artificial) case studies show that the average values of multiple runs of the GA model can deliver very good estimates of the water demand multipliers, the flow rates and nodal heads at non-measured locations with a limited number of measurement sites. In addition, the effects of the location and the number of the measurement sites to the output of the demand calibration model (Objective 2) are also analyzed in the paper.

**Chapter 5** presents the second publication (Do et al. 2017a): 'A particle filter - based model for online estimation of demand multipliers in water distribution systems', accepted for publication in the *Journal of Water Resources Planning and Management*. In this paper, a predictor-corrector approach for the online (near real-time) estimation of demand multipliers is developed (Objective 3). A conventional 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 a predictor-corrector approach. Furthermore, the paper proposes a first order approximation method to quantify the uncertainties of the model outputs caused by measurement errors (Objective 4). The impacts of the measurement placements to the results of the estimation model (Objective

3) are also discussed. Two synthetic case studies are presented to demonstrate the effectiveness of the proposed particle filter model.

**Chapter 6** contains the final publication (Do et al. 2017b): ‘Localization of inadvertently partially closed valves in water distribution systems’, submitted to the *Journal of Water Resources Planning and Management*. This paper introduces an innovative methodology for the identification of unknown partially/fully closed valves in a water distribution network (Objective 6). Three sequentially applied methods are proposed in the methodology, which include: a local sensitivity analysis (Objective 5), an application of the Genetic Algorithms and an application of the Levenberg-Marquardt algorithm. In the first method, the sensitivity of the flow rates and nodal heads at measurement locations with respect to the change in the minor losses of the valves is computed. This computation is used to remove the valves that are unable to be localized by the measurement data. The second method applies a genetic algorithm in an extended period simulation in order to preliminarily identify the locations of the partially/fully closed valves and their setting values, i.e. the degree of opening of the valve. Finally, the application of the Levenberg-Marquardt algorithm is implemented to correct the results from the GA model. Results and discussions from two synthetic case studies show that the proposed methodologies are capable of solving real world problems.

In the last chapter, **Chapter 7**, a summary of conclusions and important contributions is presented. This chapter also offers a number of possible future research direction.

## Chapter 2 Literature review

Hydraulic simulation models are commonly used by engineers, consultants and water utilities. The use of hydraulic models covers many application areas in the field of water distribution system engineering: design of water infrastructure, optimization of system operation, prediction and detection of abnormal events, or risk management and decision making purposes. Such hydraulic models usually fall into one of two categories, either unsteady state models or steady state models.

Unsteady state models consider the variation of the flow conditions with time (Junior & Vatavuk 2006). The term “unsteady” in WDS modelling is usually referred to a hydraulic transient (i.e. a pressure wave travels through the pipe network) that occurs due to the rapid change of a flow control device or sudden stopping of a pump in the system. If the compressibility effects of fluid and the elasticity of pipe materials are neglected, the hydraulic transients can be analyzed by a rigid water column modeling approach (also known as a slow transient model), in which the fundamental rigid model equation for each time instant of the transient period is applied (Walski et al. 2003). Examples of these models are Onizuka (1986), Shimada (1989), Ahmed (1997) and Nault and Karney (2016). On the other hand, hydraulic transients can be studied by an elastic model if the pipelines are considered to be deformable and the liquid is compressible. According to Walski et al. (2003), the main objectives of hydraulic transient analysis are: (1) to determine the magnitude of transient pressures that can result from flow control operations (e.g. Pinder and Cooper (1970), Liggett and Chen (1994) and Wood et al. (2005)) and (2) to formulate the design criteria for water system devices, so as to provide a tolerable level of protection against system failure (e.g. Boulos et al. (2005) and Soares et al. (2008)). In addition, hydraulic transient analysis is also applied to the problem of leak detection, for example: Vitkovsky et al. (2000), Kapelan et al. (2003) and Vitkovsky et al. (2007)). Because of the difficulty and long computational time of hydraulic transient evaluations, unsteady state

models are often applied to relatively small to medium sized distribution systems. These approaches remain computationally impractical for the analysis of very large systems (Filion and Karney 2002).

Steady state models consider only the steady condition of the flows. These models can simulate a water network at a specific point of time (steady state simulation) or over a period of time (extended period simulation, or a series of steady state simulations) if dynamic analysis of the network is required. Steady state and extended period analysis is fundamental for the design, operation and maintenance of complex WDSs. The ease and speed associated with this type of model gives an engineer the ability to examine many solutions under a wide range of conditions, which can bring more cost-effective and robust designs (Walski et al. 2003). In addition, steady state solutions are also required as an initial point for solving the unsteady state equations. According to Boulos et al. (2006), the relation between analyzing a WDS under steady state and transient conditions is interdependent. The transient analysis is built based on the analysis of steady state conditions and the transient problem can only be solved if the initial hydraulic grade line elevations and flow velocities are known. Finally, Sanz (2015) has pointed out that for a complex WDS, the network tends to be less variable in terms of pressure variation as the transient events are dissipated quickly in space. Accurate estimates of variables such as pressures and flows of the network can be obtained if the steady state model is well calibrated. For these reasons, the application of steady states models in WDS modeling is necessary given the noisy nature and unstable conditions of real systems.

Research in this thesis focuses on the calibration of steady state models. The following section describes the formulation of the steady state hydraulic simulation in WDS modeling.

## **2.1 Basic equations for steady state simulation of WDSs**

Hydraulic models are governed by two principles: the conservation of mass and the conservation of energy. Under the steady state condition of the network modelling, the

conservation of mass states that the fluid mass entering any junction must be equal to the mass leaving the junction, i.e. the algebraic sum of the flows at nodes or junctions is equal to zero:

$$\sum_{j=1}^{N_P} a_{ij} Q_j + DM_i = 0 \quad (i = 1, \dots, N_J) \quad (2.1)$$

where  $N_P$  is the number of links in the network.  $Q_j$  is the flow in link  $j$ .  $DM_i$  is the demand at node  $i$ .  $N_J$  is the total number of nodes or junctions.  $a_{ij}$  is an element that provides information about the connectivity of the nodes and the links in the network.  $a_{ij}$  is assigned a value of -1 or 1 if the link enters or leaves the node, respectively. Otherwise,  $a_{ij}$  is assigned a value of 0.

The conservation of energy of a network model states that energy must be conserved between any two junctions or nodes, which is presented by the head loss equation in each link:

$$H_i - H_k - r_j Q_j |Q_j|^{n-1} = 0 \quad (j = 1, \dots, N_P) \quad (2.2)$$

where  $H_i, H_k$  are the nodal heads at two ends of the link (i.e. at node  $i$  and node  $k$ ).  $r_j$  is the resistance factor for link  $j$ .  $Q_j$  is the flow in link  $j$  and  $n$  is the exponent of the head loss equation, whose value is 2 for the use of the Darcy-Weisbach equation or 1.852 for the use of the Hazen-Williams equation.

Equation (2.1) and (2.2) are the elementary equations for the steady state modeling of a WDS. When dealing with real systems, these equations must be altered in order to account for local losses, pumps and/or different types of valves, etc. (Todini and Rossman 2012). For example, if the minor losses of pipes/valves are considered (as in Chapter 6 of the thesis), Equation (2.2) will be modified as:

$$H_i - H_k - r_j Q_j |Q_j|^{n-1} - m_j Q_j^2 = 0 \quad (2.3)$$

where  $m_j$  is the minor loss coefficient for link  $j$ .

Various approaches have been developed for solving this set of  $(N_J + N_P)$  equations (i.e. Equations (2.1) and (2.2)), for example the Hardy Cross method (Cross 1936), the linear

theory method (Wood & Charles 1972), the Newton Raphson method and the global gradient algorithm (Todini & Pilati 1988). These algorithms simulate the flows distributed in pipes and pressures at nodes of a network based on underlying assumptions that the nodal demands, network characteristics (e.g. pipe roughnesses, diameters and lengths, etc.) and other boundary conditions (such as pump status, valve status) are known inputs. The outputs from steady state hydraulic models, therefore, represent the water distribution system behavior during the period at which these assumptions are verified (Preis et al. 2009).

In fact, a WDS is an environment of change and uncertainty. Water consumptions from domestic, residential, restaurants, hospitals, parklands or industrial users always change and cannot be predicted. The piping system deteriorates over time as a result of aging and corrosion, resulting in changes of pipe characteristics and also increasing the frequency of leaks or main breaks. Human interactions with the system is another factor that can also introduce uncertainty to the WDS. If a steady state model simulates the WDS based on the above assumptions, it may lead to a large approximation between the simulation outputs and the actual physical system.

Sensor technology that has recently been applied in WDSs can assist in providing the flowrates and pressures at specific locations in real-time (e.g. Shihu (2011) and Whittle et al. (2013)). In such cases, robust calibration of WDS models needs to be implemented so that the resulting model can represent important dynamic aspects of the actual system.

## **2.2 Calibration of WDS steady state models**

Calibration of a WDS model is the process of adjusting network parameters so that the output from the computer model matches the field measurements, which are usually the pressures and flow rates at particular locations in the network (Shamir & Howard 1977). The calibration procedure can generally be implemented by seven steps suggested by Ormsbee (1989): (1) Identifying the intended use of the model, (2) Determining estimates of the model parameters, (3) Collecting calibration data, (4) Evaluating the results of the



model, (5) Performing a macro level calibration of the model, (6) Performing a sensitivity analysis and (7) Performing a micro level calibration of the model.

Mathematically, the calibration problem can be formulated as a nonlinear problem:

$$z = h(x) + e \quad (2.4)$$

where  $z$  is the field measurement vector (i.e. pipe flow rates and nodal pressures).  $x$  is the unknown input parameter that is required to be calibrated/estimated (e.g. pipe roughness coefficients, nodal demands, etc.).  $h()$  is the nonlinear vector function relating the field measurements to the unknown parameters, (i.e. Equation (2.1) and (2.2)).  $e$  is the measurement vector error that is usually assumed to be normally distributed with zero mean.

Approaches for this calibration problem can be classified into three categories: ad hoc (trial and error) calibration, explicit calibration and implicit calibration models, which are comprehensively reviewed by Savic et al. (2009). Initial studies of the ad hoc calibration schemes were pioneered by Rahal et al. (1980), Walski (1983) and Bhave (1988), in which an iterative process to update unknown model parameters was implemented. Due to the slow convergence rate, this method was only assessed as being applicable to small water networks.

Explicit calibration methods were introduced and employed by Ormsbee and Wood (1986), Boulos and Wood (1990) and Boulos and Ormsbee (1991). These methods solved an even-determined set of water network equations where the number of unknown parameters was grouped to be equal to the number of measurements. As the measurement errors were also neglected, these methods usually did not adequately consider real-world practical outputs. Therefore, explicit calibration models were often used to analyze historical events in water systems (Savic et al. 2009).

The third type of calibration techniques is implicit calibration. This method considers the calibration problem as an optimization process and solves it using a hydraulic solver coupled with an optimization technique. An objective function is usually associated with

these optimization models and often formulated as a least absolute value (LAV) criterion, a least square (LS) criterion or a root mean square error (RMSE) criterion. Numerous optimization techniques have been applied for the implicit calibration (i.e. optimization) problem, from mathematical methods (e.g. modified Newton Raphson method (Shamir 1974), the Simplex method (Sterling & Bargiela 1984), Complex method (Ormsbee 1989), the generalized reduced gradient method (Lansley & Basnet 1991), sensitivity analysis using a Taylor series approach (Datta & Sridharan 1994), the Levenberg-Marquardt algorithm (Koppel & Vassiljev 2012) and the singular value decomposition (SVD) method (Cheng & He 2010; Sanz & Perez 2015)) to evolutionary optimization methods (e.g. Genetic Algorithms (Meier & Barkdoll 2000; Preis et al. 2009; Abe & Peter 2010)). More recently, some probabilistic methods have been introduced (e.g. Kalman filter method (Shang et al. 2006), the shuffled complex evolution metropolis algorithm (Kapelan et al. 2007) and the particle filter method (Hutton et al. 2012b)). These models were built in an attempt to determine the uncertainty of the calibrated parameters caused by measurement errors.

### **2.3 Selection of calibrated parameters**

There are many factors that contribute to the uncertainty of a WDS model. Pipe roughness coefficients, the water demands at nodes and control valve settings are of great interest for researchers in steady state calibration given their high impact on network uncertainty. However, roughness coefficients usually vary only in the long term due to pipe deterioration, for example the annual variation of these parameters was considered in Haddad et al. (2008) and Seifollahi-Aghmiuni et al. (2013). On the other hand, water demands are the main parameters that affect the uncertainty of the output of models in the shorter term (e.g. hourly, daily) while a change in the valve settings can cause a huge discrepancy between the actual system and its simulation model.

Because the detail of pipe roughness calibration can be found elsewhere (e.g. Ormsbee and Lingireddy (1997), Bush and Uber (1998) and (Wu et al. 2002)), this topic has been

excluded from the research presented in this thesis. The following discussion focuses on the calibration of the water demands and valve settings in WDS hydraulic models.

### 2.3.1 Unknown demand parameters

The hydraulic dynamics that happen in WDSs are driven by the consumption of water, also known as the water demand (Walski et al. 2003). This “driving force” should ideally be observed via smart demand metering system or sensor measurements. However, because of budgetary and physical constraints, the observation of the water demand is difficult to achieve. Nodal demands are therefore usually selected as the time varying parameters to be calibrated.

Modelling of water demands in a WDS model commonly uses the concept of baseline demand and a demand pattern (Shang et al. 2006). At each time step, the water demand at each node can be calculated by the multiplication of a base demand with its corresponding demand multiplier factor (DMF):

$$D_{k,t} = D_{0,k} * f_{k,t} \quad (2.5)$$

where  $D_{0,k}$  is base demand at the  $k_{th}$  node and  $f_{k,t}$  is the DMF at the  $k_{th}$  node at time step  $t$ .

The base demand can be calculated using quarter/annual water usage billing information. Thus, the calibration of water demands usually focuses on calibrating the demand multiplier factors. Examples of DMF calibration can be found in Shang et al. (2006), Kang and Lansey (2009), Preis et al. (2009) and Hutton et al. (2012b). 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. (2012b) introduced a particle filter method and an ensemble Kalman filter for the estimation of the demand of a single district meter area, which was assumed to follow a linear time series model. In Kang and Lansey (2009), two comprehensive methods 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 respect to the change in the state vector). In Preis et al. (2009), the DMFs were calibrated by the application of a genetic algorithm model, in which the measurement errors were ignored.

It should be noted that the previously mentioned models have been developed for an overdetermined problem where the DMFs parameters are grouped to be less than the number of measurements. The outcomes, therefore, rely on this assumption, which can lead to large approximations in real water distribution systems. Limited work has investigated the underdetermined systems where the number of unknown DMFs parameters are more than the number of measurements. From the theoretical point of view, the underdetermined problem should not be formulated nor solved due to the presence of non-unique solutions. Ideally, it should be formulated to be well-posed by reducing the number of calibration parameters and/or obtaining additional observed information. However, this issue is a reality in engineering practice and need to be solved if simplification of the problem cannot be implemented. Previous research includes the proposal of a proportional demand method (Davidson & Bouchart 2006) and the application of a singular value decomposition (SVD) method (Cheng & He 2010; Kun et al. 2015; Sanz & Pérez 2015). The application of mathematical methods (e.g. local linearization methods such as QR decomposition, SVD or using the Moore-Penrose pseudoinverse matrix in the Newton-Raphson method) can find a local solution of the problem. However, due to the possibility of non-uniqueness of the solutions, the results from these mathematical methods may be far from the actual solution. Providing a reliable method for the underdetermined problem of demand calibration is a challenging task for hydraulic researchers.

There are several other issues that need to be addressed for the problem of demand calibration. First of all, the accuracy of demand calibration models not only relies on the

use of calibration methods (mathematical, optimization or probabilistic methods) but also depends on the locations of the measurements and the number of measurement sites in the network. In calibration models, the measurement location problem has been investigated using sampling design (SD) methodologies. Piller (1995) used a SD method to minimize the influence of measurement errors in the state vector estimation. Bush and Uber (1998) developed three SD methods derived from D-optimality criteria: max-sum, weighted sum and max-min methods in order to select measurement locations based on the analysis of the Jacobian matrix. Meier and Barkdoll (2000) used a GA for the optimal SD problem with the aim of finding a set of measurement locations which maximizes the presence of non-negligible pipe velocities. De Schaetzen (2000) proposed three SD approaches for the optimal measurement locations based on shortest path algorithm, rank measurement locations and maximization of Shannon's entropy. Kapelan et al. (2005) developed two SD models using a GA to find the optimal set of pressure locations, where the first model is formulated as a single objective GA and the second is modelled as a multi-objective optimization problem. On the other hand, no research has been found that evaluates the impact of the number of measurement sites on the accuracy of the outputs from a calibration model. The question of "how many measurements are required to have a good calibration of the water demands" is, therefore, a research question that needs to be answered.

The second aspect of this research is to select a suitable technique for real-time (or near real-time) demand calibration/estimation. In a real-time context, the measurement data can be recorded in a relatively short time period (e.g. every few minutes). In this case, the temporal variation of consumer demands can only be obtained by fast and robust calibration models. Few attempts have focused on calibrating the nodal demands in real-time (e.g. Shang et al. (2006), Kang and Lansey (2009), Hutton et al. (2012b) and Okeya et al. (2014)). However, these models have either been tested on relatively simple case study (e.g. in Hutton et al. (2012b), particle filter model was applied to estimate a single DMF) or required large number of flow measurements to achieve good results (e.g. in

Shang et al. (2006), the model was tested on a 59-node network with 40 flow measurements, while in Kang and Lansey (2009) the model was tested on a 90-node network with 19 flow measurements and 5 pressure measurements). It has been noted that the cost of flow measurement devices is expensive and the installation of these devices is more difficult compared to pressure measurement devices. As a result, flow measurement devices are usually not as commonly used as pressure measurement devices in real world situations. Consequently, given all of these issues the development of a fast and robust demand calibration model that can be applied for real and complex WDS networks is still needed.

The final issue is related to the uncertainty quantification methods. According to Hutton et al. (2012a), the uncertainty in WDS modeling can be divided into three categories: (1) structural uncertainty, (2) parameter uncertainty and (3) measurement uncertainty. Structural uncertainty derives from the mathematical representation of the real system, such as network skeletonization and model aggregation. The second category, parameter uncertainty, refers to the errors of the parameters used to represent system components. 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 water demands. If the demand parameter is selected to be calibrated, the uncertainty of the demands caused by other sources of uncertainty (i.e structural uncertainty, uncertainty from other parameters and measurement uncertainty) needs to be quantified. In the literature, only a few uncertainty quantification methods can be found. Three methods have been presented in Bargiela and Hainsworth (1989): Monte Carlo simulation (uncertainty is calculated by a series of simulations), an optimization-based approach (uncertainty is computed by means of an optimization problem) and a sensitivity-based method (uncertainty is defined by an analysis of the sensitivity matrix). The most common method is the first-order second-moment method (FOSM), which computes the uncertainty based on linear regression theory as can be seen in Reddy et al. (1996), Bush and Uber (1998), Ahmed et al.(1999), Lansey et al. (2001),

Kapelan et al. (2003b, 2005), Behzadian et al. (2009) and Kang and Lansey (2009). Despite these efforts, there is a need to find a reliable method that can effectively quantify the uncertainty of the calibrated distributed demand caused by different sources of uncertainty.

### **2.3.2 Unknown valve settings**

The use of valves within distribution piping networks is aimed at improving the reliability in operation as well as maintenance of the systems. Two types of valves are usually considered in WDS modeling: control valves and isolation valves.

Control valves are automated devices that are used to regulate the flows or pressures of water throughout the distribution piping system. The hydraulic behavior in local regions of the network where these valves are placed can be adjusted by the valve settings. For example, a flow control valve enables the flow passing through the valve to be less than or equal to a specific flow setting value, or a pressure reducing valve (PRV) is set to prevent the downstream pressure from exceeding a value that could cause damage to the system (Walski et al. 2003). Because of their important role in the operation of WDSs, calibration of control valve settings is usually considered in WDS modeling (e.g. Piller and Bremond (2001), Deuerlein et al. (2005) and Piller and van Zyl (2014)). Piller and Bremond (2001) formulated a least squares method that minimizes the difference between targeted settings and computed values to find the control valve states. Deuerlein et al. (2005) applied a gradient-based algorithm to estimate the correct setting of pressure control valves. Piller and van Zyl (2014) introduced a method associated with the Karush-Kuhn-Tucker equations for the modeling of control valves in extended period simulations.

Isolation valves, the most common type of valves in WDSs, are used to close off and block any flow through pipes (Van Zyl 2014). These valves are usually placed at the ends of a pipe, around junctions or at critical locations of a WDS. The primary role of isolation valves is to isolate some portions of the system for inspection, replacement or maintenance. Therefore, research on this type of valves has mainly focused on the design of the isolation valve system to ensure the connectivity of the network when some of the isolation valves

are required to be closed. Various studies have formulated the valve placement problem as an optimization problem (including single objective optimization and multi-objectives optimization) and solved it with different techniques such as Simulated Annealing (Ozger and May 2004), Genetic Algorithms (Giustolisi and Savic 2010) or NSGAI (Creaco et al. 2010). In Jun and Loganathan (2007), a mathematical method was introduced, which represents the presence of isolation valves by a valve location matrix and a valve deficiency matrix. This method was used to evaluate the connectivity of the network as well as to detect unintended isolations when some of the valves in the system were closed. Some other studies also looked at the connectivity of the WDS such as Davidson et al. (2005) and Ostfeld (2005).

It should be noted that most of the aforementioned studies focused on the fully closed status of the valve. Given the thousands of isolation valves in a network and despite the fact that these valves are suggested to be inspected and maintained at least once a year (van Zyl 2014), there is always a possibility that one or some of the valves are closed or partially closed. Possible problems related to these valves come from a number of reasons such as: missing valves due to poor or not existing documentations, errors in data transfer, valve mechanical failure or temporary closed during inspection/rehabilitation time without adequately being reported. This issue might cause a huge disagreement of the hydraulic behavior between the real system and its simulation model. Little effort has investigated the problem of calibrating and identifying unknown valve settings and their corresponding locations. Previous research includes Delgado and Lansey (2008) and Wu et al (2012). In the Delgado and Lansey (2008) paper, a transient analysis method was used to detect closed/partially closed valve in a single pipeline, while Wu et al (2012) applied Genetic Algorithms to find unknown partially closed valve locations and to estimate their settings. The problem of calibrating fully/partially closed valve settings is, therefore, still a research gap that needs to be filled.



## 2.4 Summary

In this chapter, the state of the art in calibration of WDS steady state models has been reviewed. Generally, calibration has been previously implemented by trial and error, explicit or implicit methods. Implicit methods, which consider the calibration problem as an optimization process and solve it using a hydraulic solver coupled with an optimization technique, have gained a prominence and have drawn a high degree of attention from researchers. Nodal demands are usually selected as the time varying parameters to be calibrated because these parameters have high impact on model uncertainty during short periods of time (or in real-time). In addition, the valve setting parameters also need to be considered because they can cause a huge disagreement between the actual system and its simulation model. Although previous work has covered almost all aspects of calibration problem, there are some knowledge gaps that need to be addressed, which include: (1) the requirement for a reliable method to calibrating water demand under ill-posed conditions; (2) a comprehensive evaluation on the impacts of location and the number of the measurement sites to the outputs of a demand calibration model; (3) the necessity of a fast and robust demand calibration model that can be applied for real and complex WDS networks; (4) the need of a method that can be applied to effectively quantify the uncertainty of the calibrated demand caused by different sources of uncertainty and (5) the lack of a methodology to calibrate fully/partially closed valve settings and identify their locations.

These knowledge gaps have been translated into six research objectives as presented in Section 1.1. Solutions for each research gap and more detailed review of the literature can be found in Chapters 4, 5, and 6, which are the main contributions of this thesis to the body of knowledge.

This page has been intentionally left blank

## Chapter 3 Synopsis of Publications

This chapter gives a summary of three publications presented in this thesis and how they address the objectives of the work.

**Publication 1** proposes a methodology, which is the application of multiple runs of a GA model, for the calibration of the demand multiplier factors under ill-posed conditions in which the number of measurement locations is less than the number of unknown parameters (Objective 1). The methodology uses the EPANET toolkit to solve the system of water network equations while GAs are applied to find the best match between known measurement inputs and their simulated values (i.e. the simulated flow rates and nodal pressures at measurement locations). The average output values of multiple GA runs are proposed to use as the estimates of the demands in a system. Different scenarios of measurement availability in water networks are tested to evaluate the reliability of the model. Furthermore, this study also investigates the use of a sampling design technique for the selection of optimal measurement locations in order to improve the quality of the calibration model (Objective 2). The approach for estimation of demand multiplier factors using GA optimization has been tested on three synthetic case studies. The first case study shows that the average values of multiple runs of the GA model can deliver a very good approximation of the water demand multipliers with little information from the SCADA system. The first case study also shows that the location of the measurement sites does influence the performance of the GA model. The second case study demonstrates the advantage of the GA model in comparison to the singular value decomposition model that has been developed for the same calibration problem. It also confirms the conclusions made from the first case study about the sensitivity of the GA model to the measurement locations. Therefore, the GA model is suggested to be implemented in combination with a supporting tool for the selection of optimal measurement locations such as the sampling design (SD) greedy algorithm model. The third case study validates the approach for a

slightly larger sized network, which again exhibits the superior performance of the GA model.

**Publication 2** presents a particle filter model for the online (near real-time) demand estimation of a WDS, which is named the DMFLive model (Objective 3). A predictor-corrector methodology is adopted in the DMFLive model to predict the hydraulic behavior 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: a systematic resampling (SR) algorithm and a systematic resampling method integrated with a GA process (SRGA). Uncertainties of model outputs are quantified and evaluated in terms of confidence intervals by the first-order approximation formula (Objective 4). 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 problems. Within the first case study, the benefits of having additional information about the tank level at the next time step have been explored. If the estimation of the demand multipliers can be delayed by 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 have been developed to be used for 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 (Objective 2). Results showed that the uncertainty can be used to identify which measurement locations need to be improved.

**Publication 3** develops a methodology to solve the problem of calibrating and localizing partially closed valves in a water distribution network based on integration of

measurement data and the demand driven model EPANET (Objective 6). Three sequentially applied methods have been proposed in the methodology, which includes: (1) a sensitivity analysis, which calculates the sensitivity of flow rates and pressure heads at measurement locations with respect to (wrt) the variations of the minor loss in the valves (Objective 5). These sensitivity values were used to identify the valves that are insensitive to the measurement locations; (2) an application of a GA model during an extended period simulation (e.g. 24, 48 hours) to reduce the size of the search region as well as provide a preliminary estimate of the settings of the valves; and (3) an application of the Levenberg-Marquardt algorithm to localize the regions of partially closed valves and correct the settings of the valves. The performance of the methodology has been evaluated for two case studies. The first case study has shown that the exact solution can be found, in terms of both statuses and minor loss setting values, if the necessary conditions for the Levenberg-Marquardt are satisfied. The second case study showed the applicability of the proposed methodology when applied to a realistically sized problem. Although the valve problem of the second case study contains non-unique solutions, the methodology still can identify very well the locations of partially closed valves in the network. The proposed methodology is, therefore, an efficient tool for dealing with the problem of finding unknown partially/ fully closed valves in water distribution systems

This page has been intentionally left blank.

# Chapter 4 Calibration of Water Demand Multipliers in Water Distribution Systems Using Genetic Algorithms

## Publication 1

Nhu C. Do<sup>1</sup>, Angus R. Simpson<sup>1</sup>, Jochen W. Deuerlein<sup>2</sup>, Olivier Piller<sup>3</sup>

<sup>1</sup> - *School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide SA 5005, Australia.*

<sup>2</sup> - *Senior Researcher, 3S Consult GmbH, Karlsruhe, Germany.*

<sup>3</sup> - *Senior Researcher, Irstea UR ETBX, Dept. of Water, Cestas, France.*

*Journal of Water Resources Planning and Management*, 142(11)

[doi:10.1061/\(ASCE\)WR.1943-5452.0000691](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000691)

See Appendix 1 for a copy of the final published paper.

This page has been intentionally left blank.



# Statement of Authorship

Title of Paper	<b>Calibration of water demand multipliers in water distribution systems using Genetic Algorithms</b>
Publication Status	<input checked="" type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style
Publication Details	Do, N.C., Simpson, A.R., Deuerlein, J.W. & Piller, O. 2016, 'Calibration of Water Demand Multipliers in Water Distribution Systems Using Genetic Algorithms', <i>Journal of Water Resources Planning and Management</i> , vol. 142, no. 11, p. 04016044.

Although the manuscript has been reformatted in accordance with University guidelines, sections have been renumbered for inclusion within this thesis, and details of references have been updated where applicable, the paper is otherwise presented herein as published (with permission from ASCE)

## Principal Author

Name of Principal Author (Candidate)	Do, N.C.		
Contribution to the Paper	Development and implementation of methodology, design of case studies, interpretation and analysis of results, preparation of manuscript and acting as corresponding author.		
Overall percentage (%)	70%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	3 August 2017

## Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Simpson, A.R.		
Contribution to the Paper	Research supervision and manuscript evaluation		
Signature		Date	4 August 2017

Name of Co-Author	Deuerlein, J.W.		
Contribution to the Paper	Research supervision and manuscript evaluation		
Signature		Date	11 July 2017

Name of Co-Author	Piller, O.		
Contribution to the Paper	Research supervision and manuscript evaluation		
Signature		Date	11 July 2017



## Abstract

Hydraulic models have been widely used for design, analysis and operation of water distribution systems. As with all hydraulic models, water demands are one of the main parameters that cause the most uncertainty to the model outputs. However, the calibration of the water demands is usually not feasible due to the limited quantity of available measurements in most real water networks. This paper presents an approach to calibration of the demand multiplier factors under an ill-posed condition where the number of measurements is less than the number of parameter variables. The problem is solved using a Genetic Algorithm (GA). The results show that not only is the GA able to match the calibrated values at measured locations, but by using multiple runs of the GA model, the flow rates and nodal heads at non-measured locations can be estimated. Three case studies are presented as an illustration of the problem. The first case study is a small network that demonstrates the calibration model. The second case study shows a comparison between the genetic algorithm model and a singular value decomposition model. The last case study is a large network that allows for practical considerations in applying the proposed methodology to a realistic context.

**Keywords:** *Genetic algorithms, optimization, demand calibration, water distribution systems*

## 4.1 Introduction

As an indispensable component of urban infrastructure, a water distribution system (WDS) has to accommodate large water transfer volumes on a daily basis. The problem of ensuring a satisfactory and reliable service is further complicated by population growth which often leads to the need for augmentation of the system. Determination of flow rates and pressure heads in the existing system is a necessary step in the modification of a WDS. This task can be accomplished by using measurement devices, such as sensors, however, sensors can only capture the status of some component locations in the system. Calibration of the full WDS model using limited measurements from these devices is therefore a research area that requires further development.

Calibration of a WDS model is the process of adjusting network parameters so that the output from the computer model matches the field measurements, which are usually the pressures and flow rates at particular locations in the network (Shamir and Howard 1977). The calibration procedure for a water network model has been well addressed by Ormsbee and Lingireddy (1997). In their paper, the authors suggested a seven-step calibration process, which includes: (1) Identifying the intended use of the model, (2) Determining estimates of the model parameters, (3) Collecting calibration data, (4) Evaluation of the results of the model, (5) Performing a macro level calibration of the model, (6) Performing a sensitivity analysis and (7) Performing a micro level calibration of the model.

Savic et al. (2009) presented a comprehensive literature review on the calibration of water network models where the calibration methods can be classified by their dynamic (transient and static), by their calculation methods (iterative, explicit and implicit) or by the use of optimization methods (traditional and evolutionary).

In terms of calibration in transient analysis, the calibration models have been constructed mainly to detect leakage in distribution systems (Pudar and Liggett 1992, Liggett and Chen 1994, Vítkovský et al. 2000, Kapelan et al. 2003). These transient calibration procedures usually consider a pure leak or leaks combined with unknown nodal demands

and pipe friction factors as parameters to be calibrated. However, due to the complexity of the method, such as generating and measuring reflections of water hammer waves, transient analysis has not been widely used in practice for practical leak detection applications.

In static hydraulic analysis, an iterative procedure was applied in the early use of calibration models (Walski 1983, 1986, Bhave 1988). This procedure was implemented to update the unknown model parameters using heads/flows obtained by solving the water network equations. Due to slow convergence rate, these models are only suitable for small problems or require simplification of the water network.

The second technique is explicit calibration, which was employed in Ormsbee and Wood (1986), Boulos and Wood (1990), and Boulos and Ormsbee (1991). This technique involves solving an extended set of continuity and head-loss equations where the calibration problem is required to be even-determined (i.e. the number of calibrated parameters must be equal to the number of measurements). Measurement errors are also neglected. When the number of unknown parameters is larger than the number of measurements, calibration parameters are often grouped that may result in potentially impractical outputs. Therefore, explicit calibration models are often used for the purpose of system analysis.

Implicit calibration is the third type of static calibration model. This method considers the calibration problem as an optimization process and solves it using a hydraulic solver combined with either a traditional or an evolutionary optimization technique. The implicit method has been investigated by a majority of the previous research (e.g. Ormsbee (1989), Lansley and Basnet (1991), Datta and Sridharan (1994), Ormsbee and Lingireddy (1997), Andersen and Powell (2000), Nagar and Powell (2002), Shang et al. (2006), Preis et al. (2009), Kang and Lansley (2009), Piller et al. (2011) and Hutton et al. (2012)).

The uncertainty of results from water network models is caused by many influences, such as pipe roughness, nodal demands or valve states (e.g. Martínez et al. (2003), Liberatore and Sechi (2009)). Pipe roughness coefficients and the water demands at nodes are

normally used for the static calibration given their high impact on network uncertainty. However, roughness coefficients usually vary only in the long term, for example the annual variation of these parameters was considered in Haddad et al. (2008) and Seifollahi-Aghmiuni et al. (2013). Therefore, water demands are the main parameters that affect the uncertainty of the output of models in the shorter term (e.g. hourly, daily).

The calibration of water demand has been studied using various techniques, for instance a Predictor-Corrector algorithm (Shang et al. 2006, Preis et al. 2009), Bayesian recursive approach (Kapelan et al. 2007), Kalman filtering and tracking state estimator (Kang and Lansey 2009) and particle filter modeling (Hutton et al. 2012). Most of these models have been developed based on given frameworks where the measurement locations were predetermined and the calibration parameters are grouped to be less than the number of measurements. The outcomes, therefore, rely on these additional assumptions, which can lead to large approximations in real water distribution systems. Only few papers have directly dealt with underdetermined systems such as a proportional demand method (Davidson and Bouchart 2006) and singular value decomposition (SVD) (Cheng and He 2010, Kun et al. 2015).

Mathematically, the calibration of the demand in water distribution systems in which the number of measurements is less than the number of calibrated variables, is a nonlinear underdetermined problem. A local solution of the problem can be found by local linearization methods such as QR decomposition, SVD or using the Moore-Penrose pseudoinverse matrix in the Newton-Raphson method. However, due to the possibility of non-uniqueness of the solutions, the results from mathematical methods are either far from the actual solution or result in negative demands at some nodes. Apparently, if the data from measurement devices are considered as the only known inputs, the quest for a reliable demand calibration model is still a challenge for hydraulic researchers.

This paper proposes a methodology for the calibration of water demand multipliers for an underdetermined system where the number of measurements is less than the number of demand parameter variables. The EPANET toolkit is used to solve the system of water

network equations while Genetic Algorithms (GAs) are applied to find the best match between known measurement inputs and their calibrated values. The average values of multiple GA runs have been found to give the best estimates of the flow rates and nodal heads as well as the calibration of the demands in a system. Different scenarios of measurement availability in water networks are tested to evaluate the reliability of the model. Furthermore, this study also investigates the use of a sampling design technique (Piller 1995) for the selection of optimal measurement locations in order to improve the quality of the calibration model. This is followed by application of the model to three case studies. Finally, conclusions and suggestions for future work are given.

## 4.2 GA Calibration model

The proposed model applies an implicit technique for the steady state hydraulic simulation where the calibration process is formulated as an optimization problem. The objective function is the minimization of the differences between simulated values from the model and their corresponding measured values. The decision variables are the demand multiplier factors (DMFs) for the nodal demands as described in the *Decision variables* sub-section.

### 4.2.1 Objective function

In this study, a least-squares method is applied for the objective function. The method minimizes the sum of squared residuals between the measured and computed values of pipe flow rates and nodal heads at the measurement locations. The objective function is given by:

$$MinF = \sum_{i=1}^{N_H} w_i (H_i^{Meas} - H_i^{Sim})^2 + \sum_{j=1}^{N_Q} w_j (Q_j^{Meas} - Q_j^{Sim})^2 \quad (4.1)$$

where  $H_i^{Sim}$ ,  $Q_j^{Sim}$  are the simulated nodal head and flow rate for the  $i_{th}$  node and  $j_{th}$  pipe, respectively;  $H_i^{Meas}$ ,  $Q_j^{Meas}$  are the measured head and flow rate at the  $i_{th}$  node and  $j_{th}$  pipe;  $N_H$ ,  $N_Q$  are the number of head and flow measurement sites in the network

respectively and  $w_i, w_j$  are the weighting factors applied to different terms to ensure they have similar magnitude and unit.

Flows and heads as well as nodal demands are time dependent. However, they can be considered constant during short periods of time, e.g. 30 minutes or one hour. Given that the proposed approach is applied to each steady state step during an extended period simulation, the time dependency symbol is not explicitly given in Equation (4.1). Measured values are obtained from field measurement devices or, for testing the methodology as in our case, these values can be generated by running a hydraulic simulation toolkit such as EPANET. Weighting factors  $w_i$  and  $w_j$  can be computed by taking the inverse of the square of the observed values  $w_i = 1/(H_i^{Meas})^2$  and  $w_j = 1/(Q_j^{Meas})^2$ , respectively), which is the approach that has been used in Nardo et al. (2015).

#### 4.2.2 Decision variables

In a WDS model, the water demand at each node is calculated by the multiplication of a base demand with its corresponding DMF at each time step  $t$ :

$$D_{k,t} = D_{0,k} * f_{k,t} \quad (4.2)$$

where  $D_{0,k}$  is base demand at the  $k_{th}$  node, which is calculated using quarter/annual water usage billing information; and  $f_{k,t}$  is the DMF at the  $k_{th}$  node at time step  $t$ . The decision variables for the optimization problem, therefore, are the demand multiplier factors  $f_{k,t}$  ( $k = 1, \dots, N_{DM}$ ) at nodal demands at each time step. A bounded range of demand factors may apply as:

$$f_k^{min} \leq f_k \leq f_k^{max} \quad (4.3)$$

where  $(f_k^{min}, f_k^{max})$  are the bounds of decision variables. The value of  $f_k^{min}$  must be equal to or larger than zero, while  $f_k^{max}$  can be selected based on typical values of peaking demand factors such as those reported in Beal and Stewart (2014). In this paper, a value of  $f_k^{max} = 1.5$  has been selected for all case studies, which also guarantees that it is much larger than the “true” multiplier used to generate the calibration data.



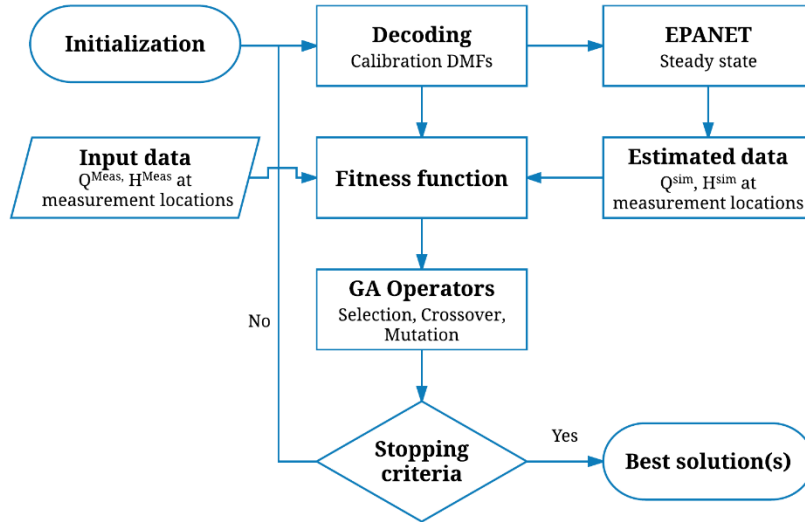
Figure 4.1 shows an example of a GA solution (namely a chromosome) for six demand multiplier factors at one time step, which are chosen from a lower bound of 0.00 and an upper bound of 1.50 with the increment step of 0.02 (these values would be problem dependent and selected accordingly). Each demand multiplier factor is coded by an integer number, ranging from 0 to 75. By using this coding information, the chromosome from GA process is decoded into a set of demand multiplier factors that are multiplied by the base demand and can be used for the hydraulic simulation process.

Chrom.	3	8	11	56	32	1
Decode	0.06	0.16	0.22	1.12	0.64	0.02
DMF No.	f2	f3	f4	f5	f6	f7

**Figure 4.1:** Example of GA chromosome and decoding for the demand calibration problem

#### 4.2.3 GA process and operators

The genetic algorithm (GA) calibration model implemented for this research has been written in the C-sharp language. The flowchart of the algorithm is shown in Figure 4.2. An initial population of chromosomes is randomly generated and decoded into corresponding DMF values for each member. To each node of the network exactly one of these DMFs is assigned and EPANET is subsequently called to simulate the steady state hydraulics of the system. Simulated flows and heads ( $Q^{Sim}$ ,  $H^{Sim}$ ) at the measurement locations are obtained and compared with their measured values via the calculation of the objective function. The inverse of the objective function is applied to define the fitness function for each member of the GA population. This is the measure for the quality of each member, and is used to determine its opportunity for survival.



**Figure 4.2:** Flowchart for the GA calibration of demand multiplier factors

By applying GA selection, crossover and mutation, new generations that inherit features of previous generations are created, and the calibration process is then repeated until the stopping criteria is met.

For the selection operator, a study by Goldberg and Deb (1991) recommended the use of tournament selection because of its better convergence compared to proportionate selection or ranking selection (see Nicklow et al. (2010) for a review of GAs). In addition, Goldberg (1989) also suggested that the two-point crossover operator with a relatively high probability ( $P_c = 0.6$  to  $1.0$ ) and the bitwise mutation with a probability of  $P_m \approx 1/str$  ( $str$  is the length of the string) can be used to improve the performance of GA models. Therefore in the study presented in this paper, tournament selection, two-point crossover and bitwise mutation have been applied in the GA calibration model.

### 4.3 Selection of measurement locations

The accuracy of demand calibration models not only depends on the number of site measurements but also on the locations of the measurements. The optimal measurement location problem has been investigated by a number of researchers using various mathematical and statistical methods, such as: Yu and Powell (1994), Vítkovský et al. (2003), Berry et al. (2005), Propato et al. (2006), Krause et al. (2008) and Giustolisi and

Ridolfi (2014). In calibration models, sampling design (SD) methodologies have been applied for the selection of the observation locations. Piller (1995) used a SD method to minimize the influence of measurement errors in the state vector estimation. Bush and Uber (1998) developed three SD methods derived from D-optimality criteria: max-sum, weighted sum and max-min methods in order to select measurement locations based on the analysis of the Jacobian matrix. Meier and Barkdoll (2000) used a GA for the optimal SD problem with the aim of finding a set of calibration locations which maximizes the presence of non-negligible pipe velocities. De Schaetzen (2000) proposed three SD approaches for the optimal measurement locations based on shortest path algorithm, rank measurement locations and maximization of Shannon's entropy. Most recently, Kapelan et al. (2005) developed two SD models using a GA to find the optimal set of pressure locations, where the first model is formulated as a single objective GA and the second is modelled as a multi-objective optimization problem.

In this paper, the SD method proposed by Piller (1995) based on a greedy algorithm is applied to select the best measurement locations for the GA calibration model of the demand multiplier factors. Influences of the measurement locations on the calibration results, thereafter, are examined by evaluating the convergence of the GA model.

#### 4.3.1 Sampling design method

The hydraulic steady state of a water network solves a non-linear problem of the continuity equations at nodes and the energy equations for pipes. The sensitivity of the nodal heads and flow rates with respect to the demand parameters  $\theta$  at nodes can be computed as:

$$\begin{cases} A^T \frac{\partial q}{\partial \theta} + G_D = 0_{ns,nd} \\ D \frac{\partial q}{\partial \theta} - A \frac{\partial h}{\partial \theta} = 0_{np,nd} \end{cases} \quad (4.4)$$

where  $A$  and  $A^T$  are the unknown head node incidence matrix and its transposed matrix, which provides information about the connectivity of the nodes and the links in the network;  $q$  and  $h$  are the flow rate and nodal head vectors of  $np$  pipes and  $ns$  nodes;  $G_D$

is the  $ns$  by  $nd$  matrix of nodal demand allocation. As demand aggregation is not considered in this study,  $G_D$  is the diagonal matrix of the base demands; and  $D$  is the  $np$  by  $np$  diagonal matrix where the diagonal elements are the derivatives of the head-loss equations of the flows in pipes. The solution for Equation (4.4) is the Jacobian matrix  $J$  of flows and heads with respect to water nodal demands:

$$J(y) = \begin{cases} \frac{\partial q}{\partial \theta} = -[D^{-1}A][A^T D^{-1}A]^{-1}G_D \\ \frac{\partial h}{\partial \theta} = -[A^T D^{-1}A]^{-1}G_D \end{cases} \quad (4.5)$$

Given an estimate of the unknown demand parameters ( $f_0$ ), the SD greedy algorithm method from Piller (1995) iteratively selects the measurement locations ( $S$  matrix) by minimizing the influence of measurement errors on the state vector estimation. The selection matrix  $S$  of the measurement locations is chosen so that the matrix  $(ST_0)$  is full rank and the infinity norm of its pseudo inverse matrix,  $\|(ST_0)^\dagger\|_\infty$ , has a minimum value. The matrix  $T_0 = E_1 J_0 E_2$  is the equilibrium matrix of the Jacobian matrix, where  $E_1$  is the pre-multiplied diagonal matrix to ensure the precision of the measurements at links and nodes,  $E_2$  is the post-multiplied matrix corresponding to the change of the parameter variables, and  $J_0 = J(y_{f_0})$  is the Jacobian matrix computed by Equation (4.5) at  $f_0$ . In this study,  $E_1$  is the identity matrix given that all measurements are assumed to have the same precision, and  $E_2$  is computed by Equation (4.6) given that only the demand parameters need to be calibrated.

$$E_2 = \text{diag}(f_0) \quad (4.6)$$

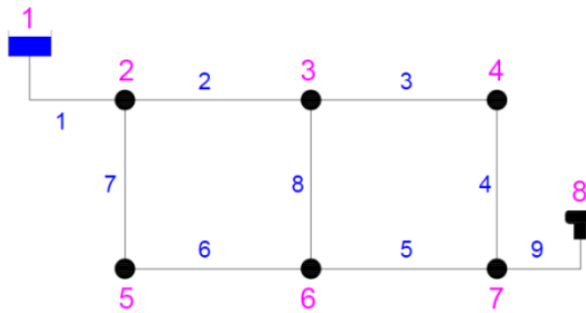
Table 4.1 shows an example of a selection matrix where the elements are assigned a value of 1 if the pipes/nodes are selected as measurement sites, and 0 for non-selected pipes/nodes. A detailed explanation of the method is described in Piller (1995).

**Table 4.1:** Example of selection matrix  $S$  for 4 measurements in the 9 pipe network in Figure 4.3

L 2	L 3	L 4	L 5	L 6	L 7	L 8	L 9	L 1	N2	N3	N4	N5	N6	N7
0	<b>1</b>	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	<b>1</b>	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	<b>1</b>	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	<b>1</b>	0	0

## 4.4 Case study 1

The first case study used to evaluate the methodology is shown in Figure 4.3. The network consists of 9 pipes, 6 nodes with unknown demands (from 2 to 7) and a 5.0 m-diameter tank at node 8 with a water surface elevation of 15.0 m. The system is fed by a reservoir at node 1 with the head of 31.5 m. All nodes are assumed to have an elevation of 0.0 m and base demands of 15.1; 10.3; 11.8; 15.6; 11.3; 8.4 (L/s) for node 2 to node 7, respectively. The demand multiplier factors assigned for the system (and which will need to be calibrated by the GA) are  $X_0 = [0.5; 0.6; 0.8; 0.7; 0.6; 0.9]$  at nodes 2 to 7, respectively. Table 4.2 shows the pipe characteristics for the test network.



**Figure 4.3:** Case study 1 for calibration problem

**Table 4.2:** Pipe characteristics for Case study 1

Pipe	L[m]	D[mm]	$\epsilon$ [mm]
2	1609	254	0.25
3	1609	254	0.25
4	1609	203	0.25
5	1609	203	0.25
6	1609	203	0.25
7	1609	254	0.25
8	1609	203	0.25
9	643.7	254	0.25
1	828	356	0.25

### 4.4.1 Input for calibration model

In practice, input data for the calibration process are usually collected from a supervisory control and data acquisition (SCADA) system. In this research, input data is generated using EPANET toolkit as follows: (1) known demand multiplier factors are assigned to

nodal demands; (2) run EPANET to retrieve the corresponding “true” pipe flow rates and nodal heads (these are then the “measured” data); (3) select the flows and heads at the locations chosen by the SD model as input for the calibration model. The output flows and heads for selected pipes and nodes based on the simulation of the “measured” values are used for the calibration process.

The selection of measurement locations for case study 1 using the SD greedy algorithm method is shown in Table 4.3. For one available measurement device, the method selects pipe 1 as the most sensitive location with respect to the demands, given that this pipe provides the main flow to the system from the source reservoir. When two measurement sites need to be selected, pipe 1 and pipe 9 are chosen as the two most sensitive places for the measurement of the flow. The availability of these two measurements provides the information of the total inflow to the network, which is equivalent to the total water demand of the system. It should be noted that the proposed calibration method is to be applied to cases where the number of measurement locations is less than the number of unknowns. Because there are six unknown demand multiplier factors that need to be calibrated, only up to five available measurement sites are considered for this test network. From a practical point of view, having so many measurement devices (relative to the total number of nodes of 6) in a water network would be unreasonable due to their cost. However, the main purpose of this section is to evaluate the ability of the GA model to calibrate the demands at the nodes with acceptable accuracy. Thus, the problems related to device costs are ignored in this context.

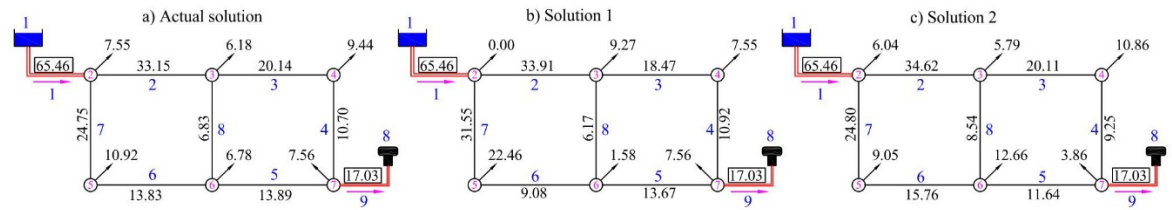
**Table 4.3:** Selection of the measurement locations for case study 1

No. of measurements	Location(s)	$\ (ST_0)^\dagger\ _\infty$
1	<b>P1</b>	0.054
2	<b>P1, P9</b>	0.133
3	<b>P1, P3, P9</b>	0.229
4	<b>P1, P3, P7, P9</b>	0.299
5	<b>P1, P3, P7, P8, P9</b>	0.456

\* Bold values are selected based on the minimum value of  $\|(ST_0)^\dagger\|_\infty$  for the SD greedy algorithm (Piller 1995)

#### 4.4.2 Non-uniqueness of the solutions

Figure 4.4 shows an example of alternative solutions when two measurement locations at pipe 1 and pipe 9 are available. Figure 4.4a presents the true solution of the problem, where the set of 6 nodal demands has a total of 48.43 (L/s) and results in the flow rates of 65.46 (L/s) and 17.09 (L/s) for pipe 1 and pipe 9, respectively. In Figure 4.4b and Figure 4.4c, two different sets of nodal demands also cause the same values of flow rates at the measurement locations.



**Figure 4.4:** Example of non-uniqueness of solutions with 2 available measurement locations

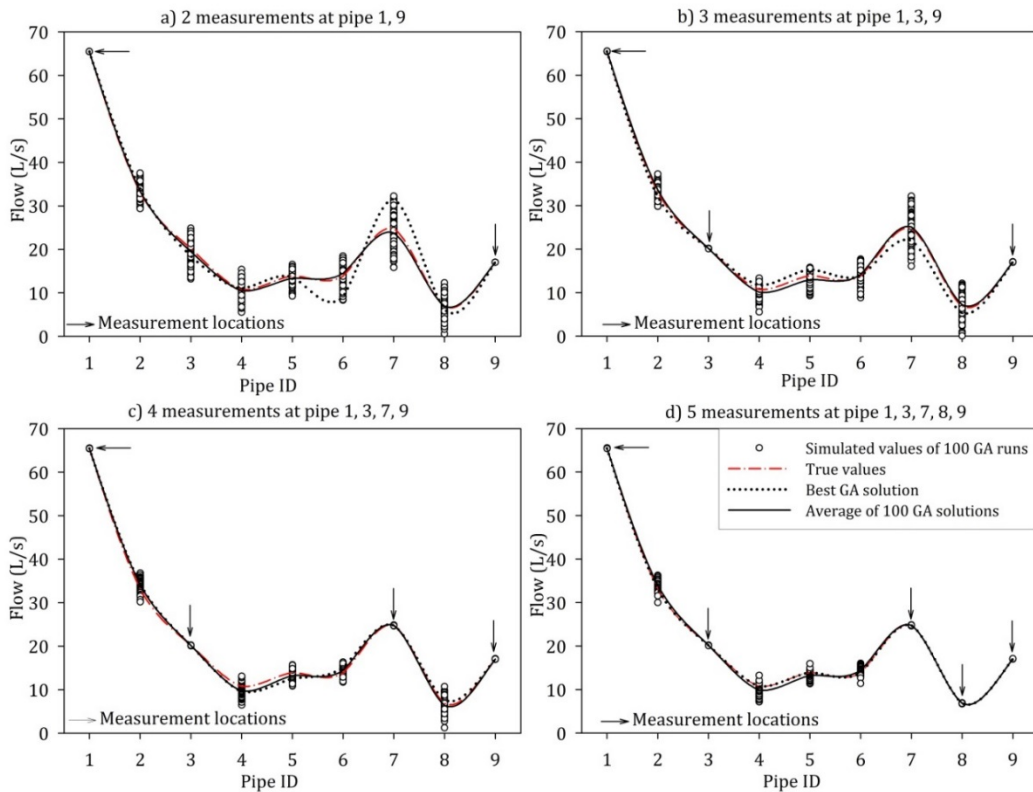
Apparently, for this network, different sets of 6 nodal demands with the total of 48.43 (L/s) that satisfy the water network equations can be a possible solution of the problem. Moreover, the number of possible solutions that match the measured flows will increase if the constraint of the total demand (given by the measurement in pipes 1 and 9) is released, i.e. the two measurements are located in different pipes of the network. A single run of the GA model, therefore, might converge to any of the non-unique solutions or be trapped at a local optimal solution where the simulated values cannot perfectly match the known values at measurement locations. As a result, it appears that a good approximation of the demand multiplier factors calibration problem can only be obtained if multiple runs of the GA model are implemented. The following section shows the results of the multiple runs of the proposed GA model.

#### 4.4.3 Results of GA calibration model

The GA calibration model has been tested with four different scenarios of measurement locations, from 2 to 5 available measurement sites. In order to evaluate the influence of the number of measurement sites on the GA model, the GA parameters are kept constant during all experiments with the size of population  $N = 100$ , probability of crossover  $P_c =$

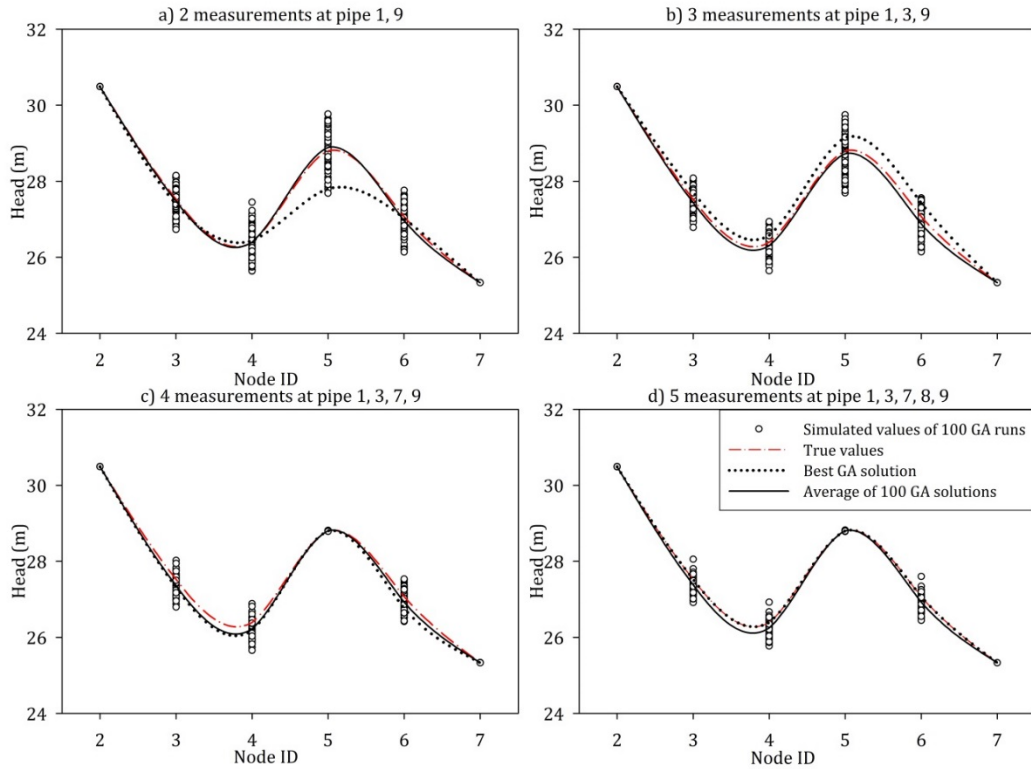
0.8, probability of mutation  $P_m = 0.3$  and the number of generations is 1000. The range of decision variables is selected from 0 to 1.5 with the increment of  $\Delta\theta = 0.02$ , corresponding to a search space size of  $76^6 = 1.927 \times 10^{11}$  possible solutions. In addition, due to the non-uniqueness and the stochastic behaviour of the problem, for each GA application, 100 runs with 100 different seeds were implemented. The results of different GA runs to the case study 1 are presented in Figure 4.5, Figure 4.6 and Figure 4.7.

Figure 4.5a, Figure 4.6a and Figure 4.7a plot the results of flow rates, nodal heads and nodal demands, respectively, from 100 GA runs where two measurement locations at pipe 1 and pipe 9 are available. As can be seen from Figure 4.5a, the calibrated flows at the measurement locations are well matched with the actual known or measured values for all 100 GA runs. Specifically, the simulated values at pipe 9 are  $17.03 \pm 0.02$  (L/s) (the actual value is 17.03 (L/s)), the simulated values at pipe 1 are  $65.46 \pm 0.03$  (L/s) (the actual value is 65.46 (L/s)). On the other hand, large variations of simulated flow rates are observed in all other pipes of the network for the individual 100 GA runs, for instance, the range of flow at pipe 6 is from 8.28 to 18.46 (L/s) while the actual value is 13.83 (L/s) in Figure 4.5a.

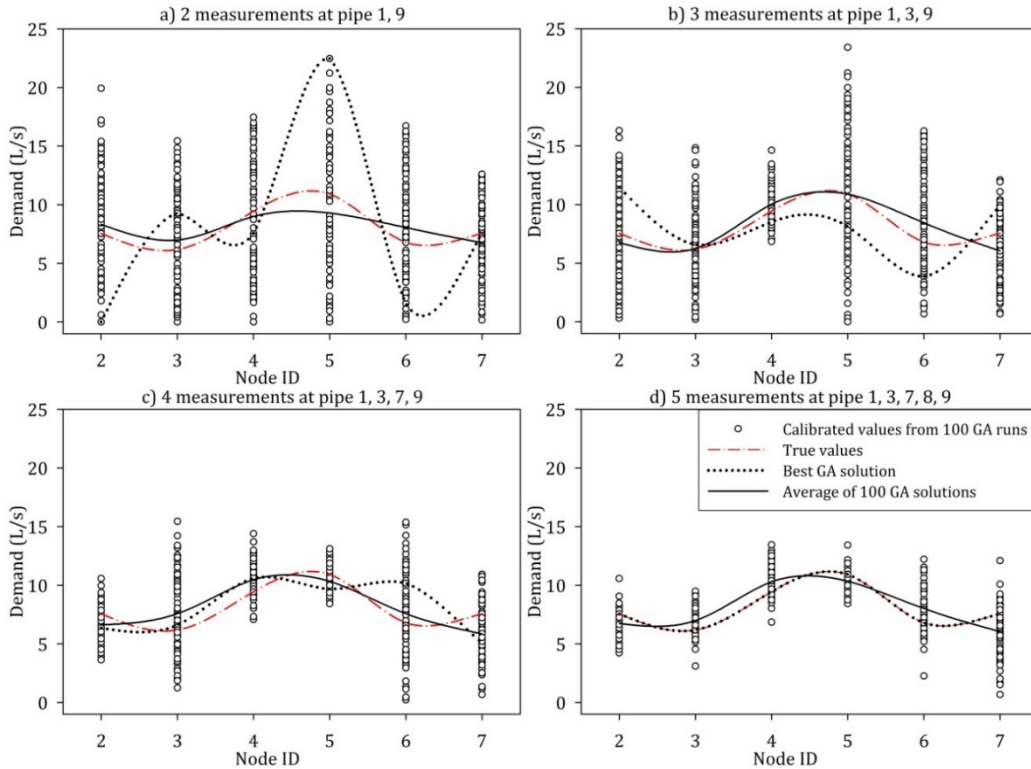


**Figure 4.5:** Results of flow rates from GA calibration model for case study 1





**Figure 4.6:** Results of nodal heads from GA calibration model for case study 1



**Figure 4.7:** Results of nodal demands from GA calibration model for case study 1

Similar to the variation of the simulated flow rates is the variation of nodal heads, as shown in Figure 4.6a. Given the similarity of the simulated results for the flow rates at

pipe 1 and pipe 9, the estimation of nodal heads for the individual 100 GA runs at node 2 and node 7 are also well matched with the actual values. Meanwhile, the simulated heads at the remaining nodes (node 3, 4, 5 and 6), compared to actual heads, vary approximately  $\pm 0.7$ ,  $\pm 0.9$ ,  $\pm 1.0$  and  $\pm 0.8$  (m), respectively. In Figure 4.7a, the calibrated nodal demands for the individual 100 GA runs show quite a large variation at all nodes, starting from 0 (L/s) up to approximately twice the magnitude of the actual nodal demands.

It can be observed in Figure 4.5b, 4.5c, 4.5d, Figure 4.6b, 4.6c, 4.6d and Figure 4.7b, 4.7c, 4.7d that the addition of measurement sites to the network increases the accuracy of the calibration model. As an example, consider the effect of adding a measurement at a pipe in the network, from 3 measurements (at pipe 1, 3 and 9) to 4 measurements (at pipes 1, 3, 7 and 9). The simulated flow rates and nodal heads at all non-measured locations are slightly improved. If with 3 measurement locations, the simulated flows and the simulated heads at pipe 6 and node 6 vary for the individual 100 GA runs in the ranges of [8.79, 17.67] (L/s) and [26.14, 27.56] (m), respectively, with four measurement locations, the simulated values are improved, varying in smaller ranges of [11.71, 16.35] (L/s) and [26.41, 27.53] (m). On the other hand, the calibrated demands show a significant improvement when the number of measurement sites increased. At node 5, for instance, the variation of the demand is improved from the maximum allowable range [0, 23.4] (L/s) to [8.42, 13.01] (L/s).

The best GA solution and the average values of 100 GA runs are also plotted in order to compare them with the actual solution. The results show that increasing number of measurement sites leads to a better accuracy of the best GA result. Due to the large search space size, the GA calibration model cannot find the exact solution in any of the 100 GA runs unless five measurement sites are provided. However, the optimal solution is only found 3 times out of 100 GA runs in the scenario of five available measurement sites.

The average values of 100 GA runs, on the contrary, yield a very good match with the actual values of the problem. In all measurement scenarios, the mean flow rates, nodal

heads and nodal demands, which are tabulated in Table 4.4, Table 4.5 and Table 4.6 respectively, are slightly different to the actual values, with less than 10% error for the flow rates, less than 1% error for the nodal heads and less than 20% error for the nodal demands. The results clearly indicate that the nodal heads are estimated more accurately than the flows. In addition, the standard deviations computed based on the 100 GA runs, which represent the variation of the calibrated results, show that the change of the nodal demands causes greater variation of the flow rates rather than the nodal heads in this looped network. As a result, the flow rates are more sensitive to the demands than the nodal heads.

**Table 4.4:** Comparison of mean estimated pipe flows and actual flows for case study 1

Measurements	Pipe	P2	P3	P4	P5	P6	P7	P8	P9	P1
	Actual flows	<b>33.15</b>	<b>20.14</b>	<b>10.70</b>	<b>13.89</b>	<b>13.83</b>	<b>24.75</b>	<b>6.83</b>	<b>17.03</b>	<b>65.46</b>
2 measurements (P1, P9)	Average flows of 100 GA runs	33.38	19.46	10.46	13.33	14.47	23.76	6.93	<b>17.03</b>	<b>65.46</b>
	Standard deviation	1.92	2.95	2.41	1.71	2.26	3.94	2.40	<b>0.01</b>	<b>0.01</b>
	% $\Delta_{(\text{Average flows, actual flows})}$	0.67%	3.35%	2.25%	4.01%	4.57%	4.01%	1.38%	<b>0.01%</b>	<b>0.00%</b>
3 measurements (P1, P3, P9)	Average flows of 100 GA runs	33.63	<b>20.14</b>	10.10	12.97	14.18	25.07	7.24	<b>17.02</b>	<b>65.46</b>
	Standard deviation	1.61	<b>0.02</b>	1.60	1.49	2.05	3.79	2.55	<b>0.03</b>	<b>0.03</b>
	% $\Delta_{(\text{Average flows, actual flows})}$	1.44%	<b>0.02%</b>	5.60%	6.62%	2.50%	1.28%	5.95%	<b>0.04%</b>	<b>0.00%</b>
4 measurements (P1, P3, P7, P9)	Average flows of 100 GA runs	34.09	<b>20.14</b>	9.66	13.21	14.41	<b>24.75</b>	6.37	<b>17.02</b>	<b>65.46</b>
	Standard deviation	1.43	<b>0.04</b>	1.47	1.08	1.00	<b>0.04</b>	2.08	<b>0.03</b>	<b>0.03</b>
	% $\Delta_{(\text{Average flows, actual flows})}$	2.82%	<b>0.00%</b>	9.69%	4.87%	4.18%	<b>0.02%</b>	6.76%	<b>0.02%</b>	<b>0.00%</b>
5 measurements (P1, P3, P7, P8, P9)	Average flows of 100 GA runs	33.97	<b>20.14</b>	9.84	13.23	14.41	<b>24.75</b>	<b>6.84</b>	<b>17.03</b>	<b>65.46</b>
	Standard deviation	1.12	<b>0.04</b>	1.14	0.86	0.80	<b>0.06</b>	<b>0.05</b>	<b>0.03</b>	<b>0.04</b>
	% $\Delta_{(\text{Average flows, actual flows})}$	2.46%	<b>0.01%</b>	8.06%	4.70%	4.19%	<b>0.00%</b>	<b>0.15%</b>	<b>0.00%</b>	<b>0.00%</b>

\* Bold-Calibrated values at measurement locations.  $\Delta$ -Differences between calibrated values and actual values.

**Table 4.5:** Comparison of mean estimated nodal heads and actual heads for case study 1

Measurements	Node	N2	N3	N4	N5	N6	N7
	Actual head	<b>30.49</b>	<b>27.53</b>	<b>26.39</b>	<b>28.80</b>	<b>27.07</b>	<b>25.33</b>
2 measurements (P1, P9)	Average head of 100 GA runs	30.49	27.48	26.39	28.89	26.97	25.33
	Standard deviation	0.00	0.33	0.44	0.50	0.39	0.00
	% $\Delta_{(\text{Average head, actual head})}$	0.00%	0.17%	0.01%	0.31%	0.39%	0.00%
3 measurements (P1, P3, P9)	Average head of 100 GA runs	30.49	27.44	26.30	28.72	26.88	25.33
	Standard deviation	0.00	0.28	0.28	0.50	0.33	0.00
	% $\Delta_{(\text{Average head, actual head})}$	0.00%	0.32%	0.34%	0.27%	0.72%	0.00%
4 measurements (P1, P3, P7, P9)	Average head of 100 GA runs	30.49	27.36	26.22	28.80	26.92	25.33
	Standard deviation	0.00	0.25	0.25	0.00	0.25	0.00
	% $\Delta_{(\text{Average head, actual head})}$	0.00%	0.61%	0.64%	0.00%	0.55%	0.00%
5 measurements (P1, P3, P7, P8, P9)	Average head of 100 GA runs	30.49	27.38	26.25	28.80	26.93	25.33
	Standard deviation	0.00	0.20	0.20	0.01	0.20	0.00
	% $\Delta_{(\text{Average head, actual head})}$	0.00%	0.53%	0.55%	0.00%	0.54%	0.00%

**Table 4.6:** Comparison of mean calibrated nodal demands and actual demands for case study 1

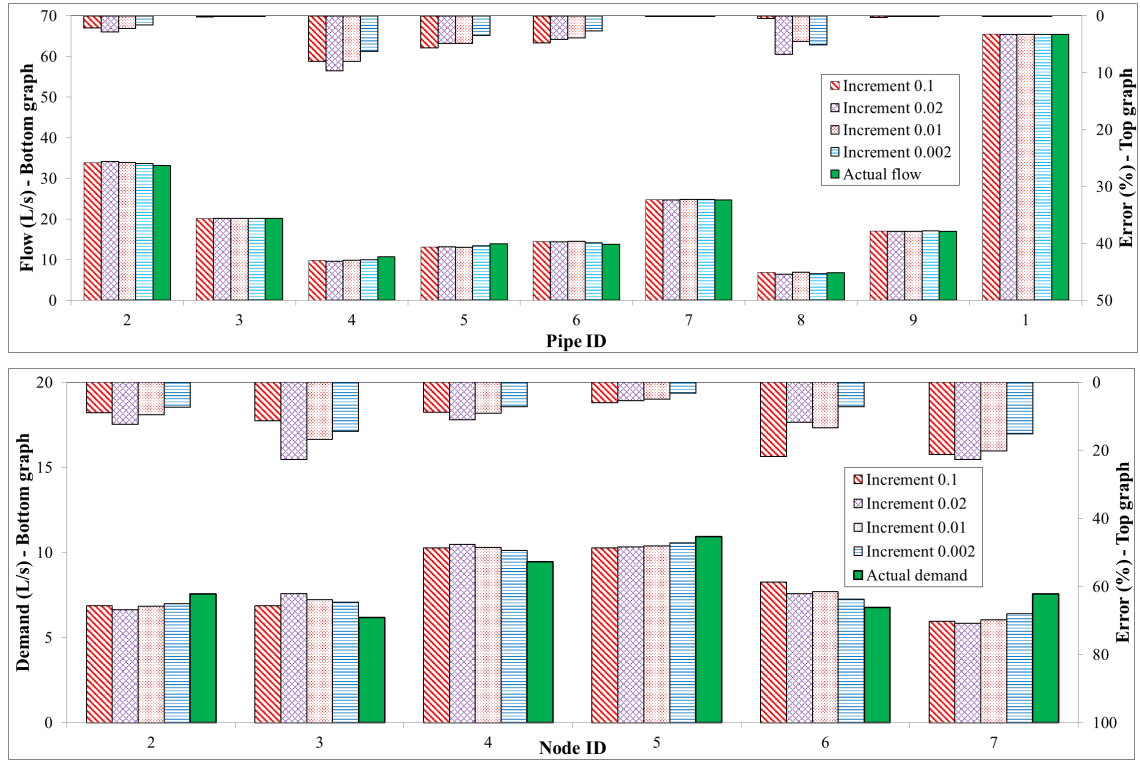
Measurements	Node	N2	N3	N4	N5	N6	N7
	Actual demand	7.55	6.18	9.44	10.92	6.78	7.56
2 measurements (P1, P9)	Average demand of 100 GA runs	8.32	6.98	9.01	9.29	8.06	6.76
	Standard deviation	4.53	3.99	4.86	5.78	4.73	3.30
	% $\Delta_{(\text{Average demand, actual demand})}$	10.20%	13.00%	4.60%	14.89%	18.93%	10.53%
3 measurements (P1, P3, P9)	Average demand of 100 GA runs	6.76	6.26	10.03	10.89	8.45	6.05
	Standard deviation	3.93	3.51	1.61	5.51	4.25	2.63
	% $\Delta_{(\text{Average demand, actual demand})}$	10.52%	1.23%	6.30%	0.26%	24.63%	19.98%
4 measurements (P1, P3, P7, P9)	Average demand of 100 GA runs	6.62	7.58	10.48	10.34	7.57	5.85
	Standard deviation	1.43	3.06	1.48	1.00	3.54	2.24
	% $\Delta_{(\text{Average demand, actual demand})}$	12.32%	22.63%	10.98%	5.34%	11.70%	22.62%
5 measurements (P1, P3, P7, P8, P9)	Average demand of 100 GA runs	6.73	6.98	10.30	10.34	8.02	6.04
	Standard deviation	1.10	1.11	1.15	0.81	1.67	2.01
	% $\Delta_{(\text{Average demand, actual demand})}$	10.80%	13.00%	9.15%	5.31%	18.33%	20.04%

#### *Effects of increment of decision variables on GA calibration results*

One of the factors that may affect the accuracy of the GA calibration model is the increment ( $\Delta\theta$ ) of the demand multiplier factors. The selection of a large increment for the decision variables leads to faster convergence of the GA model although it may result in a coarser approximation of the calibrated demands. Alternatively, the GA model can give better results if smaller increment steps are selected. However, the model requires more computational effort to converge due to the larger search space size. In order to evaluate the effect of increment steps on GA calibration results, different increment steps of 0.1, 0.02, 0.01 and 0.002 were selected and tested in the case of four available measurement locations at pipe 1, 3, 7 and 9. The average calibrated demands and estimated flows are shown in Figure 4.8.

The results of the GA model for different increment steps of the decision variables are approximately comparable. Consider the average calibrated demand at node 7 as an example. The largest demand error (22.62%) occurs at  $\Delta\theta = 0.02$ , followed by 21.22% at  $\Delta\theta = 0.1$ , 20.23% at  $\Delta\theta = 0.01$  and 15.2% at  $\Delta\theta = 0.002$ . With regard to the simulated flows, the estimation errors for different increment steps are relatively small. The maximum errors all occur at pipe 4 and are of almost the same magnitude, 6.28%, 8.02%, 8.03% and 9.69%, corresponding to  $\Delta\theta = 0.002, 0.01, 0.1$  and  $0.02$ , respectively.

Hence, for this case study, it can be concluded that the GA model is not particularly sensitive to the increment of the decision variable in the range of 0.002 to 0.1.



**Figure 4.8:** Calibrated demands & estimated flows with different increment steps for case study 1

#### *Effects of measurement locations on GA calibration results*

The GA calibration results showed that, even with a very large search space, the GA model is able to find the exact solution of the problem for at least one of the 100 GA runs. Thus, the effects of measurement locations on the GA calibration results can be evaluated by the number of times the algorithm converges to the true solution when different combinations of measurement sites are tested.

In this investigation, it is assumed that there are four measurement sites available. Four different combinations of measurement sites are tested to evaluate the convergence of the GA model, which is shown in Table 4.7.

**Table 4.7:** Effects of measurement locations on GA calibration results

Experiment	Measurement locations	Number of convergences	Note
1	P1, P3, P7, P9	16/100	SD method
2	P1, P6, P8, P9	7/100	Random
3	P1, P9, N3, N6	4/100	Random
4	N2, N5, N4, N7	5/100	Random

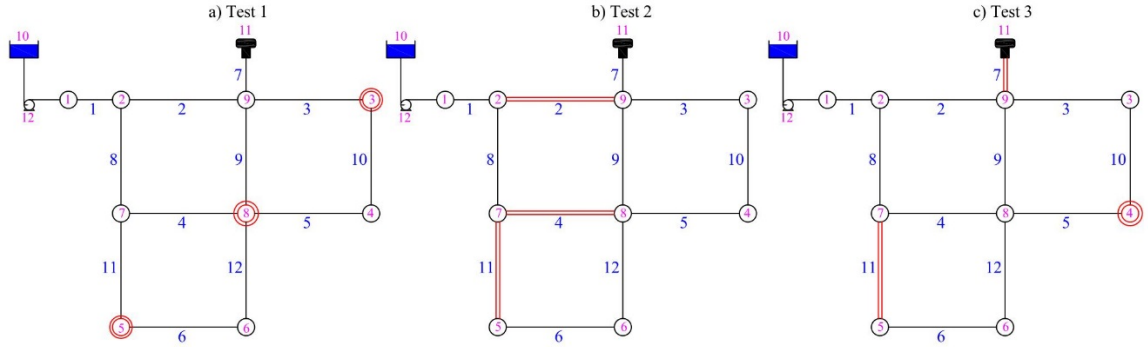
The measurement locations in the first experiment are selected by the SD greedy algorithm method and the four flow measurement sites are placed at pipes 1, 3, 7 and 9. In the second and the third experiments, two flow measurements are kept at the same locations at pipe 1 and pipe 9, while the two other measurement locations are selected randomly. This leads to two flow measurement sites being located at pipe 6 and pipe 8 for experiment 2 and two pressure measurement sites located at node 3 and node 6 for experiment 3. The last experiment involves the presence of four pressure measurement sites at nodes 2, 4, 5, and 7.

The GA parameters of the GA model were kept constant and set equal to the values presented in the previous section, except for the increment of the decision variables. The increment in this test was selected  $\Delta\theta = 0.1$ , so that it was possible to fully enumerate the search space ( $16.77 \times 10^6$  possible solutions). The enumeration of the problem shows that with this relatively small search space size, the problem has a unique optimal solution for all four experiments. In this case, the objective function reaches exactly zero when the calibrated demand multiplier factors are identical to the actual demand multiplier factors assigned to the network. The number of convergences to the optimal solution therefore, is the number of times out of 100 GA runs in which the GA model can find the exact solution of the problem. As seen in Table 4.7, the first experiment results in the highest number of convergences with 16 times out of 100 runs. The second experiment has 7 times, followed by the fourth and the third experiment where the number of convergences is 5 times and 4 times out of 100 runs, respectively. The output of the GA calibration model in this case study, therefore, seems to be sensitive to the locations and the types of the measurement in the network.

## 4.5 Case study 2

The second case study is considered to compare the performance of the GA model with the SVD model from Cheng and He (2010). The network has 9 nodes, 12 pipes, one tank, one pump and one reservoir. The network topology and all information including the pipe

parameters, length, roughness and pump characteristic can be found from the EPANET example (Rossman 2000), namely the Net1 network (Figure 4.9).



**Figure 4.9:** Case study 2 network for calibration problem

The GA calibration model (with a population size of  $N = 100$ , probability of crossover  $P_c = 0.8$  and probability of mutation  $P_m = 0.15$ ) was tested for three scenarios of measurement sites according to Cheng and He (2010). In the first test, three pressure sensors are assumed to be located at nodes 3, 5 and 8. The second test assumes the flow meters are set at pipes 2, 4 and 11, while the third test assumes two flow sensors and a pressure sensor are placed at pipes 7, 11 and node 4, respectively. The calibrated demands and the differences between real demands and calibrated demands of the SVD model and GA model are shown in Table 4.8.

**Table 4.8:** Comparison of SVD model (Cheng and He 2010) and GA model for case study 2

Node number	Real DM (GPM)	Test 1				Test 2				Test 3			
		SVD (GPM)	GA (GPM)	$\Delta_{\text{SVD}}$ (%)	$\Delta_{\text{GA}}$ (%)	SVD (GPM)	GA (GPM)	$\Delta_{\text{SVD}}$ (%)	$\Delta_{\text{GA}}$ (%)	SVD (GPM)	GA (GPM)	$\Delta_{\text{SVD}}$ (%)	$\Delta_{\text{GA}}$ (%)
		2	185.78	<b>118.67</b>	23.85	<b>20.89</b>	145.67	157.94	2.89	5.29	158.49	<b>150.09</b>	5.66
3	151.09	<b>93.15</b>	51.09	<b>6.85</b>	112.80	<b>101.88</b>	12.80	<b>1.88</b>	122.65	<b>85.76</b>	22.65	<b>14.24</b>	
4	300	216.44	<b>327.15</b>	27.85	<b>9.05</b>	200.54	<b>265.88</b>	33.15	<b>11.37</b>	279.36	<b>314.48</b>	6.88	<b>4.83</b>
5	50	49.26	<b>50.34</b>	1.48	<b>0.69</b>	49.49	51.63	1.02	3.25	77.64	<b>52.81</b>	55.28	<b>5.62</b>
6	50	64.01	35.90	28.02	28.20	62.11	<b>38.90</b>	24.22	<b>22.20</b>	58.43	38.92	16.86	22.16
7	150	159.26	164.15	6.17	9.43	151.05	142.06	0.70	5.29	127.40	<b>153.85</b>	15.07	<b>2.57</b>
8	150	221.00	<b>140.58</b>	47.33	<b>6.28</b>	209.63	<b>96.45</b>	39.75	<b>35.70</b>	157.52	141.97	5.01	5.35
9	150	129.21	<b>131.45</b>	13.86	<b>12.36</b>	164.80	168.36	9.87	12.24	114.76	<b>162.12</b>	23.49	<b>8.08</b>
$\Sigma$ DM	1100	1198.0	1061.4	8.91	<b>3.51</b>	1099.9	1043.1	0.00	5.17	1100	1100	0.00	0.00

\* Bold - Cases where the GA model finds better calibrated results

It can be seen that for all three tests, the average of 100 runs of the GA model gives more reasonable results for the nodal demands. In test 1, while the SVD model leads to large differences between actual and calibrated values at node 3 and node 8 (51.09% and 47.33%

respectively), the average GA results in relatively small errors (6.85% and 6.28%) at the corresponding nodes. In test 2, the largest error of the two models occurs at node 8, with an error of 39.75% for the SVD model and a slightly smaller error of 35.7% for the GA model. The last test presents the best performance of the GA model with the maximum error of 22.16% at node 6 while the SVD model still remains a large error of 55.28% at node 5. For estimating the total water demands of the system (the last row of Table 4.8), both models achieve reasonable results, especially for the SVD model in test 2 and 3. The total calibrated demand of the GA model only matches with the real total demand in test 3, where one of the flow sensors is placed at pipe 7 to measure the flow from the tank to the system. This highlights the importance of the selection of the optimal measurement locations. Reasonably accurate results can be achieved if information related to the total flow is provided.

### 4.6 Case study 3

The third test network aims at testing the performance of the proposed methodology for a larger network. This network is provided by EPANET, namely the Net2 network, which consists of 40 pipes and 35 nodes, one tank and one pump station. The pump station is modelled as a node with negative demand, which feeds water into the network. It is assumed that up to 3 measurement devices are able to be installed in the network. In addition, the pump flow is also assumed to be known. The network diagram is shown in Figure 4.10.

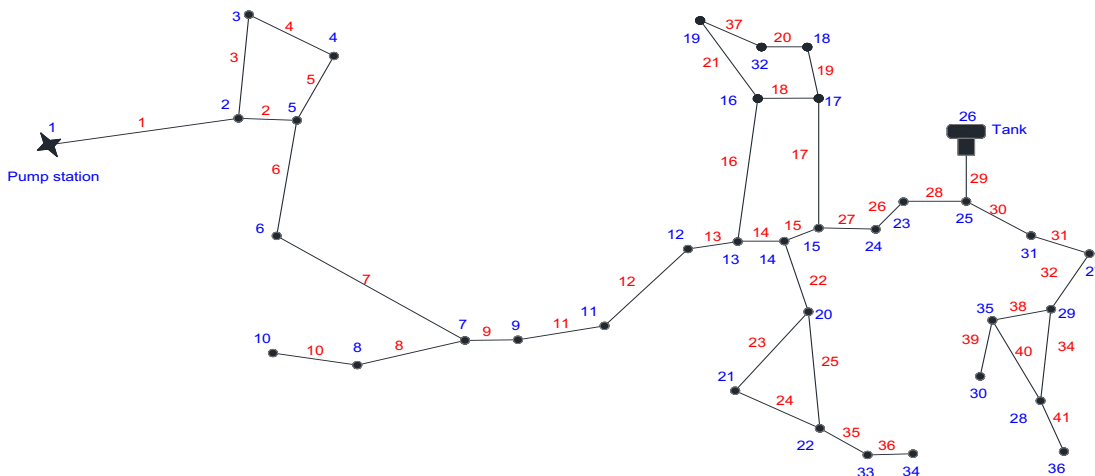


Figure 4.10: Case study 3 water distribution network (Net2 network)



The SD greedy algorithm model for the selection of measurement locations found that, for this network, the flows in pipes are much more sensitive to the demands than the heads at nodes. Three flow measurement sites are suggested to be located at pipes 12, 22 and 29. The GA calibration model of the Net2 water network was implemented with the following characteristics: (1) the number of decision variables is 32 corresponding to a total of 32 nodal demand multiplier factors of the network; (2) the size of the choice table for the decision variables is selected from  $f_k^{min} = 0$  to  $f_k^{max} = 1.5$ ; (3) a population of  $N = 500$ , probability of crossover  $P_c = 0.8$ , probability of mutation  $P_m = 0.04$  and the number of generations  $N = 1000$  were selected for the GA model parameters; (4) in order to evaluate the effects of the increment of the decision variables, different increment steps of  $\Delta\theta = 0.0005, 0.005, 0.02$  and  $0.1$ , respectively, were examined.

A hundred runs of the GA model were undertaken on the Intel® Core™ i5 (2.9GHz) computer. The total elapsed run time was approximately 16.3 hours, or about 10 minutes for one run. Table 4.9 summarizes the average errors of the nodal demands, flow rates and nodal heads in comparison with the true value at nodes and pipes in the network. It can be seen that the errors are approximately equal for four different increment steps of the decision variables. The average errors fluctuate slightly around 15% for the demands, 10% for the flows and only 0.01% for the nodal heads. Hence, also this case study shows that the GA calibration model is not sensitive to the increment selected for the decision variables.

**Table 4.9:** Summary of average GA output errors for Net2 network with different increment steps

Increment step	Average demand errors (%)	Average flow error (%)	Average head error (%)
0.1	16.41	10.72	0.009
0.02	14.74	10.81	0.009
0.005	15.19	9.96	0.010
0.0005	15.30	10.78	0.011

Figure 4.11 and Figure 4.12 plot the outcome from the GA model corresponding with the increment value  $\Delta\theta = 0.02$  as a demonstration of the results. The bottom graph of each figure presents the average calibrated/simulated values and the true values of the nodal

demands/flow rates. The top graph of each figure shows the error percentage of the difference between the calibrated/simulated values and their true values. Due to the relatively accurate estimation of the nodal heads, the plot of the simulated heads and actual heads is not shown here.

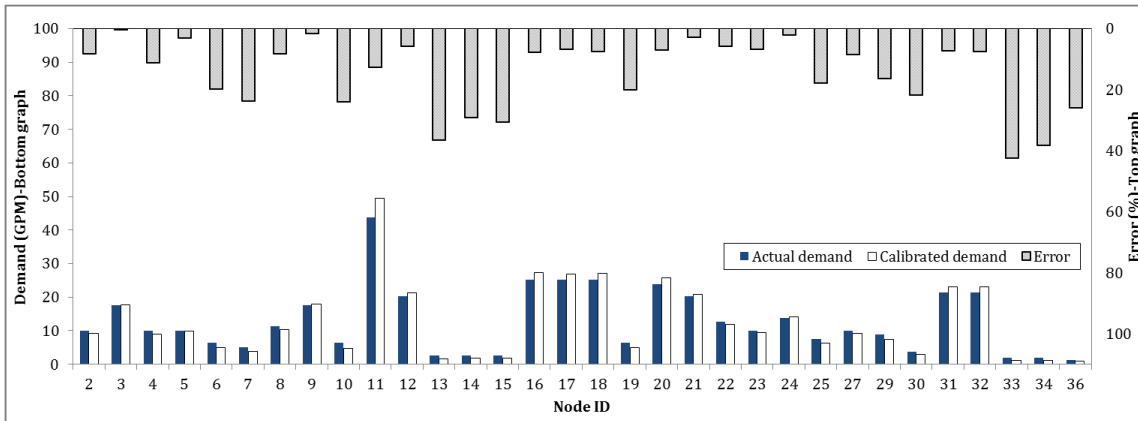


Figure 4.11: Comparison of average calibrated demands versus actual demands of Net2 network

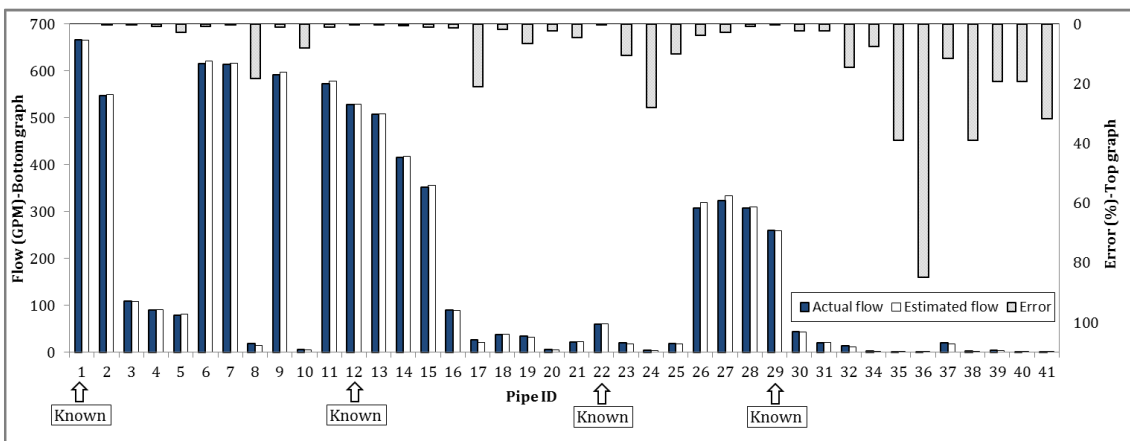


Figure 4.12: Comparison of average simulated flows versus actual flows of Net2 network

Let us consider three groups of the demands (0 to 10 GPM, 10 to 30 GPM, and > 40 GPM). Figure 4.11 shows that large calibration percentage errors only occur in the first group, where the demands are small. The errors in this group are within 16.5% and 45.48%. However, all the differences are less than 1.52 GPM in absolute value. The second group shows a very good approximation of the demand, as the percentage errors vary from 0.48% to 8.51%. The last group, which contains only node 11 with the actual demand of 43.82 GPM, also obtains a relatively accurate calibrated result of 49.47 GPM, equivalent to an error of 12.89%.

Similarly, Figure 4.12 shows that the flow estimation is generally much more accurate for the pipes with high flows than the pipes with low flows. The largest estimation error, approximately 84.8%, occurs at pipe 36 at which the actual flow is 0.63 GPM and the simulated flow is 1.16 GPM. For all the pipes that have actual flows larger than 30 GPM, the estimation errors are smaller than 6.5%.

## 4.7 Conclusions and recommendations

Calibration of water demand in real water distribution systems is complicated by the limited number of measurement sites. A GA model has been developed for the calibration of the demand multiplier factors for underdetermined water distribution systems where the number of measurement locations is less than the number of unknown parameters. The approach for estimation of demand multiplier factors using GA optimization has been tested on three case studies. The first case study has shown that the average values of multiple runs of the GA model can deliver a very good approximation of the water demand multipliers with little information from the SCADA system. The first case study also shows that the location of the measurement sites does influence the performance of the GA model. The second case study demonstrates the advantage of the GA model in comparison to the singular value decomposition model. It also confirms the conclusions made from the first case study about the sensitivity of the GA model to the measurement locations. Therefore, the GA model is suggested to be implemented in combination with a supporting tool for the selection of optimal measurement locations such as the SD greedy algorithm model. The third case study validates the approach for a slightly larger sized network, which again exhibits the superior performance of the GA model. The model run time for this last case study (approximately 16.3 hours) might be a disadvantage of the GA model. The model might, therefore, be suitable for the networks at which the SCADA data are provided on a daily basis, or if real-time calibration is required, parallel computer systems would need to be implemented for the GA model.

Future research efforts will involve finding advanced methods for the calibration of the demand to reduce the computational time. In addition, uncertainty of the calibration model is another consideration given the presence of errors (or noise) in measurement data. Finally, addressing the problem of leakage in the network is also important in achieving reliable results.

## **References**

References are included in the Bibliography section. In addition, the final published paper in Appendix 1 has the references listed.

# Chapter 5 Particle filter-based model for online estimation of demand multipliers in water distribution systems under uncertainty

## Publication 2

Nhu C. Do<sup>1</sup>, Angus R. Simpson<sup>1</sup>, Jochen W. Deuerlein<sup>2</sup>, Olivier Piller<sup>3</sup>

<sup>1</sup> - *School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide SA 5005, Australia.*

<sup>2</sup> - *Senior Researcher, 3S Consult GmbH, Karlsruhe, Germany.*

<sup>3</sup> - *Senior Researcher, Irstea UR ETBX, Dept. of Water, Cestas, France.*

*Journal of Water Resources Planning and Management, 143(11)*


[doi:10.1061/\(ASCE\)WR.1943-5452.0000841](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000841)

This page has been intentionally left blank.

# Statement of Authorship

Title of Paper	<b>A particle filter-based model for online estimation of demand multipliers in Water distribution systems under uncertainty</b>
Publication Status	<input type="checkbox"/> Published <input checked="" type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style
Publication Details	Do, N.C., Simpson, A.R., Deuerlein, J.W. & Piller, O. 2017, 'A particle filter-based model for online estimation of demand multipliers in Water distribution systems under uncertainty', <i>Journal of Water Resources Planning and Management</i> , Accepted for Publication May 2017


## Principal Author


Name of Principal Author (Candidate)	<b>Do, N.C.</b>		
Contribution to the Paper	<b>Development and implementation of methodology, design of case studies, interpretation and analysis of results, preparation of manuscript and acting as corresponding author.</b>		
Overall percentage (%)	<b>70%</b>		
Certification:	<b>This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the <u>primary</u> author of this paper</b>		
Signature	 <table border="1" style="display: inline-table; vertical-align: middle;"> <tr> <td>Date</td> <td><b>3 August 2017</b></td> </tr> </table>	Date	<b>3 August 2017</b>
Date	<b>3 August 2017</b>		


## Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	<b>Simpson, A.R.</b>		
Contribution to the Paper	<b>Research supervision and manuscript evaluation</b>		
Signature	 <table border="1" style="display: inline-table; vertical-align: middle;"> <tr> <td>Date</td> <td><b>4 August 2017</b></td> </tr> </table>	Date	<b>4 August 2017</b>
Date	<b>4 August 2017</b>		

Name of Co-Author	<b>Deuerlein, J.W.</b>		
Contribution to the Paper	<b>Research supervision and manuscript evaluation</b>		
Signature	 <table border="1" style="display: inline-table; vertical-align: middle;"> <tr> <td>Date</td> <td><b>11 July 2017</b></td> </tr> </table>	Date	<b>11 July 2017</b>
Date	<b>11 July 2017</b>		

Name of Co-Author	<b>Piller, O.</b>		
Contribution to the Paper	<b>Research supervision and manuscript evaluation</b>		
Signature	 <table border="1" style="display: inline-table; vertical-align: middle;"> <tr> <td>Date</td> <td><b>11 July 2017</b></td> </tr> </table>	Date	<b>11 July 2017</b>
Date	<b>11 July 2017</b>		





## 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*

## 5.1 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 analyze 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 analyzed 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.

## 5.2 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):

$$x_k = f_{k-1}(x_{k-1}, v_{k-1}) \quad (5.1)$$

where  $k$  is the index of discrete time steps;  $f_{k-1}$  is a known, possibly nonlinear function of the previous state and  $v$  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:

$$z_k = h_k(x_k, w_k) \quad (5.2)$$

where  $h$  is a known implicit or explicit, possibly nonlinear function and  $w$  is the measurement noise sequence. The noise terms  $v_k$  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:

$$p(x_k|Z_{k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|Z_{k-1})dx_{k-1} \quad (5.3)$$

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, i.e. Equation (5.1), with the known statistics of  $v_{k-1}$  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:

$$p(x_k|Z_k) = \frac{p(z_k|x_k)p(x_k|Z_{k-1})}{\int p(z_k|x_k)p(x_k|Z_{k-1})dx_k} \quad (5.4)$$

where  $p(z_k|x_k)$  is the likelihood function, defined by the measurement equation (Equation (5.2)) with the known statistics of  $w_k$ .

According to Ristic et al. (2004), the recursive propagation of the posterior pdf shown in Equation (5.3) and Equation (5.4) is only a conceptual solution that cannot be analytically solved. The solution requires the storage of a fully non-Gaussian pdf, corresponding to an infinite 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.

### 5.3 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 Equation (5.4), 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:

$$p(x_k | Z_k) \approx \sum_{i=1}^{N_P} w_k^i \delta(x_k - x_k^i) \quad (5.5)$$

$$w_k^i = w_{k-1}^i \frac{p(z_k | x_k^i) p(x_k^i | x_{k-1}^i)}{p(x_k^i | x_{k-1}^i, z_k)} \quad (5.6)$$

where  $\delta$  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,  $p(x_k^i | x_{k-1}^i, z_k) = p(x_k^i | x_{k-1}^i)$ , which yields with scaling:

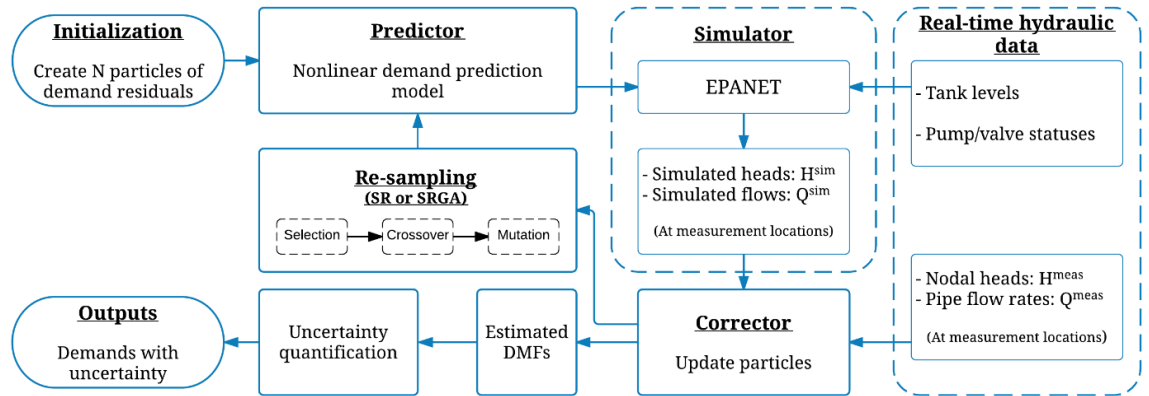
$$w_k^i = \frac{p(z_k | x_k^i)}{\sum_{i=1}^{N_P} p(z_k | x_k^i)} \quad (5.7)$$

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.



## 5.4 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 5.1.



**Figure 5.1** Process of particle filter model for real-time demand estimation in WDS

### 5.4.1 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  $1/N_p$ .

### 5.4.2 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):

$$\ln x_k^j = \sum_{i=1}^m \phi_i^j \ln x_{k-i}^j + \ln v_k^j \quad (5.8)$$

where  $x_k^j$  is the demand residual state at time step  $k$  of the  $j^{\text{th}}$  DMF;  $i$  is the lag counter;  $m$  is the number of autocorrelation lags (for the state estimation problem  $m = 1$  as referred to Equation (5.1));  $\phi_i$  is the auto-regression coefficient for lag  $i$  and  $v_k(0, \sigma_h)$  is the white noise with mean zero and standard deviation  $\sigma_h$ .

The  $j^{\text{th}}$  DMF is calculated as:

$$DMF_k^j = C_k^j x_k^j \quad (5.9)$$

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

### 5.4.3 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.

#### 5.4.4 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.

#### 5.4.5 Corrector

The weights of the particles are corrected/updated by associating the simulated heads and flows with the actual observations via Equation (5.7) where the likelihood function is assumed to be Gaussian:

$$p(z_k | x_k^i) = \frac{1}{\sqrt{2\pi|R|}} e^{(-\frac{1}{2}[z_k - h(x_k^i)]^T R^{-1} [z_k - h(x_k^i)])} \quad (5.10)$$

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,

$$\Delta^{true} = Z^{true} - h(x^{true}) \quad (5.11)$$

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:

$$Z = Z^{true} + \Delta^{meas} \quad (5.12)$$

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.

#### 5.4.6 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  $U[0, 1/N_P]$ , and consequently creates  $N_P$  ordered numbers (Hol et al. 2006):

$$u^i = \frac{i-1}{N_P} + u_s \quad (i = 1, \dots, N_P) \quad (5.13)$$

New particles are then selected that satisfy Equation (5.14):

$$x_{new}^i = x(F^{-1}(u^i)) \quad (5.14)$$

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 Equation (5.10). 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  $N_{gen}$ .

#### 5.4.7 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):

$$\hat{x}_k \approx \frac{1}{N_p} \sum_{i=1}^{N_p} x_k^{i*} \quad (5.15)$$

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:

$$var(x_k) \approx \frac{1}{N_p} \sum_{i=1}^{N_p} (x_k^{i*} - \hat{x}_k)(x_k^{i*} - \hat{x}_k)^T \quad (5.16)$$

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 Equation (5.10) 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:

$$p(z_k|x_k^i) = \frac{1}{\sqrt{2\pi|R^*|}} e^{(-\frac{1}{2}[z_k-h(x_k^i)]^T(R^*)^{-1}[z_k-h(x_k^i)])} \quad (5.17)$$

where  $R^* = R + \Sigma$ ,  $\Sigma$  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 Equation (5.16).

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 Equation (5.15) 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:

$$\|\Delta DMF_k\| \leq 1.96(W^{\frac{1}{2}}J)^\dagger \quad (5.18)$$

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

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 Equation (5.18) 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.

#### 5.4.8 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

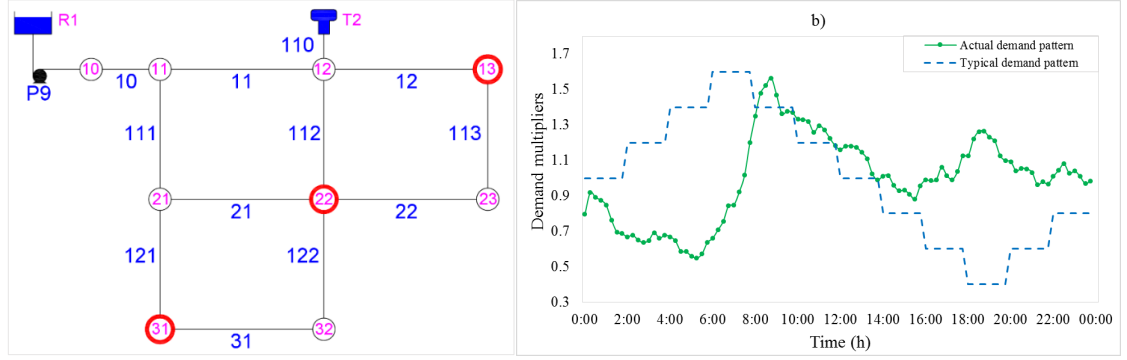
operators ( $P_c$ ,  $P_m$  and  $N_{gen}$ ). The prediction sub-model requires the data of typical demand patterns, the auto-regression coefficient ( $\phi_i$ ) and the variance of noise of demand residuals ( $\sigma_h^2$ ). These parameters are calibrated independently based on historical demand data 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.

## 5.5 Case study 1

The first case study used to evaluate the model is shown in Figure 5.2a. 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 5.2b). 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 5.2b) was selected as the typical demand pattern. Different values of the auto-regression coefficient ( $\phi$ ) as well as variance of noise ( $\sigma_h^2$ ) were applied for the demand prediction sub-model.





**Figure 5.2:** (a) Case study 1 network; (b) typical and actual demand patterns case study 1 network

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 5.1.

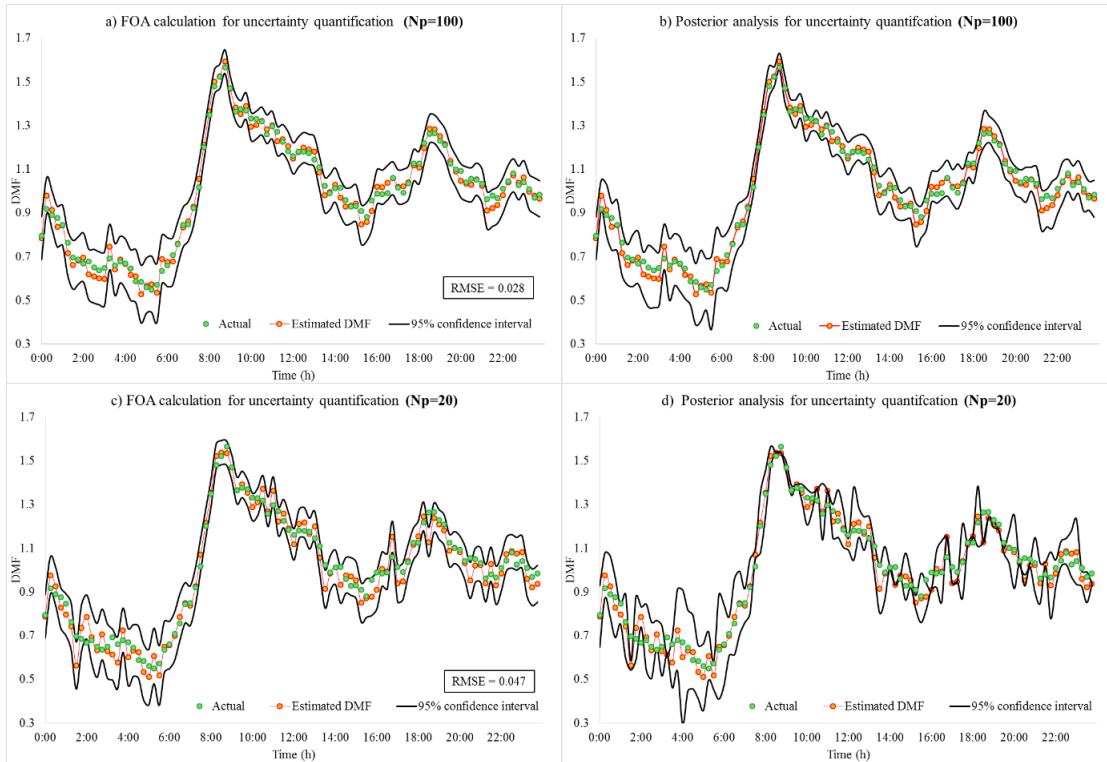
**Table 5.1:** Coefficient of determination ( $R^2$ ) and root mean squared error (RMSE) of demand estimates corresponding to different parameter values of the demand prediction model for case study 1

No	Auto-regression coefficient ( $\phi$ )	Variance of demand residual ( $\sigma_h^2$ )	$R^2$	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	<b>0.988</b>	<b>0.028</b>
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

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  $0.25 \leq \sigma_h^2 \leq 0.64$ . Due to the large difference between the typical demand value and the actual demand value at each time step (Figure 5.2b), 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 5.3a.

In Figure 5.3a, 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.



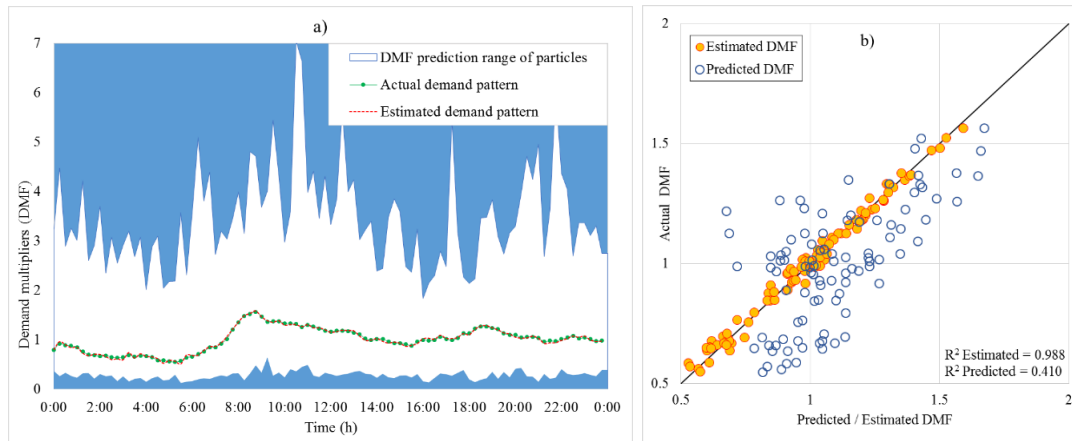
**Figure 5.3:** Estimated demand pattern and confidence intervals: (a), (c) uncertainty quantification based on first-order approximation (FOA) method (Equation (5.18)) for  $N_p = 100$  and  $N_p = 20$ ; (b), (d) uncertainty quantification based on variance of the particle samples (Equation (5.16)) for  $N_p = 100$  and  $N_p = 20$

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. Equation (5.18)) and the posterior analysis (i.e. Equation (5.16)) for uncertainty quantification, as shown in Figures 5.3b, 5.3c and 5.3d. Figures 5.3a and 5.3c show the uncertainty quantified by the FOA method while Figures 5.3b and 5.3d 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 5.3c, and 5.3d. The uncertainty boundary calculated by the FOA method in Figure 5.3c 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 Equation (5.16) in Figure 5.3d are collapsed into single value at some time steps due to an insufficient number of the particles. Application of Equation (5.16) 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 5.4a. The predicted values range from  $DMF_{min} = 0.1$  to  $DMF_{max} = 7.0$ , 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 5.4b 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.



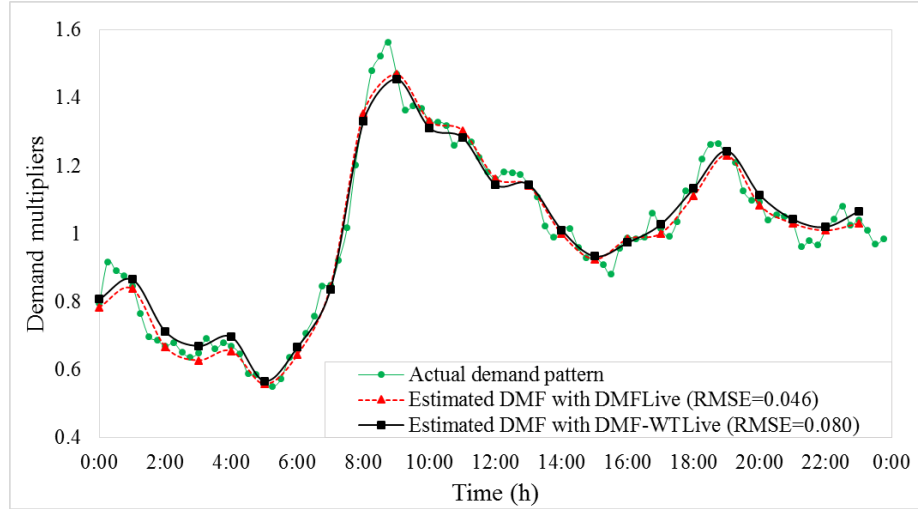
**Figure 5.4:** (a) Prediction range of demand multipliers during simulation period; (b) Predicted demand multipliers and estimated demand multipliers

### 5.5.1 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 5.5 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.



**Figure 5.5:** Estimated demand patterns with and without tank level updated

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 5.2 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  $5942.43 \text{ m}^3/\text{day}$ ,  $46.81 \text{ m}^3/\text{day}$  (or  $0.78\%$ ) less than the actual water usage. On the other hand, total estimated water usage from DMF-WTLive model was  $6007.31 \text{ m}^3/\text{day}$ , only  $18.07 \text{ m}^3/\text{day}$  (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.

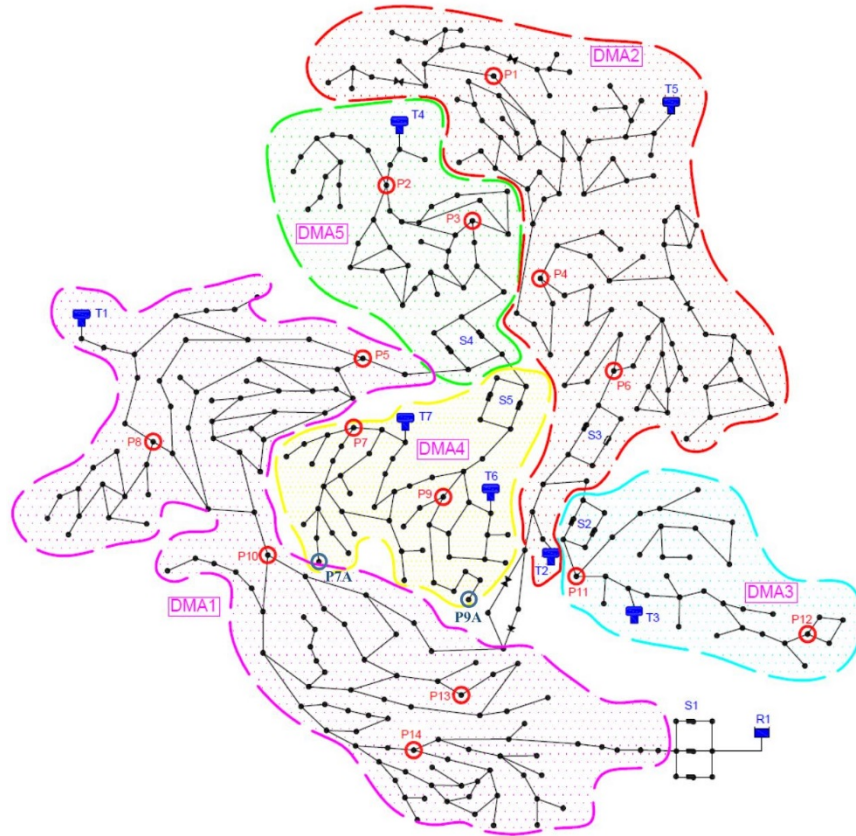
**Table 5.2:** Actual and estimated total volume of water usage during calculated period

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

## 5.6 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 5.6. 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.

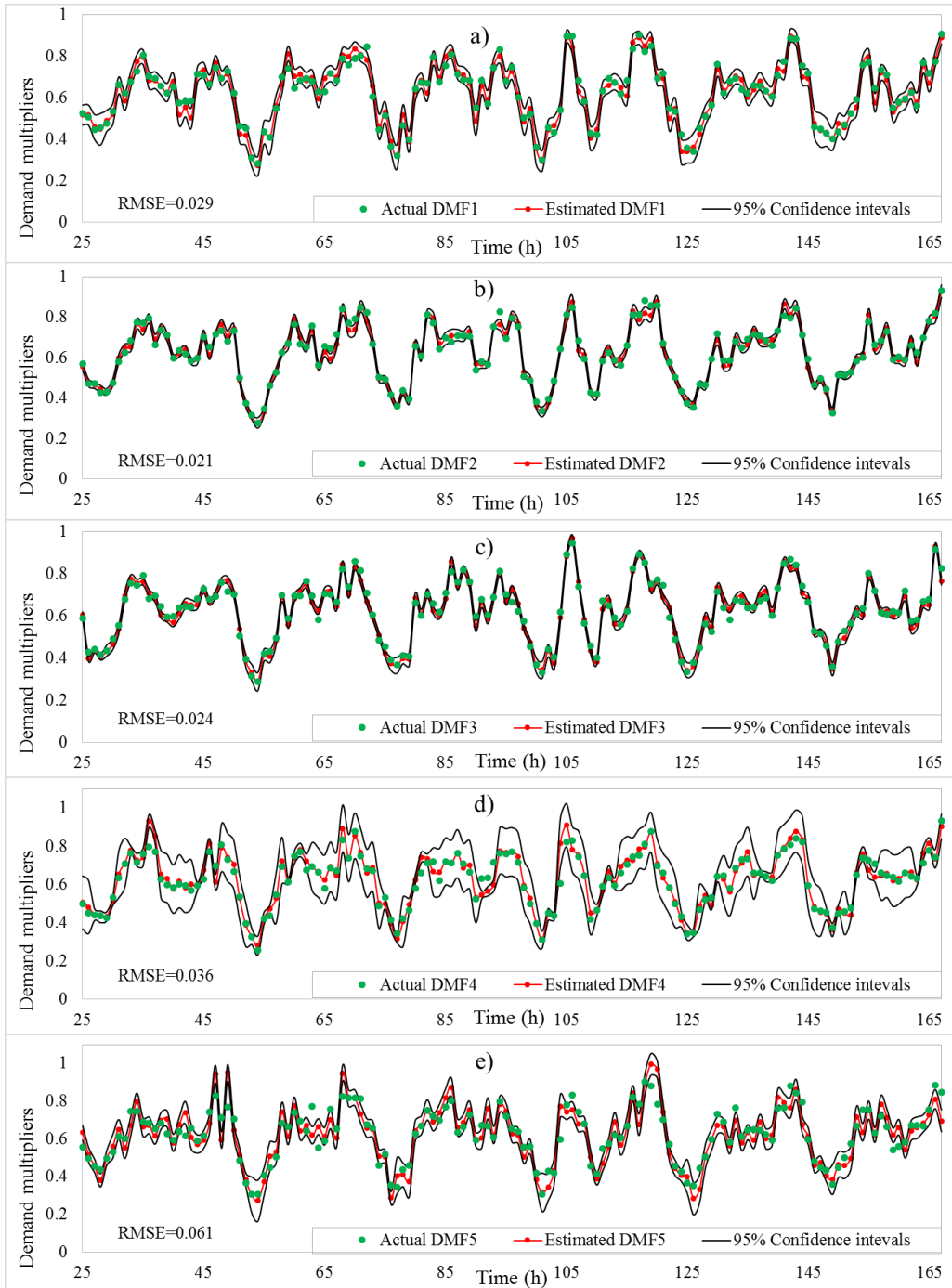


**Figure 5.6:** Case study 2 network (the C-town network)

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 5.7. 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 5.7 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 5.7a, 5.7b and 5.7c, 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 5.7d) for most of the time. However, due to the locations of the measurements (P7 and P9 - Figure 5.6), 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 5.7e,

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.



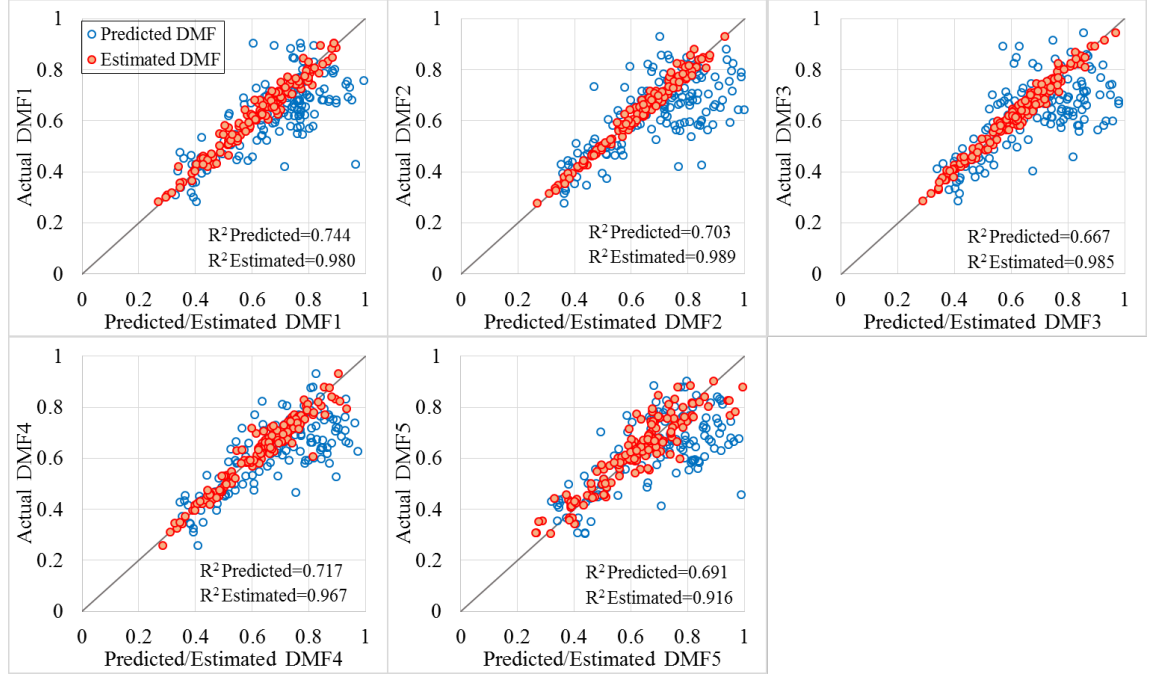
**Figure 5.7:** Five estimated demand patterns for case study 2 network ( $N_p = 25,000$ ) using DMFLive-I

Figure 5.8 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.



**Figure 5.8:** Scattergrams and coefficients of determination for five estimated demand patterns in case study 2

### 5.6.1 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 (Equation (5.17)); and (2) by the integration of a GA process into the systematic resampling of the model, herein referred as DMFLive-III model.

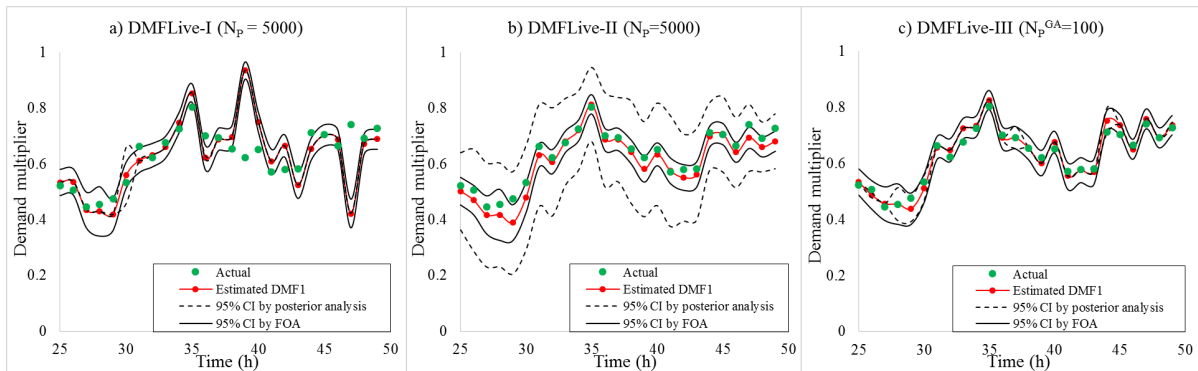
Table 5.3 presents results (in terms of the RMSE of each demand pattern) of running these models with  $N_P = 1000$  and  $N_P = 5000$  for DMFLive-I, II and with  $N_P^{GA} = 20$

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  $1.43 \cdot 10^5$  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  $3.575 \cdot 10^6$ ). This means the computation can be reduced by approximately a factor of five times.

**Table 5.3:** Performance of DMFLive model with SR (I), modified likelihood function (II) and SRGA (III)

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

Figure 5.9 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 Equation (5.16). As can be seen from Figures 5.9a and 5.9c, 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 5.9b computed by this method is wide due to the incorporation of the forecasting nodal heads into the likelihood function.



**Figure 5.9:** DMF 1 uncertainty ranges from 25 to 49 hours computed by FOA method and posterior analysis

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 5.9b and 5.9c) these ranges can cover the actual values most of the time.

### **5.6.2 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  $J$  (Equation (5.18)), 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.

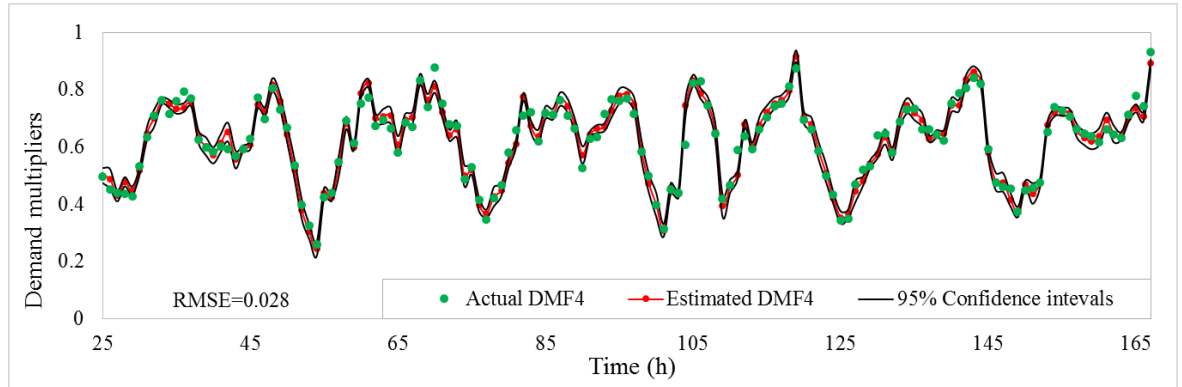
Figure 5.10 shows the sensitivity matrixes  $J_0$  (for the original test) and  $J_0^A$  (for the modified test) corresponding to a set of estimated values DMFs = [0.46; 0.54; 0.65; 0.47; 0.62]

a)						b)					
DMF 1	DMF 2	DMF 3	DMF 4	DMF 5	Measurements	DMF 1	DMF 2	DMF 3	DMF 4	DMF 5	Measurements
0	22.37	0	0	0	P1	0	22.37	0	0	0	P1
0	0	0	0	3.72	P2	0	0	0	0	3.72	P2
0	0	0	0	10.81	P3	0	0	0	0	10.81	P3
0	18.80	0	0	0	P4	0	18.80	0	0	0	P4
4.31	0	0	0	0	P5	4.31	0	0	0	0	P5
0	18.80	0	0	0	P6	0	18.80	0	0	0	P6
0	0	0	2.59	0	P7	0	0	0	5.31	0	P7A
2.95	0	0	0	0	P8	2.95	0	0	0	0	P8
0	0	0	2.55	0	P9	0	0	0	11.76	0	P9A
3.20	0	0	0	0	P10	3.20	0	0	0	0	P10
0	0	23.32	0	0	P11	0	0	23.32	0	0	P11
0	0	22.39	0	0	P12	0	0	22.39	0	0	P12
9.75	0	0	0	0	P13	9.75	0	0	0	0	P13
4.77	0	0	0	0	P14	4.77	0	0	0	0	P14

**Figure 5.10:** Sensitivity matrixes of nodal heads at measurement locations wrt demand multipliers at t=0

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 5.11, are much narrower than the confidence intervals of the estimated DMF 4 for the original test presented in Figure 5.7d. 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 Perez (2014).



**Figure 5.11:** Estimated DMF 4 and its confidence interval with the relocated measurement 7A and 9A

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.

## 5.7 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.

## 5.8 Appendix A

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:

$$y_i^{Meas} = y_i(x) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_i) \quad (\text{A5.1})$$

where  $x$  is the  $nd$  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  $x$ ;  $\varepsilon_i$  is the residual between model prediction and observation, which is assumed to be Gaussian with mean of zero and standard deviation of  $\sigma_i$ ;  $y_i^{Meas}$  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:

$$f(x) = \frac{1}{2} \sum_{i=1}^m \left( \frac{y_i(x) - y_i^{Meas}}{\sigma_i} \right)^2 = \frac{1}{2} \sum_{i=1}^m \varepsilon_R^2 \quad (\text{A5.2})$$

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

where  $m$  is the number of measurement sites,  $\varepsilon_R$  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:

$$\nabla f_0 = J(x^0)^T W (y(x^0) - y^{Meas}) \quad (\text{A5.3})$$

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:

$$H_0 = J(x^0)^T W J(x^0) \quad (\text{A5.4})$$

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 Equation (A5.2) by a quadratic function at  $x^0$  leads to the approximation of  $x$ :

$$x = x^0 - (H_0)^{-1} \nabla f_0 \quad (\text{A5.5})$$

By replacing Equation (A5.3) and Equation (A5.4) into Equation (A5.5), the approximation of  $x$  can be expressed as:

$$x = x^0 - (J(x^0)^T W J(x^0))^{-1} J(x^0)^T W (y(x^0) - y^{Meas}) \quad (\text{A5.5}')$$

Using Equation (A5.1):

$$x(\varepsilon) = x^0 + (J(x^0)^T W J(x^0))^{-1} J(x^0)^T W \varepsilon \quad (\text{A5.6})$$

The influence of the measurement errors with regards to the parameter estimates, therefore, can be obtained at the first-order of Equation (A5.6):

$$\Delta x = (J(x^0)^T W J(x^0))^{-1} J(x^0)^T W \varepsilon = (W^{\frac{1}{2}} J(x^0))^{\dagger} W^{\frac{1}{2}} \varepsilon = (W^{\frac{1}{2}} J(x^0))^{\dagger} \varepsilon_R \quad (\text{A5.7})$$

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

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

$$\begin{aligned} \|\Delta x\| &\leq 2.58 \left\| \left( J(x^0)^T W J(x^0) \right)^{-1} J(x^0)^T W^{\frac{1}{2}} \right\| = 2.58 \left\| (W^{\frac{1}{2}} J(x^0))^{\dagger} \right\| \\ |\Delta x_i| &\leq 2.58 \sum_{j=1}^m |S_{ij}|, \text{ with } S = (W^{\frac{1}{2}} J)^{\dagger} \end{aligned} \quad (\text{A5.8})$$

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

$$\begin{aligned} \|\Delta x\| &\leq 1.96 \left\| (W^{\frac{1}{2}} J(x^0))^{\dagger} \right\| \\ |\Delta x_i| &\leq 1.96 \sum_{j=1}^m |S_{ij}|, \text{ with } S = (W^{\frac{1}{2}} J)^{\dagger} \end{aligned} \quad (\text{A5.9})$$

## References

References listed below are also included in Bibliography. In addition, the final published paper in Appendix 2 has the references listed.



# Chapter 6    Localization of inadvertently partially closed valves in water distribution systems

## Publication 3

Nhu C. Do<sup>1</sup>, Angus R. Simpson<sup>1</sup>, Jochen W. Deuerlein<sup>2</sup>, Olivier Piller<sup>3</sup>

<sup>1</sup> - *School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide SA 5005, Australia.*

<sup>2</sup> - *Senior Researcher, 3S Consult GmbH, Karlsruhe, Germany.*

<sup>3</sup> - *Senior Researcher, Irstea UR ETBX, Dept. of Water, Cestas, France.*

*Journal of Water Resources Planning and Management*

Submitted July 2017

This page has been intentionally left blank.

# Statement of Authorship

Title of Paper	Localization of inadvertently partially closed valves in Water distribution systems
Publication Status	<input type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input checked="" type="checkbox"/> Submitted for Publication <input type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style
Publication Details	Do, N.C., Simpson, A.R., Deuerlein, J.W. & Piller, O. 2017, 'Localization of inadvertently partially closed valves in Water distribution systems', <i>Journal of Water Resources Planning and Management</i> , Submitted July 2017

## Principal Author

Name of Principal Author (Candidate)	Do, N.C.		
Contribution to the Paper	Development and implementation of methodology, design of case studies, interpretation and analysis of results, preparation of manuscript and acting as corresponding author.		
Overall percentage (%)	70%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature	<table border="1"> <tr> <td>Date</td> <td>3 August 2017</td> </tr> </table>	Date	3 August 2017
Date	3 August 2017		

## Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Simpson, A.R.		
Contribution to the Paper	Research supervision and manuscript evaluation		
Signature	<table border="1"> <tr> <td>Date</td> <td>4 August 2017</td> </tr> </table>	Date	4 August 2017
Date	4 August 2017		

Name of Co-Author	Deuerlein, J.W.		
Contribution to the Paper	Research supervision and manuscript evaluation		
Signature	<table border="1"> <tr> <td>Date</td> <td>11 July 2017</td> </tr> </table>	Date	11 July 2017
Date	11 July 2017		

Name of Co-Author	Piller, O.		
Contribution to the Paper	Research supervision and manuscript evaluation		
Signature	<table border="1"> <tr> <td>Date</td> <td>11 July 2017</td> </tr> </table>	Date	11 July 2017
Date	11 July 2017		



## Abstract

The reliability of a water distribution system is highly dependent on the management of its pipeline network. A pipe or a portion of the network can be isolated for inspection, maintenance and replacement by the installation of isolation valves along the pipelines.

However, the presence of isolation valves may cause a large discrepancy in the hydraulic behavior between the real system and results from a simulation model if the statuses of some of the valves in the system are unknown. Possible problems related to these valves are missing valves in the model due to poor or non-existent documentation, errors in data transfer or valve mechanical failure.

This paper introduces an innovative methodology for the identification of unknown partially/fully closed valves in a water distribution network. An optimization problem is formulated for the unknown valve issue and solved by application of three sequentially applied methods, which include: a local sensitivity analysis, an application of genetic algorithms and an application of the Levenberg-Marquardt algorithm. In the first method, the sensitivity of the flow rates and nodal heads at measurement locations with respect to the change in the minor losses of the valves is computed. This computation is used to identify the valves that are unable to be localized by the measurement data. The second method applies a genetic algorithm combined with an extended period simulation in order to preliminarily identify the locations of the partially/fully closed valves and their setting values, i.e. the degree of opening of the valve. Finally, the application of the Levenberg-Marquardt (LM) algorithm was implemented to correct the results from the GA model. Results and discussions from two case studies show that the proposed methodologies can solve real world problems.

**Keywords:** *Partially/fully closed valves, unknown valve status, Genetic Algorithms, Sensitivity analysis, Levenberg- Marquardt Algorithm, Water distribution systems.*

## 6.1 Introduction

Water distribution systems (WDS), which are built to provide adequate quantity and quality of drinking water to customers, are complex infrastructure consisting of a number of diverse components such as pumping stations, junctions, links, tanks and valves. The use of valves in distribution networks is aimed at improving the reliability in operations as well as maintenance of the systems. Various types of valves are installed with two main purposes: either controlling pressures and flows or isolating pipes from the supply system for replacement, modification and repair.

Control valves are automated devices that are used to regulate the flows or pressures of water through the distribution piping. The hydraulic behavior at local regions of the network where these valves are placed can be adjusted by the valve settings. For example, a flow control valve limits the flow passing through the valve to a specific flow setting value, or a pressure reducing valve (PRV) is set to prevent the downstream pressure from exceeding a value that could cause damage to the system (Walski et al. 2003). Because of their important role in the operation of WDS, control valves are usually considered in water distribution modeling. In various hydraulic software models, such as EPANET (Rossman 2000), modeling of these valves requires their settings or statuses (open/closed) to be known. The hydraulic models, based on these settings, find the hydraulic conditions of the system by solving the hydraulic network equations, simultaneously computing the flow rates and head losses of the valves.

Another approach for modelling of control valves is the determination of the valve settings. Piller and Bremond (2001) formulated a least squares method that minimizes the difference between targeted settings and computed values to find the control valve states. Deuerlein et al. (2005) applied a game-theoretic algorithm to calculate the correct minor loss coefficients of interacting pressure and flow control valves. In another paper Deuerlein et al. (2009) dealt with the presence of flow control and check valves in water network models by applying content and co-content theory (to ensure the existence and uniqueness

of the solution), sub-differential analysis (to handle the non-differential flow versus head loss relationship of flow control and check valves) and then solving the water network equations as a constrained nonlinear programming problem. Piller and van Zyl (2014) introduced a method associated with the Karush-Kuhn-Tucker equations for the modeling of control valves in extended period simulations.

Numerous studies have also focused on the optimal regulation and location of control valves to manage water losses by leakages or pipe bursts in WDS. PRVs are usually considered in this research since PRVs can maintain the targeted pressures at downstream of the valves regardless how large the upstream pressures are. The problem is usually formulated as a nonlinear optimization problem where the objective is to minimize the leak flows by minimizing surplus pressure at leak locations, and the decision variables are the valve locations and settings of valve operating strategies. Examples are Sterling and Bargiela (1984); Savic and Walters (1995); Liberatore and Sechi (2009) and Nicolini and Zovatto (2009). For the same pressure management purpose, an alternative to PRVs is the application of pumps as turbines (PAT) at locations that have enough difference in elevation (e.g. Carravetta et al. (2013), Barbarelli et al. (2016) and de Marchis et al. (2016)). This technical solution can provide both the system flexibility and economical benefits for the network.

Isolation valves, the most common type of valves in WDS, are used to close off and block any flow through pipes (Van Zyl 2014). These valves are usually placed at the ends of a pipe, around junctions or at critical locations of a WDS. The AWWA (1996), GLUMB (2003) Ozger and May (2004), Walski et al. (2006) and Jun and Loganathan (2007) have listed all the rules of thumb for the placement of isolation valves, which include: (1) isolation valves should be located at not more than 500 foot (152 m) intervals in commercial districts and at not more than one block of 800 foot (244 m) intervals in other districts; (2) at least  $(N_P - 1)$  valves are required around a junction to which  $N_P$  links are connected while an ideal (or fully valved) system requires two valves at the ends of each pipe; (3) valves must be placed at all critical locations such as at each city block or

at each hydrant lateral and (4) no more than four valves are required to close to isolate a pipe in order to reduce the complexity of the valving system.

The primary role of isolation valves is to isolate some portions of the system for inspection, replacement or maintenance. The reliability of the WDS, defined as the ability of the system to provide uninterrupted demands at adequate pressure to customers, is therefore, highly dependent on this type of valves and their numbers in the network. The roles of valves in determining the reliability of the WDS when a pipe breaks have been discussed in Bouchart and Goulter (1991). In their paper, the impact of isolation (due to pipe failure) was proposed to be measured by calculating a volume of deficit, i.e. the water volume of demands located in the isolated portion of the system. A methodology was also introduced to identify the number of isolation valves based on the computation of this deficit volume. Similarly, Walski (1993) described the importance of isolation valves in providing WDS reliability. The paper pointed out that the link-node model, a common representation of the WDS, failed to account for the indispensable role of isolation valves and their impact of disconnecting pipes and other components from the network. The paper then defined the term “segment” as the portion of the network isolated by valves, which was used in an approach to visually evaluate the adequacy of valves in the system.

One of the main aspects affecting the reliability of a WDS is the location of the isolation valves. Ozger and May (2004) emphasized that it is only possible to isolate a pipe break and determine the deficient performance of the network if the valve layout is known. Design of the isolation valve system to ensure the connectivity of the network, therefore, is of great interest for researchers. Various studies have investigated the valve placement problem using different techniques, for example Simulated Annealing (Ozger and May 2004), Genetic Algorithms (Giustolisi and Savic 2010) and Multi Objective Genetic Algorithms (MOGA) such as NSGAI (Creaco et al. 2010). In Jun and Loganathan (2007), a mathematical method was introduced, which represents the presence of isolation valves by a valve location matrix and a valve deficiency matrix. This method was used to evaluate the connectivity of the network as well as to detect unintended isolations when



some of the valves in the system were closed. Some other studies also looked at the connectivity of the WDS such as Davidson et al. (2005) and Ostfeld (2005).

It should be noted that most of the aforementioned studies focused on the reliability of the WDS, which solely considered the fully closed status of the valve. Given the thousands of isolation valves in a network, there is always a possibility that one or some of the valves/valve statuses are unaccounted for or unknown in the model. Possible problems related to these valves may be for a number of reasons such as: missing valves in a hydraulic model due to poor or non-existent documentation, errors in data transfer, valve mechanical failure or temporarily closed valves during inspection/rehabilitation times without adequate reporting. As a result, a huge discrepancy may result between the performance of the real system and its hydraulic simulation model. In such cases, calibration of unknown valve statuses and identification of their location is required. However, little effort has previously investigated the calibration and identification problems of unknown valve statuses and their corresponding locations. Previous research includes Delgado and Lansey (2008) and Wu et al (2012). In the Delgado and Lansey (2008) paper, a transient model was used to detect a closed/partially closed valve in a single pipeline, while Wu et al (2012) applied Genetic Algorithms for an extended period simulation of the EPANET model to calibrate partially closed valve settings.

In general, calibration of a water network model can be formulated as an optimization problem, in which the calibrated parameters are adjusted by an optimization method such as mathematical algorithms (e.g. modified Newton Raphson method (Shamir 1974), the generalized reduced gradient method (Lansey & Basnet 1991), the Levenberg-Marquardt algorithm (Koppel & Vassiljev 2012), singular value decomposition (Sanz & Perez 2015)) or evolutionary optimization algorithms (e.g. Genetic Algorithms (Abe & Peter 2010)). An objective function is usually associated with these optimization models, which is the minimization of the residuals between measured values and their corresponding simulated values from the hydraulic model. Mathematical optimization methods, or in other words, gradient-based optimization methods are robust in solving even and overdetermined

problems where the number of variables is equal or less than the number of measurements. In order to find the search directions, these methods compute a Jacobian matrix, which is relatively challenging when applied to an extended period simulation. As a result, simplification of the problem is usually required so that the problem is solvable (e.g. Lansey & Basnet (1991)). For underdetermined problems (i.e. the number of variables is more than the number of measurements) such as the unknown valve statuses problem in this study, the application of gradient methods is likely to be unstable since the Jacobian and approximated Hessian calculated during the optimization procedure are often ill-conditioned. Additional steps might be used, such as Tikhonov regularization, Choleski decomposition or QR factorization, however, their solutions are still dependent on the initial guesses.

Evolutionary optimization methods, on the other hand, are gradient free methods and can easily overcome these limitations. In fact, the search of evolutionary algorithms is based on the use of an external simulation model in order to assess the goodness of the solutions. Computing derivatives to define the direction of the search is not necessary, which allows these algorithms to easily escape local optima. However, given the random component in an evolutionary algorithm search, they usually require long computational times (e.g. Do et al. 2016) and, depending on the parameters chosen, may lack the exploitation of the information of the search in space. In this last case, they may provide a final solution that is not a local or global optimum or they may require more computational effort to achieve these solutions.

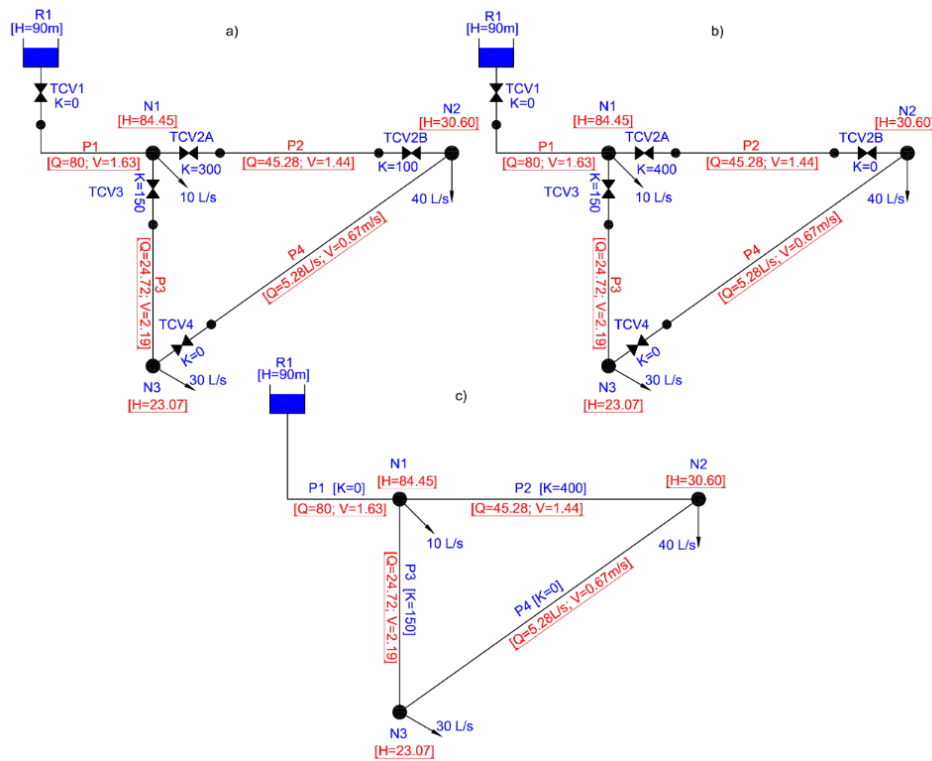
For the reasons given above, this paper proposes a hybrid methodology which takes advantage of both mathematical optimization algorithms and evolutionary optimization algorithms to calibrate and, more importantly, identify partially closed valves in a water distribution network. By integrating measurement data into the demand driven model EPANET (Rossman 2000), this problem can be solved via three sequential steps: (1) a sensitivity analysis, which calculates the sensitivity of flow rates and pressure heads at measurement locations with respect to (wrt) the variations of the minor loss in the valves,

to remove unidentifiable valves; (2) an application of a Genetic Algorithm (GA) model during an extended period simulation (e.g. 24, 48 hours) to reduce the size of the search region as well as to provide a preliminary estimate of the settings of the valves; and (3) an application of the Levenberg-Marquardt algorithm to localize the regions of partially closed valves and correct the settings of the valves.

The paper is structured as follows. First, the modeling of partially closed valves in EPANET hydraulic model is presented. Second, the methodology to solve the problem is introduced. This is followed by the application of the methodology for two case studies. Finally, conclusions are given.

## 6.2 Modeling of partially closed valves in EPANET

A throttle control valve (TCV) is usually used to model a partially closed valve in EPANET, such as in Wu et al. (2012). A TCV is modelled as a short smooth pipe that also contains a local loss. An example of a simple network with TCVs that follow the  $(N - 1)$  valves' rule is shown in Figure 6.1a and 6.1b with different TCV settings. In this network valve TCV1 is placed in pipe P1, valves TCV 2A and 2B are placed in pipe P2 while TCV3 and TCV4 are placed in pipes P3 and P4, respectively. The setting value (i.e. the dimensionless valve coefficient)  $K = 0$  corresponds to a fully opened status of the valve or, in other words, the valve is inactive. The setting value  $K > 0$  corresponds with the partially opened status of the valve.



**Figure 6.1:** Modelling of partially closed valves in EPANET, in which simulated values are underlined and in SI unit (mH<sub>2</sub>O for nodal heads, L/s for flow rates and m/s for velocities)

Simulating the two networks in Figure 6.1a and 6.1b in EPANET shows that flow rates, velocities in pipes and nodal heads can be the same with different settings of TCV2A and TCV2B (located at pipe P2). Although the settings of these two valves are different in the two cases, they result in the same total minor loss. For example, the setting of TCV2A is 300 and TCV2B is 100 (with the total of  $\Sigma K = 400$ ) in Figure 6.1a while the corresponding values in Figure 6.1b are 400 and 0 (also with the total of  $\Sigma K = 400$ ). In both cases, the flow in pipe P2 is 45.28 L/s and the pressure at the extremity nodes are 84.45 m and 30.60 m as shown in Figure 6.1.

The same hydraulic behavior can be simulated for the network in which all valves have been removed (shown in Figure 6.1c) with the minor loss setting of the pipe equal to the total setting of all TCVs belonging to that pipe. Therefore, the partially closed settings and statuses of the valves can be incorporated into the pipes by changing the minor loss in the pipe. The problem of identifying partially closed valve locations and their settings using EPANET now becomes the problem of finding the minor losses in each of the pipes. Correct locations of partially closed valves are required to be searched for along the pipe

where the valves are located. Fully closed valve can be simulated by changing the fixed status of the pipe from “Open” into “Closed” in EPANET.

The minor loss in a pipe caused by a partially closed valve is computed as:

$$h_v = mQ|Q| \quad (6.1)$$

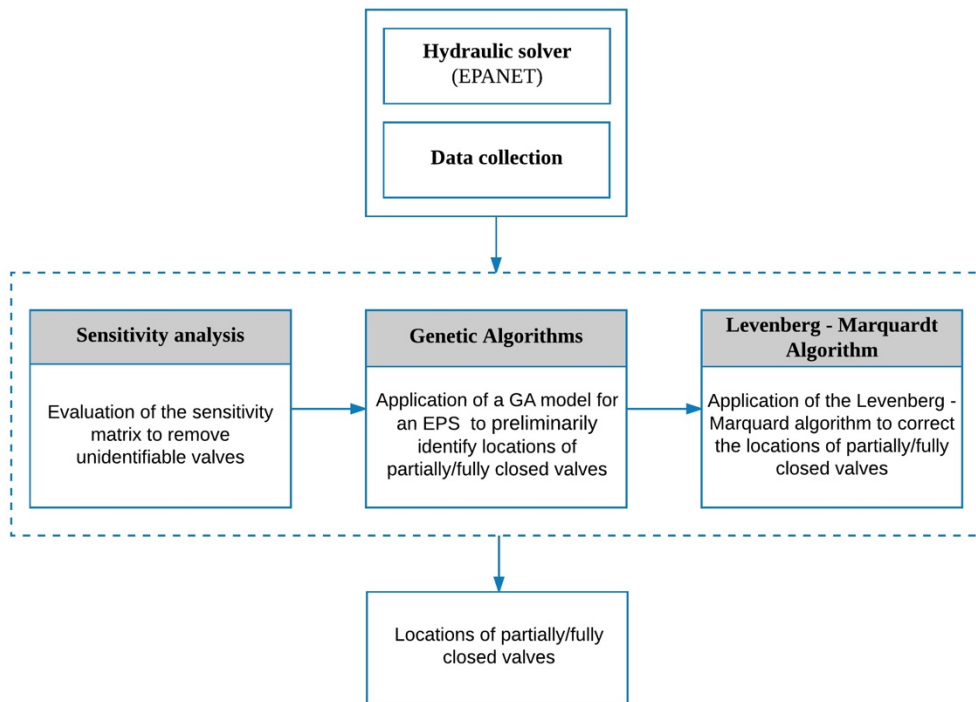
where  $Q$  is the flow through the pipe and  $m$  is the minor loss coefficient which is expressed as:

$$m = \frac{K}{2gA^2} = \frac{8K}{g\pi^2 D^4} \quad (6.2)$$

where  $K$  is dimensionless valve coefficient, which represents different opening states of the valve.  $D$  and  $A$  are the pipe diameter (in SI units) and its corresponding area, respectively.

### 6.3 Methodology

As previously mentioned, the problem of finding partially closed valves of a WDS in this paper is considered as the problem of finding cumulated minor losses in the pipes where in reality the valves are located. The proposed solution approach is based on the minimization of the difference between observed and calculated values for flows and pressures and consists of three sequential stages: (1) observability of valve statuses by the computation of the sensitivity of the flow rates and nodal heads at measurement locations wrt the change in the minor loss coefficients of the pipes, (2) the application a of genetic algorithm to an extended period simulation and (3) the application of the Levenberg-Marquardt (LM) algorithm to a steady state simulation. The flowchart of the methodology is shown in Figure 6.2, in which all stages involve the use of the EPANET model to simulate the behavior of the system.



**Figure 6.2:** Process for identifying partially closed valves in WDS

### 6.3.1 Observability based on sensitivity analysis

In WDSs, measurement devices are usually installed to observe system flows and pressures or to detect abnormal situations such as pipe bursts or hydraulic transients. Usually, these devices have not been installed for the purpose of finding partially/fully closed valves (herein, refers as pipes with minor losses) and the sensor network may not cover the entire distribution system. As a result, a sensitivity analysis step is required in order to identify “unobservable pipes”. An unobservable pipe is defined as a pipe for which any change in the status or minor loss has no impact on the measured values. This analysis requires the computation of the sensitivity of the pipe flow rates and nodal heads in the network wrt variation of minor loss coefficients. Pipes that are insensitive to the measured locations will be removed from the problem as their potentially closed valves will not be able to be identified. The sensitivity matrix can be computed by considering the hydraulic steady state of the water network that solves a partially nonlinear problem of the continuity equations at nodes and energy equations for pipes taking into account minor losses (adapted from Piller et al. (2016)):

$$-A^T q - d = 0_{nj} \quad (6.3)$$

$$g(r(q), q) + Bm - Ah - A_f e_l = 0_{np} \quad (6.4)$$

where  $A$  and  $A^T$  are the unknown head node incidence matrix and its transpose matrix, which provide information about the connectivity of the nodes and the links in the network (i.e. an element  $a_{ij}$  of the matrix  $A$  is assigned a value of -1 or 1 if link  $j$  enters or leaves node  $i$ , respectively. Otherwise,  $a_{ij}$  is assigned a value of 0);  $q$  and  $h$  are the flow rate and nodal head vectors of  $np$  links and  $nj$  nodes;  $d \in R^{nj}$  is the vector whose elements are the values of water consumption at nodes;  $g(r(q), q)$  is the vector of the friction head loss function values, which depends on the link resistance factors  $r(q)$  and flow rates  $q$ ;  $B$  is the diagonal matrix with diagonal elements  $q_i |q_i|$  and  $m$  is the column vector of minor loss coefficients for the  $np$  links (where some  $m_i$  are zeros);  $A_f$  is the fixed-head nodes incidence matrix and  $e_l$  is the elevation vector of fixed-head nodes.

Denote  $\nabla_x = \left( \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right)$  the partial derivative of the function  $f$  wrt the parameter  $x$ .

Differentiating Equation (6.3) and Equation (6.4) wrt the minor loss coefficients, gives:

$$F \nabla_m q + B + M \nabla_m q - A \nabla_m h = D \nabla_m q - A \nabla_m h + B = 0_{np, np} \quad (6.5)$$

$$-A^T \nabla_m q = 0_{nj, np} \quad (6.6)$$

where  $F = \nabla_q g(r(q), q)$ . Formulae of  $F$  for different flow regimes (laminar, transitional and turbulent) can be found in detail in Piller et al. (2016);  $M$  is the diagonal matrix whose diagonal elements are  $2m_i |q_i|$  and  $D = F + M$ .

Multiplying Equation (6.5) by  $A^T D^{-1}$  on the left and adding to Equation (6.6)

$$-A^T D^{-1} A \nabla_m h + A^T D^{-1} B = 0_{nj, np} \quad (6.7)$$

The sensitivity of the heads  $h$  wrt  $m$  is, therefore, given by:

$$\nabla_m h = (A^T D^{-1} A)^{-1} A^T D^{-1} B \quad (6.8)$$

The sensitivity of the flows  $q$  wrt  $m$  follows by:

$$\nabla_m q = D^{-1} A (A^T D^{-1} A)^{-1} A^T D^{-1} B - D^{-1} B \quad (6.9)$$

Finally, the sensitivity matrix of size  $(N_H + N_Q, np)$  is the matrix selected from  $[\nabla_m h; \nabla_m q]$  where the number of rows is the total number of measured locations  $[N_H + N_Q]$

( $N_H$  is the number of pressure measurements and  $N_Q$  is the number of flow measurements) and the number of columns is equal to the number of pipes  $np$  in the network.

A pipe flow rate or a nodal head that is not sensitive to the variation of the minor loss of a pipe in the network results in a zero value at its corresponding element in the sensitivity matrix. Therefore, the pipes that are insensitive to all of the measured locations of the network will have all zero values in their column of the sensitivity matrix. These corresponding pipes are unobservable and need to be removed from the set of unknowns in order to reduce the search space size of the problem. The number of remaining pipes, which may contain minor losses, are selected as the number of decision variables in the GA model.

### 6.3.2 Application of the GA model

Genetic algorithms have been applied in WDS modeling and analysis for over three decades. This optimization technique has been used for various purposes such as optimal network design (e.g. Simpson et al. (1994), Dandy et al. (1996)), optimal operation of WDS (e.g. Marchi et al. (2016), Blinco et al. (2016)), event detection (e.g. Vitkovsky et al. (1999), Preis and Ostfeld (2008)) or for parameter calibration in WDS (e.g. Lingireddy (2002), Nhu et al. (2016)). In this paper, a GA model is introduced, which integrates a genetic algorithm, the EPANET toolkit and measurement data during an extended period simulation to determine possible status values and locations of partially/fully closed valves in the network in a preliminary way (i.e. the statuses of the valves can only be discrete values, which are based on an integer coding scheme of the GA model). The GA model is mainly used to reduce the size of the search space of the valve problem, which enables the application of a Levenberg-Marquardt mathematical method in the third stage.

The objective function of the GA model has been formulated by the sum of squared residuals between the measured and simulated values of pipe flow rates and nodal heads at measurement locations during a selected period of time (24, 48 hours,...), expressed as:



$$\text{Min } G(K) = \frac{1}{2} \sum_{t=1}^{N_T} \left[ \sum_{i=1}^{N_Q} w_i (Q_i^{\text{Sim}} - Q_i^{\text{Meas}})^2 + \sum_{j=1}^{N_H} w_j (H_j^{\text{Sim}} - H_j^{\text{Meas}})^2 \right] + \text{pen} \quad (6.10)$$

where  $Q_i^{\text{Sim}}, H_j^{\text{Sim}}$  are the simulated flow rate and nodal head for the  $i^{\text{th}}$  pipe and  $j^{\text{th}}$  node, respectively;  $Q_i^{\text{Meas}}, H_j^{\text{Meas}}$  are the measured flow rate and head for the  $i^{\text{th}}$  pipe and  $j^{\text{th}}$  node;  $N_Q, N_H$  are number of head and flow measurement sites in the network;  $w_i, w_j$  are weighting factors applied to different terms to ensure they have similar units as well as to reflect the actual accuracy of measurement devices.  $N_T$  is the number of time steps during the simulation period. Finally, a penalty term ( $\text{pen}$ ) is added into the objective function to constrain cases in which the EPANET solver generates negative pressures at some nodes of the network. The penalty term is defined by the multiplication of a constant penalty value ( $k = 10$ ) with the number of times that EPANET gives negative pressure warnings during the extended period simulation.

The decision variables for the GA model are the minor loss setting values  $K$  (the loss coefficients of pipes in the EPANET model) for the remaining pipes as selected from the sensitivity analysis. The range of the decision variables ( $0 \leq K \leq K_{max}^{GA}$ ) that defines the size of the search space of the valve problem can be selected based on the boundary conditions given by Wu et al. (2012), which is  $K_{max}^{GA} = 15,000$ . The value of  $K = 0$  represents the situation with no minor loss in the pipe, i.e. all the valves in that pipe are fully opened. A value of  $K = 15,000$  has been chosen to represent the situation for a closed valve in the pipe. In this case, the status of the pipe in the EPANET solver is switched to “Closed”, which does not allow any flow through the pipe. This relatively small value of  $K_{max}^{GA}$  (compared to maximum value  $K_{max}^{LM}$  used in the LM method, which will be explained later in the paper) has been selected in order to reduce the search space of the GA and improve its convergence speed.

Tournament selection, two-point crossover and bitwise mutation were chosen for the GA model. In addition, an integer coding scheme has been applied. The values of the loss setting  $K$ , therefore, are taken as discrete values, with the increment of  $\Delta K$  being selected based on the size of specific network. The pipes with positive  $K$  values from the GA model

are used as initial values for and will be refined by the LM algorithm presented in the following section. Pipes with  $K = 0$  found from the GA model are assumed to be fully opened and not considered in the next step.

### 6.3.3 Application of the Levenberg-Marquardt algorithm

The preliminary search of the loss coefficient settings of the partially/fully closed valves in a water network using a GA model during an extended period simulation can be refined by the application of the LM algorithm, which is applied for a steady state simulation. At each time step, the valve problem can also be formulated as a nonlinear differentiable least squares minimization:

$$\text{Min } g(K) = \frac{1}{2} \sum_{i=1}^{N_Q} w_i (Q_i^{\text{Sim}} - Q_i^{\text{Meas}})^2 + \sum_{j=1}^{N_H} w_j (H_j^{\text{Sim}} - H_j^{\text{Meas}})^2 = \frac{1}{2} \|w_\alpha^{0.5} (y^{\text{Sim}}(K) - y^{\text{Meas}})\|^2 \quad (6.11)$$

where  $y^{\text{Sim}}$  and  $y^{\text{Meas}}$  are the simulated flow rates (or nodal heads) and their corresponding measured values at measurement locations, respectively.

A solution of the nonlinear least squares problem can be found by applying a modification of the LM method that considers all the constraints. The minor losses in the pipes are estimated with the iterative formula:

$$\begin{cases} m^{i+1} = m^i - (J_i^T J_i + \lambda_i S_i)^{-1} J_i^T r_i; & K^{i+1} = 2gA^2 m^{i+1} \\ \text{if } K^{i+1} \leq 0, & \text{set } K^{i+1} = 0 \rightarrow m^{i+1} = 0 \\ \text{if } K^{i+1} \geq K_{max}^{LM}, & \text{set } K^{i+1} = K_{max}^{LM} \rightarrow m^{i+1} = m_{max}^{LM} \end{cases} \quad (6.12)$$

where:  $m^i$  is the minor loss coefficient associated with the minor loss setting  $K^i$  via Equation (6.12).  $J_i$  and  $J_i^T$  are the Jacobian matrix and its transpose at iteration  $i$ . The Jacobian matrix is a selection of the rows of  $\nabla_m h$  (in Equation 6.8) and  $\nabla_m q$  (in Equation 6.9) corresponding to the measurement locations.  $\lambda_i$  is the damping factor for the LM algorithm;  $S = \text{diag}(J_i^T J_i)$ ;  $(r_i)_\alpha = \sqrt{w_\alpha} (y_\alpha^{\text{Sim}}(K) - y_\alpha^{\text{Meas}})$  is the weighted residual between the  $\alpha^{\text{th}}$  simulated values and measured values at iteration  $i$ . Note that for GA model, a maximum value of  $K_{max}^{GA} = 15,000$  is used above, at which the status of the pipe is switched from "Open" to "Closed". In this case, no flow is allowed through the pipe. A zero flow pipe has no impact on the GA model as it does not require computation of the Jacobian matrix. However, a zero flow in pipes may cause singularity in the Jacobian

matrix and prematurely stops the LM method. Therefore, a large value  $K_{max}^{LM}$  is assigned to a pipe instead of closing it. This value allows a small flow through the pipe but only causes a small difference on the simulated values at measurement locations compared to having fully closed pipes. For the two case studies in this paper, a maximum value of  $K_{max}^{LM} = 500,000$  was selected, which has been shown to be sufficient to represent an almost closed pipe in both case study networks.

A number of necessary conditions, based on the gradient of  $g(K)$  and its Hessian, are required in order to achieve a local optimal and a unique solution  $\hat{K}$  of the problem, which includes: (1) the problem needs to be overdetermined, i.e. the number of pipes selected from the GA model is less than the number of measurements; (2) the residuals  $(r_i)_\alpha$  are small or equal to zero at the optimum; and (3) the Jacobian matrix  $J(\hat{K})$  is of full rank (see Piller (1995) for the detailed derivation of the gradient and Hessian of  $g$ ). In complex networks, these conditions are usually not satisfied due to a limited number of measurements, which means that the correct setting of the partially closed valves cannot be determined by the LM method. However, the methodology is still useful to identify the locations of the partially closed valves in the networks. This will be shown in the second case study of the paper where the condition of full rank for the Jacobian matrix is not satisfied.

For the LM method, two stopping criteria are applied: either the change of the K values after every iteration is less than  $\|\Delta K_i\| = 0.01$  or the number of iterations reaches  $N_{iter} = 200$ .  $N_{iter}$  is used to terminate the method if the function in Equation (6.11) cannot converge to a solution.

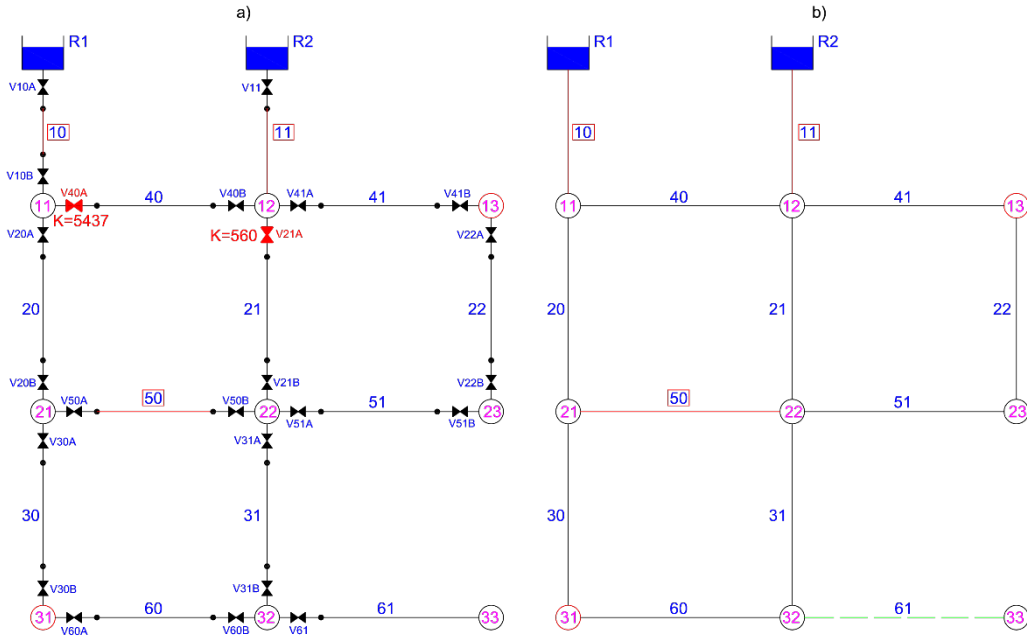
## 6.4 Case studies

A relatively small network and a much larger network of realistic size are used to demonstrate the possible use of the proposed methodology to find partially/ fully closed valves. The exact solutions (i.e. the valve setting values) in both cases are given and used in the networks to create synthetic measurement data. The problems associated with noisy

data, which results from measurement devices, demand aggregation or caused by incomplete system knowledge, are not considered in this paper.

#### 6.4.1 Case study 1

The first case study is shown in Figure 6.3. The network has two reservoirs, 9 nodes, 13 pipes that contain 24 valves. The following assumptions are made for the network: (1) five measurement devices (three flow measurements and two pressure measurements) are placed at pipe P10, pipe P11, pipe P50, Node 13 and Node 31, respectively; (2) two partially closed valves V40A ( $K = 5437$ ) and V21A ( $K = 560$ ) are located at pipe P40 and pipe P21 and (3) all nodal demands follow a single demand pattern. The proposed methodology is applied to locate these partially closed valves given a collection of measurement data during 48 hours with a time step  $t = 1$  hour.



**Figure 6.3:** (a) Case study 1 network; (b) Results from sensitivity analysis and GA model

The total inflow ( $\sum Q_t^{In}$ ) and outflow ( $\sum Q_t^{Out}$ ) to the system at a time step  $t$  is identified based on two measured flows at pipes P10 and P11. The demand multiplier of the demand pattern at each time step is required to be known in the demand driven model EPANET, therefore, can be computed as:

$$DMF_t = \frac{\sum Q_t^{In} - \sum Q_t^{Out}}{\sum_{i=1}^{nj} D_i^0} \quad (6.13)$$

where  $D_i^0$  is the base demand of node  $i^{th}$  in the network.

Table 6.1 shows the transpose of the sensitivity matrix at a steady state  $T = 0$  ( $DMF_0 = 0.616$ ) of flow rates and nodal heads at measured locations wrt the variation of minor losses in all pipes of the network. It can be seen that the sensitivity for pipe P61 results in zeros at all measured locations, which means that this pipe is insensitive to the measured values. As a result, the pipe is required to be removed from the search process of the further stages. Note that fully closed status of the valves are not considered in the sensitivity analysis due to the fact that this status can change the topology of the network, which may cause inaccurate removal of pipes. For example, if a valve in pipe P31 is fully closed, pipe P60 becomes insensitive wrt the measurements and would need to be removed.

**Table 6.1:** Sensitivity of flow rates ( $m^3/s$ ) and nodal heads (m) at measurement locations wrt variation of minor losses ( $s^2/m^5$ )

Pipe ID	$\partial h / \partial m$		$\partial q / \partial m$		
	Node 13	Node 31	Pipe 10	Pipe 50	Pipe 11
40	$-6.2*10^{-8}$	$2.0*10^{-5}$	$-2.1*10^{-7}$	$7.1*10^{-9}$	$2.1*10^{-7}$
41	$-1.0*10^{-3}$	$-1.4*10^{-5}$	$3.8*10^{-8}$	$2.4*10^{-8}$	$-3.8*10^{-8}$
50	$-2.8*10^{-6}$	$1.7*10^{-5}$	$-1.9*10^{-7}$	$-3.0*10^{-7}$	$1.9*10^{-7}$
51	$-2.3*10^{-5}$	$1.9*10^{-5}$	$-5.1*10^{-8}$	$-3.2*10^{-8}$	$5.1*10^{-8}$
60	$-1.7*10^{-6}$	$8.6*10^{-5}$	$-1.2*10^{-7}$	$6.6*10^{-8}$	$1.2*10^{-7}$
61	0	0	0	0	0
20	$-1.1*10^{-4}$	$-1.4*10^{-2}$	$-7.7*10^{-6}$	$-4.8*10^{-6}$	$7.7*10^{-6}$
21	$-4.3*10^{-5}$	$-7.1*10^{-4}$	$1.9*10^{-6}$	$1.2*10^{-6}$	$-1.9*10^{-6}$
22	$5.8*10^{-7}$	$-4.8*10^{-7}$	$1.3*10^{-9}$	$8.2*10^{-10}$	$-1.3*10^{-9}$
30	$-1.3*10^{-5}$	$-3.6*10^{-3}$	$-8.7*10^{-7}$	$4.9*10^{-7}$	$8.7*10^{-7}$
31	$1.9*10^{-6}$	$-9.5*10^{-5}$	$1.3*10^{-7}$	$-7.3*10^{-8}$	$-1.3*10^{-7}$
10	$-4.5*10^{-4}$	$-4.7*10^{-2}$	$-9.3*10^{-5}$	$-1.7*10^{-5}$	$9.3*10^{-5}$
11	$-1.5*10^{-2}$	$-3.8*10^{-3}$	$2.1*10^{-5}$	$3.8*10^{-6}$	$-2.1*10^{-5}$

The application of the GA model, after removing all unidentifiable pipes, has 12 decision variables. The objective function of the GA model follows Equation (6.10), with  $N_T = 48$  hours,  $N_Q = 3$  and  $N_H = 2$ . The following parameters are set for the GA model: the probability of crossover  $P_c = 0.75$ , the probability of mutation  $P_m \approx 1/L = 0.085$  ( $L = 12$  is the length of a chromosome), population size  $N = 100$  and the number of generations was selected as  $N_{stop} = 1000$ . In Wu et al. (2012), the range of the decision variables was selected from 0 to 15,000 with the increment of 1000. With this relatively

small network, the increment is selected  $\Delta_K = 500$  in order to improve the accuracy of the GA model.

The GA model were undertaken on the Intel ® Core™ i5 (2.9GHz) computer. The total elapsed run time for this case study was approximately 40 minutes. The model finds that three pipes P40, P21 and P22 contain minor losses with the  $K$  values respectively equal to  $K_{P40} = 5500$  (actual  $K_{P40} = 5437$ ),  $K_{P21} = 500$  (actual  $K_{P21} = 560$ ) and  $K_{P22} = 500$  (actual  $K_{P22} = 0$ ), as shown in Figure 6.3b. The minor losses of the other pipes are all equal to zero, i.e. all the valves in these pipes are fully opened. In this case, it can be seen that the GA model does not find the  $K$  values appropriately. Thus, a refinement step (i.e an application of the LM algorithm) for the three possible partially closed valves is implemented.

Table 6.2 summarizes the results of the LM method given three possible partially closed valves obtained from the GA model. The acronyms RSS and GRAD represent the residual sum of squares Equation (6.11) and the norm of the gradient along the search direction, respectively. The damping factor  $\lambda_i$  starts at  $\lambda_0 = 0.0001$  and is changed throughout the iterations based on the least squares criterion as follows: if RSS decreases,  $\lambda_i$  is set to be decreased 60% (i.e. multiplied by a factor of 0.4) for the next iteration. Otherwise,  $\lambda_i$  is forced to be increased by a factor of 10 and a new least squares value is recalculated based on the new value of  $\lambda_i$ . Within the iteration, this process is repeated until the least squares value decreases.

**Table 6.2:** Convergence of case study 1 network with two partially closed valves

Iteration #	RSS	GRAD	$\lambda_i$	$K_{P40}$	$K_{P21}$	$K_{P22}$
0	0.0523	$5.2158 \times 10^{-6}$	0.0001	5500	500	500
1	$4.213 \times 10^{-5}$	$5.2276 \times 10^{-8}$	$4.00 \times 10^{-5}$	4933.201	545.863	0
2	$6.561 \times 10^{-6}$	$8.1804 \times 10^{-8}$	$1.60 \times 10^{-5}$	5264.362	555.559	0
3	$1.822 \times 10^{-7}$	$1.2480 \times 10^{-8}$	$6.40 \times 10^{-6}$	5409.255	559.307	0
4	$2.343 \times 10^{-9}$	$3.3520 \times 10^{-9}$	$2.56 \times 10^{-6}$	5435.21	559.959	0
5	$1.840 \times 10^{-14}$	$3.3516 \times 10^{-9}$	$1.02 \times 10^{-6}$	5436.919	559.997	0
6	$1.225 \times 10^{-14}$	$3.3515 \times 10^{-9}$	0.1024	<b>5436.991</b>	<b>559.999</b>	<b>0</b>
<i>Actual values</i>				<b>5437</b>	<b>560</b>	<b>0</b>

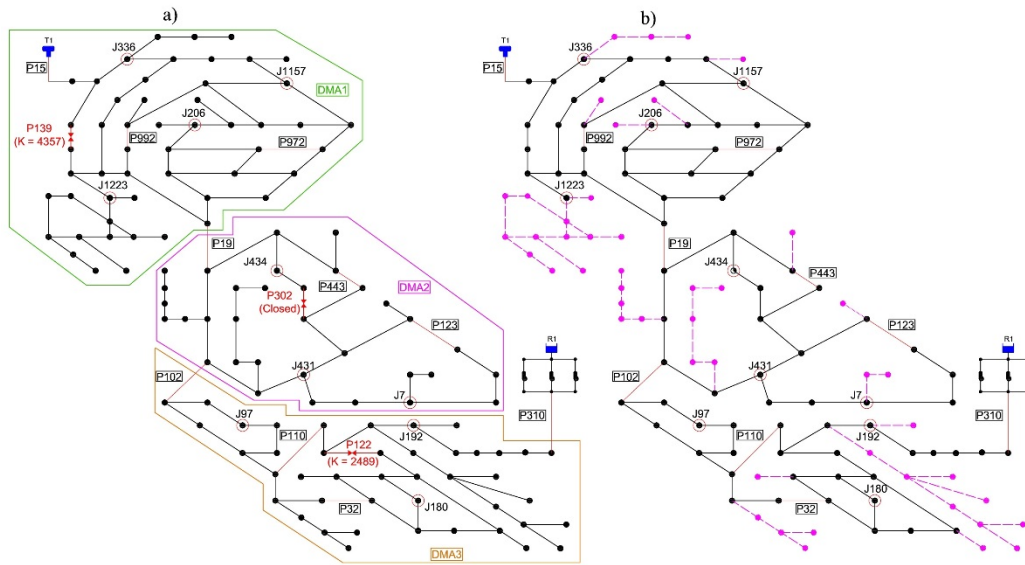
The method is terminated after six iterations (with the total run time of 5.5 seconds), which found very accurately that two valves, one in pipe P40 and the other in pipe P21, are partially closed with the loss setting values  $K_{P40} = 5436.991$  (actual  $K_{P40} = 5437$ ) and  $K_{P21} = 559.999$  (actual  $K_{P21} = 560$ ), respectively. The status of the valve in pipe P22 is corrected by the LM method, which becomes fully opened with  $K_{P22} = 0$ . The Jacobian matrix is checked after each iteration to verify that the requirement of full rank is satisfied. The residual of the observations and the simulated values represented by the RSS values is approximately zero. As a result, the LM method converges to the true solution as highlighted in Table 6.2.

#### 6.4.2 Case study 2

The second case study (shown in Figure 6.4a) is used to describe the complexity of the problem in practice as well as to evaluate the applicability of the proposed methodology. This realistic size network is part of the C-town network (Ostfeld et al. 2011), which has 147 pipes, 135 nodes, 1 reservoir, one tank and a pump station with 3 pumps. The network is divided into 3 DMAs, where the demand multipliers factors can be computed by measuring the inflows and outflows to each DMA. The valving system of the network follows the  $2*N$  valve rule (Jun and Loganathan 2007), corresponding to 294 valves across the network. Among them, three valves are assumed to be both partially closed (a valve in Pipe 122 and another in Pipe 139) and fully closed (a valve in Pipe 302). These valves are required to be localized by the proposed methodology. For this case study, the setting values of  $K_{P122} = 2489$ ,  $K_{P139} = 4357$  and fully closed state of the valve in Pipe 302 are used to synthesize the measurement data. Figure 6.4a also displays the locations of 20 measurement devices (10 flow measurements and 10 pressure measurements), which are assumed to be placed within the network in order to record data at every hour. Data for a period of 48 hours are available to be used in the GA model.

The sensitivity analysis shows that 50 out of 147 pipes (shown in Figure 6.4b as the dashed lines) are not sensitive to the measured locations and hence, are removed from the localization process. The number of the remaining pipes is selected as the number of

decision variables in the GA model, which is 97 variables. The GA model is implemented with the following parameters: the probability of crossover  $P_c = 0.75$ , the probability of mutation  $P_m = 0.01$ , population size  $N = 1000$  and the number of generations  $N_{stop} = 2000$ . Similar to Wu et al. (2012), the minor loss setting  $K$  is selected from 0 to 15,000 with an increment of 1000. The value of  $K = 15,000$  corresponds to fully closed status of the pipe. The total run time of the GA model for this case study was approximately 4.6 hours.

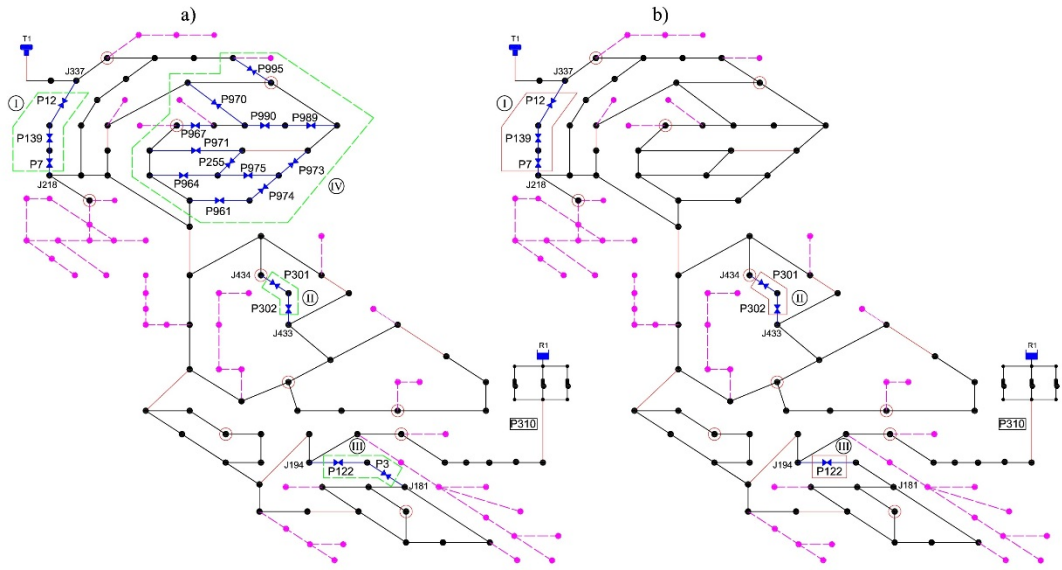


**Figure 6.4:** (a) Case study 2 Network; (b) Results from sensitivity analysis

Figure 6.5a shows the output from the GA model. It is seen that the model has significantly reduced the size of the search space, from 97 pipes to 19 pipes that possibly contain minor losses. These 19 pipes can be divided into 4 areas. Area I is located between Node J337 and Node J218 where three pipes P12, P139 and P7 are connected in series. Area II and area III both have two links P301, P302 and P3, P122, respectively. These three areas cover the locations of the partially/fully closed valves that need to be determined, which are two partially closed valves in P139 (area I) and in P122 (area III) and a fully closed valve in P302 (area II). The last area is the region of the remaining pipes, i.e. 12 out of 19 pipes. This large set of pipes is selected by the GA model because the small flow rates in these pipes make them almost insensitive to the variation of the minor losses.



Figure 6.5b shows an example of the results from the LM model at time step  $T = 2$ , where the LM method converges to a very good solution (this is not always the case as it will be shown in Table 6.4). It took only approximately 10.2 seconds to complete a run of the LM model. The results point out that all pipes in area IV no longer have minor losses, which means all the valves in this area are fully opened. On the other hand, all pipes in areas I and II contain minor losses while in area III, the LM method indicates that only pipe 122 contains a minor loss.



**Figure 6.5:** (a) Results from GA model; (b) Results from LM model at time step  $T=2$

The values of the minor losses for each pipe, RSS, GRAD and  $\lambda$  are shown in Table 6.3. In order to obtain smaller value of RSS at each iteration, the value of  $\lambda$  is altered within the iteration. For example, the first iteration required four modifications of  $\lambda$  (from 0.0001 to 0.1) before moving to the second iteration. The method stops after 55 iterations, with the sum of square of the residuals approaching zero. It is observed that the Jacobian matrix for this case study problem does not have full rank ( $r = 8$ ). As a result, the LM method does not converge to the true solution. In area I, the result from the LM method showed that all the pipes have minor losses ( $K_{P12} = 1643.45$ ,  $K_{P139} = 1643.60$  and  $K_{P7} = 2575.53$ ), while actually only a valve in P139 is partially closed with  $K_{P139} = 4357$ . In area II, the LM model finds that pipe 302 is fully closed (represented by the maximum value of  $K = 500,000$ ), which matches the actual status

of the valve in this pipe. However, the minor loss setting in pipe P301 identified by the model is  $K_{P301} = 67,826.89$  while the actual value in this case is  $K = 0$ . Only in area III, the minor loss setting value of pipe P122 from the model is found to be close to the actual value with  $K_{P122} = 2490.51$  ( $K_{actual} = 2489$ ).

The results for case study 2 are to be expected due to the non-uniqueness of solutions for the problem. Non-unique solutions are caused by two main reasons. The first issue is caused by pipes connected in series and the second is the observability problem due to the limited number of measurement devices.

**Table 6.3:** Correction of valve setting values and statuses by the LM model at time step T=2

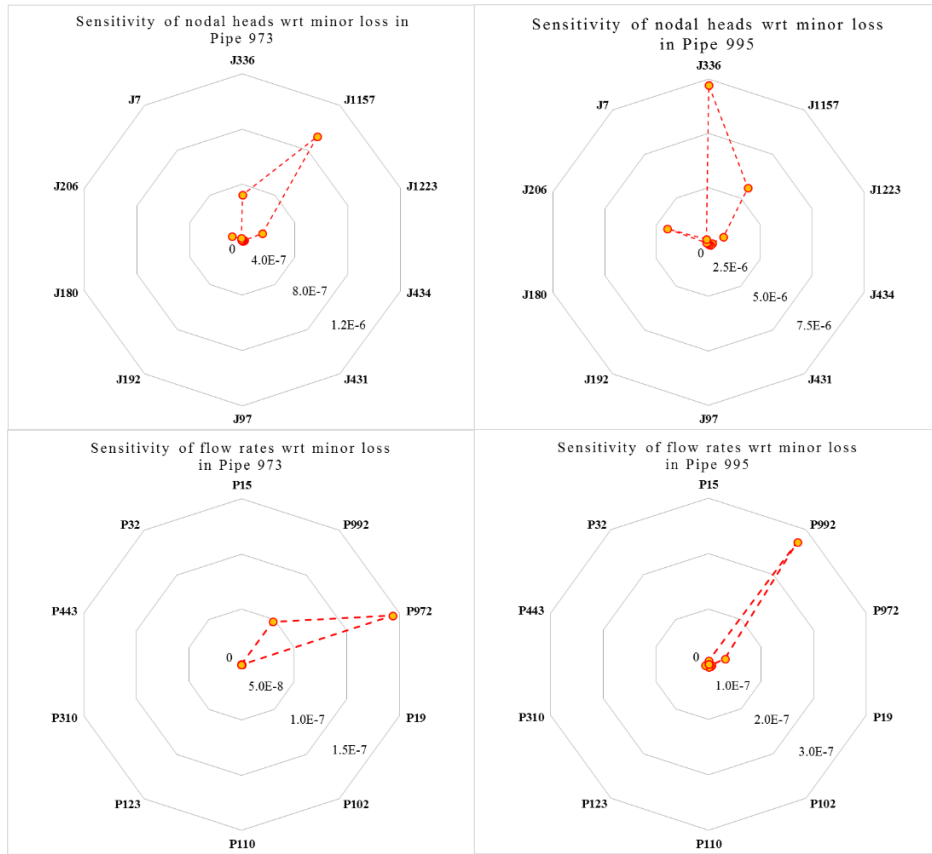
Areas	Pipe	Iteration 0	Iteration 1	Iteration 2	...	Iteration 55	Actual values
I	P12	1000	2966.74	2607.06	...	1643.45	<b>0</b>
	P139	1000	0.00	76.42	...	1643.60	<b>4357</b>
	P7	1000	2941.48	2589.31	...	2575.53	<b>0</b>
II	P301	2000	0	820.37	...	67,826.89	<b>0</b>
	P302	15,000	0	951.11	...	500,000	<b>Closed</b>
III	P3	1000	0	81.79	...	0	<b>0</b>
	P122	2000	5599.47	4782.58	...	2490.51	<b>2489</b>
IV	P255	9000	0	0	...	0	<b>0</b>
	P961	8000	423,735.00	500,000	...	0	<b>0</b>
	P964	8000	73,575.14	500,000	...	0	<b>0</b>
	P967	15,000	0	0	...	0	<b>0</b>
	P970	5000	500,000	0	...	0	<b>0</b>
	P971	2000	0	0	...	0	<b>0</b>
	P973	1000	0	24,898.93	...	0	<b>0</b>
	P974	1000	500,000	0	...	0	<b>0</b>
	P975	1000	398,411.88	0	...	0	<b>0</b>
	P989	1000	0	2092.74	...	0	<b>0</b>
	P990	15,000	389,393.48	0	...	0	<b>0</b>
P995	1000	9297.22	20,494.59	...	0	<b>0</b>	
Final	RSS	13.28	11.34	7.29	...	$9.92 \cdot 10^{-10}$	
	Grad	0.0022	0.0004	0.0003	...	$3.35 \cdot 10^{-9}$	
	Lamda	0.0001	0.1	100	...	1.93	

The first issue of the valve problem is the non-unique solution of minor loss settings for pipes in series. As discussed previously (see Figure 6.1), multiple combination of  $K$  values can result in the same total energy loss. This is valid also for pipes with different flows and velocities, if the sum of their minor losses ( $K_i V_i^2 / 2g$ ) is the same. Hence, if there are no intermediate measurements along a set of pipes in series, it is not possible to define

exactly the minor loss setting of each single valve. The non-uniqueness of solutions of the valve problem can be seen more clearly if there is a closed valve in the series. An example is area III for case study 2. Since a valve in pipe P302 is closed, pipe P301 disconnects from its loop and becomes insensitive to the nodal heads or flow rates at measurement locations. Therefore, the loss setting value in P301 can accept any value without changing the values at measurement locations. In other words, the problem does not have a unique solution.

A possible way to overcome this problem is to consider the minor loss for only one single pipe of the series where the others are assumed to have no minor loss (i.e. these pipes no longer need to be evaluated). This consideration can be implemented as a preprocessing step in order to reduce the search space and regularize the problem. If the method finds that there exists a minor loss in that pipe, it will be necessary to physically check all the valves along the series of the pipes in this area.

The second issue is related to the observability of the network. Although the number of the measurements is larger than the number of the pipes with unknown valve statuses, the problem may still be underdetermined due to limited number of the measurements locations at some parts of the system. Specifically, the sensitivity matrix shows that only the nodal heads and flow rates at 6 measurement locations in the DMA1 of the network are sensitive to the minor losses of 12 pipes in area IV. Figure 6.6 plots an example of the sensitivity in pipe P973 and pipe P995.



**Figure 6.6:** Sensitivity of nodal heads and flow rates at measurement locations wrt minor loss of pipe P973 and P995

It is seen that the nodal heads at nodes J1157, J336, J1123 and J206 are sensitive to the variation of the minor losses in pipes P973 and P995. The sensitivity of nodal heads at other locations is trivial, which results in approximately zero values in the sensitivity matrix. Similarly, only the flow rates for pipes P992 and P972 are sensitive to the change in minor losses of these two pipes. Apparently, the determination of 12 unknown minor loss settings based on these 6 measurement values cannot obtain a unique solution.

Due to the lack of observability within area IV of the network, the results for these pipes in this area are relatively unstable when the LM method is applied for different time steps. Table 6.4 presents the results of the LM model applied for 11 time steps from  $T = 0$  to  $T = 10$ .

It is seen that the fully opened statuses of all the pipes in area IV is found at  $T = 2, 3, 4$ . For other time steps, one or several pipes in this area are found to contain minor losses. This shows the importance of choosing the number of the sensors and their locations so

that an observable system with high sensitivity can be obtained. As the main aim of the methodology presented in this paper is to identify the locations of partially/fully closed valves, the results from the LM model are acceptable in this case. With the application of the proposed methodology, the partially/fully closed valve problem has been simplified from the search of 147 pipes (i.e. 304 valves) to 7 pipes (as at  $T = 2, 3, 4$ ) or to a maximum of 14 pipes (as at  $T = 9$ ), which is equivalent to the search of only 14 valves or maximum of 28 valves in the network.

**Table 6.4** Minor loss setting from LM method for 11 time steps from T=0 to T=10 (Case study 2)

Area	Pipe	T=0	T=1	T=2	T=3	T=4	T=5	T=6	T=7	T=8	T=9	T=10
I	P12	1663.20	1642.71	1643.46	1644.59	1645.35	1644.95	11.14	1666.50	1675.03	1662.68	1506.98
	P139	1615.64	1643.07	1643.60	1644.81	1645.55	1645.15	781.55	1645.76	1642.81	1592.55	1595.94
	P7	2248.78	2575.13	2575.53	2580.08	2582.24	2580.95	8416.09	2458.38	2390.91	2107.50	3412.72
II	P301	62,239.8	73,956.0	67,826.9	76,379.6	85,827.9	74,886.8	69,546.6	75,072.6	63,970.9	67,278.7	57,920.4
	P302	500,000	454,150	500,000	500,000	500,000	500,000	500,000	0	500,000	132,701	500,000
III	P122	1677.90	2504.00	2490.51	2501.67	2505.67	2494.48	0	1375.77	1378.43	2031.57	2130.06
	P3	1565.92	0	0	0	0	0	3668.92	1958.81	2087.70	935.76	825.79
IV	P255	0	0	0	0	0	16.729	16.806	0	0	0	0
	P961	0	0	0	0	0	0	0	0	0	0	0
	P964	0	0	0	0	0	0	0	26.25	0	0	0
	P967	34.81	0	0	0	0	0	0	526.05	404.21	938.50	0
	P970	0	30.93	0	0	0	0	0	660.99	89.62	40.32	0
	P971	0	0	0	0	0	0	0	105.19	245.42	161.14	0
	P973	0	0	0	0	0	0	0	0	0	0	0
	P974	0	0	0	0	0	0	0	0	0	0	0
	P975	0	0	0	0	0	0	0	0	36.26	82,326.4	0
	P989	61.94	0	0	0	0	0	0	1168.28	500,000	272.15	3645.99
	P990	194.18	0	0	0	0	0	0	0	0	12.33	961.17
	P995	0	0	0	0	0	0	0	0	0	0	14.06
RSS		1.19*10 <sup>9</sup>	6.59*10 <sup>9</sup>	9.9*10 <sup>10</sup>	2.28*10 <sup>9</sup>	1.95*10 <sup>9</sup>	4.9*10 <sup>10</sup>	8.4*10 <sup>10</sup>	2.1*10 <sup>8</sup>	1.6*10 <sup>8</sup>	4.21*10 <sup>9</sup>	1.6*10 <sup>8</sup>

In this case study, if the correct setting values are required, this observability problem can only be solved by installing additional measurements into the underdetermined area or by running additional ad hoc tests (e.g. opening fire hydrants). Moreover, considering all the observations over an extended period simultaneously could improve the observability of the system, but it would require modifications the LM method and additional computation effort as presented in the following sub-section.

#### *Consideration of extended period simulation in LM method*

The output from the GA model can be fine-tuned by the application of the LM method for an extended period simulation. In this case, the objective of the LM model is similar to the objective function of the GA model (i.e. Equation (6.10)) without the penalty term:

$$\text{Min } G(K) = \frac{1}{2} \sum_{t=1}^{N_T} \left[ \sum_{i=1}^{N_Q} w_i (Q_i^{Sim} - Q_i^{Meas})^2 + \sum_{j=1}^{N_H} w_j (H_j^{Sim} - H_j^{Meas})^2 \right] \quad (6.14)$$

The minor losses in the pipes are estimated with the following expression:

$$\begin{cases} m^{i+1} = m^i - \left( \sum_{t=1}^{N_T} J_{i,t}^T J_{i,t} + \lambda_i S_{i,t} \right)^{-1} \sum_{t=1}^{N_T} J_{i,t}^T r_{i,t}; & K^{i+1} = 2gA^2 m^{i+1} \\ \text{if } K^{i+1} \leq 0, & \text{set } K^{i+1} = 0 \rightarrow m^{i+1} = 0 \\ \text{if } K^{i+1} \geq K_{max}^{LM}, & \text{set } K^{i+1} = K_{max}^{LM} \rightarrow m^{i+1} = m_{max}^{LM} \end{cases} \quad (6.15)$$

where  $J_{i,t}$  and  $r_{i,t}$  are the Jacobian matrix and the weighted residuals, respectively, at time  $T = t$  and at iteration  $i$ .

Table 6.5 shows the results from the LM model for an extended period ( $T = 48$  hours). The model converges after 10 iterations (114.85 seconds). Results for areas I, II and III are similar to the previous case (where the LM method was applied to each single time step) as the actual statuses of valves in pipes in series cannot be determined due to the previously mentioned problem. Results for area IV are intermediate. As seen in Table 6.5, four valves are classified as partially closed in this area, which is not as good as  $T = 2, 3$  and 4 of the LM method applied to a single time step. On the other hand, this is a clear improvement over the worst results of the LM applied to a single step ( $T = 9$ ) where 6 pipes have partially closed valves. Depending on the case study, applying the LM method to a series of single time steps may have advantages compared to applying the LM method to an extended period simulation. For example, in this case study, analyzing the results of  $T = 2, 3$  and 4 in Table 6.4 highlights that there are no partially closed valves (and hence there cannot be also closed valve in the other time steps). However, using the LM method on an extended period simulation may provide more reliable results since the Jacobian matrix is always of full rank during the computation process of the LM method.

**Table 6.5:** Minor loss setting from LM method for an extended period T=48 hours (Case study 2)

Area	Pipe	Iteration 0	Iteration 1	Iteration 2	...	Iteration 10	Actual values
I	P12	1000	1676.41	1642.75	...	1685.75	0
	P139	1000	1665.22	1631.70	...	1655.29	4357
	P7	1000	2552.08	2471.37	...	2388.68	0
II	P301	2000	0.00	43.76	...	225,784.33	0
	P302	15,000	0.00	50.55	...	500,000	Closed
III	P122	1000	0.00	27.54	...	1593.02	0
	P3	2000	0.00	30.91	...	2347.46	2489
IV	P255	9000	500,000	0.00	...	0.00	0
	P961	8000	0.00	110.27	...	1044.70	0
	P964	8000	0.00	0.00	...	0.00	0
	P967	15,000	45,355.49	22,877.10	...	4456.44	0
	P970	5000	3275.70	4403.19	...	21,721.52	0
	P971	2000	3308.39	379.96	...	0.00	0
	P973	1000	2680.95	1628.87	...	0.00	0
	P974	1000	0.00	295.40	...	0.00	0
	P975	1000	0.00	0.00	...	0.00	0
	P989	1000	8541.55	26,059.31	...	0.00	0
	P990	15,000	0.00	0.00	...	120,642.77	0
	P995	1000	0.00	0.00	...	0.00	0
	Final	RSS	2677.83	12.53	5.96		0.34
Grad		0.0199	0.0003	8.30*10 <sup>-5</sup>		2.41*10 <sup>-5</sup>	
Lamda		10	4	40		0.065	

## 6.5 Conclusions

The reliability of a WDS is highly dependent on the presence of valves in the network, which are used to isolate a pipe or a portion of pipes for inspection, maintenance and replacement. However, the unknown statuses of these valves may cause a large difference in the hydraulic behavior between the real system and its simulation model. A methodology for the localization of the partially/fully closed valves in the water distribution network has been introduced in this paper. Three sequentially applied methods have been proposed in the methodology, which includes: a local sensitivity analysis, application of a genetic algorithm and finally an application of the Levenberg-Marquardt algorithm. Within the first method, the sensitivity of the flow rates and nodal heads at measurement locations with respect to the change in the minor losses of the pipes was computed. These sensitivity values were used to remove the valves that are insensitive to the measurement locations from the search space. Within the second method, a genetic

algorithm was implemented in an extended period simulation to preliminarily identify the locations of the partially/fully closed valves and their setting values, i.e. the degree of opening of the valve. Finally, the third method applied the Levenberg-Marquardt algorithm in order to correct the results from the GA model. The performance of the methodology has been evaluated for two case studies. The first case study has shown that the exact solution can be found, in terms of both statuses and minor loss setting values, if the necessary conditions for the Levenberg-Marquardt are satisfied. The second case study showed the applicability of the proposed methodology when applied to a realistic problem. Although the valve problem of the second case study contains non-unique solutions, the methodology still can identify very well the locations of partially closed valves in the network. The proposed methodology is, therefore, an efficient tool for dealing with the problem of finding unknown partially/fully closed valves in water distribution systems.

## References

References listed below are also included in Bibliography.

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

AWWA 1986, *Introduction to Water Distribution*, American Water Works Association, Denver, Colo.

Barbarelli, S., Amelio, M. and Florio, G. 2016. Predictive model estimating the performances of centrifugal pumps used as turbines. *Energy*, 107, pp.103-121.

Blinco, L.J., Simpson, A.R., Lambert, M.F. & Marchi, A. 2016, 'Comparison of Pumping Regimes for Water Distribution Systems to Minimize Cost and Greenhouse Gases', *Journal of Water Resources Planning and Management*, vol. 142, no. 6, p. 04016010.

Bouchart, F. & Goulter, I. 1991, 'Reliability improvements in design of water distribution networks recognizing valve location', *Water Resources Research*, vol. 27, no. 12, pp. 3029-3040.



- Carravetta, A., Del Giudice, G., Fecarotta, O. and Ramos, H.M. 2013. Pump as turbine (PAT) design in water distribution network by system effectiveness. *Water*, 5(3), pp.1211-1225.
- Creaco, E., Franchini, M. & Alvisi, S. 2010, 'Optimal placement of isolation valves in water distribution systems based on valve cost and weighted average demand shortfall', *Water Resources Management*, vol. 24, no. 15, pp. 4317-4338.
- Dandy, G.C., Simpson, A.R. & Murphy, L.J. 1996, 'An improved genetic algorithm for pipe network optimization', *Water Resources Research*, vol. 32, no. 2, pp. 449-458.
- Davidson, J., Bouchart, F., Cavill, S. & Jowitt, P. 2005, 'Real-time connectivity modeling of water distribution networks to predict contamination spread', *Journal of Computing in Civil Engineering*, vol. 19, no. 4, pp. 377-386.
- De Marchis, M., Milici, B., Volpe, R. and Messineo, A. 2016. Energy saving in water distribution network through pump as turbine generators: Economic and environmental analysis. *Energies*, 9(11), p.877.
- Delgado, D. & Lansey, K. 2008, 'Detection of Closed Valves in Water Distribution Systems', *Water Distribution Systems Analysis 2008*, pp. 1-7.
- Deuerlein, J., Cembrowicz, R.G. & Dempe, S. 2005, 'Hydraulic Simulation of Water Supply Networks Under Control', *EWRI World Water and Environmental Resources Congress*, Anchorage, Alaska, US. DOI: 10.1061/40792(173)24, pp. 1-12.
- Deuerlein, J.W., Simpson, A.R. & Dempe, S. 2009, 'Modeling the Behavior of Flow Regulating Devices in Water Distribution Systems Using Constrained Nonlinear Programming', *Journal of Hydraulic Engineering*, vol. 135, no. 11, pp. 970-982.
- Do, N.C., Simpson, A.R., Deuerlein, J.W. and Piller, O. 2016. Calibration of Water Demand Multipliers in Water Distribution Systems Using Genetic Algorithms. *Journal of Water Resources Planning and Management*, 142(11), p.04016044.
- Giustolisi, O. & Savic, D. 2010, 'Identification of segments and optimal isolation valve system design in water distribution networks', *Urban Water Journal*, vol. 7, no. 1, pp. 1-15.

Great Lakes and Upper Mississippi River Board of State Public health and Environmental Managers (GLUMB) 2003, *Recommended Standards for Water Works*, Health Research, Inc. Albany, NY.

Jun, H. & Loganathan, G.V. 2007, 'Valve-Controlled Segments in Water Distribution Systems', *Journal of Water Resources Planning and Management*, vol. 133, no. 2, pp. 145-155.

Lansey, K.E. and Basnet, C. 1991. Parameter estimation for water distribution networks. *Journal of Water Resources Planning and Management*, 117(1), pp.126-144.

Liberatore, S. & Sechi, G. 2009, 'Location and calibration of valves in water distribution networks using a scatter-search meta-heuristic approach', *Water Resources Management*, vol. 23, no. 8, pp. 1479-1495.

Lingireddy, S. & Ormsbee, L.E. 2002, 'Hydraulic network calibration using genetic optimization', *Civil Engineering and Environmental Systems*, vol. 19, no. 1, pp. 13-39.

Marchi, A., Simpson, A.R. & Lambert, M.F. 2016, 'Optimization of Pump Operation Using Rule-Based Controls in EPANET2: New ETTAR Toolkit and Correction of Energy Computation', *Journal of Water Resources Planning and Management*, vol. 142, no. 7, p. 04016012.

Nicolini, M. & Zovatto, L. 2009, 'Optimal location and control of pressure reducing valves in water networks', *Journal of Water Resources Planning and Management*, vol. 135, no. 3, pp. 178-187.

Ostfeld, A. 2005, 'Water distribution systems connectivity analysis', *Journal of Water Resources Planning and Management*, vol. 131, no. 1, pp. 58-66.

Ostfeld, A., Salomons, E., Ormsbee, L., Uber, J.G., Bros, C.M., Kalungi, P., Burd, R., Zazula-Coetzee, B., Belrain, T., Kang, D. and Lansey, K. 2011, 'Battle of the water calibration networks', *Journal of Water Resources Planning and Management*, 138(5), pp.523-532.

Ozger, S., & Mays, L.W. 2004, "Optimal Location of Isolation Valves: A Reliability Approach", in Mays LW, *Water Supply Systems Security*, McGraw Hill, New York, New York.

Piller, O. 1995, 'Modeling the behavior of a network-Hydraulic analysis and sampling procedures for parameter estimation', University of Bordeaux, PhD Thesis.

Piller, O. & Bremond, B. 2001, 'Modeling of Pressure Regulating Devices: A Problem Now Solved', *Bridging the Gap*, DOI: 10.1061/40569(2001)394, pp. 1-10.

Piller, O. & van Zyl, J.E. 2014, 'Modeling Control Valves in Water Distribution Systems Using a Continuous State Formulation', *Journal of Hydraulic Engineering*, vol. 140, no. 11, p. 04014052.

Piller, O., Elhay, S., Deuerlein, J.W. & Simpson, A.R. 2016, 'Local sensitivity of pressure-driven modeling and demand-driven modeling steady-state solutions to variations in parameters', *Journal of Water Resources Planning & Management*, DOI: 10.1061/(ASCE)WR.1943-5452.0000729.

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

Preis, A. & Ostfeld, A. 2008, 'Genetic algorithm for contaminant source characterization using imperfect sensors', *Civil Engineering and Environmental Systems*, vol. 25, no. 1, pp. 29-39.

Sanz, G. & Perez, 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, D.A. & Walters, G.A. 1995, 'An evolution program for optimal pressure regulation in water distribution networks', *Engineering Optimization*, vol. 24, no. 3, pp. 197-219.

Shamir, U., 1974. Optimal design and operation of water distribution systems. *Water resources research*, 10(1), pp.27-36.

Simpson, A.R., Dandy, G.C. & Murphy, L.J. 1994, 'Genetic algorithms compared to other techniques for pipe optimization', *Journal of Water Resources Planning and Management*, vol. 120, no. 4, pp. 423-443.

Sterling, M. & Bargiela, A. 1984, 'Leakage reduction by optimised control of valves in water networks', *Transactions of the Institute of Measurement and Control*, vol. 6, no. 6, pp. 293-298.

Van Zyl, J.E. 2014, Introduction to operation and maintenance of water distribution systems, Water Research Commission, South Africa

Vassiljev, A. and Koppel, T., 2012. Estimation of real-time demands on the basis of pressure measurements. In *Proceedings of the Eighth International Conference on Engineering Computational Technology*, BHV Topping, Civil-Comp Press, Stirlingshire, United Kingdom, vol. 54.

Vikovsky, J.P., Simpson, A.R. & Lambert, M.F. 2000, 'Leak detection and calibration using transients and genetic algorithms', *Journal of Water Resources Planning and Management*, vol. 126, no. 4, pp. 262-265.

Walski, T.M. 1993, 'Water distribution valve topology for reliability analysis', *Reliability engineering & system safety*, vol. 42, no. 1, pp. 21-27.

Walski, T.M., Chase, D.V., Savic, D.A., Grayman, W., Beckwith, S. & Koelle, E. 2003, 'Advanced water distribution modeling and management', Haestad Method Inc., USA

Walski, T.M., Weiler, J.S. & Culver, T. 2006, 'Using Criticality Analysis to Identify Impact of Valve Location', paper presented at Water Distribution Systems Analysis Symposium 2006.

Wu, Z.Y., Song, Y. & Syed, J.L. 2012, 'Optimization model for identifying unknown valve statuses and settings', paper presented at 14<sup>th</sup> *Water Distribution Systems Analysis Conference (WDSA 2012)*, Adelaide, South Australia.

## Chapter 7 Conclusions and future work

Water distribution systems 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. Conventional demand driven models simulate flows and pressures of a water distribution system requiring assumptions of known demands and known valve statuses. Due to the stochastic behavior of the water demands, the complexity of the piping network as well as the existence of systematic errors in measurements, these assumptions usually lead 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, calibration of the water demands and identification of valves statuses are still either not feasible or very difficult. This is attributable to the usual limited quantity of available measurement devices in most real water networks. This thesis has addressed these gaps in **Chapter 2** and formulated the six objectives in **Chapter 1**.

The work in this thesis has been implemented to achieve these objectives and is shown in Chapters 4, 5 and 6. Three methodologies have been developed for calibration of water demands as well as the unknown valve statuses, which partially fulfilled those addressed research gaps.

In this last chapter, the thesis is concluded with a short summary of key contributions and recommendations for future research.

### 7.1 Research contributions

The overall aim of the research has been to develop reliable and effective methodologies for calibrating WDS steady state models, which include calibration of water demand

multiplier factors and calibration of valve settings. From three publications presented in Chapter 4 to 6 of this thesis, the following major contributions have been made.

The first contribution has been the development of a novel approach to calibrate the water demand multipliers under ill-posed conditions where the number of measurement sites is less than the number of unknown water demands. The novel approach, which is the application of multiple runs of a GA model, has been found to effectively calibrate the water demands as well as to give relatively good estimates of the nodal heads and flow rates at non-measured locations in a WDS. Details of this approach are presented in Chapter 4.

The second contribution has been a comprehensive evaluation of the impacts of location and number of the measurement sites to the outputs of a water demand calibration model, which can be found in Chapters 4 and 5. In Chapter 4, the use of a sampling design technique for the selection of optimal measurement locations for the calibration of the water demands has been investigated. The GA calibration model was tested with different simulations of measurement locations (i.e. measurement sites are placed at optimal locations versus random placement of measurements). Results from case study 1 and 2 have highlighted the importance of both the selection of the optimal measurement site locations and the type of measurements on the calibration of the demand multiplier factors. Moreover, the GA model was also tested with various scenarios of measurement availability in water networks, showing that the addition of measurement sites to the network significantly increases the accuracy of the calibration model. In Chapter 5, an additional evaluation of the measurement location problem is presented, in which the effect of measurement location on the uncertainty quantification has been evaluated.

The third contribution is presented in Chapter 5, which has been the development of a particle filter model for the online estimation of water demands under uncertainty. The proposed model, namely the DMFLive model, implements a predictor-corrector process, in which: (1) a demand forecasting model has been used to predict the water demand multiplier factors, (2) the EPANET hydraulic solver has been applied to simulate the flow

rates and nodal heads at measurement locations and (3) real-time observation has been integrated via a Bayesian formulation of the particle filter model to correct the demand predictions.

The fourth contribution has been the development of an improved particle filtering model for the online estimation of water demands. In order to prevent degeneracy, impoverishment and convergence problems that often occur in the particle filter model, an improved resampling process has been developed. This resampling process incorporates an evolutionary scheme from genetic algorithms. Three GA operators of selection, crossover and mutation are responsible for improving the prediction of the water demands. The application of the improved particle filter model to case study 2 in Chapter 5 shows a significant improvement of the demand calibration model in terms of both accuracy and computational effectiveness.

The fifth contribution has been the investigation of the first order approximation (FOA) method for quantifying the demand uncertainty. This is presented in Chapter 5. The uncertainty of the calibrated demands caused by measurement errors is quantified and evaluated in terms of confidence intervals. Through its application to the two case studies in Chapter 5, the FOA method has been shown to be a very reliable method. In addition, the uncertainty of the output model from FOA method can also provide meaningful information about the sensitivity of the pressure with respect to the changes in the nodal demands.

The sixth contribution in this thesis has been the formulation of a sensitivity analysis method, which is the computation of the sensitivity of the flow rates and nodal heads at measurement site locations with respect to the changes in the minor loss coefficients of the valves. This sensitivity analysis has been used to evaluate whether a valve is sensitive to the measurements. Valves that are insensitive to the measured locations need to be removed from the problem as their potentially partially/fully closed settings will not be able to be identified. For the problem of calibration and localization of partially/fully

closed valves, this step is very important as it can significantly reduce the size of the search space. Details of the method may be found in Chapter 6.

The last, but not least, contribution has been the development of an innovative methodology to calibrate and localize partially/fully closed valves in a water network. This is the application of three sequentially applied methods: a local sensitivity analysis, the application of a genetic algorithm model and the application of the Levenberg-Marquardt algorithm. These sensitivity values were used to remove from the problem the valves that are insensitive to the measurement locations. The genetic algorithm was implemented in an extended period simulation to preliminarily identify the locations of the partially/fully closed valves and their setting values. Finally, the third method applied the Levenberg-Marquardt algorithm in order to correct the settings and locations of the potential partially/fully closed valves found by the GA model. This methodology is presented in Chapter 6.

## 7.2 Recommendations for future research

The issue of WDS model calibration is known to be an exceptionally difficult problem in WDS analysis. Although the research presented here has made a considerable advancement to this field, several important issues remain unresolved. A number of possible future research directions are suggested as follows:

In Chapter 4, a GA model was proposed to calibrate the water demands under ill-posed conditions. The GA model was tested with three different case studies. However, the measurement data in these case studies was assumed to be perfect, which does not represent the practicality of the actual systems. Uncertainty of the GA calibration model given the presence of errors (or noise) in measurement data is, therefore, an aspect that needs to be considered. As also seen in Chapter 4, one of the conclusions made for the GA model was that the model run time might be a disadvantage of the GA approach. Hence, future research efforts could involve finding advanced methods for the calibration of water demands to reduce the computational time.



The particle filter model developed in Chapter 5 has been tested for two case studies. Nevertheless, in the second case study where a real network is simulated, the demand patterns are assumed to be perfectly geographically distributed. This assumption should be discussed as in real cases the demand patterns may be non-geographically distributed all over the network. Moreover, some of the other assumptions made in the studies (e.g. the model of the water distribution network perfectly represents the real system with known network characteristics and the errors of the measurement devices are assumed to be known and to follow a Gaussian distribution) might not be valid in real networks. It would be interesting to evaluate the most robust model for an actual system in the future.

In Chapter 6, the problem of calibration and localization of inadvertently partially/fully closed valves is solved based on assumptions of perfect measurement data, aggregated demand groups and applied to synthetic networks that have relatively high density of measurement devices compared to current reality in most networks. In practice different measurements may have very different error scales and the demands in each group may not be homogeneous. These issues need to be addressed and especially, the proposed methodology needs to be tested with real networks to evaluate its performance. In addition, the second case study shows that it took almost 5 hours for the model to identify the locations of the partially/fully closed valves. Meanwhile, blockages caused by unknown closed valves are critical in some emergency cases (e.g. in the case of fire), in which immediate identification of the unknown closed valves are required. Therefore, advanced methods have to be investigated to improve the computational time of the proposed model.

Further study for the identification of partially/fully closed valve in Chapter 6 also needs to be implemented as in real-life situations it is luxury to only calibrate the valve statuses. The detection of partially closed valve should, therefore, be performed in combination with other problems, e.g. detecting leakage in the pipe network.

Addressing the problem of leakage in the networks is also important issue in achieving comprehensive results of the demand estimation models. Although leakage has not been

considered, the demand calibration models could be used to calibrate the total demand, i.e. the sum of user demands and leakage. This can be useful in emergency situations due to water contamination or in situations where the leakage level has not changed. However, further research has to be done in order to distinguish between real demand and leakage. Additional effort also needs to be implemented so that these demand calibration models can be used to detect and manage leakage in the systems.

Finally, the proposed models are developed based on the common assumption of having a series of steady state conditions in the system. Nevertheless, real systems usually experience many simultaneous transient events at any point in time. This can affect the magnitude of the measurements and, consequently, the estimated demands and valve statuses. Future work using real data in a controlled environment and in real systems needs to be conducted in order to assess the effects of these transient events on the estimates.

## Bibliography

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.

Ahmed, I. 1997, 'Application of the gradient method for the analysis of unsteady flow in water networks', M.Sc. thesis, Univ. of Arizona, Tucson, AZ.

Andersen, J.H. and Powell, R.S. 2000 'Implicit state-estimation technique for water network monitoring', *Urban Water*, 2(2), 123-130.

AWWA 1986, *Introduction to Water Distribution*, American Water Works Association, Denver, Colo.

AWWA 2005. *Computer modeling of water distribution systems*, American Water Works Association, Denver, Colo.

Barbarelli, S., Amelio, M. and Florio, G., 2016. Predictive model estimating the performances of centrifugal pumps used as turbines. *Energy*, 107, pp.103-121.

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.

Berry, J.W., Fleischer, L., Hart, W.E., Phillips, C.A. and Watson, J.P. 2005 'Sensor placement in municipal water networks', *Journal of Water Resources Planning and Management*, 131(3), 237-243.

- Bhave, P.R. 1988, 'Calibrating water distribution network models', *Journal of Environmental Engineering*, vol. 114, no. 1, pp. 120-136.
- Blinco, L.J., Simpson, A.R., Lambert, M.F. & Marchi, A. 2016, 'Comparison of Pumping Regimes for Water Distribution Systems to Minimize Cost and Greenhouse Gases', *Journal of Water Resources Planning and Management*, vol. 142, no. 6, p. 04016010.
- Bouchart, F. & Goulter, I. 1991, 'Reliability improvements in design of water distribution networks recognizing valve location', *Water Resources Research*, vol. 27, no. 12, pp. 3029-3040.
- Boulos, P.F., Karney, B.W., Wood, D.J. and Lingireddy, S. 2005. Hydraulic transient guidelines for protecting water distribution systems. *Journal (American Water Works Association)*, 97(5), pp.111-124.
- Boulos, P.F., Lansey, K.E. and Karney, B.W. 2006. *Comprehensive water distribution systems analysis handbook for engineers and planners*. American Water Works Association, MWH soft, Pasadena, California, USA.
- Boulos, P.F. & Ormsbee, L.E. 1991, 'Explicit network calibration for multiple loading conditions', *Civil Engineering Systems*, vol. 8, no. 3, pp. 153-160.
- Boulos, P.F. & Wood, D.J. 1990, 'Explicit calculation of pipe-network parameters', *Journal of Hydraulic Engineering*, vol. 116, no. 11, pp. 1329-1344.
- Bush, C.A. & Uber, J.G. 1998, 'Sampling design methods for water distribution model calibration', *Journal of Water Resources Planning and Management*, vol. 124, no. 6, pp. 334-344.
- Carravetta, A., Del Giudice, G., Fecarotta, O. and Ramos, H.M. 2013, 'Pump as turbine (PAT) design in water distribution network by system effectiveness' *Water*, 5(3), pp.1211-1225.
- Cheng, W. and He, Z. 2010 'Calibration of nodal demand in water distribution systems', *Journal of Water Resources Planning and Management*, 137(1), 31-40.
- Chin, D.A., Mazumdar, A. & Roy, P.K. 2000, *Water Resources Engineering*, vol. 12, Prentice Hall Englewood Cliffs.

Creaco, E., Franchini, M. & Alvisi, S. 2010, 'Optimal placement of isolation valves in water distribution systems based on valve cost and weighted average demand shortfall', *Water Resources Management*, vol. 24, no. 15, pp. 4317-4338.

Cross, H. 1936, *Analysis of flow in networks of conduits or conductors*, University of Illinois at Urbana Champaign, College of Engineering. Engineering Experiment Station.

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.

Dandy, G.C., Simpson, A.R. & Murphy, L.J. 1996, 'An improved genetic algorithm for pipe network optimization', *Water Resources Research*, vol. 32, no. 2, pp. 449-458.

Davidson, J. and Bouchart, F.C. 2006, 'Adjusting nodal demands in SCADA constrained real-time water distribution network models', *Journal of Hydraulic Engineering*, 132(1), 102-110.

Davidson, J., Bouchart, F., Cavill, S. & Jowitt, P. 2005, 'Real-time connectivity modeling of water distribution networks to predict contamination spread', *Journal of Computing in Civil Engineering*, vol. 19, no. 4, pp. 377-386.

De Marchis, M., Milici, B., Volpe, R. and Messineo, A., 2016, 'Energy saving in water distribution network through pump as turbine generators: Economic and environmental analysis' *Energies*, 9(11), p.877.

De Schaetzen, W.B.F. 2000, *Optimal Calibration and Sampling Design for Hydraulic Network Models*, University of Exeter.

Delgado, D. & Lansey, K. 2008, 'Detection of closed valves in Water Distribution Systems', paper presented at WDSA 2008, *Water Distribution Systems Analysis 2008*, American Society of Civil Engineers, pp. 1-7.

Deuerlein, J., Cembrowicz, R.G. & Dempe, S. 2005, 'Hydraulic Simulation of Water Supply Networks Under Control', *Impacts of Global Climate Change*, DOI: 10.1061/40792(173)24, pp. 1-12.

Deuerlein, J.W., Simpson, A.R. & Dempe, S. 2009, 'Modeling the Behavior of Flow Regulating Devices in Water Distribution Systems Using Constrained Nonlinear Programming', *Journal of Hydraulic Engineering*, vol. 135, no. 11, pp. 970-982.

Di Nardo, A., Di Natale, M., Gisonni, C. and Iervolino, M. 2015 'A genetic algorithm for demand pattern and leakage estimation in a water distribution network', *Journal of Water Supply: Research and Technology - Aqua*, 64(1), 35-46.

Do, N.C., Simpson, A.R., Deuerlein, J.W. & Piller, O. 2016, 'Calibration of Water Demand Multipliers in Water Distribution Systems Using Genetic Algorithms', *Journal of Water Resources Planning and Management*, p. 04016044.

Do, N.C., Simpson, A.R., Deuerlein, J.W. & Piller, O. 2017a, 'A particle filter-based model for online estimation of demand multipliers in water distribution systems under uncertainty', *Journal of Water Resources Planning and Management*, (Accepted for publication)

Do, N.C., Simpson, A.R., Deuerlein, J.W. & Piller, O. 2017b, 'Localization of inadvertently partially closed valves in water distribution systems', *Journal of Water Resources Planning and Management*, (Submitted July 2017)

Do, N.C., Simpson, A.R., Deuerlein, J.W. & Piller, O. 2017c 'Demand Estimation In Water Distribution Systems: Solving Underdetermined Problems Using Genetic Algorithms', *Procedia Engineering*, vol. 186, pp. 193-201, ISSN 1877-7058,

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.

Filion, Y.R. and Karney, B.W. 2002. Extended-period analysis with a transient model. *Journal of Hydraulic Engineering*, 128(6), pp.616-624.

Giustolisi, O. and Ridolfi, L. 2014 'New Modularity-Based Approach to Segmentation of Water Distribution Networks', *Journal of Hydraulic Engineering*, 140(10), 04014049.

Giustolisi, O. & Savic, D. 2010, 'Identification of segments and optimal isolation valve system design in water distribution networks', *Urban Water Journal*, vol. 7, no. 1, pp. 1-15.

Goldberg, D.E. 1989 *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Longman Publishing Co., Inc.

Goldberg, D.E. and Deb, K. 1991 'A comparative analysis of selection schemes used in genetic algorithms', *Foundations of Genetic Algorithms*, 1, 69-93.

Great Lakes and Upper Mississippi River Board of State Public Health and Environmental Managers (GLUMB) 2003, *Recommended Standards for Water Works*, Health Research, Inc. Albany, NY.

Haddad, O., Adams, B. and Marino, M. 2008 'Optimum rehabilitation strategy of water distribution systems using the HBMO algorithm', *Journal of Water Supply: Research and Technology-AQUA*, 57(5), 337-350.

Hol, J.D., Schon, T.B. & Gustafsson, F. 2006, 'On resampling algorithms for particle filters', *Nonlinear Statistical Signal Processing Workshop*, 2006 IEEE, pp. 79-82.

Hutton, C., Kapelan, Z., Vamvakeridou-Lyroudia, L. & Savic, 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, C.J. & 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, C.J., Kapelan, Z., Vamvakeridou-Lyroudia, L. & Savic, D.A. 2012a, 'Real-time demand estimation in water distribution systems under uncertainty', paper presented at WDSA 2012: 14<sup>th</sup> *Water Distribution Systems Analysis Conference*, Adelaide, South Australia.

Hutton, C.J., Kapelan, Z., Vamvakieridou-Lyroudia, L. & Savić, D.A. 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.

Jun, H. & Loganathan, G.V. 2007, 'Valve-Controlled Segments in Water Distribution Systems', *Journal of Water Resources Planning and Management*, vol. 133, no. 2, pp. 145-155.

Junior E.L. & Vataavuk P. 2006, 'A new procedure to solve steady and slow transient flows in pip networks', *7<sup>th</sup> International Conference on Hydroinformatics*, HIC 2006, Nice, France

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, Z.S., Savic, D.A. and Walters, G.A. 2003 'A hybrid inverse transient model for leakage detection and roughness calibration in pipe networks', *Journal of Hydraulic Research*, 41(5), 481-492.

Kapelan, Z.S., Savic, D.A. and Walters, G.A. 2005 'Optimal sampling design methodologies for water distribution model calibration', *Journal of Hydraulic Engineering*, 131(3), 190-200.

Kapelan, Z.S., Savic, D.A. & Walters, G.A. 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.

Krause, A., Leskovec, J., Guestrin, C., VanBriesen, J. and Faloutsos, C. 2008 'Efficient sensor placement optimization for securing large water distribution networks', *Journal of Water Resources Planning and Management*, 134(6), 516-526.



Kun, D., Tian-Yu, L., Jun-Hui, W. and Jin-Song, G. 2015 'Inversion model of water distribution systems for nodal demand calibration', *Journal of Water Resources Planning and Management*, 141(9), p.04015002.

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

Lansley, 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.

Liberatore, S. & Sechi, G. 2009, 'Location and calibration of valves in water distribution networks using a scatter-search meta-heuristic approach', *Water Resources Management*, vol. 23, no. 8, pp. 1479-1495.

Liggett, J.A. and Chen, L.C. 1994. Inverse transient analysis in pipe networks. *Journal of Hydraulic Engineering*, 120(8), pp.934-955.

Lingireddy, S. & Ormsbee, L.E. 2002, 'Hydraulic network calibration using genetic optimization', *Civil Engineering and Environmental Systems*, vol. 19, no. 1, pp. 13-39.

Marchi, A., Simpson, A.R. & Lambert, M.F. 2016, 'Optimization of pump operation using rule-based controls in EPANET2: New ETTAR toolkit and correction of energy computation', *Journal of Water Resources Planning and Management*, vol. 142, no. 7, p. 04016012.

Martinez, F., Bartolin, H., Bou, V. and Kapelan, Z. 2003 'Calibration of valve-driven water distribution systems using jointly GIS and Genetic Algorithms. Application to the WDS of Valencia (Spain)', in 7<sup>th</sup> *International Conference on Computing and Control for the Water Industry (CCWI)*.

Meier, R.W. and Barkdoll, B.D. 2000 'Sampling design for network model calibration using genetic algorithms', *Journal of Water Resources Planning and Management*, 126(4), 245-250.

Moradkhani, H., Hsu, K.L., 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, A.K. & Powell, R.S. 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.

Nault, J. & Karney, B. 2016, 'Improved rigid water column formulation for simulating slow transients and controlled operations', *Journal of Hydraulic Engineering*, vol. 142, no. 9, p. 04016025.

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.

Nicolini, M. & Zovatto, L. 2009, 'Optimal location and control of pressure reducing valves in water networks', *Journal of Water Resources Planning and Management*, vol. 135, no. 3, pp. 178-187.

Okeya, I., Kapelan, Z., Hutton, C. and Naga, D. 2014, 'Online modelling of water distribution system using data assimilation', *Procedia Engineering*, 70, 1261-1270

Onizuka, K. 1986, 'System dynamics approach to pipe network analysis', *Journal of Hydraulic Engineering*, vol. 112, no. 8, pp. 728-749.

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

Ormsbee, L.E. and Lingireddy, S. (1997) 'Calibration of hydraulic network models', *Journal AWWA*, 89(2), 42-50.

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

Ostfeld, A. 2005, 'Water distribution systems connectivity analysis', *Journal of Water Resources Planning and Management*, vol. 131, no. 1, pp. 58-66.

Ostfeld, A., Salomons, E., Ormsbee, L., Uber, J.G., Bros, C.M., 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.

Ozger, S., & Mays, L.W. 2004, "Optimal Location of Isolation Valves: A Reliability Approach", *Water Supply Systems Security*, McGraw Hill, New York.

Perelman, L., Maslia, M.L., Ostfeld, A. & Sautner, J.B. 2008, 'Using aggregation/skeletonization network models for water quality simulations in epidemiologic studies', *Journal of 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.

Piller, O. & Bremond, B. 2001, 'Modeling of Pressure Regulating Devices: A Problem Now Solved', *Bridging the Gap: Meeting the World's Water and Environmental Resources Challenges*, American Society of Civil Engineers, doi:10.1061/40569(2001)394, pp. 1-10.

Piller, O., Elhay, S., Deuerlein, J.W. & Simpson, A.R. 2016, 'Local sensitivity of pressure-driven modeling and demand-driven modeling steady-state solutions to variations in parameters', *Journal of Water Resources Planning & Management*, DOI: 10.1061/(ASCE)WR.1943-5452.0000729.

Piller, O., Gilbert, D., and Van Zyl, J.E. 2010 'Dual calibration for coupled flow and transport models of water distribution systems', *Water Distribution Systems Analysis WDSA 2010*, Tucson, Arizona, United States, pp. 722-731.

- Piller, O. & van Zyl, J.E. 2014, 'Modeling Control Valves in Water Distribution Systems Using a Continuous State Formulation', *Journal of Hydraulic Engineering*, vol. 140, no. 11, p. 04014052.
- Pinder, G.F. and Cooper, H.H., 1970. A numerical technique for calculating the transient position of the saltwater front. *Water Resources Research*, 6(3), pp.875-882.
- Preis, A. & Ostfeld, A. 2008, 'Genetic algorithm for contaminant source characterization using imperfect sensors', *Civil Engineering and Environmental Systems*, vol. 25, no. 1, pp. 29-39.
- 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, A.J., Ostfeld, A. & Perelman, L. 2010, 'On-line hydraulic state estimation in urban water networks using reduced models', 10<sup>th</sup> *International Conference on Computing and Control for the Water Industry: Integrating Water Systems, CCWI 2009*, CRC Press, Sheffield, United kingdom, pp. 319-324.
- Preis, A., Whittle, A.J., 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.
- Propato, M., Cheung, P.B., and Piller, O. 2006 'Sensor location design for contaminant source identification in water distribution systems', in *Proceedings of the 8<sup>th</sup> Annual Water Distribution System Analysis Symposium*, Cincinnati, Ohio, United States.
- Pudar, R.S. and Liggett, J.A. 1992 'Leaks in pipe networks', *Journal of Hydraulic Engineering*, 118(7), 1031-1046.
- Rahal, C.M., Sterling, M.J.H. & 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., Arulampalam, S. & Gordon, N. 2004, *Beyond the Kalman filter: particle filters for tracking applications*, Artech House Boston, Ma., London.
- Rossman, L.A. 2000, 'EPANET 2: users manual', US Environmental Protection Agency (EPA), USA.
- Salmond, D. & Gordon, N. 2005, 'An introduction to particle filters', *State space and unobserved component models theory and applications*, Cambridge University Press, pp. 1-19.
- Sanz, G. & Perez, R. 2014, 'Comparison of demand pattern calibration in water distribution networks with geographic and non-geographic parameterization', 11<sup>th</sup> *International Conference on Hydroinformatics*, pp. 1-8.
- Sanz, G. & Perez, 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, D.A., Kapelan, Z.S. & Jonkerkouw, P.M. 2009, 'Quo vadis water distribution model calibration?', *Urban Water Journal*, vol. 6, no. 1, pp. 3-22.
- Savic, D.A. & Walters, G.A. 1995, 'An evolution program for optimal pressure regulation in water distribution networks', *Engineering Optimization*, vol. 24, no. 3, pp. 197-219.
- Seifollahi-Aghmiuni, S., Bozorg Haddad, O., Omid, M.H. and Marino, M.A. 2013 'Effects of pipe roughness uncertainty on water distribution network performance during its operational period', *Water Resources Management*, 27(5), 1581-1599.
- Shamir, U. and Howard, C.D. (1977) 'Engineering analysis of water distribution systems', *Journal of American Water Works Association*, 69(9), 510-514.
- Shang, F., Uber, J.G., Van Bloemen Waanders, B.G., Boccelli, D. & Janke, R. 2006, 'Real time water demand estimation in water distribution system', 8<sup>th</sup> *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.

Shihu, S., 2011. 'Multi-sensor remote sensing technologies in water system management' *Procedia Environmental Sciences*, vol. 10, pp.152-157.

Simpson, A.R., Dandy, G.C. & Murphy, L.J. 1994, 'Genetic algorithms compared to other techniques for pipe optimization', *Journal of Water Resources Planning and Management*, vol. 120, no. 4, pp. 423-443.

Shimada, M. 1989, 'Graph-theoretical model for slow transient analysis of pipe networks', *Journal of Hydraulic Engineering*, vol. 115, no. 9, pp. 1165-1183.

Soares, A.K., Covas, D.I. and Reis, L.F., 2008 'Analysis of PVC pipe-wall viscoelasticity during water hammer' *Journal of Hydraulic Engineering*, 134(9), pp.1389-1394.

Sterling, M. & Bargiela, A. 1984, 'Leakage reduction by optimised control of valves in water networks', *Transactions of the Institute of Measurement and Control*, vol. 6, no. 6, pp. 293-298.

Todini, E. & Pilati, S. 1988, 'A gradient algorithm for the analysis of pipe networks', *Computer applications in water supply: vol. 1 - systems analysis and simulation*, Research Studies Press Ltd., pp. 1-20.

Todini, E. and Rossman, L.A., 2012. Unified framework for deriving simultaneous equation algorithms for water distribution networks. *Journal of Hydraulic Engineering*, 139(5), pp.511-526.

Thyer, M.A., Micevski, T, Kuczera, G.A. & 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, P.J. 2009, 'Particle filtering in geophysical systems', *Monthly Weather Review*, vol. 137, no. 12, pp. 4089-4114.

van Leeuwen, P.J. 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, J.E. 2014, Introduction to operation and maintenance of water distribution systems, Water Research Commission, South Africa, ISBN 978-1-4312-0556-1

van Zyl, J.E., 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.

Vassiljev, A. and Koppel, T., 2012. Estimation of real-time demands on the basis of pressure measurements. In *Proceedings of the Eighth International Conference on Engineering Computational Technology*, BHV Topping, Civil-Comp Press, Stirlingshire, United Kingdom, vol. 54.

Vitkovsky, J.P., Simpson, A.R. & Lambert, M.F. 2000, 'Leak detection and calibration using transients and genetic algorithms', *Journal of Water Resources Planning and Management*, vol. 126, no. 4, pp. 262-265.

Vitkovsky, J.P., Liggett, J.A., Simpson, A.R. and Lambert, M.F. (2003) 'Optimal measurement site locations for inverse transient analysis in pipe networks', *Journal of Water Resources Planning and Management*, 129(6), 480-492.

Waller, J.A. 2013, 'Using observations at different spatial scales in data assimilation for environmental prediction', PhD Thesis, Department of Mathematics and Statistics, University of Reading.

Walski, T.M. 1983, 'Technique for calibrating network models', *Journal of Water Resources Planning and Management*, vol. 109, no. 4, pp. 360-372.

Walski, T.M. (1986) 'Case study: Pipe network model calibration issues', *Journal of Water Resources Planning and Management*, 112(2), 238-249.

- Walski, T.M. 1993, 'Water distribution valve topology for reliability analysis', *Reliability engineering & system safety*, vol. 42, no. 1, pp. 21-27.
- Walski, T.M., Chase, D.V., Savic, D.A., Grayman, W., Beckwith, S. & Koelle, E. 2003, 'Advanced water distribution modeling and management', Haestad Method Inc., USA
- Walski, T.M., Weiler, J.S. & Culver, T. 2006, 'Using Criticality Analysis to Identify Impact of Valve Location', paper presented at *Water Distribution Systems Analysis Symposium 2006*, Cincinnati, Ohio, United States, <https://doi.org/10.1061/9780784409411>.
- Weerts, A.H. & El Serafy, G.Y. 2006, 'Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models', *Water Resources Research*, vol. 42, no. 9.
- Whittle, A.J., M. Allen, A. Preis, and Iqbal M. 2013 "Sensor Networks for Monitoring and Control of Water Distribution Systems." 6<sup>th</sup> *International Conference on Structural Health Monitoring of Intelligent Infrastructure (SHMII 2013)*, Hong Kong.
- Wood, D.J. & Charles, C.O. 1972, 'Hydraulic network analysis using linear theory', *Journal of the Hydraulics division*, vol. 98, no. 7, pp. 1157-1170.
- Wood, D.J., Lingireddy, S., Boulos, P.F., Karney, B.W. and McPherson, D.L., 2005. Numerical methods for modeling transient flow in distribution systems. *Journal (American Water Works Association)*, 97(7), pp.104-115.
- Wu, Z.Y., Song, Y. & Syed, J.L. 2012, 'Optimization model for identifying unknown valve statuses and settings', paper presented at 14<sup>th</sup> *Water Distribution Systems Analysis Conference (WDSA 2012)*, Adelaide, South Australia.
- Xu, C. & Goulter, I.C. 1998, 'Probabilistic model for water distribution reliability', *Journal of Water Resources Planning and Management*, vol. 124, no. 4, pp. 218-228.
- Yu, G.P. and Powell, R.S. 1994 'Optimal design of meter placement in Water Distribution Systems', *International journal of systems science*, 25(12), 2155-2166.



# Appendix 1      Final Published Version of Publication 1 (Chapter 4)

## Calibration of Water Demand Multipliers in Water Distribution Systems Using Genetic Algorithms

Nhu C. Do<sup>1</sup>, Angus R. Simpson<sup>1</sup>, Jochen W. Deuerlein<sup>2</sup>, Olivier Piller<sup>3</sup>

<sup>1</sup> - *School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide SA 5005, Australia.*

<sup>2</sup> - *Senior Researcher, 3S Consult GmbH, Karlsruhe, Germany.*

<sup>3</sup> - *Senior Researcher, Irstea UR ETBX, Dept. of Water, Cestas, France.*

*Journal of Water Resources Planning and Management 142(11)*

[doi:10.1061/\(ASCE\)WR.1943-5452.0000691](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000691)

This page has been intentionally left blank.

# Calibration of Water Demand Multipliers in Water Distribution Systems Using Genetic Algorithms

Nhu C. Do<sup>1</sup>; Angus R. Simpson, M.ASCE<sup>2</sup>; Jochen W. Deuerlein<sup>3</sup>; and Olivier Piller<sup>4</sup>

**Abstract:** Hydraulic models have been widely used for design, analysis, and operation of water distribution systems. As with all hydraulic models, water demands are one of the main parameters that cause the most uncertainty to the model outputs. However, the calibration of the water demands is usually not feasible attributable to the limited quantity of available measurements in most real water networks. This paper presents an approach to calibration of the demand multiplier factors under an ill-posed condition where the number of measurements is less than the number of parameter variables. The problem is solved using a genetic algorithm (GA). The results show that not only is the GA able to match the calibrated values at measured locations, but by using multiple runs of the GA model, the flow rates and nodal heads at nonmeasured locations can be estimated. Three case studies are presented as an illustration of the problem. The first case study is a small network that demonstrates the calibration model. The second case study shows a comparison between the genetic algorithm model and a singular value decomposition model. The last case study is a large network that allows for practical considerations in applying the proposed methodology to a realistic context. DOI: 10.1061/(ASCE)WR.1943-5452.0000691. © 2016 American Society of Civil Engineers.

**Author keywords:** Genetic algorithms; Optimization; Demand calibration; Water distribution systems.

## Introduction

As an indispensable component of urban infrastructure, a water distribution system (WDS) has to accommodate large water transfer volumes on a daily basis. The problem of ensuring a satisfactory and reliable service is further complicated by population growth which often leads to the need for augmentation of the system. Determination of flow rates and pressure heads in the existing system is a necessary step in the modification of a WDS. This task can be accomplished by using measurement devices, such as sensors; however, sensors can only capture the status of some component locations in the system. Calibration of the full WDS model using limited measurements from these devices is therefore a research area that requires further development.

Calibration of a WDS model is the process of adjusting network parameters so that the output from the computer model matches the field measurements, which are usually the pressures and flow rates at particular locations in the network (Shamir and Howard 1977). The calibration procedure for a water network model has been well addressed by Ormsbee and Lingireddy (1997). In their paper, the authors suggested a seven-step calibration process, which includes: (1) Identifying the intended use of the model, (2) Determining estimates of the model parameters, (3) Collecting calibration data, (4) Evaluation of the results of the model, (5) Performing a macro

level calibration of the model, (6) Performing a sensitivity analysis, and (7) Performing a micro level calibration of the model.

Savic et al. (2009) presented a comprehensive literature review on the calibration of water network models where the calibration methods can be classified by their dynamic (transient/static), by their calculation methods (iterative/explicit/implicit), or by the use of optimization methods (traditional/evolutionary).

In terms of calibration in transient analysis, the calibration models have been constructed primarily to detect leakage in distribution systems (Pudar and Liggett 1992; Liggett and Chen 1994; Vítkovský et al. 2000; Kapelan et al. 2003). These transient calibration procedures usually consider a pure leak or leaks combined with unknown nodal demands and pipe friction factors as parameters to be calibrated. However, because of the complexity of the method, such as generating and measuring reflections of water hammer waves, transient analysis has not been widely used in practice for practical leak detection applications.

In static hydraulic analysis, an iterative procedure was applied in the early use of calibration models (Walski 1983, 1986; Bhawe 1988). This procedure was implemented to update the unknown model parameters using heads/flows obtained by solving the water network equations. Owing to slow convergence rate, these models are only suitable for small problems or require simplification of the water network.

The second technique is explicit calibration, which was employed in Ormsbee and Wood (1986), Boulos and Wood (1990), and Boulos and Ormsbee (1991). This technique involves solving an extended set of continuity and head-loss equations where the calibration problem is required to be even-determined (i.e., the number of calibrated parameters must be equal to the number of measurements). Measurement errors are also neglected. When the number of unknown parameters is larger than the number of measurements, calibration parameters are often grouped that may result in potentially impractical outputs. Therefore, explicit calibration models are often used for the purpose of system analysis.

Implicit calibration is the third type of static calibration model. This method considers the calibration problem as an optimization

<sup>1</sup>Ph.D. Candidate, School of Civil, Environmental and Mining Engineering, Univ. of Adelaide, Adelaide, SA 5005, Australia (corresponding author). E-mail: nhu.do@adelaide.edu.au

<sup>2</sup>Professor, School of Civil, Environmental and Mining Engineering, Univ. of Adelaide, Adelaide, SA 5005, Australia.

<sup>3</sup>Senior Researcher, 3S Consult GmbH, D-76137 Karlsruhe, Germany.

<sup>4</sup>Senior Researcher, Dept. of Water, Irstea UR ETBX, F-33612 Cestas, France.

Note. This manuscript was submitted on November 3, 2015; approved on April 11, 2016; published online on June 22, 2016. Discussion period open until November 22, 2016; separate discussions must be submitted for individual papers. This paper is part of the *Journal of Water Resources Planning and Management*, © ASCE, ISSN 0733-9496.

process and solves it using a hydraulic solver combined with either a traditional or an evolutionary optimization technique. The implicit method has been investigated by a majority of the previous research (e.g., Ormsbee 1989; Lansley and Basnet 1991; Datta and Sridharan 1994; Ormsbee and Lingireddy 1997; Andersen and Powell 2000; Nagar and Powell 2002; Shang et al. 2006; Preis et al. 2009; Kang and Lansley 2009; Piller et al. 2010; Hutton et al. 2012).

The uncertainty of results from water network models is caused by many influences, such as pipe roughness, nodal demands or valve states (e.g., Martínez et al. 2003; Liberatore and Sechi 2009). Pipe roughness coefficients and the water demands at nodes are normally used for the static calibration because of their high impact on network uncertainty. However, roughness coefficients usually vary only in the long term; for example, the annual variation of these parameters was considered in Haddad et al. (2008) and Seifollahi-Aghmiuni et al. (2013). Therefore, water demands are the main parameters that affect the uncertainty of the output of models in the shorter term (e.g., hourly, daily).

The calibration of water demand has been studied using various techniques, for instance a Predictor–Corrector algorithm (Shang et al. 2006; Preis et al. 2009), Bayesian recursive approach (Kapelán et al. 2007), Kalman filtering and tracking state estimator (Kang and Lansley 2009), and particle filter modeling (Hutton et al. 2012). Most of these models have been developed based on given frameworks where the measurement locations were predetermined and the calibration parameters are grouped to be less than the number of measurements. The outcomes, therefore, rely on these additional assumptions, which can lead to large approximations in real water distribution systems. Only a few papers have directly dealt with underdetermined systems such as a proportional demand method (Davidson and Bouchart 2006) and singular value decomposition (SVD) (Cheng and He 2010; Kun et al. 2015).

Mathematically, the calibration of the demand in water distribution systems in which the number of measurements is less than the number of calibrated variables, is a nonlinear underdetermined problem. A local solution of the problem can be found by a local linearization methods such as QR decomposition, SVD, or using the Moore–Penrose pseudoinverse matrix in the Newton–Raphson method. However, because of the possibility of nonuniqueness of the solutions, the results from mathematical methods are either far from the actual solution or result in negative demands at some nodes. Apparently, if the data from measurement devices are considered as the only known inputs, the quest for a reliable demand calibration model is still a challenge for hydraulic researchers.

This paper proposes a methodology for the calibration of water demand multipliers for an underdetermined system where the number of measurements is less than the number of demand parameter variables. The EPANET toolkit (Rossman 2000) is used to solve the system of water network equations, whereas genetic algorithms (GAs) are applied to find the best match between known measurement inputs and their calibrated values. The average values of multiple GA runs have been found to give the best estimates of the flow rates and nodal heads and the calibration of the demands in a system. Different scenarios of measurement availability in water networks are tested to evaluate the reliability of the model. Furthermore, this study also investigates the use of a sampling design technique (Piller 1995) for the selection of optimal measurement locations to improve the quality of the calibration model. This is followed by application of the model to three case studies. Finally, conclusions and suggestions for future work are given.

## GA Calibration Model

The proposed model applies an implicit technique for the steady-state hydraulic simulation where the calibration process is formulated as an optimization problem. The objective function is the minimization of the differences between simulated values from the model and their corresponding measured values. The decision variables are the demand multiplier factors (DMFs) for the nodal demands as described in the “Decision Variables” subsection.

### Objective Function

In this study, a least-squares method is applied for the objective function. The method minimizes the sum of squared residuals between the measured and computed values of pipe flow rates and nodal heads at the measurement locations. The objective function is given by

$$\text{Min}F = \sum_{i=1}^{N_H} w_i (H_i^{\text{Meas}} - H_i^{\text{Sim}})^2 + \sum_{j=1}^{N_Q} w_j (Q_j^{\text{Meas}} - Q_j^{\text{Sim}})^2 \quad (1)$$

where  $H_i^{\text{Sim}}$ ,  $Q_j^{\text{Sim}}$  = simulated nodal head and flow rates for the  $i$ th node and  $j$ th pipe, respectively;  $H_i^{\text{Meas}}$ ,  $Q_j^{\text{Meas}}$  = measured head and flow rate at the  $i$ th node and  $j$ th pipe;  $N_H$ ,  $N_Q$  = number of head and flow measurement sites in the network respectively and  $w_i$ ,  $w_j$  = weighting factors applied to different terms to ensure they have similar magnitude and unit of measurement.

Flows and heads and nodal demands are time dependent. However, they can be considered constant during short periods of time, e.g., 30 min or 1 h. Because the proposed approach is applied to each steady state step during an extended period simulation, the time dependency symbol is not explicitly given in Eq. (1). Measured values are obtained from field measurement devices or, for testing the methodology as in our case, these values can be generated by running a hydraulic simulation toolkit such as EPANET. Weighting factors  $w_i$  and  $w_j$  can be computed by taking the inverse of the square of the observed values [ $w_i = 1/(H_i^{\text{Meas}})^2$  and  $w_j = 1/(Q_j^{\text{Meas}})^2$ , respectively], which is the approach that has been used in Di Nardo et al. (2015).

### Decision Variables

In a WDS model, the water demand at each node is calculated by the multiplication of a base demand with its corresponding DMF at each time step  $t$

$$D_{k,t} = D_{0,k} \times f_{k,t} \quad (2)$$

where  $D_{0,k}$  = base demand at the  $k$ th node, which is calculated using quarter/annual water usage billing information; and  $f_{k,t}$  = demand multiplier factor at the  $k$ th node at time step  $t$ . The decision variables for the optimization problem, therefore, are the demand multiplier factors  $f_{k,t}$  ( $k = 1, \dots, N_{\text{DM}}$ ) at nodal demands at each time step. A bounded range of demand factors may apply as

$$f_k^{\text{min}} \leq f_k \leq f_k^{\text{max}} \quad (3)$$

where  $(f_k^{\text{min}}, f_k^{\text{max}})$  = bounds of decision variables. The value of  $f_k^{\text{min}}$  must be equal to or larger than zero, whereas  $f_k^{\text{max}}$  can be selected based on typical values of peaking demand factors such as those reported in Beal and Stewart (2014). In this paper, a value of  $f_k^{\text{max}} = 1.5$  has been selected for all case studies, which also guarantees that it is much larger than the true multiplier used to generate the calibration data.

Chrom.	3	8	11	56	32	1
Decode	0.06	0.16	0.22	1.12	0.64	0.02
DMF No.	f2	f3	f4	f5	f6	f7

Fig. 1. Example of GA chromosome and decoding for the demand calibration problem

Fig. 1 shows an example of a GA solution (namely a chromosome) for six demand multiplier factors at one time step, which are chosen from a lower bound of 0.00 and an upper bound of 1.50 with the increment step of 0.02 (these values would be problem dependent and selected accordingly). Each demand multiplier factor is coded by an integer number, ranging from 0 to 75. By using this coding information, the chromosome from GA process is decoded into a set of demand multiplier factors that are multiplied by the base demand and can be used for the hydraulic simulation process.

### GA Process and Operators

The GA calibration model implemented for this research has been written in the C-sharp language. The flowchart of the algorithm is shown in Fig. 2. An initial population of chromosomes is randomly generated and decoded into corresponding DMF values for each member. To each node of the network exactly one of these DMFs is assigned, and EPANET is subsequently called to simulate the steady state hydraulics of the system. Simulated flows and heads ( $Q^{Sim}$ ,  $H^{Sim}$ ) at the measurement locations are obtained and compared with their measured values through the calculation of the objective function. The inverse of the objective function is applied to define the fitness function for each member of the GA population. This is the measure for the quality of each member, and is used to determine its opportunity for survival.

By applying GA selection, crossover and mutation, new generations that inherit features of previous generations are created, and the calibration process is then repeated until the stopping criteria are met.

For the selection operator, a study by Goldberg and Deb (1991) recommended the use of tournament selection because of its better convergence compared with proportionate selection or ranking selection [Nicklow et al. (2010) for a review of GAs]. In addition, Goldberg (1989) also suggested that the two-point crossover

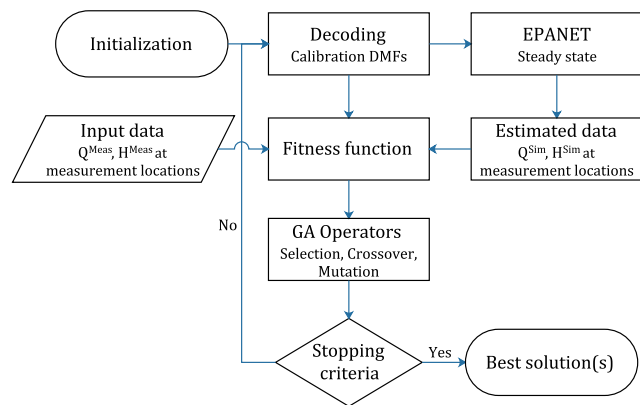


Fig. 2. Flowchart for the genetic algorithm calibration of demand multiplier factors

operator with a relatively high probability ( $P_c = 0.6$  to 1.0) and the bitwise mutation with a probability of  $P_m \approx 1/str$  ( $str$  is the length of the string) can be used to improve the performance of GA models. Therefore, in the study presented in this paper, tournament selection, two-point crossover, and bitwise mutation have been applied in the GA calibration model.

### Selection of Measurement Locations

The accuracy of demand calibration models not only depends on the number of site measurements but also on the locations of the measurements. The optimal measurement location problem has been investigated by a number of researchers using various mathematical and statistical methods, such as Yu and Powell (1994), Vítkovský et al. (2003), Berry et al. (2005), Propato et al. (2006), Krause et al. (2008), and Giustolisi and Ridolfi (2014). In calibration models, sampling design (SD) methodologies have been applied for the selection of the observation locations. Piller (1995) used a SD method to minimize the influence of measurement errors in the state vector estimation. Bush and Uber (1998) developed three SD methods derived from D-optimality criteria: max-sum, weighted sum, and max-min methods to select measurement locations based on the analysis of the Jacobian matrix. Meier and Barkdoll (2000) used a GA for the optimal SD problem with the aim of finding a set of calibration locations which maximizes the presence of nonnegligible pipe velocities. De Schaezen (2000) proposed three SD approaches for the optimal measurement locations based on shortest path algorithm, rank measurement locations, and maximization of Shannon's entropy. Most recently, Kapelan et al. (2005) developed two SD models using a GA to find the optimal set of pressure locations, where the first model is formulated as a single objective GA, and the second is modeled as a multiobjective optimization problem.

In this paper, the SD method proposed by Piller (1995) based on a greedy algorithm is applied to select the best measurement locations for the GA calibration model of the demand multiplier factors. Influences of the measurement locations on the calibration results, thereafter, are examined by evaluating the convergence of the GA model.

### Sampling Design Method

The hydraulic steady-state of a water network solves a nonlinear problem of the continuity equations at nodes and the energy equations for pipes. The sensitivity of the nodal heads and flow rates with respect to the demand parameters  $\theta$  at nodes can be computed as

$$\begin{cases} A^T \frac{\partial q}{\partial \theta} + G_D = 0_{ns,nd} \\ D \frac{\partial q}{\partial \theta} - A \frac{\partial h}{\partial \theta} = 0_{np,nd} \end{cases} \quad (4)$$

**Table 1.** Example of Selection Matrix S for Four Measurements in the Nine Pipe Network in Fig. 3

L 2	L 3	L 4	L 5	L 6	L 7	L 8	L 9	L 1	N2	N3	N4	N5	N6	N7
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

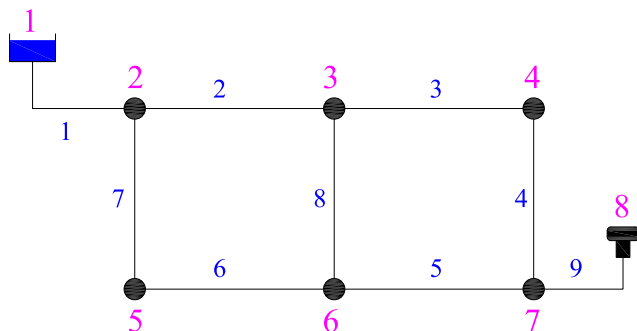
where  $A$  and  $A^T$  = unknown head node incidence matrix and its transposed matrix, which provides information about the connectivity of the nodes and the links in the network;  $q$  and  $h$  = flow rate and nodal head vectors of  $np$  pipes and  $ns$  nodes;  $G_D = ns$  by  $nd$  matrix of nodal demand allocation. As demand aggregation is not considered in this study,  $G_D$  = diagonal matrix of the base demands; and  $D = np$  by  $np$  diagonal matrix where the diagonal elements are the derivatives of the head-loss equations of the flows in pipes. The solution for Eq. (4) is the Jacobian matrix  $J$  of flows and heads with respect to water nodal demands

$$J(y) = \begin{cases} \frac{\partial q}{\partial \theta} = -[D^{-1}A][A^T D^{-1}A]^{-1}G_D \\ \frac{\partial h}{\partial \theta} = -[A^T D^{-1}A]^{-1}G_D \end{cases} \quad (5)$$

Because of an estimate of the unknown demand parameters ( $f_0$ ), the SD greedy algorithm method from Piller (1995) iteratively selects the measurement locations (S matrix) by minimizing the influence of measurement errors on the state vector estimation. The selection matrix  $S$  of the measurement locations is chosen so that the matrix  $(ST_0)$  is full rank and the infinity norm of its pseudoinverse matrix,  $\|((ST_0)^\dagger)\|_\infty$ , has a minimum value. The matrix  $T_0 = E_1 J_0 E_2$  = equilibrium matrix of the Jacobian matrix, where  $E_1$  = premultiplied diagonal matrix to ensure the precision of the measurements at links and nodes,  $E_2$  = postmultiplied matrix corresponding to the change of the parameter variables, and  $J_0 = J(y_{f_0})$  = Jacobian matrix computed by Eq. (5) at  $f_0$ . In this study,  $E_1$  = identity matrix given that all measurements are assumed to have the same precision, and  $E_2$  is computed by Eq. (6) because only the demand parameters need to be calibrated

$$E_2 = \text{diag}(f_0) \quad (6)$$

Table 1 shows an example of a selection matrix where the elements are assigned a value of 1 if the pipes/nodes are selected as measurement sites, and 0 for nonselected pipes/nodes. A detailed explanation of the method is described in Piller (1995).



**Fig. 3.** Case study 1 for calibration problem

### Case Study 1

The first case study used to evaluate the methodology is shown in Fig. 3. The network consists of nine pipes, six nodes with unknown demands (from two to seven) and a 5.0 m-diameter tank at node 8 with a water surface elevation of 15.0 m. The system is fed by a reservoir at node 1 with the head of 31.5 m. All nodes are assumed to have an elevation of 0.0 m and base demands of 15.1, 10.3, 11.8, 15.6, 11.3, and 8.4 L/s for nodes 2 to 7, respectively. The demand multiplier factors assigned for the system (and which will need to be calibrated by the GA) are  $X_0 = [0.5; 0.6; 0.8; 0.7; 0.6; \text{and } 0.9]$  at nodes 2 to 7, respectively. Table 2 shows the pipe characteristics for the test network.

#### Input for Calibration Model

In practice, input data for the calibration process are usually collected from a supervisory control and data acquisition (SCADA) system. In this research, input data are generated using the EPANET toolkit as follows: (1) known demand multiplier factors are assigned to nodal demands; (2) run EPANET to retrieve the corresponding true pipe flow rates and nodal heads (these are then the measured data); (3) select the flows and heads at the locations chosen by the SD model as input for the calibration model. The output flows and heads for selected pipes and nodes based on simulation of the measured values are used for the calibration process.

The selection of measurement locations for case study 1 using the SD greedy algorithm method is shown in Table 3. For one available measurement device, the method selects pipe 1 as the most sensitive location with respect to the demands, because this pipe provides the main flow to the system from the source reservoir.

**Table 2.** Pipe Characteristics for Case Study 1

Pipe	L (m)	D (mm)	$\epsilon$ (mm)
2	1,609	254	0.25
3	1,609	254	0.25
4	1,609	203	0.25
5	1,609	203	0.25
6	1,609	203	0.25
7	1,609	254	0.25
8	1,609	203	0.25
9	643.7	254	0.25
1	828	356	0.25

**Table 3.** Selection of the Measurement Locations for Case Study 1

Number of measurements	Location(s)	$\ ((ST_0)^\dagger)\ _\infty$
1	<b>P1</b>	0.054
2	P1, <b>P9</b>	0.133
3	P1, <b>P3</b> , P9	0.229
4	P1, P3, <b>P7</b> , P9	0.299
5	P1, P3, P7, <b>P8</b> , P9	0.456

Note: Bold values are selected based on the minimum value of  $\|((ST_0)^\dagger)\|_\infty$  for the SD greedy algorithm (Piller 1995).

When two measurement sites need to be selected, pipes 1 and 9 are chosen as the two most sensitive places for the measurement of the flow. The availability of these two measurements provides the information of the total inflow to the network, which is equivalent to the total water demand of the system. It is noted that the proposed calibration method is to be applied to cases where the number of measurement locations is less than the number of unknowns. Because there are six unknown demand multiplier factors that need to be calibrated, only up to five available measurement sites are considered for this test network. From a practical point of view, having so many measurement devices (relative to the total number of nodes of six) in a water network would be unreasonable owing to their cost. However, the main purpose of this section is to evaluate the ability of the GA model to calibrate the demands at the nodes with acceptable accuracy. Thus, the problems related to device costs are ignored in this context.

### Nonuniqueness of the Solutions

Fig. 4 shows an example of alternative solutions when two measurement locations at pipes 1 and 9 are available. Fig. 4(a) presents the true solution of the problem, where the set of six nodal demands has a total of 48.43 L/s and results in the flow rates of 65.46 L/s and 17.09 L/s for pipe 1 and pipe 9, respectively. In Figs. 4(b and c), two different sets of nodal demands also cause the same values of flow rates at the measurement locations.

Apparently, for this network, different sets of six nodal demands with the total of 48.43 L/s that satisfy the water network equations can be a possible solution of the problem. Moreover, the number of possible solutions that match the measured flows will increase if the constraint of the total demand (given by the measurement in pipes 1 and 9) is released, i.e., the two measurements are located in different pipes of the network. A single run of the GA model, therefore, might converge to any of the nonunique solutions or be trapped at a local optimal solution where the simulated values cannot perfectly match the known values at measurement locations. As a result, it appears that a good approximation of the demand multiplier factors calibration problem can only be obtained if multiple runs of the GA model are implemented. The following section shows the results of the multiple runs of the proposed GA model.

### Results of GA Calibration Model

The GA calibration model has been tested with four different scenarios of measurement locations, from two to five available measurement sites. To evaluate the influence of the number of measurement sites on the GA model, the GA parameters are kept constant during all experiments with the size of population  $N = 100$ , probability of crossover  $P_c = 0.8$ , probability of mutation  $P_m = 0.3$  and the number of generations is 1,000. The range of decision variables is selected from 0 to 1.5 with the increment of

$\Delta\theta = 0.02$ , corresponding to a search space size of  $76^6 = 1.927 \times 10^{11}$  possible solutions. In addition, owing to the nonuniqueness and the stochastic behavior of the problem, for each GA application, 100 runs with 100 different seeds were implemented. The results of different GA runs to the case study 1 are presented in Figs. 5–7.

Figs. 5(a), 6(a), and 7(a) plot the results of flow rates, nodal heads, and nodal demands, respectively, from 100 GA runs where two measurement locations at pipes 1 and 9 are available. As seen from Fig. 5(a), the calibrated flows at the measurement locations are well matched with the actual known or measured values for all 100 GA runs. Specifically, the simulated values at pipe 9 are  $17.03 \pm 0.02$  L/s (the actual value is 17.03 L/s), the simulated values at pipe 1 are  $65.46 \pm 0.03$  L/s (the actual value is 65.46 L/s). On the other hand, large variations of simulated flow rates are observed in all other pipes of the network for the individual 100 GA runs, for instance, the range of flow at pipe 6 is from 8.28 to 18.46 L/s, whereas the actual value is 13.83 L/s in Fig. 5(a).

Similar to the variation of the simulated flow rates is the variation of nodal heads, as shown in Fig. 6(a). Because of the similarity of the simulated results for the flow rates at pipes 1 and 9, the estimation of nodal heads for the individual 100 GA runs at nodes 2 and 7 are also well matched with the actual values. Meanwhile, the simulated heads at the remaining nodes (node 3, 4, 5, and 6), compared with actual heads, vary approximately  $\pm 0.7$ ,  $\pm 0.9$ ,  $\pm 1.0$ , and  $\pm 0.8$  m, respectively. In Fig. 7(a), the calibrated nodal demands for the individual 100 GA runs show a large variation at all nodes, starting from 0 L/s up to approximately twice the magnitude of the actual nodal demands.

It is observed in Figs. 5(b–d), 6(b–d), and 7(b–d) that the addition of measurement sites to the network increases the accuracy of the calibration model. As an example, consider the effect of adding a measurement at a pipe in the network, from three measurements (at pipe 1, 3, and 9) to four measurements (at pipes 1, 3, 7, and 9). The simulated flow rates and nodal heads at all nonmeasured locations are slightly improved. If with three measurement locations, the simulated flows and the simulated heads at pipe 6 and node 6 vary for the individual 100 GA runs in the ranges of [8.79, 17.67] L/s and [26.14, 27.56] m, respectively, with four measurement locations, the simulated values are improved, varying in smaller ranges of [11.71, 16.35] L/s and [26.41, 27.53] m. On the other hand, the calibrated demands show a significant improvement when the number of measurement sites increased. At node 5, for instance, the variation of the demand is improved from the maximum allowable range [0, 23.4] L/s to [8.42, 13.01] L/s.

The best GA solution and the average values of 100 GA runs are also plotted to compare them with the actual solution. The results show that an increasing number of measurement sites leads to better accuracy of the best GA result. Owing to the large search space size, the GA calibration model cannot find the exact solution in

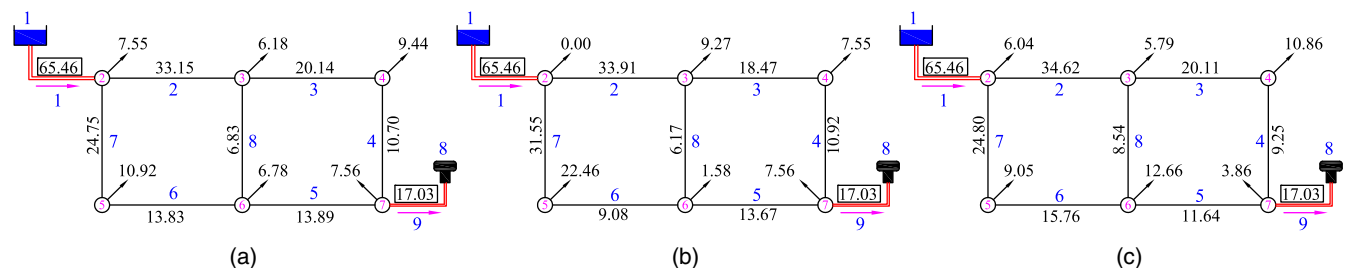
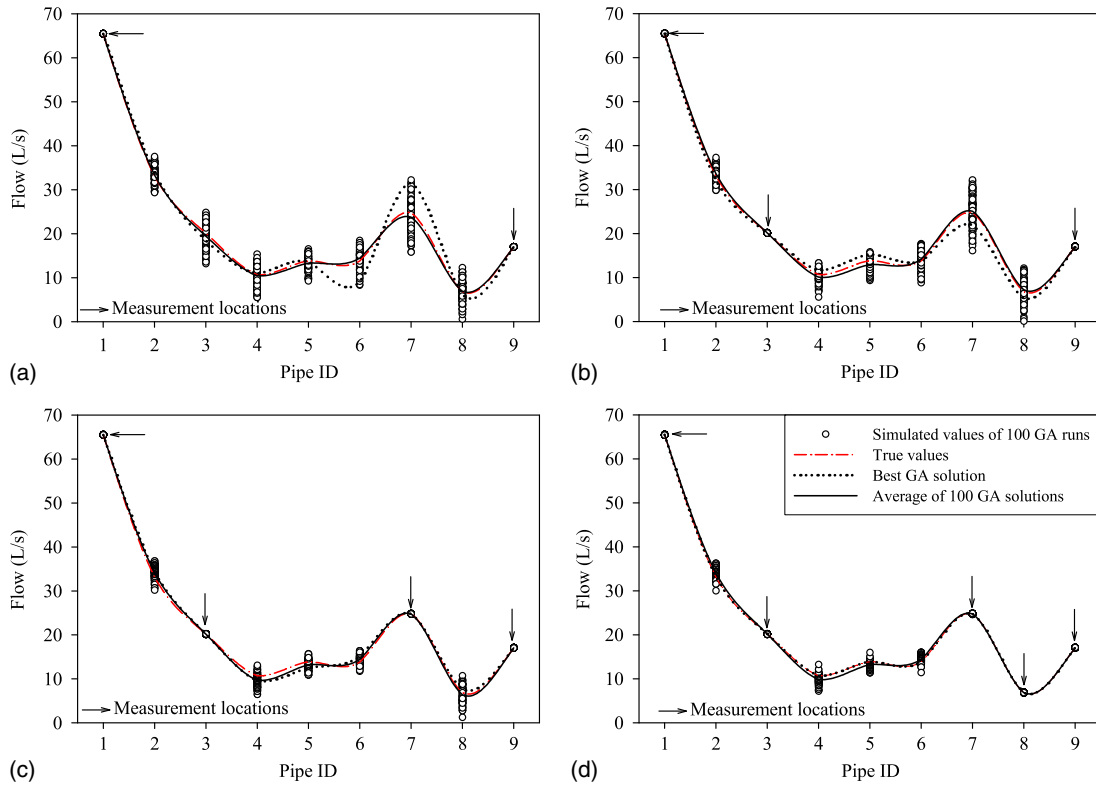
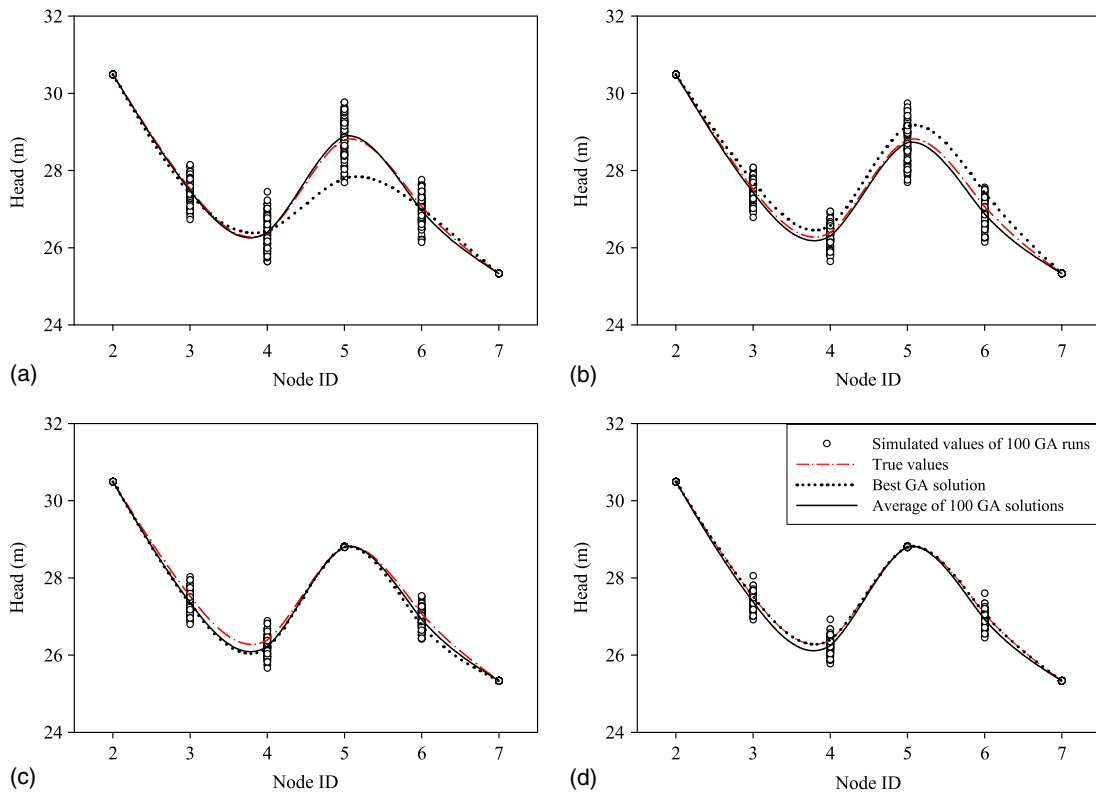


Fig. 4. Example of nonuniqueness of solutions with two available measurement locations: (a) actual solution; (b) solution 1; (c) solution 2

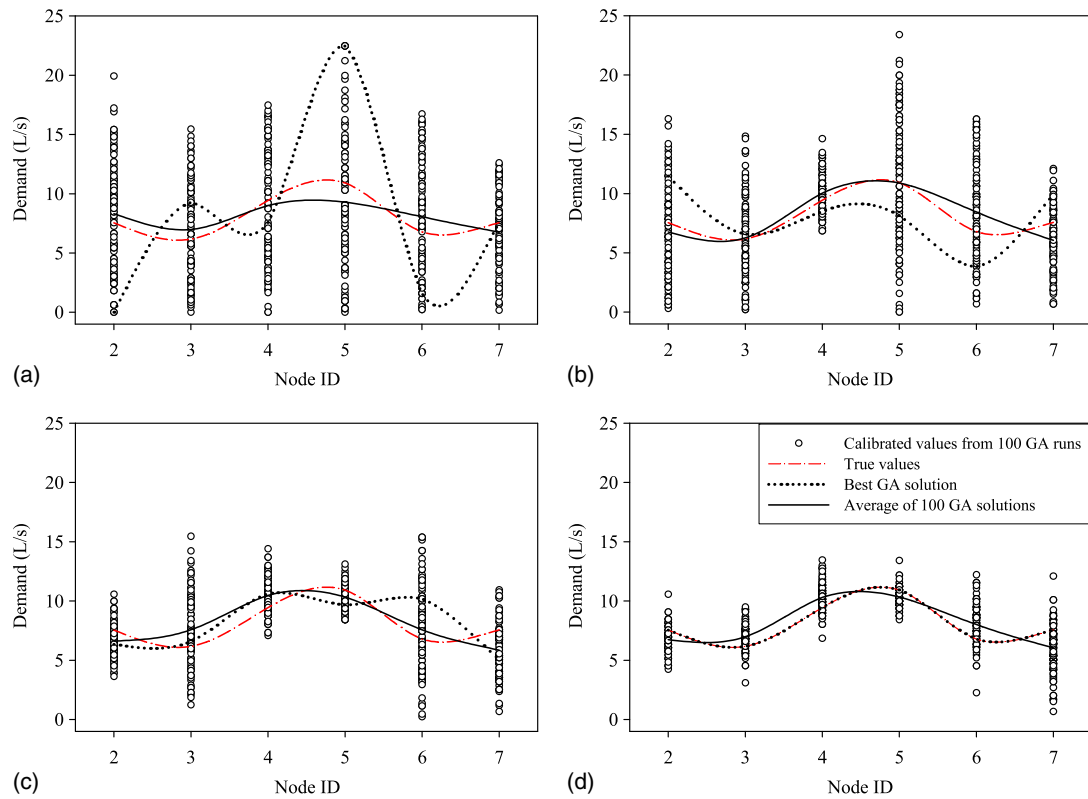


**Fig. 5.** Results of flow rates from GA calibration model for case study 1: (a) two measurements at pipes 1 and 9; (b) three measurements at pipes 1, 3, and 9; (c) four measurements at pipes 1, 3, 7, and 9; (d) five measurements at pipes 1, 3, 7, 8, and 9



**Fig. 6.** Results of nodal heads from GA calibration model for case study 1: (a) two measurements at pipes 1 and 9; (b) three measurements at pipes 1, 3, and 9; (c) four measurements at pipes 1, 3, 7, and 9; (d) five measurements at pipes 1, 3, 7, 8, and 9





**Fig. 7.** Results of nodal demands from GA calibration model for case study 1: (a) two measurements at pipes 1, and 9; (b) three measurements at pipes 1, 3, and 9; (c) four measurements at pipes 1, 3, 7, and 9; (d) five measurements at pipes 1, 3, 7, 8, and 9

**Table 4.** Comparison of Mean Estimated Pipe Flows and Actual Flows for Case Study 1

	Pipe	P2	P3	P4	P5	P6	P7	P8	P9	P1
Measurements	Actual flows (L/s)	33.15	20.14	10.70	13.89	13.83	24.75	6.83	17.03	65.46
2 measurements (P1, P9)	Average flows of 100 GA runs (L/s)	33.38	19.46	10.46	13.33	14.47	23.76	6.93	<b>17.03</b>	<b>65.46</b>
	Standard deviation	1.92	2.95	2.41	1.71	2.26	3.94	2.40	<b>0.01</b>	<b>0.01</b>
	% $\Delta_{(Average\ flows, actual\ flows)}$	0.67%	3.35%	2.25%	4.01%	4.57%	4.01%	1.38%	<b>0.01%</b>	<b>0.00%</b>
3 measurements (P1, P3, P9)	Average flows of 100 GA runs (L/s)	33.63	<b>20.14</b>	10.10	12.97	14.18	25.07	7.24	<b>17.02</b>	<b>65.46</b>
	Standard deviation	1.61	<b>0.02</b>	1.60	1.49	2.05	3.79	2.55	<b>0.03</b>	<b>0.03</b>
	% $\Delta_{(Average\ flows, actual\ flows)}$	1.44%	<b>0.02%</b>	5.60%	6.62%	2.50%	1.28%	5.95%	<b>0.04%</b>	<b>0.00%</b>
4 measurements (P1, P3, P7, P9)	Average flows of 100 GA runs (L/s)	34.09	<b>20.14</b>	9.66	13.21	14.41	<b>24.75</b>	6.37	<b>17.02</b>	<b>65.46</b>
	Standard deviation	1.43	<b>0.04</b>	1.47	1.08	1.00	<b>0.04</b>	2.08	<b>0.03</b>	<b>0.03</b>
	% $\Delta_{(Average\ flows, actual\ flows)}$	2.82%	<b>0.00%</b>	9.69%	4.87%	4.18%	<b>0.02%</b>	6.76%	<b>0.02%</b>	<b>0.00%</b>
5 measurements (P1, P3, P7, P8, P9)	Average flows of 100 GA runs (L/s)	33.97	<b>20.14</b>	9.84	13.23	14.41	<b>24.75</b>	<b>6.84</b>	<b>17.03</b>	<b>65.46</b>
	Standard deviation	1.12	<b>0.04</b>	1.14	0.86	0.80	<b>0.06</b>	<b>0.05</b>	<b>0.03</b>	<b>0.04</b>
	% $\Delta_{(Average\ flows, actual\ flows)}$	2.46%	<b>0.01%</b>	8.06%	4.70%	4.19%	<b>0.00%</b>	<b>0.15%</b>	<b>0.00%</b>	<b>0.00%</b>

Note: Bold = calibrated values at measurement locations;  $\Delta$  = differences between calibrated values and actual values.

any of the 100 GA runs unless five measurement sites are provided. However, the optimal solution is only found three times out of 100 GA runs in the scenario of five available measurement sites.

The average values of 100 GA runs, on the contrary, yield a very good match with the actual values of the problem. In all measurement scenarios, the mean flow rates, nodal heads, and nodal demands, which are tabulated in Tables 4–6, respectively, are slightly different from the actual values, with less than 10% error for the flow rates, less than 1% error for the nodal heads, and less than 20% error for the nodal demands. The results clearly indicate that the nodal heads are estimated more accurately than the flows. In addition, the standard deviations computed based on the 100 GA runs, which represent the variation of the calibrated results,

show that the change of the nodal demands causes greater variation of the flow rates rather than the nodal heads in this looped network. As a result, the flow rates are more sensitive to the demands than the nodal heads.

#### Effects of Increment of Decision Variables on GA Calibration Results

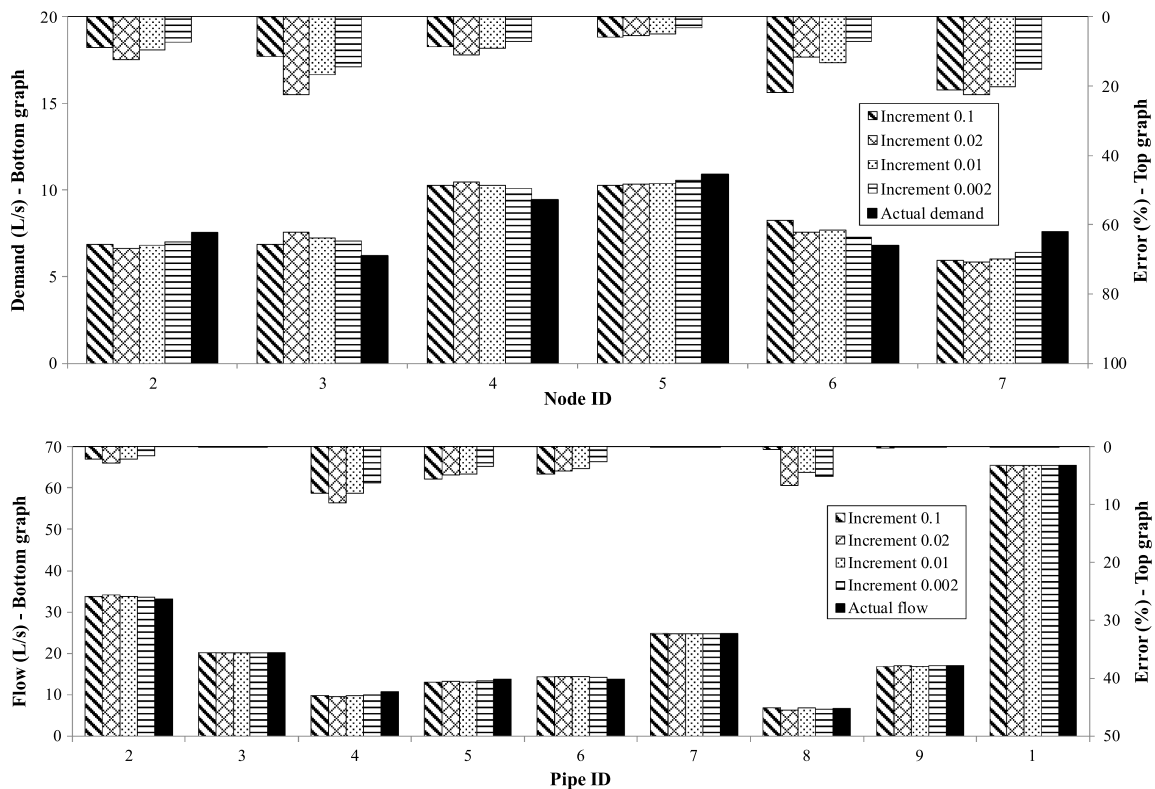
One of the factors that may affect the accuracy of the GA calibration model is the increment ( $\Delta\theta$ ) of the demand multiplier factors. The selection of a large increment for the decision variables leads to faster convergence of the GA model, although it may result in a coarser approximation of the calibrated demands. Alternatively, the GA model can give better results if smaller increment steps

**Table 5.** Comparison of Mean Estimated Nodal Heads and Actual Heads for Case Study 1

	Node	N2	N3	N4	N5	N6	N7
Measurements	Actual head (m)	30.49	27.53	26.39	28.80	27.07	25.33
2 measurements (P1, P9)	Average head of 100 GA runs(m)	30.49	27.48	26.39	28.89	26.97	25.33
	Standard deviation	0.00	0.33	0.44	0.50	0.39	0.00
	% $\Delta_{(Average\ head,\ actual\ head)}$	0.00%	0.17%	0.01%	0.31%	0.39%	0.00%
3 measurements (P1, P3, P9)	Average head of 100 GA runs(m)	30.49	27.44	26.30	28.72	26.88	25.33
	Standard deviation	0.00	0.28	0.28	0.50	0.33	0.00
	% $\Delta_{(Average\ head,\ actual\ head)}$	0.00%	0.32%	0.34%	0.27%	0.72%	0.00%
4 measurements (P1, P3, P7, P9)	Average head of 100 GA runs (m)	30.49	27.36	26.22	28.80	26.92	25.33
	Standard deviation	0.00	0.25	0.25	0.00	0.25	0.00
	% $\Delta_{(Average\ head,\ actual\ head)}$	0.00%	0.61%	0.64%	0.00%	0.55%	0.00%
5 measurements (P1, P3, P7, P8, P9)	Average head of 100 GA runs(m)	30.49	27.38	26.25	28.80	26.93	25.33
	Standard deviation	0.00	0.20	0.20	0.01	0.20	0.00
	% $\Delta_{(Average\ head,\ actual\ head)}$	0.00%	0.53%	0.55%	0.00%	0.54%	0.00%

**Table 6.** Comparison of Mean Calibrated Nodal Demands and Actual Demands for Case Study 1

	Node	N2	N3	N4	N5	N6	N7
Measurements	Actual demand (L/s)	7.55	6.18	9.44	10.92	6.78	7.56
2 measurements (P1, P9)	Average demand of 100 GA runs (L/s)	8.32	6.98	9.01	9.29	8.06	6.76
	Standard deviation	4.53	3.99	4.86	5.78	4.73	3.30
	% $\Delta_{(Average\ demand,\ actual\ demand)}$	10.20%	13.00%	4.60%	14.89%	18.93%	10.53%
3 measurements (P1, P3, P9)	Average demand of 100 GA runs (L/s)	6.76	6.26	10.03	10.89	8.45	6.05
	Standard deviation	3.93	3.51	1.61	5.51	4.25	2.63
	% $\Delta_{(Average\ demand,\ actual\ demand)}$	10.52%	1.23%	6.30%	0.26%	24.63%	19.98%
4 measurements (P1, P3, P7, P9)	Average demand of 100 GA runs (L/s)	6.62	7.58	10.48	10.34	7.57	5.85
	Standard deviation	1.43	3.06	1.48	1.00	3.54	2.24
	% $\Delta_{(Average\ demand,\ actual\ demand)}$	12.32%	22.63%	10.98%	5.34%	11.70%	22.62%
5 measurements (P1, P3, P7, P8, P9)	Average demand of 100 GA runs (L/s)	6.73	6.98	10.30	10.34	8.02	6.04
	Standard deviation	1.10	1.11	1.15	0.81	1.67	2.01
	% $\Delta_{(Average\ demand,\ actual\ demand)}$	10.80%	13.00%	9.15%	5.31%	18.33%	20.04%



**Fig. 8.** Calibrated nodal demands and estimated pipe flows with different increment steps for case study 1

are selected. However, the model requires more computational effort to converge attributable to the larger search space size. To evaluate the effect of increment steps on GA calibration results, different increment steps of 0.1, 0.02, 0.01, and 0.002 were selected and tested in the case of four available measurement locations at pipes 1, 3, 7, and 9. The average calibrated demands and estimated flows are shown in Fig. 8.

The results of the GA model for different increment steps of the decision variables are approximately comparable. Consider the average calibrated demand at node 7 as an example. The largest demand error (22.62%) occurs at  $\Delta\theta = 0.02$ , followed by 21.22% at  $\Delta\theta = 0.1$ , 20.23% at  $\Delta\theta = 0.01$ , and 15.2% at  $\Delta\theta = 0.002$ . With regard to the simulated flows, the estimation errors for different increment steps are relatively small. The maximum errors all occur at pipe 4 and are of almost the same magnitude, 6.28, 8.02, 8.03, and 9.69%, corresponding to  $\Delta\theta = 0.002, 0.01, 0.1,$  and  $0.02$ , respectively. Hence, for this case study, it can be concluded that the GA model is not particularly sensitive to the increment of the decision variable in the range of 0.002 to 0.1.

### Effects of Measurement Locations on GA Calibration Results

The GA calibration results showed that, even with a very large search space, the GA model is able to find the exact solution of the problem for at least one of the 100 GA runs. Thus, the effects of measurement locations on the GA calibration results can be evaluated by the number of times the algorithm converges to the true solution when different combinations of measurement sites are tested.

In this investigation, it is assumed that there are four measurement sites available. Four different combinations of measurement sites are tested to evaluate the convergence of the GA model, which is shown in Table 7.

**Table 7.** Effects of Measurement Locations on GA Calibration Results

Experiment	Measurement locations	Number of convergences to optimal solution out of 100 runs	Note
1	P1, P3, P7, P9	16	SD method
2	P1, P6, P8, P9	7	Random
3	P1, P9, N3, N6	4	Random
4	N2, N5, N4, N7	5	Random

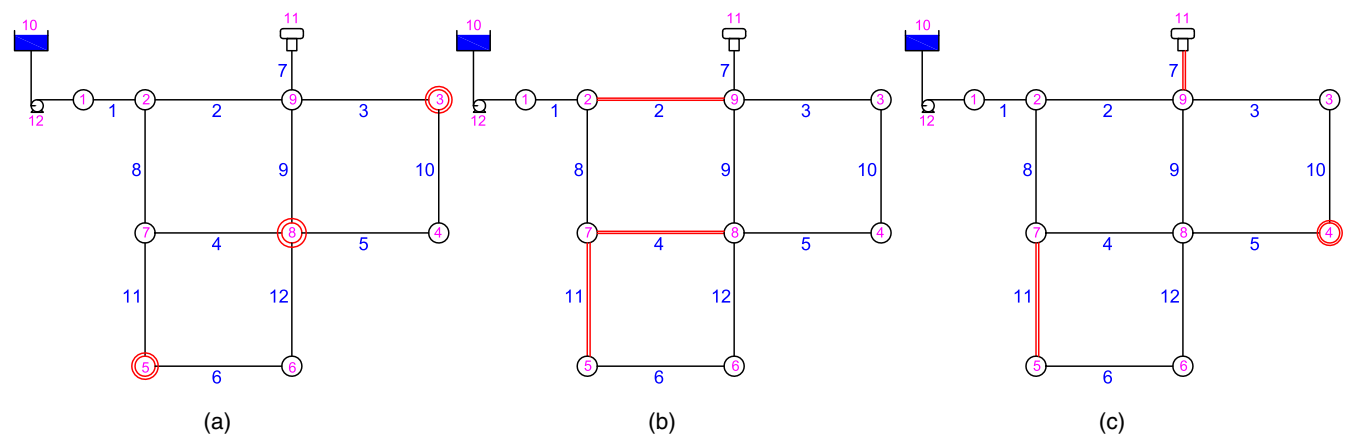
The measurement locations in the first experiment are selected by the SD greedy algorithm method, and the four flow measurement sites are placed at pipes 1, 3, 7, and 9. In the second and the third experiments, two flow measurements are kept at the same locations at pipes 1 and 9, whereas the two other measurement locations are selected randomly. This leads to two flow measurement sites being located at pipes 6 and 8 for experiment 2 and two pressure measurement sites located at nodes 3 and 6 for experiment 3. The last experiment involves the presence of four pressure measurement sites at nodes 2, 4, 5, and 7.

The GA parameters of the GA model were kept constant and set equal to the values presented in the previous section, except for the increment of the decision variables. The increment in this test was selected  $\Delta\theta = 0.1$ , so that it was possible to fully enumerate the search space ( $16.77 \times 10^6$  possible solutions). The enumeration of the problem shows that with this relatively small search space size, the problem has a unique optimal solution for all four experiments. In this case, the objective function reaches exactly zero when the calibrated demand multiplier factors are identical to the actual demand multiplier factors assigned to the network. The number of convergences to the optimal solution therefore, is the number of times out of 100 GA runs in which the GA model can find the exact solution of the problem. As seen in Table 7, the first experiment results in the highest number of convergences with 16 times out of 100 runs. The second experiment has seven times, followed by the fourth and the third experiment where the number of convergences is five times and four times out of 100 runs, respectively. The output of the GA calibration model in this case study, therefore, seems to be sensitive to the locations and the types of the measurement in the network.

### Case Study 2

The second case study is considered to compare the performance of the GA model with the SVD model from Cheng and He (2010). The network has nine nodes, 12 pipes, one tank, one pump, and one reservoir. The network topology and all information including the pipe parameters, length, roughness, and pump characteristic can be found from the EPANET example (Rossman 2000), namely the Net1 network (Fig. 9).

The GA calibration model (with a population size of  $N = 100$ , probability of crossover  $P_c = 0.8$  and probability of mutation  $P_m = 0.15$ ) was tested for three scenarios of measurement sites



**Fig. 9.** Case study 2 network for calibration problem: (a) test 1 (three heads measured); (b) test 2 (three flows measured); (c) test 3 (two flows, one head measured)

**Table 8.** Comparison of SVD Model (Cheng and He 2010) and GA Model for Case Study 2

Node number	Test 1					Test 2				Test 3			
	Real DM (GPM)	SVD (GPM)	GA (GPM)	$\Delta_{SVD}$ (%)	$\Delta_{GA}$ (%)	SVD (GPM)	GA (GPM)	$\Delta_{SVD}$ (%)	$\Delta_{GA}$ (%)	SVD (GPM)	GA (GPM)	$\Delta_{SVD}$ (%)	$\Delta_{GA}$ (%)
2	150	185.78	<b>118.67</b>	23.85	<b>20.89</b>	145.67	157.94	2.89	5.29	158.49	<b>150.09</b>	5.66	<b>0.06</b>
3	100	151.09	<b>93.15</b>	51.09	<b>6.85</b>	112.80	<b>101.88</b>	12.80	<b>1.88</b>	122.65	<b>85.76</b>	22.65	<b>14.24</b>
4	300	216.44	<b>327.15</b>	27.85	<b>9.05</b>	200.54	<b>265.88</b>	33.15	<b>11.37</b>	279.36	<b>314.48</b>	6.88	<b>4.83</b>
5	50	49.26	<b>50.34</b>	1.48	<b>0.69</b>	49.49	51.63	1.02	3.25	77.64	<b>52.81</b>	55.28	<b>5.62</b>
6	50	64.01	35.90	28.02	28.20	62.11	<b>38.90</b>	24.22	<b>22.20</b>	58.43	38.92	16.86	22.16
7	150	159.26	164.15	6.17	9.43	151.05	142.06	0.70	5.29	127.40	<b>153.85</b>	15.07	<b>2.57</b>
8	150	221.00	<b>140.58</b>	47.33	<b>6.28</b>	209.63	<b>96.45</b>	39.75	<b>35.70</b>	157.52	141.97	5.01	5.35
9	150	129.21	<b>131.45</b>	13.86	<b>12.36</b>	164.80	168.36	9.87	12.24	114.76	<b>162.12</b>	23.49	<b>8.08</b>
$\Sigma DM$	1,100	1,198.0	1,061.4	8.91	<b>3.51</b>	1,099.9	1,043.1	0.00	5.17	1,100	1,100	0.00	0.00

Note: Bold = cases where the GA model finds better calibrated results.

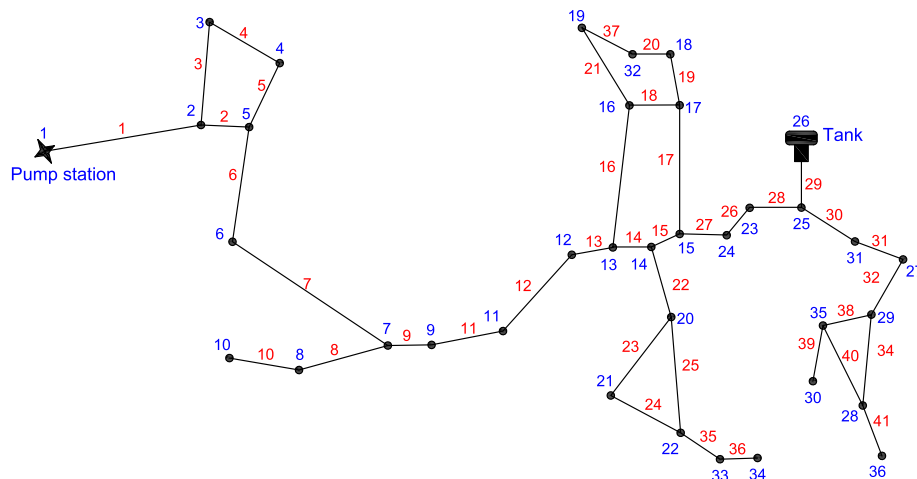
according to Cheng and He (2010). In the first test, three pressure sensors are assumed to be located at nodes 3, 5, and 8. The second test assumes the flow meters are set at pipes 2, 4, and 11, whereas the third test assumes two flow sensors and a pressure sensor are placed at pipes 7 and 11 and node 4, respectively. The calibrated demands and the differences between real demands and calibrated demands of the SVD model and GA model are shown in Table 8.

For all three tests, the average of 100 runs of the GA model gives more reasonable results for the nodal demands. In test 1, whereas the SVD model leads to large differences between actual and calibrated values at nodes 3 and 8 (51.09 and 47.33%, respectively), the average GA results in relatively small errors (6.85 and 6.28%) at the corresponding nodes. In test 2, the largest error of the two models occurs at node 8, with an error of 39.75% for the SVD model and a slightly smaller error of 35.7% for the GA model. The last test presents the best performance of the GA model with the maximum error of 22.16% at node 6, whereas the SVD model still remains a large error of 55.28% at node 5. For estimating the total water demands of the system (the last row of Table 8), both models achieve reasonable results, especially for the SVD model in tests 2 and 3. The total calibrated demand of the GA model only matches with the real total demand in test 3, where one of the flow sensors is placed at pipe 7 to measure the flow from the tank to the system. This highlights the importance of the selection of the optimal measurement locations. Reasonably accurate results can be achieved if information related to the total flow is provided.

### Case Study 3

The third test network aims at testing the performance of the proposed methodology for a larger network. This network is provided by EPANET, namely the Net2 network, which consists of 40 pipes and 35 nodes, one tank and one pump station. The pump station is modeled as a node with negative demand, which feeds water into the network. It is assumed that up to three measurement devices are able to be installed in the network. In addition, the pump flow is also assumed to be known. The network diagram is shown in Fig. 10.

The SD greedy algorithm model for the selection of measurement locations found that, for this network, the flows in pipes are much more sensitive to the demands than the heads at nodes. Three flow measurement sites are suggested to be located at pipes 12, 22, and 29. The GA calibration model of the Net2 water network was implemented with the following characteristics: (1) the number of decision variables is 32 corresponding to a total of 32 nodal demand multiplier factors of the network; (2) the size of the choice table for the decision variables is selected from  $f_k^{\min} = 0$  to  $f_k^{\max} = 1.5$ ; (3) a population of  $N = 500$ , probability of crossover  $P_c = 0.8$ , probability of mutation  $P_m = 0.04$  and the number of generations  $N = 1,000$  were selected for the GA model parameters; (4) to evaluate the effects of the increment of the decision variables, different increment steps of  $\Delta\theta = 0.0005, 0.005, 0.02, \text{ and } 0.1$ , respectively, were examined.



**Fig. 10.** Case study 3 water distribution network (Net2 network)

**Table 9.** Summary of Average GA Output Errors for Net2 Network with Different Increment Steps

Increment step	Average demand errors (%)	Average flow error (%)	Average head error (%)
0.1	16.41	10.72	0.009
0.02	14.74	10.81	0.009
0.005	15.19	9.96	0.010
0.0005	15.30	10.78	0.011

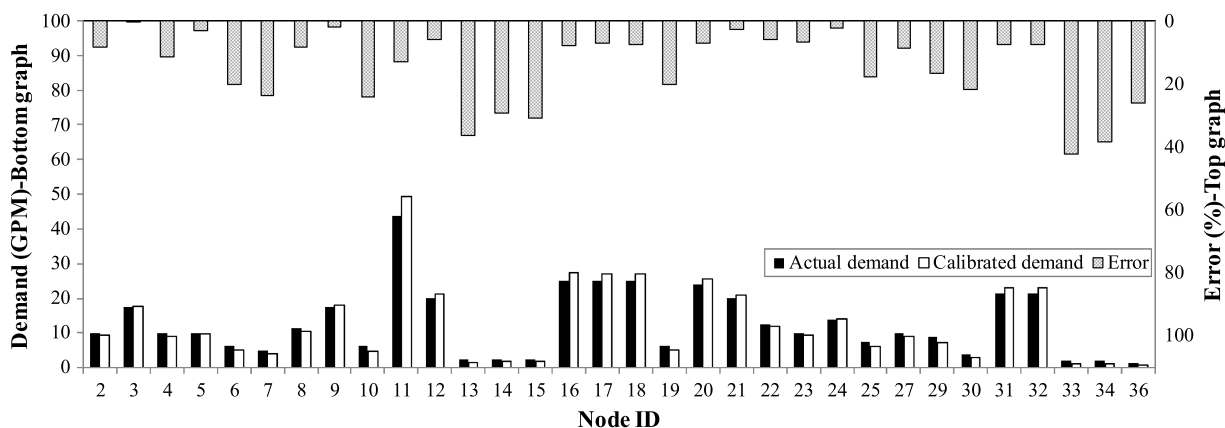
A hundred runs of the GA model were undertaken on the Intel Core i5 (2.9 GHz) computer. The total elapsed run time was approximately 16.3 h, or approximately 10 min for one run. Table 9 summarizes the average errors of the nodal demands, flow rates, and nodal heads in comparison with the true value at nodes and pipes in the network. The errors are approximately equal for four different increment steps of the decision variables. The average errors fluctuate slightly approximately 15% for the demands, 10% for the flows, and only 0.01% for the nodal heads. Hence, this case study also shows that the GA calibration model is not sensitive to the increment selected for the decision variables.

Figs. 11 and 12 plot the outcome from the GA model corresponding with the increment value  $\Delta\theta = 0.02$  as a demonstration

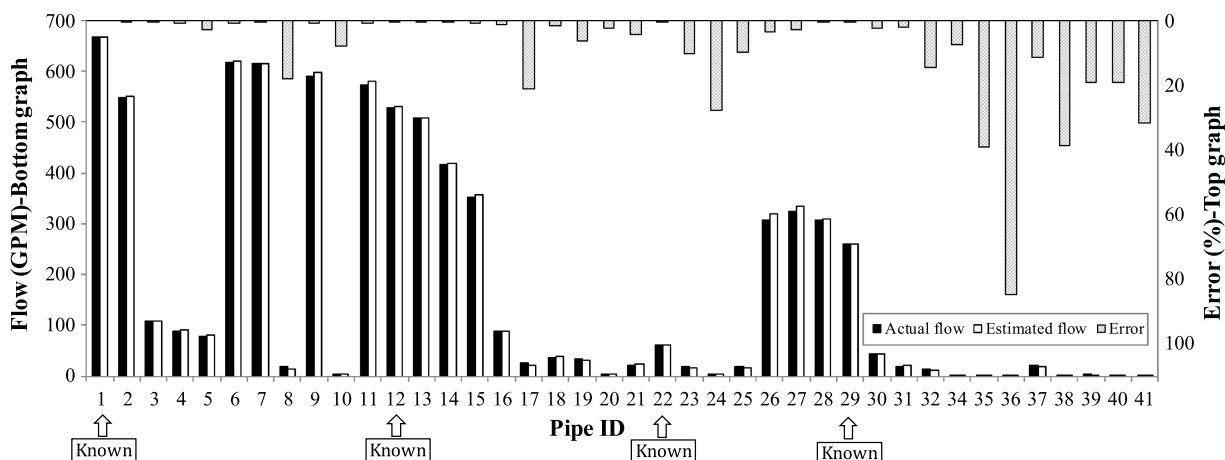
of the results. The bottom graph of each figure presents the average calibrated/simulated values and the true values of the nodal demands/flow rates. The top graph of each figure shows the error percentage of the difference between the calibrated/simulated values and their true values. Because of the relatively accurate estimation of the nodal heads, the plot of the simulated heads and actual heads is not shown here.

Let us consider three groups of the demands (0 to 10, 10 to 30, and >40 GPM). Fig. 11 shows that large calibration percentage errors only occur in the first group, where the demands are small. The errors in this group are within 16.5 and 45.48%. However, all the differences are less than 1.52 GPM in absolute value. The second group shows a very good approximation of the demand, as the percentage errors vary from 0.48 to 8.51%. The last group, which contains only node 11 with the actual demand of 43.82 GPM, also obtains a relatively accurate calibrated result of 49.47 GPM, equivalent to an error of 12.89%.

Similarly, Fig. 12 shows that the flow estimation is generally much more accurate for the pipes with high flows than the pipes with low flows. The largest estimation error, approximately 84.8%, occurs at pipe 36 at which the actual flow is 0.63 GPM and the simulated flow is 1.16 GPM. For all the pipes that have actual flows larger than 30 GPM, the estimation errors are smaller than 6.5%.



**Fig. 11.** Comparison of average calibrated demands (100 GA runs) with actual demands of Net2 network



**Fig. 12.** Comparison of average simulated flows (100 GA runs) with actual flows of Net2 network

## Conclusions and Recommendations

Calibration of water demand in real water distribution systems is complicated by the limited number of measurement sites. A GA model has been developed for the calibration of the demand multiplier factors for underdetermined water distribution systems where the number of measurement locations is less than the number of unknown parameters. The approach for estimation of demand multiplier factors using GA optimization has been tested on three case studies. The first case study has shown that the average values of multiple runs of the GA model can deliver a very good approximation of the water demand multipliers with little information from the SCADA system. The first case study also shows that the location of the measurement sites does influence the performance of the GA model. The second case study demonstrates the advantage of the GA model in comparison with the singular value decomposition model. It also confirms the conclusions made from the first case study about the sensitivity of the GA model to the measurement locations. Therefore, the GA model is suggested to be implemented in combination with a supporting tool for the selection of optimal measurement locations such as the SD greedy algorithm model. The third case study validates the approach for a slightly larger sized network, which again exhibits the superior performance of the GA model. The model run time for this last case study (approximately 16.3 h) might be a disadvantage of the GA model. The model might, therefore, be suitable for the networks at which the SCADA data are provided on a daily basis, or if real-time calibration is required, parallel computer systems would need to be implemented for the GA model.

Future research efforts will involve finding advanced methods for the calibration of the demand to reduce the computational time. In addition, uncertainty of the calibration model is another consideration because of the presence of errors (or noise) in measurement data. Finally, addressing the problem of leakage in the network is also important in achieving reliable results.

## References

Andersen, J. H., and Powell, R. S. (2000). "Implicit state-estimation technique for water network monitoring." *Urban Water*, 2(2), 123–130.

Beal, C., and Stewart, R. (2014). "Identifying residential water end uses underpinning peak day and peak hour demand." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000357, 04014008.

Berry, J. W., Fleischer, L., Hart, W. E., Phillips, C. A., and Watson, J.-P. (2005). "Sensor placement in municipal water networks." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2005)131:3(237), 237–243.

Bhave, P. R. (1988). "Calibrating water distribution network models." *J. Environ. Eng.*, 10.1061/(ASCE)0733-9372(1988)114:1(120), 120–136.

Boulos, P. F., and Ormsbee, L. E. (1991). "Explicit network calibration for multiple loading conditions." *Civ. Eng. Syst.*, 8(3), 153–160.

Boulos, P. F., and Wood, D. J. (1990). "Explicit calculation of pipe-network parameters." *J. Hydraul. Eng.*, 10.1061/(ASCE)0733-9429(1990)116:11(1329), 1329–1344.

Bush, C. A., and Uber, J. G. (1998). "Sampling design methods for water distribution model calibration." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1998)124:6(334), 334–344.

Cheng, W., and He, Z. (2010). "Calibration of nodal demand in water distribution systems." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000093, 31–40.

Datta, R., and Sridharan, K. (1994). "Parameter estimation in water-distribution systems by least squares." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1994)120:4(405), 405–422.

Davidson, J., and Bouchart, F.-C. (2006). "Adjusting nodal demands in SCADA constrained real-time water distribution network models." *J. Hydraul. Eng.*, 10.1061/(ASCE)0733-9429(2006)132:1(102), 102–110.

De Schaezen, W. B. F. (2000). "Optimal calibration and sampling design for hydraulic network models." Univ. of Exeter, Exeter, U.K.

Di Nardo, A., Di Natale, M., Gissoni, C., and Iervolino, M. (2015). "A genetic algorithm for demand pattern and leakage estimation in a water distribution network." *J. Water Supply: Res. Technol.-Aqua*, 64(1), 35–46.

Giustolisi, O., and Ridolfi, L. (2014). "New modularity-based approach to segmentation of water distribution networks." *J. Hydraul. Eng.*, 10.1061/(ASCE)HY.1943-7900.0000916, 04014049.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning*, Addison-Wesley Longman, Menlo Park, CA.

Goldberg, D. E., and Deb, K. (1991). "A comparative analysis of selection schemes used in genetic algorithms." *Foundations of genetic algorithms*, G. J. E. Rawlins, ed., Vol. 1, Morgan Kaufmann, San Francisco, 69–93.

Haddad, O., Adams, B., and Marino, M. (2008). "Optimum rehabilitation strategy of water distribution systems using the HBMO algorithm." *J. Water Supply: Res. Technol.-AQUA*, 57(5), 337–350.

Hutton, C., Kapelan, Z., Vamvakieridou-Lyroudia, L., and Savic, D. (2012). "Real-time demand estimation in water distribution systems under uncertainty." *WDSA 2012: 14th Water Distribution Systems Analysis Conf.*, Engineers Australia, Barton, ACT, Australia.

Kang, D., and Lansey, K. (2009). "Real-time demand estimation and confidence limit analysis for water distribution systems." *J. Hydraul. Eng.*, 10.1061/(ASCE)HY.1943-7900.0000086, 825–837.

Kapelan, Z. S., Savic, D. A., and Walters, G. A. (2003). "A hybrid inverse transient model for leakage detection and roughness calibration in pipe networks." *J. Hydraul. Res.*, 41(5), 481–492.

Kapelan, Z. S., Savic, D. A., and Walters, G. A. (2005). "Optimal sampling design methodologies for water distribution model calibration." *J. Hydraul. Eng.*, 10.1061/(ASCE)0733-9429(2005)131:3(190), 190–200.

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.*, 10.1061/(ASCE)0733-9429(2007)133:8(927), 927–936.

Krause, A., Leskovec, J., Guestrin, C., VanBriesen, J., and Faloutsos, C. (2008). "Efficient sensor placement optimization for securing large water distribution networks." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2008)134:6(516), 516–526.

Kun, D., Tian-Yu, L., Jun-Hui, W., and Jin-Song, G. (2015). "Inversion model of water distribution systems for nodal demand calibration." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000506, 04015002.

Lansey, K. E., and Basnet, C. (1991). "Parameter estimation for water distribution networks." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1991)117:1(126), 126–144.

Liberatore, S., and Sechi, G. M. (2009). "Location and calibration of valves in water distribution networks using a scatter-search Meta-heuristic approach." *Water Resour. Manage.*, 23(8), 1479–1495.

Liggett, J. A., and Chen, L.-C. (1994). "Inverse transient analysis in pipe networks." *J. Hydraul. Eng.*, 10.1061/(ASCE)0733-9429(1994)120:8(934), 934–955.

Martínez, F., Bartolín, H., Bou, V., and Kapelan, Z. (2003). "Calibration of valve-driven water distribution systems using jointly GIS and genetic algorithms: Application to the WDS of Valencia (Spain)." *7th Int. Conf. on Computing and Control for the Water Industry (CCWI)*, Instituto de Ingeniería del Agua y Medio Ambiente (IIAMA), Valencia, Spain.

Meier, R. W., and Barkdoll, B. D. (2000). "Sampling design for network model calibration using genetic algorithms." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2000)126:4(245), 245–250.

Nagar, A. K., and Powell, R. S. (2002). "LFT/SDP approach to the uncertainty analysis for state estimation of water distribution systems." *Control Theory and Applications, IEE Proc.*, Vol. 149, Institution of Engineering and Technology (IET), Stevenage, U.K., 137–142.

Nicklow, J., et al. (2010). "State of the art for genetic algorithms and beyond in water resources planning and management." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000053, 412–432.

Ormsbee, L. E. (1989). "Implicit network calibration." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1989)115:2(243), 243–257.

- Ormsbee, L. E., and Lingireddy, S. (1997). "Calibration of hydraulic network models." *J. AWWA*, 89(2), 42–50.
- Ormsbee, L. E., and Wood, D. J. (1986). "Explicit pipe network calibration." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1986)112:2(166), 166–182.
- Piller, O. (1995). "Modeling the behavior of a network—Hydraulic analysis and sampling procedures for parameter estimation." Ph.D. thesis, Univ. of Bordeaux, Bordeaux Cedex, France.
- Piller, O., Gilbert, D., and Van Zyl, J. E. (2010). "Dual calibration for coupled flow and transport models of water distribution systems." *Water Distribution Systems Analysis 2010*, ASCE, Reston, VA, 722–731.
- Preis, A., Whittle, A., and Ostfeld, A. (2009). "Online hydraulic state prediction for water distribution systems." *World Environmental and Water Resources Congress*, ASCE, Kansas City, MO.
- Propato, M., Cheung, P. B., and Piller, O. (2006). "Sensor location design for contaminant source identification in water distribution systems." *Proc., 8th Annual Water Distribution System Analysis Symp.*, ASCE, Reston, VA.
- Pudar, R. S., and Liggett, J. A. (1992). "Leaks in pipe networks." *J. Hydraul. Eng.*, 10.1061/(ASCE)0733-9429(1992)118:7(1031), 1031–1046.
- Rossman, L. A. (2000). "EPANET2." *EPA/600/R-00/057*, Water Supply and Water Resources Division, National Risk Management Research Laboratory, Office of Research and Development, U.S. EPA, Cincinnati.
- Savic, D. A., Kapelan, Z. S., and Jonkergouw, P. M. (2009). "Quo vadis water distribution model calibration?" *Urban Water J.*, 6(1), 3–22.
- Seifollahi-Aghmiuni, S., Bozorg Haddad, O., Omid, M. H., and Mariño, M. A. (2013). "Effects of pipe roughness uncertainty on water distribution network performance during its operational period." *Water Resour. Manage.*, 27(5), 1581–1599.
- Shamir, U., and Howard, C. D. (1977). "Engineering analysis of water distribution systems." *J. Am. Water Works Assoc.*, 69(9), 510–514.
- Shang, F., Uber, J., van Bloemen Waanders, B., Boccelli, D., and Janke, R. (2006). "Real time water demand estimation in water distribution systems." *Proc., 8th Annual Water Distribution Systems Analysis Symp.*, ASCE, Reston, VA.
- Vítkovský, J. P., Liggett, J. A., Simpson, A. R., and Lambert, M. F. (2003). "Optimal measurement site locations for inverse transient analysis in pipe networks." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2003)129:6(480), 480–492.
- Vítkovský, J. P., Simpson, A. R., and Lambert, M. F. (2000). "Leak detection and calibration using transients and genetic algorithms." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2000)126:4(262), 262–265.
- Walski, T. M. (1983). "Technique for calibrating network models." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1983)109:4(360), 360–372.
- Walski, T. M. (1986). "Case study: Pipe network model calibration issues." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1986)112:2(238), 238–249.
- Yu, G. P., and Powell, R. S. (1994). "Optimal-design of meter placement in water distribution-systems." *Int. J. Syst. Sci.*, 25(12), 2155–2166.

This page has been intentionally left blank.



## Appendix 2 Final Published Version of Publication 2 (Chapter 5)

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

Nhu C. Do<sup>1</sup>, Angus R. Simpson<sup>1</sup>, Jochen W. Deuerlein<sup>2</sup>, Olivier Piller<sup>3</sup>

<sup>1</sup> - *School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide SA 5005, Australia.*

<sup>2</sup> - *Senior Researcher, 3S Consult GmbH, Karlsruhe, Germany.*

<sup>3</sup> - *Senior Researcher, Irstea UR ETBX, Dept. of Water, Cestas, France.*

*Journal of Water Resources Planning and Management 143(11)*

[doi:10.1061/\(ASCE\)WR.1943-5452.0000841](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000841)

This page has been intentionally left blank.

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

Nhu C. Do<sup>1</sup>; Angus R. Simpson, M.ASCE<sup>2</sup>; Jochen W. Deuerlein<sup>3</sup>; and Olivier Piller<sup>4</sup>

**Abstract:** Accurate modeling of water distribution systems is fundamental for planning and operating decisions in any water network. One important component that directly affects model accuracy is knowledge of nodal demands. Conventional models simulate flows and pressures of a water distribution network either assuming constant demand at nodes or using a short-term sample of demand data. Given the stochastic behavior of water demand, 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 over 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 that (1) predicts the hydraulic behaviors of the water network based on a nonlinear demand prediction model; and (2) corrects the prediction by integrating online observation data. The standard particle filter and an improved particle-filter method that 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. **DOI: 10.1061/(ASCE)WR.1943-5452.0000841.** © 2017 American Society of Civil Engineers.

**Author 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 demand and operational conditions), for example a day or a week (USEPA 2005). The outputs from hydraulic models, therefore, usually represent the distribution system's 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, use of these additional data is currently limited to computing gross differences between the model outputs and reality (Kang and Lansley 2009). Modification of the hydraulic models to maintain 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 pressures and flow rates at particular locations in the water network, with 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 Bhawe (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. Because of the slow convergence rate, this method is only applicable to small water networks. Later, explicit calibration methods were introduced (Ormsbee and Wood 1986; Boulos and Wood 1990; Boulos and Ormsbee 1991). These methods solved an even-determined set of water network equations where the number of unknown parameters was grouped to be equal to the number of measurements. Because the measurement errors were also neglected, these methods usually did not represent real-world practical outputs. Therefore, explicit calibration models

<sup>1</sup>Ph.D. Candidate, School of Civil, Environmental and Mining Engineering, Univ. of Adelaide, Adelaide, SA 5005, Australia (corresponding author). ORCID: <https://orcid.org/0000-0001-7809-7832>. E-mail: [nhu.do@adelaide.edu.au](mailto:nhu.do@adelaide.edu.au)

<sup>2</sup>Professor, School of Civil, Environmental and Mining Engineering, Univ. of Adelaide, Adelaide, SA 5005, Australia. E-mail: [angus.simpson@adelaide.edu.au](mailto:angus.simpson@adelaide.edu.au)

<sup>3</sup>Senior Researcher, 3S Consult GmbH, Albtalstraße 13, D-76137 Karlsruhe, Germany. E-mail: [deuerlein@3sconsult.de](mailto:deuerlein@3sconsult.de)

<sup>4</sup>Senior Research Scientist, Dept. of Waters, Irstea, UR ETBX, 50 Ave. De Verdun, Gazinet, F-33612 Cestas Cedex, Bordeaux, France. E-mail: [olivier.piller@irstea.fr](mailto:olivier.piller@irstea.fr)

Note. This manuscript was submitted on August 11, 2016; approved on May 17, 2017; published online on August 29, 2017. Discussion period open until January 29, 2018; separate discussions must be submitted for individual papers. This paper is part of the *Journal of Water Resources Planning and Management*, © ASCE, ISSN 0733-9496.

were often used to analyze historic events in water systems (Savic et al. 2009).

Subsequently, implicit methods were developed using either mathematical techniques or evolutionary optimization techniques, for example, the complex method (Ormsbee 1989), weighted least-squares approaches (Lansey and Basnet 1991; Datta and Sridharan 1994), singular-value decomposition (SVD) method (Sanz and Pérez 2015), or genetic algorithms (GA) (Preis et al. 2009; Abe and Peter 2010; Do et al. 2016). These methods have drawn a high degree of attention from researchers. However, these models are mostly impractical because of 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 semidefinite programming method to estimate pressure heads and their confidence bounds. In addition, some probabilistic methods (Xu and Goulter 1998; Kapelan et al. 2007; Hutton et al. 2013) have also been investigated for the estimation of model parameters. Given 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 by Hutton et al. (2012b), who divided uncertainty 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 analyzed, as well as to increase computational speed. It has been shown that skeletonized/aggregated network models can closely resemble the behavior of full-sized systems under steady-state conditions (e.g., Perelman et al. 2008; 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 or 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 demand. 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 has been considered in some recent studies. Short-term demand forecasting and demand estimation are two different problems; the former focuses on predicting future demand (e.g., Cutore et al. 2008; Hutton and Kapelan 2015; Alvisi and Franchini 2015). The latter focuses on estimation of the current demand, which is also the main interest of this paper. This is useful because demand estimation can be used at regular time steps to verify the accuracy of a 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 autoregressive integrated moving average (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 has been offered by 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, water demand patterns were also assumed to follow a linear time-series model, whereas the TSE model involved recursively computing the sensitivity matrix (i.e., 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. 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, given 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 demand in WDS studies is 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 unpredictable variation in water demand. Some efforts have been focused on 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 demand, 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, named the DMFLive model. A predictor-corrector methodology is adopted in the DMFLive model to (1) predict the hydraulic behaviors of the water network based on a nonlinear demand prediction submodel; and (2) 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: (1) systematic resampling (SR) algorithm; and (2) 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)

$$[x]_k = f_{k-1}([x]_{k-1}, v_{k-1}) \quad (1)$$

where  $k$  = index of discrete time steps;  $f_{k-1}$  = known, possibly nonlinear function of the previous state; and  $v$  = process noise sequence. The value of  $[x]_k$  can be found from measurements  $z_k$ , which are related to  $[x]_k$  via the following measurement equation:

$$z_k = h_k([x]_k, w_k) \quad (2)$$

where  $h$  = known implicit or explicit, possibly nonlinear function; and  $w$  = measurement noise sequence. The noise terms  $v_k$  and  $w_k$  are usually assumed to be white noise and independent.

From 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  $p(x_k|x_{k-1})$  = 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  $v_{k-1}$ ; and  $p(x_{k-1}|Z_{k-1})$  = 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  $p(z_k|x_k)$  = 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 Eqs. (3) and (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 infinite dimensional vector. Because the true solution is too complex and almost impossible to compute, an implementation of approximation techniques or suboptimal Bayesian algorithms has been 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

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

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

where  $\delta$  = Dirac delta function;  $i$  = particle index; and  $p(x_k^i|x_{k-1}^i, z_k)$  = 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,  $p(x_k^i|x_{k-1}^i, z_k) = p(x_k^i|x_{k-1}^i)$ , which yields with scaling

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

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. Because 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 with samples having high importance weights, is added to the procedure of particle-filter models. In this paper, the systematic resampling method, also called stochastic universal resampling, introduced by Kitagawa (1996), was selected for the resampling procedure of the particle-filter model. A comprehensive explanation of the systematic resampling and a full review of particle-filtering methods have been provided by van Leeuwen (2009). In addition, an improved resampling method that 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 paper, 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 submodel presented by van Zyl et al. (2008) has been applied to predict the water demand multipliers (DMF) in a WDS. The hydraulic EPANET toolkit, which solves the hydraulic equations, was used to compute the model equivalent of the measurement data (i.e., nodal pressures, flow rates at measurement locations, or 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

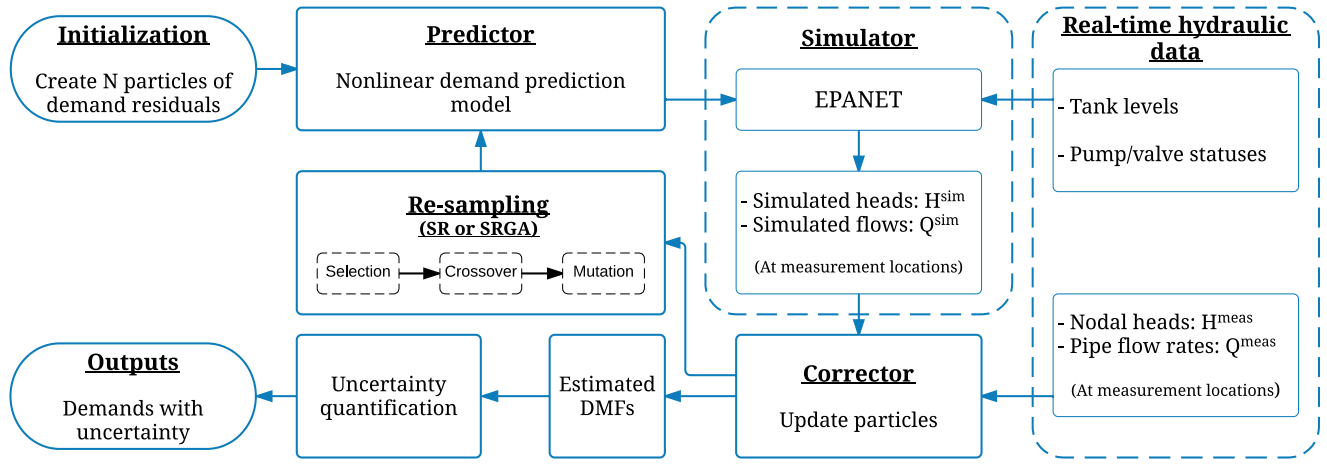


Fig. 1. Process of particle-filter model for real-time demand estimation in WDS

were computed using the first-order approximation formula. The flowchart of the DMFLive model is shown in Fig. 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 Submodel

The initial particles (for the first iteration) or the particles after re-sampling (from the second iteration onwards) are transferred to the demand prediction submodel. 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):

$$\ln x_k^j = \sum_{i=1}^m \phi_i \ln x_{k-i}^j + \ln v_k^j \quad (8)$$

where  $x_k^j$  = demand residual state at time-step  $k$  of the  $j$ th DMF;  $i$  = lag counter;  $m$  = number of autocorrelation lags [for the state estimation problem,  $m = 1$  as referred to Eq. (1)];  $\phi_i$  = autoregression coefficient for lag  $i$ ; and  $v_k$  ( $0, \sigma_h$ ) = white noise with mean zero and standard deviation  $\sigma_h$ .

The  $j$ th DMF is calculated

$$\text{DMF}_k^j = C_k^j x_k^j \quad (9)$$

where  $C_k^j$  = 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., as discussed by 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: (1) tank levels and pump and valve statuses; and (2) nodal heads and pipe flow rates at measurement locations. Tank levels and pump and valve statuses are used as boundary conditions for the hydraulic simulation of the water network model whereas 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 paper were assumed to have pressure measurements only. The input data sets to evaluate the DMFLive model were synthetically generated based on deterministic models, where the water network parameters are fully known: (1) known demand patterns were assigned to nodal demands; (2) EPANET was run to record tank levels and pump statuses and pressures at selected measurement locations; and (3) to introduce the measurement errors, a normal distributed random error in an allowable range ( $\pm \Delta^{\text{meas}}$ ) was added to each nodal pressure.

### Simulator

The hydraulic behavior of the water distribution network at each time step was simulated using an EPANET steady-state simulation. The inputs were 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., were assumed to be known and constant. The outputs from the EPANET hydraulic solver are the model equivalent of the observations, i.e., simulated nodal heads and pipe flow rates at measurement locations.

### Corrector

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

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

where  $h(x_k^i)$  = model equivalent of the observations  $z_k$  (simulated nodal heads and flow rates); and  $[R]$  = 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

$$\Delta^{\text{true}} = [Z]^{\text{true}} - h(x^{\text{true}}) \quad (11)$$

is the difference between the true observation vector,  $[Z]^{\text{true}}$ , and the corresponding vector output from the hydraulic simulation model *EPANET* using the true state  $x^{\text{true}}$ . 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 following expression:

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

The observation error covariance matrix, therefore, can be estimated as  $[R] = R^{\text{true}} + R^{\text{meas}}$ , where  $R^{\text{true}}$  and  $R^{\text{meas}}$  denote the covariance of the forward model error and covariance of measurement error, respectively [Waller (2013) has provided 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, 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 zero. The covariance matrix  $[R]$ , therefore, is the diagonal matrix where the diagonal elements are the variances of the measurement errors, because observations are independently measured at different locations of the network by different measurement devices. The measurement errors with specified ranges were assumed to be known so that the covariance matrix  $[R]$  could be identified.

### Resampling

Resampling was applied to create new ensembles of particles from the posterior PDF of the previous step. In this paper, two alternatives of resampling were tested: SR and SRGA.

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

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

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

$$x_{\text{new}}^i = x[F^{-1}(u^i)] \quad (14)$$

where  $F^{-1}$  = 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 was 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 with 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. Details of GA can have been given by Nicklow et al. (2010); here, it is important to know that new parameters need to be introduced: the probability of crossover  $P_c$ , probability of mutation  $P_m$ , and number of generations  $N_{\text{gen}}$ .

### Demand Multiplier Outputs and Uncertainty Quantification

The estimate of the state  $x_k$  was obtained by taking the mean of the particle-filter sample set (Salmond and Gordon 2005)

$$\hat{x}_k \approx \frac{1}{N_p} \sum_{i=1}^{N_p} x_k^{i*} \quad (15)$$

where  $x_k^{i*}$  = 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

$$\text{var}(x_k) \approx \frac{1}{N_p} \sum_{i=1}^{N_p} (x_k^{i*} - \hat{x}_k)(x_k^{i*} - \hat{x}_k)^T \quad (16)$$

For the demand multiplier estimation problem, 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 nonconvergence. 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

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

where  $[R^*] = R + \Sigma$ ,  $\Sigma$  = 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 nonconvergence and degeneracy, this would increase the noise of the output model. The variance of the model output (i.e., 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, which will be shown in the case studies. The model outputs, therefore, were 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., range in which the true demand multipliers are expected to be 95% of the time) can be obtained by the following expression:

$$\|\Delta DMF_k\| \leq 1.96([W]^{1/2}[J])^\dagger$$

$$|\Delta DMF_k^j| \leq 1.96 \sum_{j=1}^m |S_{ij}|, \text{ with } S = ([W]^{1/2}[J])^\dagger \quad (18)$$

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

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

The inputs required for the DMFLive model consist of the number of particles, the inputs for the demand prediction submodel, 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 submodel requires the data of typical demand patterns, autoregression coefficient ( $\phi_i$ ), and variance of noise of demand residuals ( $\sigma_h^2$ ). 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 and pump and valve statuses. The correction step requires the observation data at measurement sites. The particle-filter model associated with the GA process can only be applied to networks with multiple demand patterns (e.g., second case study in this paper). A 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 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 Fig. 2. The network has 9 nodes (8 nodes with demand), 12 pipes, 1 tank, and 1 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) were assumed to be placed at three random locations (Nodes 13, 22, and 31). All nodal demands were assumed to follow a single demand pattern that varies every 15 min [represented by the continuous line in Fig. 2(b)]. The demand pattern is a random daily demand pattern (from a yearly demand pattern) for 100 households obtained from the behavioural end-use stochastic simulator (BESS) model (Thyer et al. 2011). The DMFLive model was required to track this demand pattern using the three pressure measurements, which were also obtained every 15 min.

In this case study, the default demand pattern given in the Net1 example [represented by the dashed line in Fig. 2(b)] was selected as the typical demand pattern. Different values of the autoregression coefficient ( $\phi$ ) as well as variance of noise ( $\sigma_h^2$ ) were applied for the demand prediction submodel.

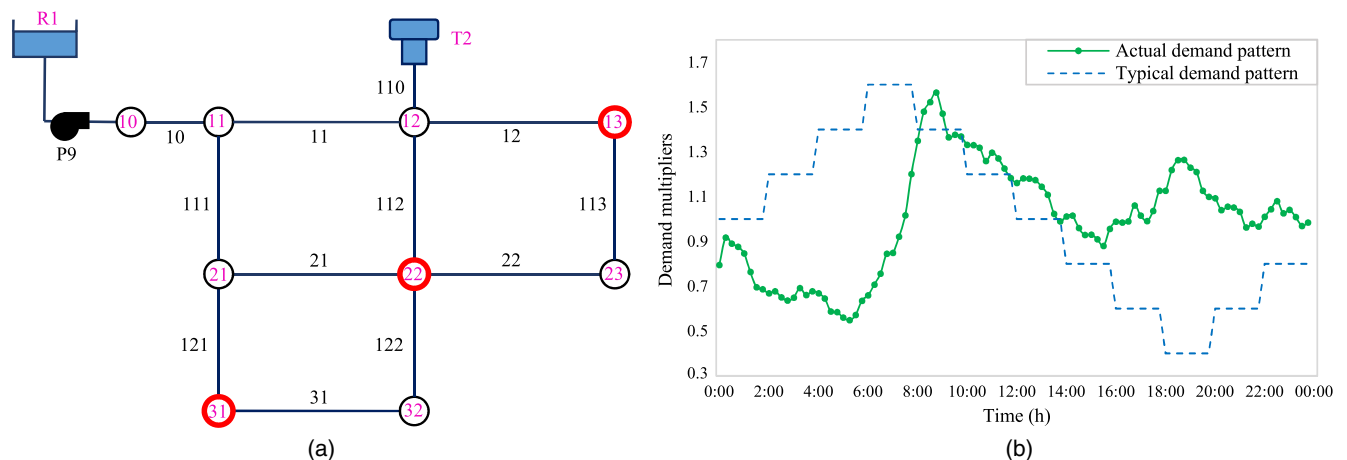


Fig. 2. (a) Case Study 1 network; (b) typical and actual demand patterns for Case Study 1 network



**Table 1.** Coefficient of Determination ( $R^2$ ) and RMSE of Demand Estimates Corresponding to Different Parameter Values of the Demand Prediction Model for Case Study 1

Number	Autoregression coefficient ( $\phi$ )	Variance of demand residual ( $\sigma_h^2$ )	$R^2$	RMSE
1	0.3	0.04	0.465	0.198
2		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	<b>0.988</b>	<b>0.028</b>
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

Note: Bold = best estimated result.

The accuracy of the demand estimates from the DMFLive model were evaluated in terms of the coefficient of determination ( $R^2$ ) and 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.

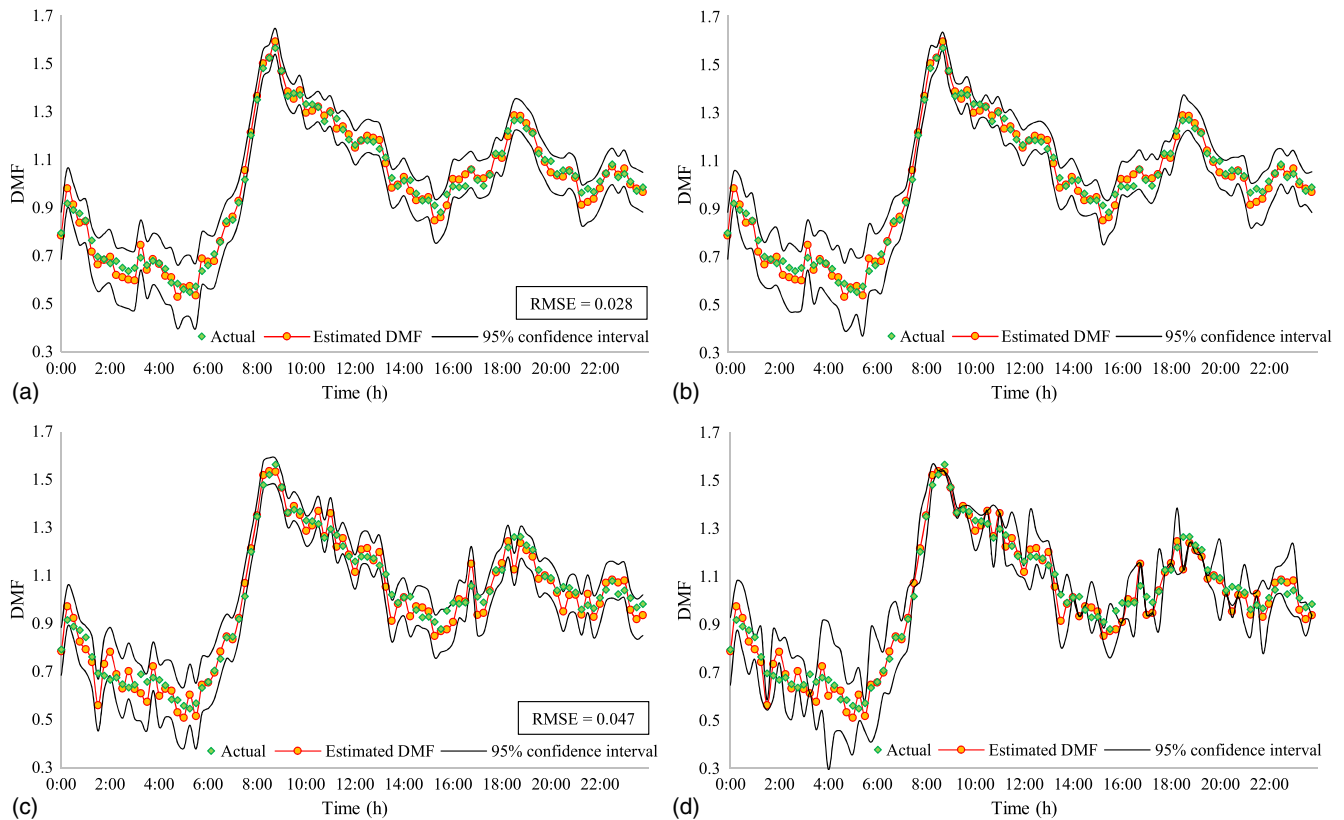
The DMFLive model performed very well when the autoregression coefficient was selected in the range of  $0.3 \leq \phi \leq 0.9$  and the noise variance was selected in the range of  $0.25 \leq \sigma_h^2 < 0.64$ . Due to the large difference between the typical demand value and

actual demand value at each time step [Fig. 2(b)], the selection of small values of the autoregression 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.

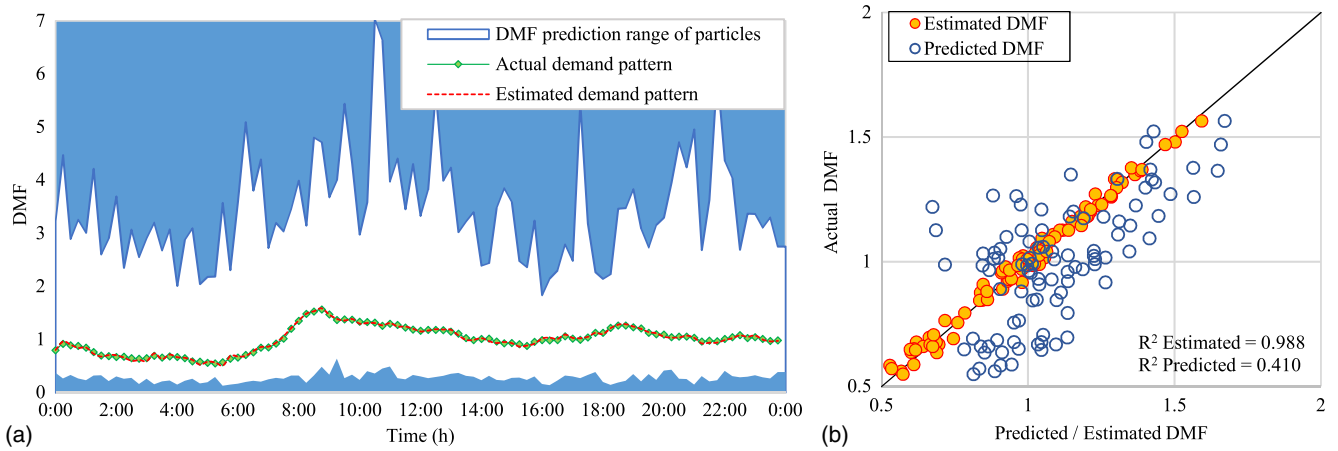
For this best estimated demand pattern, the confidence intervals and the scattergram between actual demand multipliers and estimated demand multipliers are plotted in Fig. 3(a).

In Fig. 3(a), the estimated demand pattern yields a very good match with the actual demand pattern during the time period (24 h, 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 was also 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 Figs. 3(b–d). Figs. 3(a–c) show the uncertainty quantified by the FOA method whereas Figs. 3(b and d) show 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 with the variance of the particle-filter samples.



**Fig. 3.** Estimated demand pattern and confidence intervals: (a and c) uncertainty quantification based on first-order approximation method [Eq. (18)] for  $N_p = 100$  and  $N_p = 20$ ; (b and d) posterior analysis for uncertainty quantification based on variance of the particle samples [Eq. (16)] for  $N_p = 100$  and  $N_p = 20$



**Fig. 4.** (a) Prediction range of demand multipliers during simulation period; (b) predicted demand multipliers and estimated demand multipliers

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 Figs. 3(c and d). The uncertainty boundary calculated by the FOA method in Fig. 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 Fig. 3(d) are collapsed into single value at some time steps because of 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 Fig. 4(a). The predicted values range from  $DMF_{min} = 0.1$  to  $DMF_{max} = 7.0$ , indicating that the demand prediction submodel can predict a large range of demand multipliers, and cover the range  $0 \leq DMF \leq 4$  suggested by Chin et al. (2000). Fig. 4(b) plots the scattergram of the actual demand multipliers versus the predicted demand multipliers (i.e., 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 the large difference between the typical demand pattern and actual demand pattern, the forecasting model does not provide a good prediction, resulting in weak and skewed correlation between the actual values and predicted values. Despite this, the DMFLive model is still capable of providing 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 in 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 was conducted. Instead of assuming that the observations are available at every 15 min, in this test it was assumed that the data could be obtained every hour, and the model is required to estimate the demand pattern at each hour time step (whereas the actual demand pattern is varied every 15 min).

Fig. 5 plots the two estimated demand patterns with and without tank-level information (herein referred to as DMF-WTLive and DMFLive); the DMF-WTLive model is the modified version of DMFLive model in 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 indicate that the DMFLive model obtained slightly better results than the DMF-WTLive model ( $RMSE = 0.046$  compared with  $RMSE = 0.080$ , respectively). However, the total estimated water usages in Table 2 indicate that the DMF-WTLive model is more accurate in predicting the volume of water delivered to users.

The total estimated water usage during the 24-h simulation period from the DMFLive model was  $5,942.43 \text{ m}^3/\text{day}$ , which was  $46.81 \text{ m}^3/\text{day}$  (or 0.78%) less than the actual water usage. On the other hand, total estimated water usage from DMF-WTLive model was  $6,007.31 \text{ m}^3/\text{day}$ , only  $18.07 \text{ m}^3/\text{day}$  (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 in 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) was selected as the second

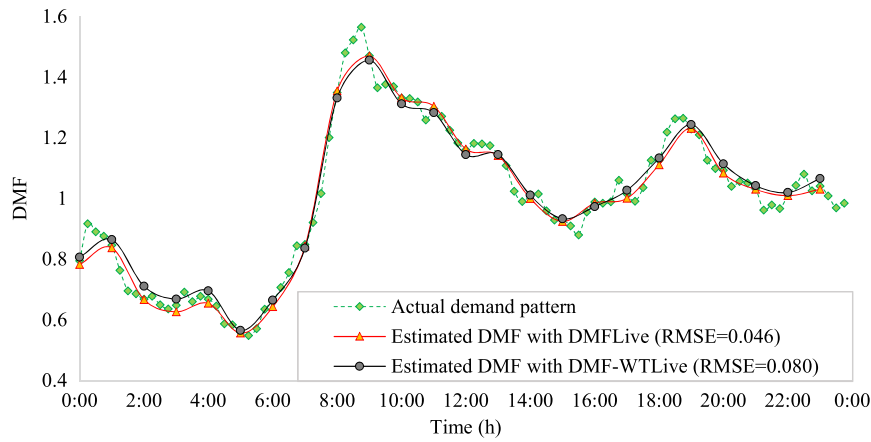


Fig. 5. Estimated demand patterns with and without tank level updated

**Table 2.** Actual and Estimated Total Volume of Water Usage during Calculated Period

Case	Total (m <sup>3</sup> /day)	Difference (m <sup>3</sup> /day)	Percentage difference
Actual daily water usage	5,989.25	—	—
Estimated water usage with DMFLive	5,942.43	46.81	0.78
Estimated water usage with DMF-WTLive	6,007.31	18.07	0.30

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 demand), which are divided into five district demand areas. Each district demand area follows a different hourly demand pattern. Because the data of the demand patterns are available for 7 days, the first 24 h of these demand patterns were assumed to be the typical demand patterns for the demand prediction submodel. The performance of the particle-filter model was evaluated by estimating the remaining 6-day hourly demand patterns.

It was assumed that there are 14 pressure measurement sites (from P1 to P14) that are randomly located at 14 places. These pressure measurements, again, were assumed to have a measurement error of  $\Delta^{\text{meas}} = \pm 0.2$  m. The inputs for the real-time demand estimation model were, therefore, the pressures at these locations, tank levels of seven tanks, and pump statuses of 11 pumps at each hour time step. The topology and measurement locations of the C-Town network are shown in Fig. 6. Five different demand prediction submodels were used to predict the five demand patterns. The parameters of the five demand prediction submodels were assumed to have the same values of  $\phi = 0.7$  for the autoregression 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 25 to 168 h) are shown in Fig. 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 RMSE = 0.061.

Fig. 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 [Figs. 7(a–c)] 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 [Fig. 7(d)] most of the time. However, because of the locations of the measurements (P7 and P9 in Fig. 6), the confidence interval of estimated DMF 4 pattern is relatively large compared with the others. The effect of measurement locations on the confidence intervals of the estimates will be discussed later in the paper. In Fig. 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.

Fig. 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, whereas 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 (6 days), because 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 resulting from particle collapse at some steps. Because increasing the number of demand patterns requires an exponentially increasing number of particles, it is necessary to improve the efficiency of the particle-filter model so that it can be applied to complex systems.

Two methods have been investigated, as mentioned previously:

- Incorporating the variance of the forecasting nodal heads into the likelihood function. The weights of particles in the model, referred to as the DMFLive-II model, are then calculated by the modified likelihood function [Eq. (17)]; and
- Integration of a GA process into the systematic resampling of the model, herein referred to as the DMFLive-III model.

Table 3 presents results (in terms of the RMSE of each demand pattern) of running these models with  $N_p = 1,000$  and  $N_p = 5,000$  for DMFLive-I and DMFLive-II, and with  $N_p^{\text{GA}} = 20$  and

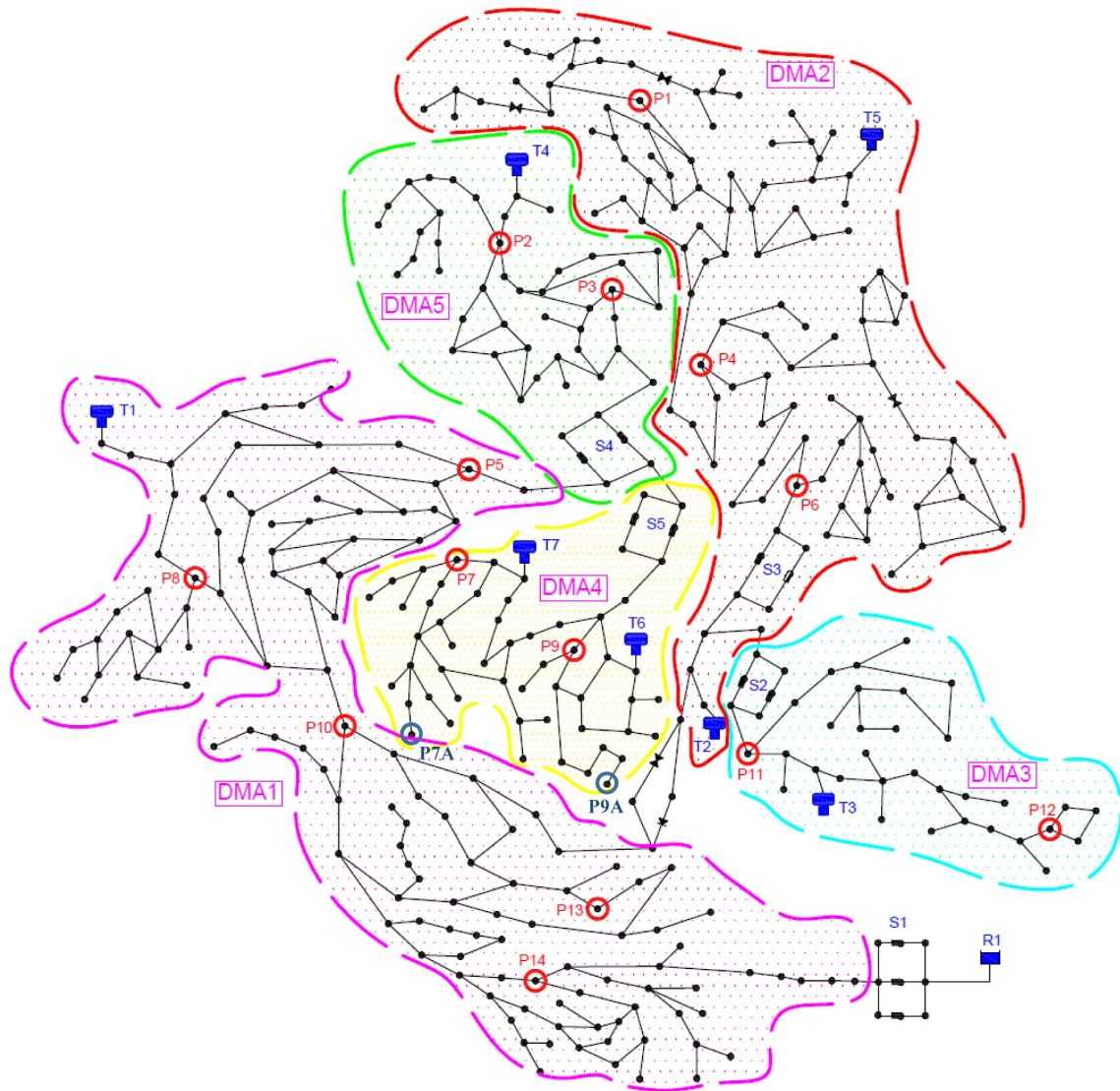


Fig. 6. Case Study 2 network (C-town network)

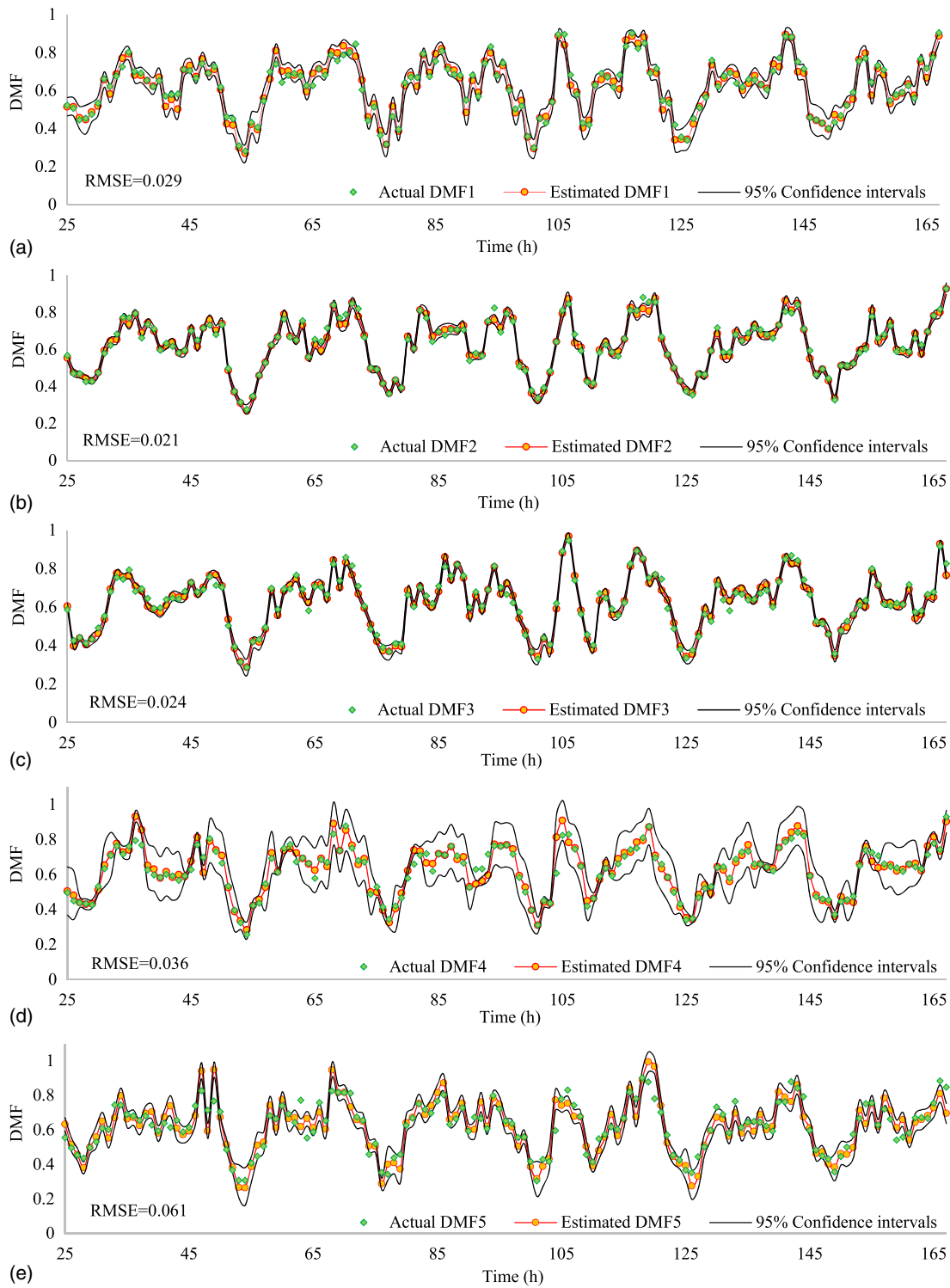
$N_p^{GA} = 100$  for DMFLive-III. It may be seen that for both  $N_p$  values, the DMFLive-I gives very poor estimates of the DMFs. On the other hand, the DMFLive-II model only requires  $N_p = 1,000$  (corresponding to  $1.43 \times 10^5$  evaluations for 143 h) to provide fairly good results, whereas the DMFLive-III performs well when  $N_p^{GA} = 100$ . The results of DMFLive-II ( $N_p = 5,000$ ) and DMFLive-III ( $N_p^{GA} = 100$ ) give similar results to those of DMFLive-I running at  $N_p = 25,000$  (corresponding to total evaluations of  $3.575 \times 10^6$ ). This means the computation can be reduced by approximately a factor of 5.

Fig. 9 shows the DMF 1 uncertainty ranges from 25 to 49 h of the three models, DMFLive-I, DMFLive-II, and DMFLive-III, computed by FOA method and by variance of the particles [Eq. (16)]. As can be seen from Figs. 9(a and c), because of particle impoverishment, the uncertainty computed by particle variance, represented by the dashed lines, is merged into a single line at almost all time steps. The uncertainty in Fig. 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 the FOA method, which are directly computed from

the sensitivity matrix and measurement errors, show consistent ranges in both cases. Given good estimates of the demand multipliers [as in Figs. 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**

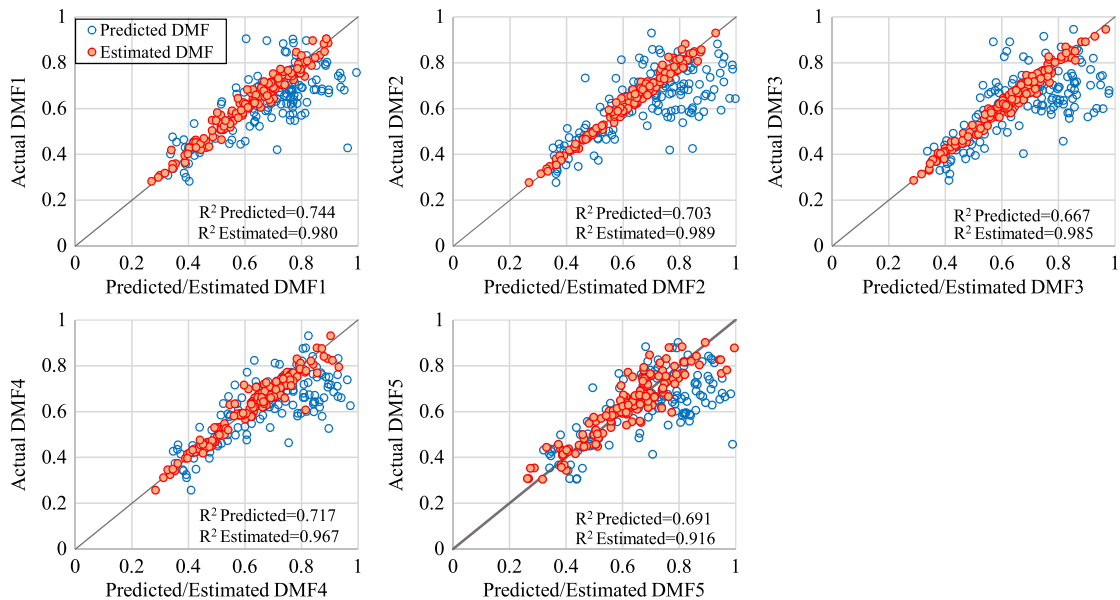
As discussed in a number of studies, such as those by Piller (1995) and Do et al. (2016), the locations of the measurements have a strong impact on the results of the demand estimation models. Furthermore, the selection of measurement locations also affects the confidence intervals of the estimation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



**Fig. 7.** Five estimated demand patterns for Case Study 2 network ( $N_p = 25,000$ ) using DMFLive-I

on two factors: (1) position of the nodes in the network; and (2) base demand 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 a change in nodal demand 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

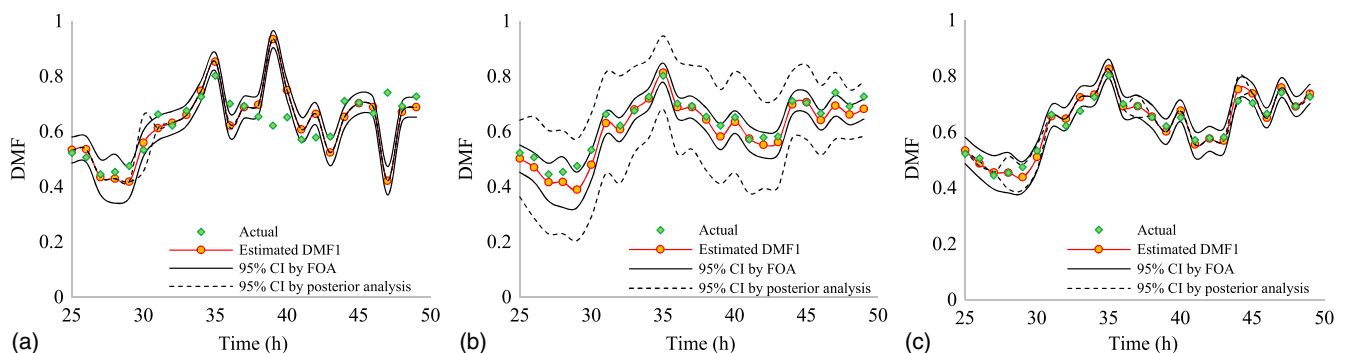
changes in pressures. Therefore, nodes selected in these regions may cause large uncertainty in demand multiplier estimation. The sensitivity matrix takes into account these two factors. Small values in the sensitivity matrix values mean that the nodes are less sensitive to the demands and the estimation might have large uncertainty. Therefore, the uncertainty of the estimated DMFs can be reduced by selecting the more sensitive locations in the network.



**Fig. 8.** Scattergrams and coefficients of determination for five estimated demand patterns in Case Study 2 network

**Table 3.** Performance of DMFLive Model with SR (DMFLive-I), Modified Likelihood Function (DMFLive-II), and SRGA (DMFLive-III)

Model type	Number of particles	Number evaluation	RMSE				
			DMF1	DMF2	DMF3	DMF4	DMF5
DMFLive-I	$N_p = 1,000$	$1.43 \times 10^5$	0.386	0.365	0.416	0.385	0.366
	$N_p = 5,000$	$7.15 \times 10^5$	0.405	0.422	0.237	0.229	0.246
DMFLive-II	$N_p = 1,000$	$1.43 \times 10^5$	0.05	0.026	0.029	0.043	0.074
	$N_p = 5,000$	$7.15 \times 10^5$	0.027	0.021	0.027	0.038	0.049
DMFLive-III ( $N_{Gen} = 50$ )	$N_p^{GA} = 20$	$1.08 \times 10^5$	0.107	0.067	0.068	0.086	0.19
	$N_p^{GA} = 100$	$5.43 \times 10^5$	0.03	0.025	0.023	0.032	0.05



**Fig. 9.** DMF 1 uncertainty ranges from 25 to 49 h computed by FOA method and posterior analysis: (a) DMFLive-I ( $N_p = 5,000$ ); (b) DMFLive-II ( $N_p = 5,000$ ); (c) DMFLive-III ( $N_{pGA} = 100$ )

An additional test was conducted 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

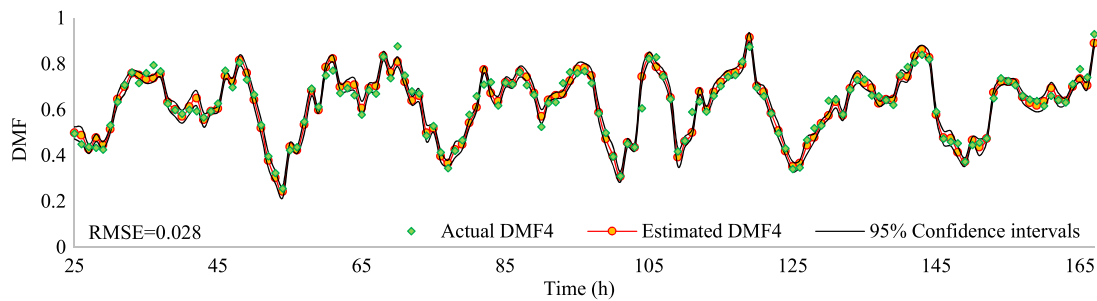
measurement locations were fixed at the same places as in the original test.

Fig. 10 shows the sensitivity matrixes  $[J]_0$  (for the original test) and  $[J]_0^A$  (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 a change in the DMF to which they

	DMF 1	DMF 2	DMF 3	DMF 4	DMF 5	Measurements		DMF 1	DMF 2	DMF 3	DMF 4	DMF 5	Measurements
$J_0 =$	0	22.37	0	0	0	P1	$J_0^A =$	0	22.37	0	0	0	P1
	0	0	0	0	3.72	P2		0	0	0	0	3.72	P2
	0	0	0	0	10.81	P3		0	0	0	0	10.81	P3
	0	18.80	0	0	0	P4		0	18.80	0	0	0	P4
	4.31	0	0	0	0	P5		4.31	0	0	0	0	P5
	0	18.80	0	0	0	P6		0	18.80	0	0	0	P6
	0	0	0	2.59	0	P7		0	0	0	5.31	0	P7A
	2.95	0	0	0	0	P8		2.95	0	0	0	0	P8
	0	0	0	2.55	0	P9		0	0	0	11.76	0	P9A
	3.20	0	0	0	0	P10		3.20	0	0	0	0	P10
	0	0	23.32	0	0	P11		0	0	23.32	0	0	P11
	0	0	22.39	0	0	P12		0	0	22.39	0	0	P12
	9.75	0	0	0	0	P13		9.75	0	0	0	0	P13
	4.77	0	0	0	0	P14		4.77	0	0	0	0	P14

**Fig. 10.** Sensitivity matrixes of nodal heads at measurement locations with respect to demand multipliers at  $t_0 = 0$



**Fig. 11.** Estimated Demand Pattern 4 and its confidence interval with the relocated Measurements 7 and 9A

belong. For example, the variation in the DMF 4 pattern only affects the sensitivity of the heads at measurement locations P7 and P9 (original test) and at measurement locations P7A and P9A (modified test). The nonzero values in the sensitivity matrices, therefore, correspond to the measurement locations. For the sensitivity of the heads, the new locations P7A ( $\partial H/\partial DMF4 = 5.31$ ) and P9A ( $\partial H/\partial DMF4 = 11.76$ ) are considerably more sensitive than the locations P7 ( $\partial H/\partial DMF4 = 2.59$ ) and P9 ( $\partial H/\partial DMF4 = 2.55$ ). As a result, the confidence intervals of the estimated DMF 4 for the modified test, as shown in Fig. 11, are much narrower than the confidence intervals of the estimated DMF 4 for the original test presented in Fig. 7(d). 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 nongeographically-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 by 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 with 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 because of 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 [Do et al. (2016) provide an example of optimal measurement location]. However, it has also been shown that the DMFLive model can be used to estimate 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 because of the unpredictable stochastic behavior of the water demand as well as the nonlinearities of hydraulic systems. This paper has introduced the DMFLive model framework, 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 was applied to predict water demand multipliers at each time step, and online pressure observations were 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 was 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 suitable for use 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. First-Order Approximation Method for Uncertainty Quantification of Water Demand Multipliers

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

$$y_i^{\text{Meas}} = y_i[x] + \varepsilon_i, \varepsilon_i \sim N(0, \sigma_i) \quad (19)$$

where  $[x] = n_d \times 1$  vector of parameters to calibrate (demand multiplier factors that depend on time);  $y_i[x]$  = scalar multivariate function of predictions from the network hydraulic model, given the parameter  $[x]$ ;  $\varepsilon_i$  = residual between model prediction and observation, which was assumed to be Gaussian with mean of zero and standard deviation of  $\sigma_i$ ; and  $y_i^{\text{Meas}}$  =  $i$ 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

$$f(x) = \frac{1}{2} \sum_{i=1}^m \left( \frac{y_i[x] - y_i^{\text{Meas}}}{\sigma_i} \right)^2 = \frac{1}{2} \sum_{i=1}^m \varepsilon_i^2 \quad \text{s.t. } x^{\min} \leq x \leq x^{\max} \quad (20)$$

where  $m$  = number of measurement sites;  $\varepsilon_R$  = 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

$$\nabla f_0 = [J](x^0)^T [W] [(y(x^0) - y^{\text{Meas}})] \quad (21)$$

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

The Hessian approximation takes the simple form of the symmetrical, positive semidefinite matrix

$$H_0 = [J](x^0)^T [W] [J](x^0) \quad (22)$$

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. (20) by a quadratic function at  $x^0$  leads to the approximation of  $x$ :

$$x = x^0 - (H_0)^{-1} \nabla f_0 \quad (23)$$

By replacing Eqs. (21) and (22) into Eq. (23), the approximation of  $x$  can be expressed

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

Using Eq. (19)

$$x(\varepsilon) = x^0 + [[J](x^0)^T [W] [J](x^0)]^{-1} [J](x^0)^T [W] \varepsilon \quad (24)$$

The influence of the measurement errors with regards to the parameter estimates, therefore, can be obtained at the first-order of Eq. (24):

$$\begin{aligned} \Delta x &= [[J](x^0)^T [W] [J](x^0)]^{-1} [J](x^0)^T [W] \varepsilon \\ &= [[W]^{1/2} [J](x^0)]^\dagger [W]^{1/2} \varepsilon = [[W]^{1/2} [J](x^0)]^\dagger \varepsilon_R \end{aligned} \quad (25)$$

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

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

$$\begin{aligned} \|\Delta x\| &\leq 2.58 \|([J](x^0)^T [W] [J](x^0)]^{-1} [J](x^0)^T [W]^{1/2}\| \\ &= 2.58 \|([W]^{1/2} [J](x^0)]^\dagger\| \end{aligned}$$

$$|\Delta x_i| \leq 2.58 \sum_{j=1}^m |S_{ij}|, \quad \text{with } S = ([W]^{1/2} [J])^\dagger \quad (26)$$

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

$$\|\Delta x\| \leq 1.96 \|([W]^{1/2} [J](x^0)]^\dagger\|$$

$$|\Delta x_i| \leq 1.96 \sum_{j=1}^m |S_{ij}|, \quad \text{with } S = ([W]^{1/2} [J])^\dagger \quad (27)$$

### References

- Abe, N., and Peter, B. (2010). "Epanet calibrator: An integrated computational tool to calibrate hydraulic models." *Proc., 10th Int. Conf. on Computing and Control for the Water Industry: Integrating Water Systems, CCWI 2009*, J. Boxall and C. Maksimovic, eds., CRC Press, Boca Raton, FL, 129–133.
- Alvisi, S., and Franchini, M. (2015). "Assessment of predictive uncertainty within the framework of water demand forecasting using the model conditional processor (MCP)." *Urban Water J.*, 14(1), 1–10.
- Bargiela, A., and Hainsworth, G. (1989). "Pressure and flow uncertainty in water systems." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1989)115:2(212), 212–229.
- Beal, C., and Stewart, R. (2014). "Identifying residential water end uses underpinning peak day and peak hour demand." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000357, 04014008.
- Bhave, P. R. (1988). "Calibrating water distribution network models." *J. Environ. Eng.*, 10.1061/(ASCE)0733-9372(1988)114:1(120), 120–136.
- Boulos, P. F., and Ormsbee, L. E. (1991). "Explicit network calibration for multiple loading conditions." *Civil Eng. Syst.*, 8(3), 153–160.
- Boulos, P. F., and Wood, D. J. (1990). "Explicit calculation of pipe-network parameters." *J. Hydraul. Eng.*, 10.1061/(ASCE)0733-9429(1990)116:11(1329), 1329–1344.
- Bush, C. A., and Uber, J. G. (1998). "Sampling design methods for water distribution model calibration." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1998)124:6(334), 334–344.
- Chin, D. A., Mazumdar, A., and Roy, P. K. (2000). *Water-resources engineering*, Vol. 12, Prentice Hall, Englewood Cliffs, NJ.



- Cutore, P., Campisano, A., Kapelan, Z., Modica, C., and Savic, D. (2008). "Probabilistic prediction of urban water consumption using the SCEM-UA algorithm." *Urban Water J.*, 5(2), 125–132.
- Datta, R., and Sridharan, K. (1994). "Parameter estimation in water-distribution systems by least squares." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1994)120:4(405), 405–422.
- Do, N. C., Simpson, A. R., Deuerlein, J. W., and Piller, O. (2016). "Calibration of water demand multipliers in water distribution systems using genetic algorithms." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000691, 04016044.
- Doucet, A., Godsill, S., and Andrieu, C. (2000). "On sequential Monte Carlo sampling methods for Bayesian filtering." *Stat. Comput.*, 10(3), 197–208.
- EPANET version 2 [Computer software]. USEPA, Cincinnati.
- Hol, J. D., Schon, T. B., and Gustafsson, F. (2006). "On resampling algorithms for particle filters." *Nonlinear Statistical Signal Processing Workshop*, IEEE, Piscataway, NJ, 79–82.
- Hutton, C. J., and Kapelan, Z. (2015). "A probabilistic methodology for quantifying, diagnosing and reducing model structural and predictive errors in short term water demand forecasting." *Environ. Modell. Software*, 66(2), 87–97.
- Hutton, C. J., Kapelan, Z., Vamvakieridou-Lyroudia, L., and Savic, D. A. (2012a). "Real-time demand estimation in water distribution systems under uncertainty." *WDSA 2012: 14th Water Distribution Systems Analysis Conf.*, Engineers Australia, Barton, Australia.
- Hutton, C. J., Kapelan, Z., Vamvakieridou-Lyroudia, L., and Savić, D. (2013). "Application of formal and informal Bayesian methods for water distribution hydraulic model calibration." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000412, 04014030.
- Hutton, C. J., Kapelan, Z., Vamvakieridou-Lyroudia, L., and Savić, D. A. (2012b). "Dealing with uncertainty in water distribution system models: A framework for real-time modeling and data assimilation." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000325, 169–183.
- Kang, D., and Lansley, K. (2009). "Real-time demand estimation and confidence limit analysis for water distribution systems." *J. Hydraul. Eng.*, 10.1061/(ASCE)HY.1943-7900.0000086, 825–837.
- 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.*, 10.1061/(ASCE)0733-9429(2007)133:8(927), 927–936.
- Kitagawa, G. (1996). "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models." *J. Comput. Graphical Stat.*, 5(1), 1–25.
- Lansley, K., El-Shorbagy, W., Ahmed, I., Araujo, J., and Haan, C. (2001). "Calibration assessment and data collection for water distribution networks." *J. Hydraul. Eng.*, 10.1061/(ASCE)0733-9429(2001)127:4(270), 270–279.
- Lansley, K. E., and Basnet, C. (1991). "Parameter estimation for water distribution networks." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1991)117:1(126), 126–144.
- Moradkhani, H., Hsu, K. L., Gupta, H., and Sorooshian, S. (2005). "Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter." *Water Resour. Res.*, 41(5), W05012.
- Nagar, A. K., and Powell, R. S. (2002). "LFT/SDP approach to the uncertainty analysis for state estimation of water distribution systems." *IEE Proc. Control Theory Appl.*, 149(2), 137–142.
- Nicklow, J., et al. (2010). "State of the art for genetic algorithms and beyond in water resources planning and management." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000053, 412–432.
- Ormsbee, L. E. (1989). "Implicit network calibration." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1989)115:2(243), 243–257.
- Ormsbee, L. E., and Wood, D. J. (1986). "Explicit pipe network calibration." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1986)112:2(166), 166–182.
- Ostfeld, A., et al. (2011). "Battle of the water calibration networks." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000191, 523–532.
- Perelman, L., Maslia, M. L., Ostfeld, A., and Sautner, J. B. (2008). "Using aggregation/skeletonization network models for water quality simulations in epidemiologic studies." *J. Am. Water Works Assoc.*, 100(6), 122–133.
- Piller, O. (1995). "Modeling the behavior of a network: Hydraulic analysis and sampling procedures for parameter estimation." Ph.D. thesis, Univ. of Bordeaux, Bordeaux, France.
- Preis, A., Whittle, A., and Ostfeld, A. (2009). "Online hydraulic state prediction for water distribution systems." *World Environmental and Water Resources Congress 2009*, Vol. 342, ASCE, Reston, VA, 323–345.
- Preis, A., Whittle, A. J., Ostfeld, A., and Perelman, L. (2010). "On-line hydraulic state estimation in urban water networks using reduced models." *10th Int. Conf. on Computing and Control for the Water Industry: Integrating Water Systems, CCWI 2009*, J. Boxall and C. Maksimovic, eds., CRC Press, Boca Raton, FL, 319–324.
- Preis, A., Whittle, A. J., Ostfeld, A., and Perelman, L. (2011). "Efficient hydraulic state estimation technique using reduced models of urban water networks." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000113, 343–351.
- Rahal, C. M., Sterling, M. J. H., and Coulbeck, B. (1980). "Parameter tuning for simulation models of water distribution networks." *Proc. Inst. Civil Eng.*, 69(3), 751–762.
- Ristic, B., Arulampalam, S., and Gordon, N. (2004). *Beyond the Kalman filter: Particle filters for tracking applications*, Artech House, Boston.
- Salmond, D., and Gordon, N. (2005). "An introduction to particle filters." *State space and unobserved component models theory and applications*, Cambridge University Press, Cambridge, U.K., 1–19.
- Sanz, G., and Pérez, R. (2014). "Comparison of demand pattern calibration in water distribution networks with geographic and non-geographic parameterization." *11th Int. Conf. on Hydroinformatics*, City Univ. of New York, New York, 1–8.
- Sanz, G., and Pérez, R. (2015). "Sensitivity analysis for sampling design and demand calibration in water distribution networks using the singular value decomposition." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000535, 04015020.
- Savic, D. A., Kapelan, Z. S., and Jonkergouw, P. M. (2009). "Quo vadis water distribution model calibration?" *Urban Water J.*, 6(1), 3–22.
- Shang, F., Uber, J. G., van Bloemen Waanders, B. G., Boccelli, D., and Janke, R. (2006). "Real time water demand estimation in water distribution system." *8th Annual Water Distribution Systems Analysis Symp. 2006*, ASCE, Reston, VA, 95.
- Thyer, M. A., Micevski, T., Kuczera, G. A., and Coombes, P. (2011). "A behavioural approach to stochastic end use modeling." *Ozwater 11, Australia's National Water Conf. and Exhibition*, Australian Water Association, NSW, Australia.
- USEPA. (2005). *Water distribution system analysis: Field studies, modeling and management. A reference guide for utilities*, Cincinnati.
- van Leeuwen, P. J. (2009). "Particle filtering in geophysical systems." *Mon. Weather Rev.*, 137(12), 4089–4114.
- van Leeuwen, P. J. (2010). "Nonlinear data assimilation in geosciences: An extremely efficient particle filter." *Q. J. R. Meteorolog. Soc.*, 136(653), 1991–1999.
- van Zyl, J. E., Piller, O., and le Gat, Y. (2008). "Sizing municipal storage tanks based on reliability criteria." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2008)134:6(548), 548–555.
- Waller, J. A. (2013). "Using observations at different spatial scales in data assimilation for environmental prediction." Ph.D. thesis, Dept. of Mathematics and Statistics, Univ. of Reading, Reading, U.K.
- Walski, T. M. (1983). "Technique for calibrating network models." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1983)109:4(360), 360–372.
- Weerts, A. H., and El Serafy, G. Y. (2006). "Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models." *Water Resour. Res.*, 42(9), W09403.
- Xu, C., and Goulter, I. C. (1998). "Probabilistic model for water distribution reliability." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1998)124:4(218), 218–228.

This page has been intentionally left blank.

## Appendix 3 Final Published Version of WDSA 2016 Paper

### Demand Estimation in Water Distribution Systems: Solving Underdetermined Problems Using Genetic Algorithms

Paper presented at the 18th Conference on Water Distribution System Analysis, WDSA 2016 in Cartagena, Colombia

Nhu C. Do<sup>1</sup>, Angus R. Simpson<sup>1</sup>, Jochen W. Deuerlein<sup>2</sup>, Olivier Piller<sup>3</sup>

<sup>1</sup> - *School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide SA 5005, Australia.*

<sup>2</sup> - *Senior Researcher, 3S Consult GmbH, Karlsruhe, Germany.*

<sup>3</sup> - *Senior Researcher, Irstea UR ETBX, Dept. of Water, Cestas, France.*

*Procedia Engineering, vol. 186, pp. 193-201, ISSN 1877-7058,*

<https://doi.org/10.1016/j.proeng.2017.03.227>.

This page has been intentionally left blank

XVIII International Conference on Water Distribution Systems Analysis, WDSA2016

## Demand Estimation In Water Distribution Systems: Solving Underdetermined Problems Using Genetic Algorithms

Nhu Do<sup>a\*</sup>, Angus Simpson<sup>b</sup>, Jochen Deuerlein<sup>c</sup>, Olivier Piller<sup>d</sup>

<sup>a</sup>PhD Candidate, The University of Adelaide, Adelaide SA 5000, Australia

<sup>b</sup>Professor, The University of Adelaide, Adelaide SA 5000, Australia

<sup>c</sup>Senior Researcher, 3S Consult GmbH, Karlsruhe, Germany

<sup>d</sup>Senior Researcher, Irstea UR ETBX, Dept. of Water, Cestas, France

---

### Abstract

Modeling of water distribution systems is fundamental for the design, analysis and operation of any water network. As with all hydraulic models, water demands are one of the most important input components in the model. However, estimation of the demand parameters is usually complicated due to the stochastic behavior of the water consumptions. Several methods have been proposed for estimating water demands. Most of them have been developed based on given frameworks where the number of unknown parameters is assumed to be equal or less than the number of measurements. The outcomes, therefore, rely on this assumption, which can lead to significant approximation errors in real water distribution systems.

The approach proposed in this paper does not require the number of known inputs to be equal to the number of variables. In fact, nodes in the model could each have a different demand pattern. The genetic algorithm approach adopted here shows that the average results of multiple GA runs can estimate the demand patterns at each node. Moreover, the model can also be used to estimate the flow rates and nodal heads at non-measured locations of the water network, although the accuracy of the estimation depends on number, type and location of the measurements. Results are shown and discussed for a literature case study tested for a 24-hour time period.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the organizing committee of the XVIII International Conference on Water Distribution Systems

---

\* Corresponding author. Tel: +61 414 503 468; Fax: +61 8-8313-4359  
Email: [nhu.do@adelaide.edu.au](mailto:nhu.do@adelaide.edu.au)

*Keywords:* Genetic algorithms; optimization; demand estimation; water distribution systems

## 1. Introduction

Water distribution system infrastructure has been constructed and developed for hundreds of years across the world. Together with the population growth and urbanization, water distribution systems (WDS) have expanded, and have become more complex and more difficult to operate. Modeling of WDS is, therefore, becoming increasingly important due to the need of understanding the behavior of these systems. Various simulation software solutions have been developed and broadly used for the design, analysis and operation of WDS, including EPANET, CWSNET or HydraulCAD. However, a new major challenge to deal with these models is the requirement for consistency between observed data of the real networks and simulated data from simulations models. Estimation of the model parameters, hence, is required so that the model is able to represent the real system.

In WDS, estimation is a process of fitting the outputs from the computer model, usually the pressures and flow rates at particular locations in the network, with the field measurements as well as calculating the parameters of interest [1]. Various model parameters, such as pipe diameters, roughness coefficients, valve resistances, or nodal demands, are required to be estimated. While most of these parameters are time invariant or vary slowly, nodal demands are the only parameter that can cause immediate changes in the model output. Estimation of model parameters during short periods of time using supervisory control and data acquisition (SCADA) systems, therefore, usually focuses on the demand parameter.

In the literature, the estimation of water demand has been studied by numerous researchers based on different methods, for example, extended Kalman filtering [2], tracking state estimation and Kalman filtering [1], Genetic Algorithms [3], and Particle Filtering [4]. However, these models have been developed based on given frameworks where the measurement locations were predetermined and the calibration parameters are grouped to be less than the number of measurements. The outcomes, therefore, rely on these additional assumptions, which can lead to large approximations in real water distribution systems. Only few papers have directly dealt with underdetermined systems such as a proportional demand method [5] and singular value decomposition (SVD) [6], [7].

This paper presents a study by [8] for the estimation of water demands in WDS over a period of 24 hours. A methodology is proposed to find the demand multiplier factors for an underdetermined system where the number of measurements is less than the number of demand parameter variables. The EPANET toolkit is used to solve the system of water network equations while Genetic Algorithms (GAs) are applied to find the best match between known measurements inputs and their estimated values. The mean values of multiple GA runs are suggested to be used as the best estimation of the flow rates and nodal heads as well as the estimation of the nodal demands in a system.

## 2. Methodology

The proposed model applies an implicit technique for the steady-state hydraulic simulation where the estimation process is formulated as an optimization problem. The objective function is a weighted least squares function in order to minimize the differences between simulated values from the model and their corresponding measured values.

The objective function at each time step is given by:

$$\text{Min}F = \sum_{i=1}^{N_H} w_i (H_i^{\text{Meas}} - H_i^{\text{Sim}})^2 + \sum_{j=1}^{N_Q} w_j (Q_j^{\text{Meas}} - Q_j^{\text{Sim}})^2 \quad (1)$$

where  $H_i^{\text{Sim}}$ ,  $Q_j^{\text{Sim}}$  are the simulated nodal head and flow rate for the  $i^{\text{th}}$  node and  $j^{\text{th}}$  pipe, respectively;  $H_i^{\text{Meas}}$ ,  $Q_j^{\text{Meas}}$  are the measured head and flow rate at the  $i^{\text{th}}$  node and  $j^{\text{th}}$  pipe (in this case these values are known exactly as they have been generated from a forward model run by a hydraulic simulator);  $N_H$ ,  $N_Q$  are the number of head and flow measurement sites in the network and  $w_i$ ,  $w_j$  are the weighting factors applied to different terms to ensure they have similar magnitude.

The decision variables for the optimization problem are the demand multiplier factors  $f_{k,t}$ , which are used to calculate the nodal demands at each time step:

$$D_{k,t} = D_{0,k} * f_{k,t} \tag{2}$$

where  $D_{0,k}$  is base demand at the  $k^{th}$  node, which is calculated using quarter/annual billing information for water usage; and  $f_{k,t}$  ( $k = 1 \dots N_{DM}$ ) is the demand multiplier factor at the  $k^{th}$  node at time step  $t$ .

A bounded range of demand factors may apply as:

$$f_k^{min} \leq f_k \leq f_k^{max} \tag{3}$$

where  $(f_k^{min}, f_k^{max})$  are the bounds of decision variables.

The genetic algorithm (GA) estimation model implemented for this research has been written in C-sharp language. The flowchart of the algorithm is shown in Figure 1. An initial population of chromosomes is randomly generated and decoded into corresponding demand multiplier factor values of each chromosome. For each node of the network exactly one of these DMFs is assigned and EPANET is subsequently called to simulate the steady state hydraulics of the system. Simulated flows and heads ( $Q^{Sim}, H^{Sim}$ ) at the measurement locations are obtained and compared with their actual measured values via the calculation of the objective function of Eq.(1). The inverse of the objective function (in the expression of  $1/(F+1)$  to avoid an indeterminate form when F equals to zero) is applied to define the fitness function for each member of the GA population. This is the measure for the quality of each member, and is used to determine its opportunity of survival.

By applying GA selection, crossover and mutation, new generations that inherit features of previous generation are created, and the estimation process is then repeated until the stopping criteria is met.

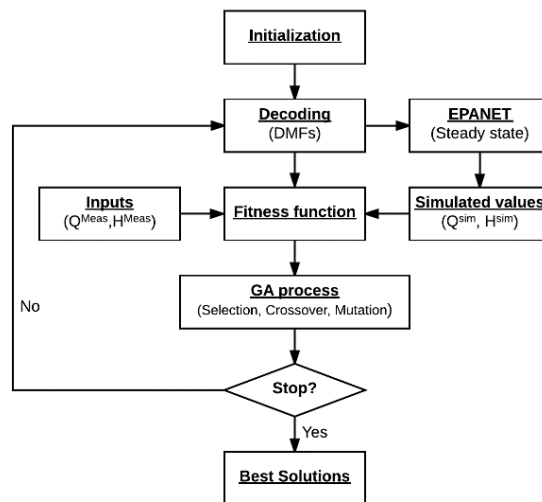


Fig. 1. Flowchart for GA estimation of demand multiplier factors

It should be noted that the proposed estimation model is to be applied to cases where the number of measurements is less than the number of unknown variables. In other words, the estimation of the demand in water distribution systems is mathematically a nonlinear underdetermined problem. A local solution of the problem can be found by local linearization methods such as QR decomposition, SVD or using the Moore-Penrose pseudoinverse matrix in Newton-Raphson method. However, due to the possibility of non-uniqueness of the solutions, the results from mathematical methods may be either far from the actual solution or result in negative demands at some nodes. A single run of the GA model, therefore, might converge to any of the non-unique solutions or be trapped in a local optimal solution where the simulated values cannot perfectly match the measured values at measurement locations. As a result, it appears that a good approximation of the demand multiplier factors estimation problem can only be obtained if multiple runs of the GA model are implemented and then averaged. The following section shows the results of the

multiple runs of the proposed GA model of a WDS that has been studied by the SVD model from [6] for the same estimation problem.

### 3. Case study

The case study used to evaluate the methodology is shown in Figure 2. The network has 9 nodes, 12 pipes, one tank, one pump and one reservoir. The network topology and all information such as the pipe parameters, length, roughness and pump characteristic can be found from EPANET example [9], namely the Net1 network.

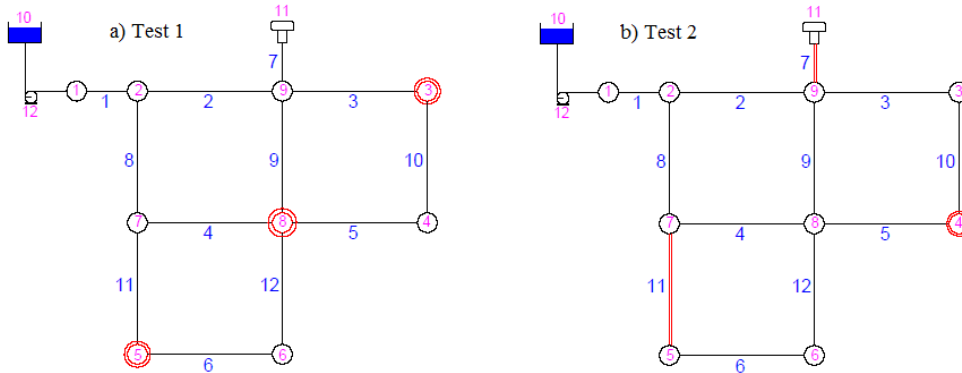


Fig. 2. Study network for the estimation problem.

The GA estimation model was tested for two scenarios of measurement sites according to [6]. In the first test, three pressure sensors are assumed to be located at nodes 3, 5 and 8. The second test assumes two flow meters are set at pipes 7 and 11, respectively and a pressure sensor is placed at node 4.

#### 3.1. Input for GA estimation model

In practice, input data for the estimation process are collected from a supervisory control and data acquisition (SCADA) system. In this research, input data is generated using EPANET toolkit as follows: (1) known demand multiplier factors are assigned to nodal demands; (2) run EPANET to retrieve the corresponding “true” or “known” pipe flow rates and nodal heads; (3) select the flows and heads at the selected locations as input for the GA model. The output flows and heads for selected pipes and nodes based on the simulation of the “true” values are used as the measured values ( $H^{Meas}$ ,  $Q^{Meas}$ ) for the estimation process.

#### 3.2. GA operators and model parameters

For the selection of GA operators, tournament selection was chosen because of its better convergence compared to proportionate selection or ranking selection; two-point crossover operator with the relatively high probability of  $P_c = 0.8$  and the bitwise mutation with the probability of  $P_m = 1/str \approx 0.13$  ( $str$  is the length of the string) are chosen, which are suggested by a study from [10].

The variation of the water demands at each node is presented by the range of the water demand multiplier factors. These values, after multiplying with their base demand, are expected to cover all the water usage throughout the day, including the lowest water use (possibly zero demand) and some extreme cases where the water demand is much larger than its average values. The range of decision variables (i.e. the demand multiplier factors), therefore, was selected to be from  $f_k^{min} = 0.0$  to  $f_k^{max} = 4.0$  which is selected based on typical values of the demand factors reported in [11].

One of the factors that may affect the accuracy of the GA estimation model is the increment ( $\Delta\theta$ ) of the DMFs. The selection of a large increment for the decision variables leads to faster convergence of the GA model although it may result in a coarser approximation of the estimated demands. Alternatively, the GA model may give better results if selecting smaller increment steps. However, the model requires more computational effort to converge due to the larger search space size. In this case study, the increment step of the decision variables was selected as  $\Delta\theta = 0.05$ , corresponding to a search space size of  $81^8 = 1.85 \cdot 10^{15}$  possible solutions for the GA model.

An integer-coding scheme was selected for the GA model. Each decision variable was coded by an integer number, ranging from 0 to 80 based on the choice table of the demand multiplier factors (corresponding to  $\Delta\theta = 0.05$ ). By



using this coding information, the chromosome in the GA process was decoded into a set of demand multiplier factors that are multiplied by the base demand and can be used for the hydraulic simulation process. An example of GA model process for the estimation problem is shown in Figure 3.

Chromosome 1	10	12	34	1	71	5	42	16
Chromosome 2	35	38	48	14	80	65	58	73
Mutation	10	12	48	14	80	49	42	16
	N	N	N	N	N	Y	N	N
	2	38	34	1	71	65	58	73
Y	N	N	N	N	N	N	N	N
Crossover	10	12	48	14	80	5	42	16
	35	38	34	1	71	65	58	73
Child 1	10	12	48	14	80	49	42	16
Decode	0.5	0.6	2.4	0.7	4	2.45	2.1	0.8
Child 2	2	38	34	1	71	65	58	73
Decode	0.1	1.9	1.7	0.05	3.55	3.25	2.9	3.65

Fig. 3. Example of a GA process for the estimation problem

The population size for GA model was selected as  $N=100$ . Finally, due to the non-uniqueness and the stochastic behavior of the problem, for each GA application, multiple runs with different seeds were implemented. The stopping criteria for each GA run is defined by the number of generations, which equals to  $N_{stop} = 1000$  generations.

#### 4. Results and discussions

##### 4.1. Comparison between two scenarios of measurement sites

A series of 100 runs (each initiated with a different random number seed) of the GA model was implemented for each scenario of measurement sites in a steady state simulation. The results are shown in Table 1 and Table 2.

Table 1. Average nodal demands and nodal pressures of the GA model for Test 1 and Test 2 (shown in Figure 2)

Node	Nodal demands (GPM)							Nodal pressure (psi)						
	Actual	Test 1 (100 runs averaged)	Test 2	$\Delta_{Test 1}$	$\Delta_{Test 2}$	%Error Test 1	%Error Test 2	Actual	Test 1 (100 runs averaged)	Test 2	$\Delta_{Test 1}$	$\Delta_{Test 2}$	%Error Test 1	%Error Test 2
2	150	260.83	146.18	110.83	3.82	73.89	2.55	119.26	119.66	119.55	0.40	0.29	0.34	0.24
3	100	114.05	135.6	14.05	35.60	14.05	35.60	118.67	<i>118.67</i>	118.08	<i>0.00</i>	0.59	<i>0.00</i>	0.50
4	150	116.93	110.03	33.07	39.97	22.05	26.65	120.74	123.81	<i>120.74</i>	3.07	<i>0.00</i>	2.54	<i>0.00</i>
5	100	83.9	91.35	16.10	8.65	16.10	8.65	115.86	<i>115.86</i>	116.51	<i>0.00</i>	0.65	<i>0.00</i>	0.56
6	100	88.15	112.65	11.85	12.65	11.85	12.65	110.79	113.43	109.39	2.64	1.40	2.38	1.26
7	150	237.3	135.23	87.30	14.77	58.20	9.85	117.66	119.39	118.3	1.73	0.64	1.47	0.54
8	200	130.9	175.1	69.10	24.90	34.55	12.45	118.76	<i>118.76</i>	118.83	<i>0.00</i>	0.07	<i>0.00</i>	0.06
9	150	241.7	162	91.70	12.00	61.13	8.00	117.02	120.06	117.02	3.04	0.00	2.60	0.00

\*italic - measured locations,  $\Delta$  - absolute differences between actual values and estimated values

Table 1 presents the average values of nodal demands and nodal pressures from 100 runs of the GA model. For the estimation of the nodal demands, Test 2 (2 pipe flows and 1 nodal pressure are measured) provides better results than Test 1 (3 nodal pressures are measured). The maximum error was found at node 3 for Test 2, of 35.6%, while for Test 1 the maximum error occurred at node 2 with the corresponding percentage of 73.88%. Test 2 also shows the best estimate at node 2 with an estimated nodal demand of 146.18 GPM (the actual value is 150 GPM, corresponding with only 2.55% error). On the other hand, the minimum estimate error of Test 1 was found at node 6 with a proportion of 11.85%. For the estimation of the nodal pressures, the GA model achieved very good results, with the maximum error less than 2.6% for all nodes in the network.

Table 2. Average pipe flow rates of the GA model for Test 1 and Test 2 (shown in Figure 2)

Pipe	Flow rates (GPM)						
	Actual	Test 1 (100 runs averaged)	Test 2 (100 runs averaged)	$\Delta_{Test 1}$	$\Delta_{Test 2}$	% error Test 1	% error Test 2
1	1866.18	1863.60	1864.32	2.58	1.86	0.14	0.10
2	1234.21	1108.63	1244.62	125.58	10.41	10.17	0.84
3	129.34	128.40	153.39	0.94	24.05	0.73	18.59
4	191.16	103.23	202.67	87.93	11.51	46.00	6.02
5	120.66	102.58	128.36	18.08	7.70	14.98	6.38
6	40.81	29.71	55.47	11.10	14.66	27.20	35.92

7	766.18	509.85	<i>766.19</i>	256.33	<b>0.01</b>	33.46	<b>0.00</b>
8	481.97	454.14	473.52	27.83	8.45	5.77	1.75
9	188.70	188.69	166.30	0.01	22.40	0.01	11.87
10	29.34	14.35	51.23	14.99	21.89	51.11	74.61
11	140.81	113.61	<i>140.80</i>	27.20	<b>0.01</b>	19.32	<b>0.00</b>
12	59.19	58.44	63.44	0.75	4.25	1.26	7.19

*\*italic - measured locations, Δ - absolute differences between actual values and estimated values*

Table 2 presents the average values of pipe flow rates from 100 runs of the GA model. It can be seen from the table that Test 2 again shows better results than for Test 1. In Test 2, the estimate of the flows is relatively accurate for the pipe with average flows (>50GPM) and high flows (>400 GPM). Large estimate errors only occur at pipes with low flows (<50GPM) such as pipe 10 (74.61%) and pipe 6 (35.92%). For Test 1, relatively large estimate errors can be found in all ranges of flow magnitude, for example, pipe 7 (the actual value is 766.18 GPM, the estimate value is 509.85 GPM, corresponding to an error of 33.46%), pipe 4 (the actual value is 191.16 GPM, the estimate value is 103.23 GPM, estimated error of 46%) or pipe 10 (the actual value is 29.34 GPM, the estimate value is 14.35 GPM, estimated error of 51.23%). This result highlights the important role of the locations and types of the measurements in the demand estimation problem.

4.2. Results of GA model for extended period simulation

The GA model is tested in a 24-hour extended period simulation for Test 2 where the measurements are placed at pipes 7, 11 (flow measurements) and node 4 (a pressure measurement), respectively. In order to evaluate the effects of the number of GA runs on the model output, the results of an individual 20 GA runs and 100 GA runs are examined. The estimated flow rates at selected pipes and estimated demands at selected nodes are plotted in Figure 4 and Figure 6. By increasing the number of GA runs from 20 to 100, the model provides slightly better estimates of flows (e.g. flow at pipe 5 – Figure 6), while giving very good approximations of the water demands at all nodes.

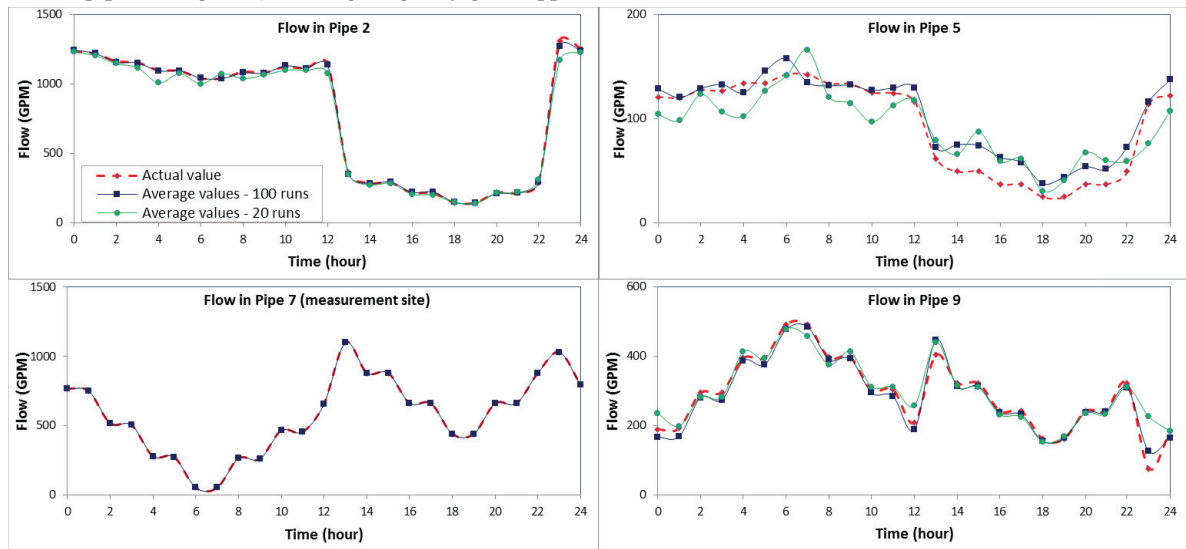


Figure 5 displays the scattergrams and correlation coefficients ( $R^2$ ) of the average estimated flows of 100 GA runs versus actual flows at all pipes of the network. Similarly, Figure 7 plots the scattergrams and correlation coefficients of the average estimated demands of 100 GA runs versus actual demands at all nodes. It can be seen that the correlations are strong and for both flow estimates and demand estimates, except the flow in pipe 10 ( $R = 0.715$ ) and the demand at node 7 ( $R=0.677$ ). The strong correlation values indicate that the estimation is consistently good over time for the entire network.

Fig. 4. Estimation of flow rates at pipe 2, 5, 7 and 9.

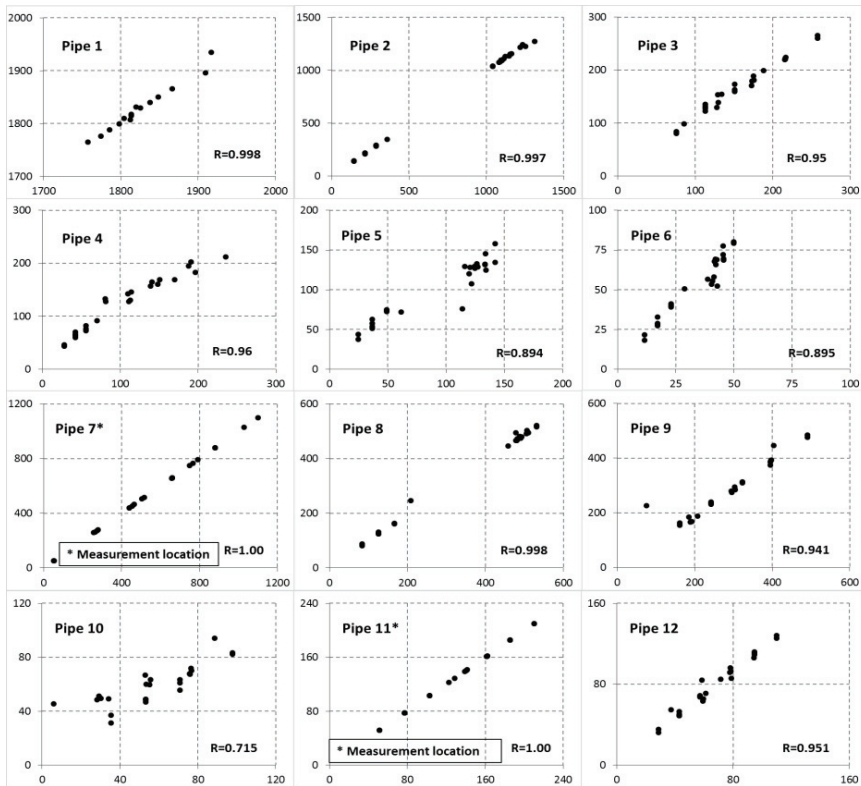


Fig. 5. Scattergrams and correlation coefficients between flow estimates (average of 100 GA runs) and actual flows (units for both axes are GPM)

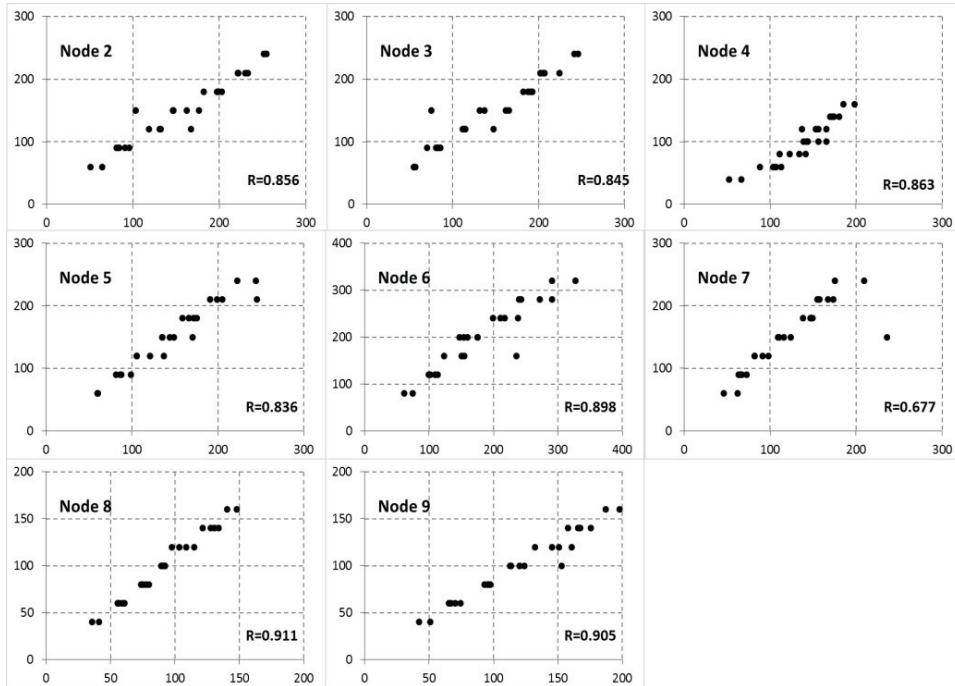
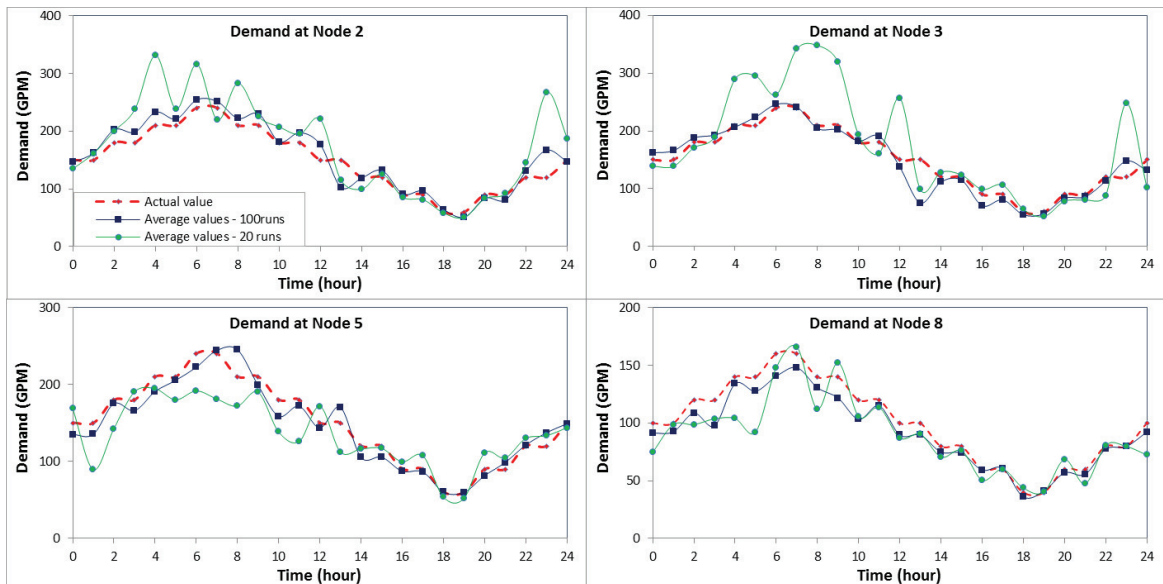


Fig. 6. Estimation of demands at node 2, 3, 5 and 8.

Fig. 7. Scattergrams and correlation coefficients between demand estimates (average of 100 GA runs) and actual demands (units for both axes are GPM)

5. Conclusions

Estimation of water demand in water distribution systems is problematical due to the limited number of



measurement sites. In this paper, a GA model based on integer coding has been introduced for the estimation of the water demand multipliers for underdetermined problems in water distribution systems where the number of measurement sites is less than the number of unknown parameters. The results provided from the case study show that

multiple runs of the GA model can produce relatively good approximation of the state in a water network. Future research efforts will involve finding advanced methods for the estimation of the demand to reduce the computational time. In addition, uncertainty of the model is another consideration given the presence of errors in measurement data. Finally, addressing the problem of leakage in the network is also important in achieving reliable results.

## References

- [1] Kang, D. and Lansey, K. , Real-time demand estimation and confidence limit analysis for water distribution systems. *Journal of Hydraulic Engineering*, 2009. 135(10): p. 825-837.
- [2] Shang, F., et al. Real time water demand estimation in water distribution system. in 8th Annual Water Distribution Systems Analysis Symposium 2006, August 27, 2006 - August 30, 2006. 2006. Cincinnati, OH, United States: American Society of Civil Engineers.
- [3] Preis, M. Allen, and A.J. Whittle. On-line hydraulic modeling of a water distribution system in Singapore. 2012. Tucson, AZ.
- [4] Hutton, C.J., et al., Real-time demand estimation in water distribution systems under uncertainty in WDSA 2012: 14th Water Distribution Systems Analysis Conference. 2012.
- [5] Davidson, J. and F.-C. Bouchart, Adjusting nodal demands in SCADA constrained real-time water distribution network models. *Journal of Hydraulic Engineering*, 2006. 132(1): p. 102-110.
- [6] Cheng, W. and Z. He, Calibration of nodal demand in water distribution systems. *Journal of Water Resources Planning and Management*, 2010. 137(1): p. 31-40.
- [7] Kun, D., et al., Inversion Model of Water Distribution Systems for Nodal Demand Calibration. *Journal of Water Resources Planning and Management*, 2015.
- [8] Do, N.C., et al., Calibration of water demand multipliers in water distribution systems using Genetic Algorithms. *Journal of Water Resources Planning & Management* (Accepted on April 2016), 2016.
- [9] Rossman, L.A., *EPANET 2: Users Manual*. 2000, US Environmental Protection Agency (EPA): USA.
- [10] Goldberg, D.E. and K. Deb, A comparative analysis of selection schemes used in genetic algorithms. *Foundations of Genetic Algorithms*, 1991. 1: p. 69-93.
- [11] Chin, D.A., A. Mazumdar, and P.K. Roy, *Water-resources engineering*. Vol. 12. 2000: Prentice Hall Englewood Cliffs.

This page has been intentionally left blank.

## Appendix 4 Final Published Version of CCWI 2017 Paper

Online Demand Estimation of Geographical and Non-Geographical Distributed Demand Pattern in Water Distribution Networks

Paper presented at the 15th Conference on Computing and Control for Water Industry, CCWI 2017 in Sheffield, UK

Nhu C. Do<sup>1</sup>, Angus R. Simpson<sup>1</sup>, Jochen W. Deuerlein<sup>2</sup>, Olivier Piller<sup>3</sup>

<sup>1</sup> - *School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide SA 5005, Australia.*

<sup>2</sup> - *Senior Researcher, 3S Consult GmbH, Karlsruhe, Germany.*

<sup>3</sup> - *Senior Researcher, Irstea UR ETBX, Dept. of Water, Cestas, France.*

<https://doi.org/10.15131/shef.data.5364133.v1>

This page has been intentionally left blank



## Online Demand Estimation of Geographical and Non-Geographical Distributed Demand Pattern in Water Distribution Networks

Nhu Cuong Do<sup>1</sup>, Angus R. Simpson<sup>2</sup>, Jochen W. Deurelein<sup>3</sup>, Olivier Piller<sup>4</sup>

<sup>1,2</sup> School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide SA 5005, Australia.

<sup>3</sup> 3S Consult GmbH, Karlsruhe, Germany

<sup>4</sup> Irstea, UR ETBX, Department of Waters, France, Bordeaux

[nhu.do@adelaide.edu.au](mailto:nhu.do@adelaide.edu.au)

### ABSTRACT

*The issue of demand calibration and estimation under uncertainty is known to be an exceptionally difficult problem in water distribution system modelling. In the context of real-time event modelling, the stochastic behaviour of the water demands and non-geographical distribution of the demand patterns makes it even more complicated.*

*This paper considers a predictor – corrector approach, implemented by a particle filter model, for solving the problem of demand multiplier factor estimation. A demand forecasting model is used to predict the water demand multiplier factors. The EPANET hydraulic solver is applied to simulate the hydraulic behaviour of a water network. Real time observations are integrated via a formulation of the particle filter model to correct the demand predictions.*

*A water distribution network of realistic size with two configurations of demand patterns (geographically distributed demand patterns and non-geographically distributed demand patterns) are used to evaluate the particle filter model. Results show that the model is able to provide good estimation of the demand multiplier factors in a near real-time context if the measurement errors are small. Large measurement errors may result in inaccurate estimates of the demand values.*

**Keywords:** Particle filtering, real time demand estimation, water distribution systems, calibration.

## 1 INTRODUCTION

Water distribution systems (WDSs) 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. Conventional demand driven models simulate flows and pressures of a WDS requiring assumptions of known demands. Sensor technology that has recently been applied in WDSs can assist in providing localised flow rates and pressures, which also enables new approaches to estimate the water consumptions within the networks in real-time or near real-time, for example [1], [2] and [3]. These water demand estimates can be used for developing a better understanding of the full range of operational states (e.g. [4]) as well as detecting abnormal events (e.g. [5]).

Water demand estimation is the process of fitting the outputs from a computer model (i.e. the pressures and flow rates at particular locations in the network) with the field measurements. Given a large number of water consumers (a.k.a. nodal demands) and a limited number of measurements in a real network, it would be infeasible for any model to estimate all of the unknown nodal demands. Instead, the demand patterns, which represent groups of similar water consumers, are usually considered to be estimated. Different criteria can be used to categorize the water consumers into groups. For better management of leakage, demands are grouped into pressure zones or into district

metered areas (DMAs), which usually results in geographically distributed demand patterns. In order to capture the changing habits of different water users, demands are grouped based on types of customers, e.g. domestic, residential, restaurants, hospitals, parklands or industrial, etc., which result in non-geographically distributed demand patterns. These categorization techniques may introduce errors to the demand estimation results. Other sources of uncertainties such as measurement inaccuracy, parameter uncertainty or unaccounted hydraulic events (e.g. leakage, transient...) also may contribute to the errors of the estimated demand values. These issues need to be examined to assess whether or not integrating the sensor data to estimate water demand in near real-time can improve the accuracy of the hydraulic models.

The research work presented here evaluates the outputs from the estimation model developed by [2]. New developments focus on the estimation of the demand multipliers of a geographically distributed demand pattern network as well as a non-geographically distributed demand pattern network, given different scenarios of measurement errors.

## 2 DEMAND ESTIMATION MODEL

Figure 1 shows the process for the estimation of water demand multipliers (DMFs) proposed by [2]. The model applies a predictor – corrector approach, which is implemented by a sequential Monte Carlo sampling technique, also known as the particle filter. The DMFs are estimated through three main steps: prediction, simulation and correction within a particle filter setting.

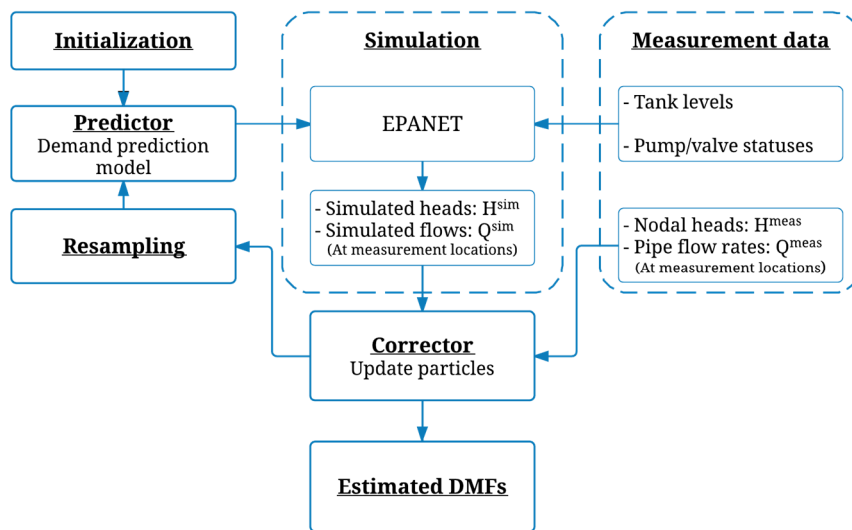


Figure 1: Particle filter model for near real-time state estimation in WDS

### 2.1 Predictor step

The model starts with a creation of an ensemble of the particles ( $N_p$ ), at which each particle is assigned an initial weight equal to  $1/N_p$ . The particles are the demand residuals, which are computed from the demand residuals of previous steps [6]:

$$\ln x_k^j = \sum_{i=1}^m \phi_i^j \ln x_{k-i}^j + \ln v_k^j \quad (1)$$

where  $x_k^j$  is the water demand residual at time step  $k$  of the  $j^{\text{th}}$  DMF,  $i$  is the lag counter,  $m$  is the number of autocorrelation lags.  $\phi_i$  is the auto-regression coefficient for lag  $i$  and  $v_k(0, \sigma_h)$  is white noise with mean zero and standard deviation  $\sigma_h$ .

Predictions of the demand multipliers based on these demand residuals, therefore, can be calculated via the following equation:

$$DMF_k^j = C_k^j x_k^j \quad (2)$$

where  $C_k^j$  is the demand multiplier value of  $x_k^j$  at time  $k$  of a typical diurnal demand pattern (or a default demand pattern) of the  $j^{\text{th}}$  DMF. The  $C$  value can be identified based on meter information of different water users (e.g. in [7]).

## 2.2 Simulation

The EPANET hydraulic solver [8] is used to simulate the hydraulic behaviour of the water distribution network at each time step. The inputs are the predicted DMF from the prediction phase and real-time hydraulic data from sensor devices (i.e. nodal heads and flow rates), which may also include: 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 solver is the model equivalent of the field observations, i.e. the simulated nodal heads and pipe flow rates at measurement locations.

## 2.3 Corrector step

The weights of the particles are corrected/updated by associating the simulated heads and flows with the actual observations via Equation (3) where the conditional probability of the observations is assumed to be Gaussian:

$$p(z_k | x_k^i) = \frac{1}{\sqrt{2\pi|R|}} e^{\left(-\frac{1}{2}[z_k - z(x_k^i)]^T R^{-1} [z_k - z(x_k^i)]\right)} \quad (3)$$

where  $z(x_k^i)$  is the simulation model equivalent of the observations  $z_k$  (nodal heads and flow rates), and  $R$  is the error covariance of the observations. The importance weight of each particle is then computed by:

$$w_k^i = \frac{p(z_k | x_k^i)}{\sum_{i=1}^{N_p} p(z_k | x_k^i)} \quad (4)$$

New ensembles of particles for subsequent time steps are created through a resampling process, which replaces samples with low importance weights by the samples with high importance weights. In this work, the systematic resampling algorithm is applied. The algorithm generates a random number  $u_s$  from the uniform density  $U[0, 1/N_p]$ , and consequently creates  $N_p$  ordered numbers [9]:

$$u^i = \frac{i-1}{N_p} + u_s \quad (i = 1, \dots, N_p) \quad (5)$$

New particles that satisfy Equation (6) are then selected:

$$x_{new}^i = x(F^{-1}(u^i)) \quad (6)$$

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

By recursively implementing the predictor – corrector approach, the demand multiplier of each group of demands can be estimated. Note that at each time step, the estimates of the demand multipliers are obtained by taking the mean of the particle filter sample set:

$$\hat{x}_k \approx \frac{1}{N_p} \sum_{i=1}^{N_p} x_k^{i*} \tag{7}$$

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

### 3 CASE STUDY

The case study used to evaluate the model is shown in Figure 2, which is the modification from an example of the EPANET software, namely the Net3 network. The network has 2 reservoirs, 3 tanks, 92 nodes, 117 pipes and 2 pumps. The demands in the network are classified into four different demand patterns. Two configurations of the demand patterns are considered in this study. In Figure 2.a, the demands are divided based on the topographic information, which results in a geographically distributed demand network. In Figure 2.b, the demands are categorized, as an example, based on the magnitudes of the base demands: DMF1 for nodes with base demands less than 10 L/s, DMF2 for nodes with base demands from 10 L/s to 20 L/s, DMF3 for nodes with base demands from 20 L/s to 30 L/s and DMF4 for nodes with base demands larger than 30 L/s. In this case, the network has non-geographically distributed demand patterns.

It is assumed that there are 12 pressure measurement sites randomly located within the network. The inputs for the near real-time demand estimation model are, therefore, the pressures at these locations, the tank levels of the three tanks and the pump statuses at each hour time step.

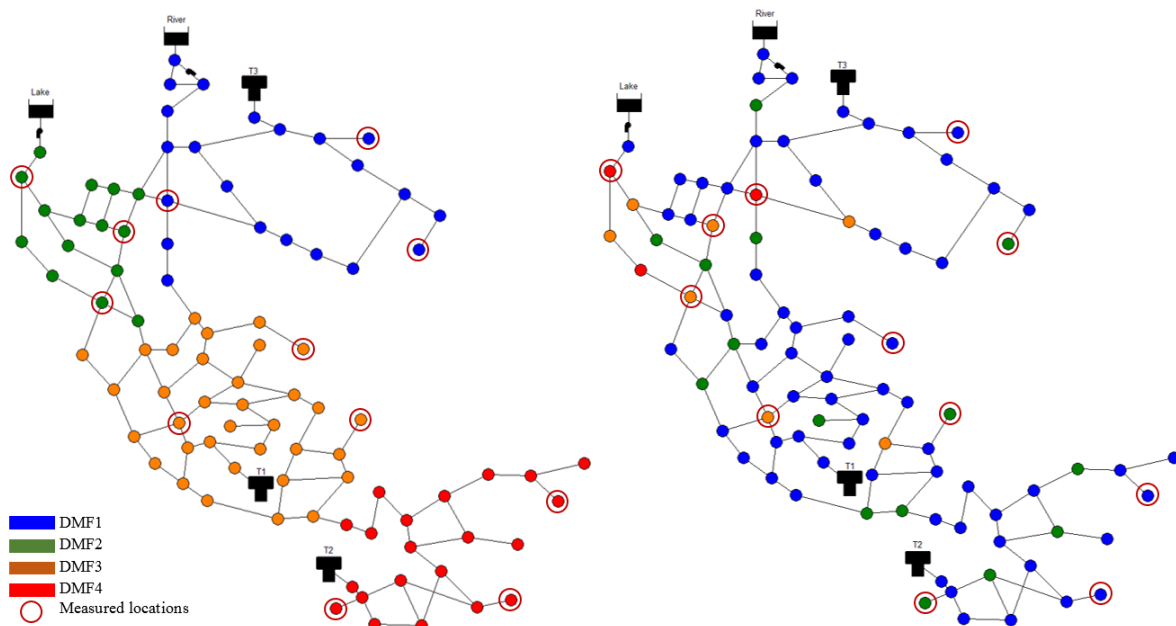


Figure 2. Case study network - (a) Geographically distributed demand patterns, (b) Non-geographically distributed demand patterns

It is also assumed that the default patterns, which are calibrated based on historical water used data, of four demand groups are known. In order to evaluate the estimation results of the PF model, measurement data sets are synthetically generated for a period of 48 hours as follows: (1) a random deviation  $N(0, 0.15)$  sampled from a normal distribution is added to each default demand pattern to create an “actual” demand pattern; (2) EPANET is run to generate two sets of nodal pressures at measured locations, corresponding with two configurations of the demand patterns; (3) a random

error is added to each nodal pressure. Three scenarios of random errors are considered:  $\Delta^{\text{meas}} = 0$  (perfect measurement),  $\Delta^{\text{meas}} = \pm 0.5$  m and  $\Delta^{\text{meas}} = \pm 1.0$  m.

Table 1 shows the parameters applied in the PF model. The estimation results for the geographically distributed demand pattern network as well as the non-geographically distributed demand pattern network associated with different level of measurement errors are summarised in the following sections.

Table 1. Particle filter model parameters

PF model parameters	Values
Auto regression coefficient	0.7
Variance of noise	0.16
Number of particles	100,000

### 3.1 Perfect measurements

The left hand side plots of Figure 3 give the default DMFs, the actual DMFs and the estimated DMFs for the geographically distributed and non-geographically distributed patterns over 48 hours. The right hand side plots of Figure 3 display the scattergrams of the default DMFs as well as the estimated DMFs versus their actual values.



Figure 3. Outputs from the PF model when the measurements are considered error free

Without considering measurement errors, the PF model provides good estimates of the DMFs for these networks. By assimilating real-time pressure information into the estimation process, the default demand patterns are adjusted, approaching the actual patterns. This adjustment can be observed in Table 2, via the values of the coefficients of determination ( $R^2$ ) and mean absolute error (MAE) of each demand pattern. The default DMFs show an average correlation to the actual DMFs with ( $R^2$ ) ranging from 0.709 to 0.814, while the estimated DMFs for both geographically distributed and non-geographically distributed demand patterns are strongly correlated to the actual ones with all  $R^2$  values being close to unity. It is also seen that the PF model gives better estimates for the network with geographically distributed demand patterns than the network with non-geographically distributed demand patterns, as the  $R^2$  values are closer to unity and the MAE values are smaller.

Table 2. Comparison of the estimation derived from the PF model when the measurements are considered error free ( $R^2$  – Coefficient of determination, MAE – Mean absolute error)

DMFs	Default DMFs		Geo. distributed DMFs		Non-Geo. distributed DMFs	
	$R^2$	MAE	$R^2$	MAE	$R^2$	MAE
DMF1	0.709	0.097	0.956	0.041	0.895	0.049
DMF2	0.703	0.101	0.975	0.023	0.868	0.053
DMF3	0.840	0.175	0.949	0.102	0.909	0.144
DMF4	0.814	0.259	0.983	0.081	0.940	0.159

### 3.2 Measurement errors

Two levels of measurement errors including  $\Delta^{\text{meas}} = \pm 0.5$  m and  $\Delta^{\text{meas}} = \pm 1.0$  m for all measurements are considered for the second test of the PF model. The accuracy of the estimates, presented by the same assessment criteria ( $R^2$  and MAE), as shown in Table 3 and Table 4.

Table 3. Comparison of the estimation from the PF model when  $\Delta^{\text{meas}} = \pm 0.5$  m

DMFs	Geographically distributed DMFs		Non-Geo. distributed DMFs	
	$R^2$	MAE	$R^2$	MAE
DMF1	0.864	0.071	0.812	0.062
DMF2	0.954	0.032	0.818	0.064
DMF3	0.934	0.120	0.905	0.141
DMF4	0.977	0.089	0.946	0.153

It is observed that with relative small measurement errors ( $\Delta^{\text{meas}} = \pm 0.5$  m), the model can still provide reasonable estimates of the DMFs in both networks. The MEA values are smaller and  $R^2$  values are larger for both networks (compared to the default DMFs in Table 2), which means that the PF model has shifted the default DMFs closer to the actual DMFs. Similar to the previous test, the estimation results of the PF model for the geographically distributed demand pattern network are more accurate than for the non-geographically distributed demand pattern network, especially for DMF2 and DMF4.

The values in Table 4, on the other hand, show that with large measurement errors ( $\Delta^{\text{meas}} = \pm 1.0$  m) the PF model cannot provide good estimates of the DMFs. For the geographically distributed demand pattern network, the estimated DMF1 has very weak correlation with the actual DMF1. The estimates of DMF2 and DMF3 show an accuracy similar to the default patterns. Better estimation results can only be achieved in DMF4, where the demand group is connected to the others by a single pipe and the pressure in this region is mainly dependent on the pressure in Tank 2.

For the non-geographically distributed demand pattern network, the estimated DMF1, DMF2 and DMF3 are even more inaccurate. The estimated DMF4, which is associated with nodes with largest base demands in the network, is slightly improved because this demand group dominates the other demand groups. Due to large base demands, small changes in this pattern can cause a large change in the demand at these nodes, which subsequently results in a large change in the pressure at measured locations.

Table 4. Comparison of the estimation from the PF model when  $\Delta^{meas} = \pm 1.0 m$

DMFs	Geographically distributed DMFs		Non-Geo. distributed DMFs	
	R <sup>2</sup>	MAE	R <sup>2</sup>	MAE
DMF1	0.527	0.113	0.458	0.136
DMF2	0.778	0.092	0.577	0.140
DMF3	0.847	0.220	0.809	0.202
DMF4	0.941	0.145	0.869	0.235

The estimated DMFs derived from the PF model that considers a measurement error of  $\Delta^{meas} = \pm 1.0 m$  for 48 hours are shown in Figure 4. Large errors can be seen at almost all the time steps during this extended period. In this case, the default demand patterns would be a better input for a hydraulic model and would provide a more accurate representation of the real network behaviour.



Figure 4. Outputs from the PF model when the measurements error  $\Delta^{meas} = \pm 1.0 m$

## 4 CONCLUSIONS

The work in this paper has evaluated the performance of the PF model proposed by [2] for the near real-time estimation of water demand multipliers. Two types of networks have been studied: a geographically distributed demand pattern network and a non-geographically distributed demand pattern network. Different level of measurement errors have also been examined. Results show that the PF model can be used for relatively large networks with multiple demand patterns. Well estimated DMFs can be obtained if the measurement errors are relatively small. The model cannot provide good estimates if large errors are contained in the measurement data. In addition, the results also show that the model performs better with the geographically distributed demand pattern network than with the non-geographically distributed demand pattern network for all scenarios.

### References

- [1] N.C. Do, A.R. Simpson, J.W. Deuerlein & O. Piller, 'Calibration of Water Demand Multipliers in Water Distribution Systems Using Genetic Algorithms', *Journal of Water Resources Planning and Management*, p. 04016044, 2016.
- [2] 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', *Journal of Water Resources Planning and Management*, accepted on 18 May 2017
- [3] D. Kang & K. Lansey, 'Real-time demand estimation and confidence limit analysis for water distribution systems', *Journal of Hydraulic Engineering*, vol. 135, no. 10, pp. 825-837, 2009.
- [4] F. Odan, L.R. Reis & Z. Kapelan, 'Real-Time Multiobjective Optimization of Operation of Water Supply Systems', *Journal of Water Resources Planning and Management*, vol. 141, no. 9, 10.1061/(ASCE)WR.1943-5452.0000515, 04015011, 2015.
- [5] G. Sanz, R. Pérez, Z. Kapelan & D. Savic, 'Leak detection and localization through demand components calibration' *Journal of Water Resources Planning and Management*, 142(2), p.04015057, 2015.
- [6] J.E. van Zyl, O. Piller & Y. le Gat, 'Sizing municipal storage tanks based on reliability criteria', *Journal of Water Resources Planning and Management*, vol. 134, no. 6, pp. 548-555, 2008.
- [7] C. Beal & R. Stewart, '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, 2014.
- [8] L.A. Rossman, 'EPANET 2: users manual', US Environmental Protection Agency (EPA), USA, 2000.
- [9] J.D. Hol, T.B. Schon & F. Gustafsson, 'On resampling algorithms for particle filters', *Nonlinear Statistical Signal Processing Workshop, 2006 IEEE, IEEE*, pp. 79-82, 2006.